Systematic Analyses of a Chemokine Family-Based Risk Model Predicting Clinical Outcome and Immunotherapy Response in Lung Adenocarcinoma

Jiarui Chen[1] ⓘ, Xingyu Liu[1], Qiuji Wu[1,2,3], Xueping Jiang[1],
Zihang Zeng[1], Jiali Li[1], Yanping Gao[1], Yan Gong[4,5],
and Conghua Xie[1,2,3] ⓘ

Abstract

Chemokines exhibited complicated functions in antitumor immunity, with their expression profile and clinical importance of lung adenocarcinoma (LUAD) patients remaining largely undetermined. This study aimed to explore the expression patterns of chemokine family in LUAD and construct a predictive chemokine family-based signature. A total of 497 samples were downloaded from the Cancer Genome Atlas (TCGA) data portal as the training set, and the combination of 4 representative Gene Expression Omnibus (GEO) datasets, including GSE30219, GSE50081, GSE37745, and GSE31210, were utilized as the validation set. A three gene-based signature was constructed using univariate and stepwise multivariate Cox regression analysis, classifying patients into high and low risk groups according to the overall survival. The independent GEO datasets were utilized to validate this signature. Another multivariate analysis revealed that this signature remained an independent prognostic factor in LUAD patients. Furthermore, patients in the low risk group featured immunoactive tumor microenvironment (TME), higher IPS scores and lower TIDE scores, and was regarded as the potential beneficiaries of immunotherapy. Finally, the role of risky CCL20 was validated by immunohistochemistry (IHC), and patients possessed higher CCL20 expression presented shorter overall survival ($P = 0.011$).

Keywords

lung adenocarcinoma, chemokine family, clinical prognosis, tumor microenvironment, immunotherapy

Introduction

Non-small cell lung cancer (NSCLC) is the leading cause of cancer-related mortality worldwide [1]. NSCLC can be further classified as squamous carcinoma, adenocarcinoma and large cell carcinoma according to the histology. Despite the spectacular progress in detection of driver gene mutations and individualized targeted therapy, especially tyrosine kinase inhibitors (TKIs), the 5-year overall survival (OS) of lung adenocarcinoma (LUAD) remains approximately 15%[2], and the standard platinum-based chemotherapy and targeted therapy seemed to reach a plateau in this setting.

Recently, multiple clinical trials provide exciting results of immune checkpoint inhibitors (ICIs) in solid tumors and ICIs are becoming the first-line treatment option for advanced NSCLCs. KEYNOTE 024 examined the efficacy of pembrolizumab monotherapy or platinum-based chemotherapy in previously untreated NSCLC patients with

[1] Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan, China
[2] Hubei Key Laboratory of Tumor Biological Behaviors, Zhongnan Hospital of Wuhan University, Wuhan, China
[3] Hubei Cancer Clinical Study Center, Zhongnan Hospital of Wuhan University, Wuhan, China
[4] Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, China
[5] Tumor Precision Diagnosis and Treatment Technology and Translational Medicine, Hubei Engineering Research Center, Zhongnan Hospital of Wuhan University, Wuhan, China

Corresponding Authors:
Conghua Xie and Yan Gong, Zhongnan Hospital of Wuhan University, 169 Donghu Road, Wuhan, Hubei 430071, China.
Emails: chxie_65@whu.edu.cn; yan.gong@whu.edu.cn

PD-L1 tumor proportion score $\geq 50\%$. The median progression-free survival was 10.3 months for the pembrolizumab group and 6.0 months for the platinum-based chemotherapy group. The estimated rate of OS was 44.8% versus 27.8%, respectively[3]. However, the objective response rate to ICIs ranged from 12% to 23% in solid tumors depending on multiple factors[4]. Current researches focus on several potential biomarkers predictive to ICIs treatment response in NSCLC, including tumor PD-L1 expression, tumor mutational burden (TMB), tumor infiltrative T cells[5]. However, these biomarkers could not fully reflect the heterogeneous tumor microenvironment (TME) of LUAD. As a matter of fact, none of these biomarkers alone or in combination could give optimal predictive value such as those biomarkers (i.e., *EGFR* mutations, *ALK* rearrangements) seen in targeted therapies. New biomarkers and prediction models are in urgent need to stratify response and to identify patients who might benefit from ICIs treatment.

The chemokine family, consisting of *CC*, *CXC*, *CX₃C*, and *XC* subfamilies, exhibit complicated functions in antitumor immunity, with the collective properties being either anti- and pro-cancer. The antitumor chemokines are known to favor the antitumor immunity by forming a concentration gradient of chemokines between tumor lesions and peripheral blood and lymph vessels, therefore recruiting immune cells expressing corresponding receptors. Chemokines also directly suppress tumor growth and metastasis[6]. The pro-cancer chemokines are associated with promoting immunosuppressive TME[7], supporting tumor growth[8], angiogenesis, and metastasis[9]. Novel agents targeting the pro-cancer chemokine receptors, such as *CXCR4* antagonists and *CCR4* antibodies, are currently under clinical investigations in different phases[10–12]. However, chemokine family-related gene expression profiling and the subsequent clinical implications in LUAD remain undetermined.

This comprehensive study aimed to explore the expression profile and clinical implications of chemokine family in LUAD. A chemokine family-based risk model was constructed and validated to predict the clinical outcome of LUAD patients. Given the complex effects of chemokine family on TME and their potential roles in ICI responses, we further investigated this signature related immune-cancer landscape, hoping to provide novel insights to optimize the delivery of immunotherapy.

Methods and Materials

Data Collection and Preprocessing

The RNA-Seq data and clinical information of 497 LUAD patients, acquired from the Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov/), were utilized to construct the training set. Data on somatic mutations were also downloaded from the TCGA data portal. Representative Gene Expression Omnibus (GEO) datasets containing large number of LUAD patients based on the platform of GPL570

(Affymetrix Human Genome U133 Plus 2.0 Array), including 85 cases from GSE30219, 128 cases from GSE50081, 106 cases from GSE37745 and 226 cases from GSE31210 were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/geo) as the validation sets. The mRNA expression profiling from GEO datasets were normalized by Robust Multi-Array average (RMA) algorithm, and experimental batch effects were corrected by limma package. The mean value was employed for genes represented by more than one probe.

Construction and Validation of the Prognostic Model

OS-related chemokine family members were sought out by the univariate Cox regression model and were enrolled in a subsequent stepwise multivariate Cox regression model. Three genes with the most powerful predictive value were finally extracted, and the risk score was calculated as follows: risk score = expression of *Gene 1* * coefficient 1 + expression of *Gene 2* * coefficient 2+ expression of *Gene 3* * coefficient 3. Patients were categorized into high or low risk groups, according to the optimal cutoff value of risk score, which was calculated by the maxstat R package, corresponding to the most significant relation with survival. A total of 4 cohorts from the GEO datasets were utilized to further validate this model.

Prognostic Meta-Analysis

In order to assess the prognostic value of chemokine-based signature, meta-analysis was performed using the meta package. A fixed-effect model was then adopted to compute the pooled hazard ratio (HR) value.

Functional Enrichment Analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted and visualized by the clusterProfiler package.

Evaluation of Tumor-Infiltrating Immune Cells

We applied CIBERSORT algorithm to infer the abundance of tumor-infiltrating immune cells in different risk groups[13]. LM22 signature file utilized in the analysis were downloaded from the website (http://cibersort.stanford.edu/). The *P*-value was estimated using the deconvolution approach and $P < 0.05$ was considered as the main criterion to further select samples enrolled in the analysis. Fractions of individual immune cells between different risk groups were visualized by the pheatmap and ggplot packages.

Prediction of Immunoreactivity

Expression of biomarkers believed or known to be associated with immunotherapy responses, including PD-1, PD-L1, PD-L2, CTLA4, CD4, CD8A, and CD8B, was compared between the high and low risk groups. Immunophenoscore
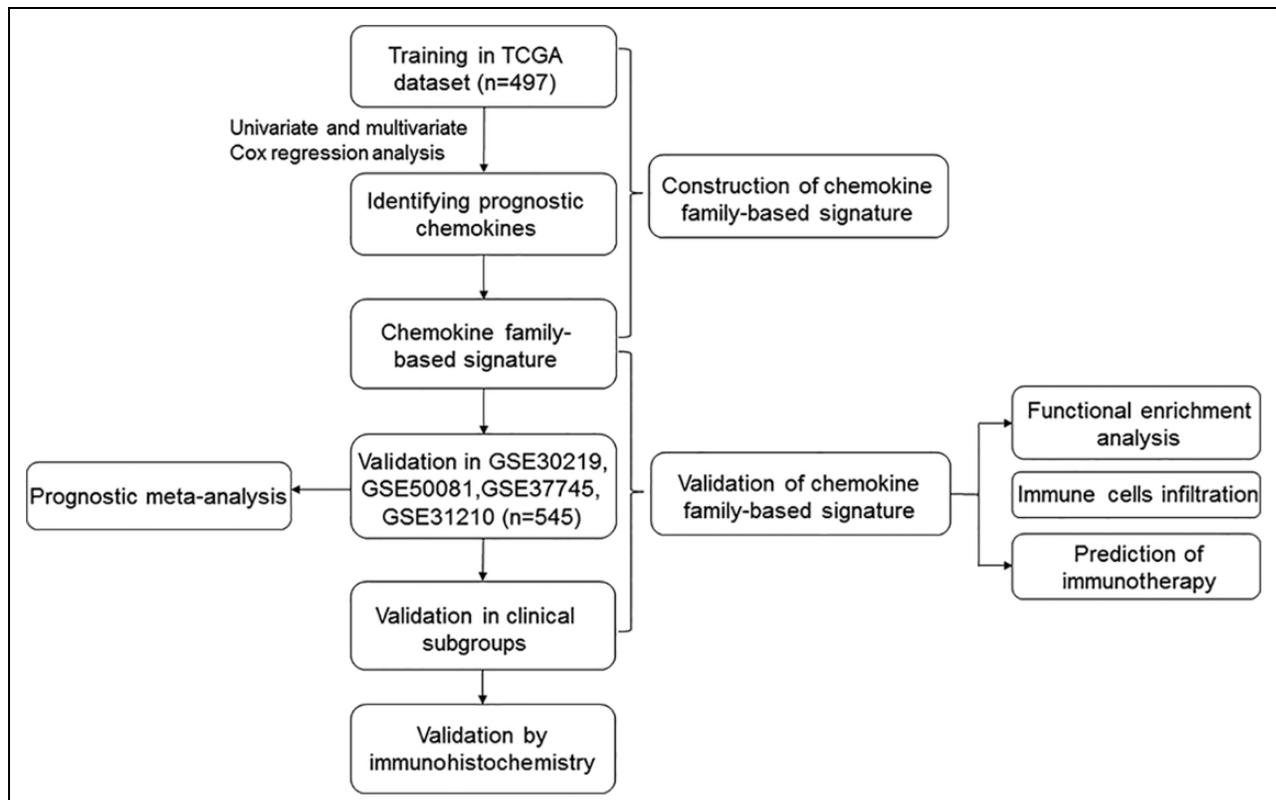
Figure 1. Flow chart of the study. The TCGA-LUAD cohort was used as the training set, and the combination of GSE30219, GSE50081, and GSE37745, GSE31210 cohort was used as the validation set. TCGA, the Cancer Genome Atlas; LUAD, lung adenocarcinoma.

(IPS) for LUAD patients, ranging from 0 to 10, was obtained from The Cancer Immunome Atlas (TCIA) (https://tcia.at/home), with higher scores correlated with stronger immunoreactivity[14]. Tumor immune dysfunction and exclusion (TIDE), a calculation method based on two different immune evasion mechanisms, including T cell dysfunction and exclusion[15], was also utilized to predict the immunotherapy response. Upon uploading transcriptome data, the TIDE scores and PD-L1 expression for LUAD patients in TCGA dataset were obtained from TIDE website (http://tide.dfci.harvard.edu/).

Tissue Microarray and Immunohistochemistry (IHC)

Tissue microarray, containing a total of 52 pairs of human lung adenocarcinoma and adjacent normal tissue, were obtained from Zhuoli Biotechnology Co, Shanghai, China. The samples come from the National Human Genetic Resources Sharing Service Platform (2005DKA21300). After the tissue was dewaxed, antigen retrieval was performed using the EDTA antigen repair buffer. Tissue section was incubated with the primary antibody for CCL20 (Affinity Biosciences, China, DF2238) at 4°C overnight. Furthermore, the section was incubated with second antibody for 50 minutes and visualized by diaminobenzidine staining. Finally, a counterstain was performed by hematoxylin. IHC scores were analyzed by two independent investigators. Staining intensity was

graded as follows: 0 (−, no staining); 1 (+, weak staining, light yellow); 2 (++, moderate staining, yellow brown); 3 (+++, strong staining, brown). Both the intensity and positive percentages of IHC were used to examine the CCL20 expression: the IHC H-score (values 0-300) = staining intensity × the percentage of positive-stained cells × 100.

Statistical Analysis

Univariate Cox regression model and a subsequent stepwise multivariate Cox regression were employed to screen the most significant survival-related genes. Kaplan-Meier method was carried out to evaluate OS for patients in different risk groups or with different CCL20 expression, and a log-rank test was performed to compare survival differences between groups. Wilcoxon test was utilized to evaluate fractions of immune cells, expression of biomarkers, TIDE scores between different risk groups and CCL20 expression between tumor tissues and adjacent tissues. All the aforementioned statistics were conducted using R software (version 3.6.3). A two-sided P value < 0.05 was considered to be statistically significant.

Results

Predictive Value of Chemokine Family Genes in LUAD

The workflow of study design was summarized in Fig. 1. A total of 71 well-specified chemokine genes, divided into four

Table 1. Clinical Characteristics of Patients Enrolled in the Study.

Characteristics	TCGA (n = 497) n (%)	GSE30219 (n = 85) n (%)	GSE50081 (n = 128) n (%)	GSE37745 (n = 106) n (%)	GSE31210 (n = 226) n (%)
Age (years)					
≤65	236 (47.5)	60 (70.6)	40 (31.2)	57 (53.8)	176 (77.9)
>65	261 (52.5)	25 (29.4)	88 (68.8)	49 (46.2)	50 (22.1)
Gender					
Female	269 (54.1)	19 (22.4)	63 (49.2)	60 (56.6)	121 (53.5)
Male	228 (45.9)	66 (77.6)	65 (50.8)	46 (43.4)	105 (46.5)
Smoking history					
Yes	412 (82.9)	–	92 (71.9)	–	111 (49.1)
No	71 (14.3)	–	23 (18.0)	–	115 (50.9)
NA	14 (2.8)	–	13 (10.1)	–	
TNM stage					
I & II	385 (77.5)	85 (100.0)	128 (100.0)	89 (84.0)	226 (100)
III & IV	105 (21.1)	0 (0.0)	0 (0.0)	17 (16.0)	–
NA	7 (1.4)	0 (0.0)	0 (0.0)	0 (0.0)	–
OS status					
Alive	317 (63.8)	40 (47.1)	76 (59.4)	29 (27.4)	191 (84.5)
Dead	180 (36.2)	45 (52.9)	52 (40.6)	77 (72.6)	35 (15.5)

*NA, not available; OS, overall survival

Table 2. Univariate Cox Regression Analysis Identifying Chemokine Family Correlated with OS.

Official Symbol	HR	95%CI	P value
CXCL17	0.86	0.80–0.92	<0.001
CCL13	0.90	0.82–0.99	0.03
CCL17	0.88	0.79–0.98	0.02
CCL20	1.10	1.02–1.19	0.01
XCL2	0.77	0.61–0.97	0.03
CX3CL1	0.89	0.80–0.98	0.02
CXCR4	0.83	0.72–0.96	0.01
CXCR6	0.76	0.62–0.94	0.01
CCR2	0.66	0.54–0.81	<0.001
CCR4	0.73	0.60–0.90	0.003
CCR7	0.82	0.70–0.96	0.02
CX3CR1	0.71	0.58–0.88	0.001
ACKR1	0.87	0.78–0.97	0.01

*HR, hazard ratio; CI, confidential interval

Table 3. Multivariate Cox Regression Analysis Identifying Chemokine Family Correlated with OS.

Official Symbol	HR	95%CI	P value
CXCL17	0.85	0.80–0.91	<0.001
CCL20	1.15	1.06–1.23	<0.001
CCR2	0.67	0.54–0.82	<0.001

*OS, overall survival; HR, hazard ratio; CI, confidential interval

reverse, only *CCL20* was recognized as a risky gene, with an HR greater than 1.

Construction of Chemokine Family-Based Signature in TCGA Dataset

To explore genes most significantly related with OS, the aforementioned 13 genes were entered into a step-wise multivariate Cox regression model. *CXCL17*, *CCL20*, and *CCR2* were screened out (Table 3) and risk scores were constructed using the following formula: risk score = (−0.158397 * expression of *CXCL17*) + (0.135678 * expression of *CCL20*) + (-0.406781 * expression of *CCR2*). Risk score for each patient was calculated based on the formula. Fig. 2a demonstrated the expression of *CXCL17*, *CCL20*, and *CCR2* in detail and displayed the corresponding risk score of each sample.

Patients were subdivided into the high ($n = 154$) and low risk ($n = 343$) groups according to the optimal cut-off value (risk score = 1.2477). As shown in Fig. 2b-d, the high risk group revealed a relatively adverse clinical outcome for the overall patients ($P < 0.001$) and patients presented with either early stages (stages I & II, $P < 0.001$) or advanced stages (stages III & IV, $p = 0.005$) diseases.
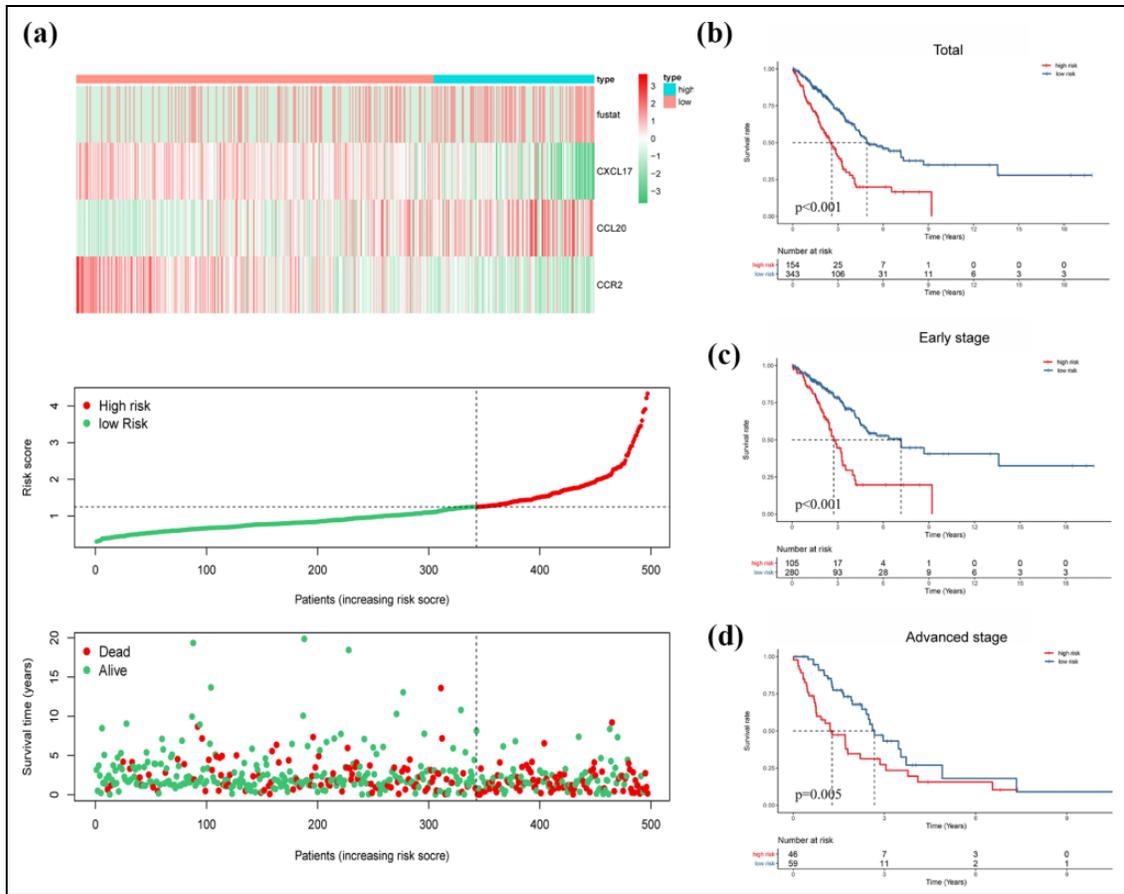
distinct subfamilies-*CC*, *CXC*, *CX₃C*, and *XC*-and their receptors were enrolled in this study. First, RNA-seq data and clinical information data of LUAD patients were downloaded from the TCGA dataset. Patients without follow-up data were excluded. Demographic characteristics of this cohort were listed in Table 1. RNA expression profile of chemokine family genes was extracted and genes with log2 expression value < 1 were excluded. A univariate Cox proportional hazards regression model was established to identify genes correlated with the OS of LUAD patients. Finally, thirteen genes were shown to be associated with OS ($P < 0.05$) (Table 2). Among these genes, there existed 12 protective genes, including *CXCL17*, *CCL13*, *CCL17*, *XCL2*, *CX3CL1*, *CXCR4*, *CXCR6*, *CCR2*, *CCR4*, *CCR7*, *CX3CR1*, and *ACKR1*, with hazard ratios (HRs) <less than 1. Quite the

Figure 2. The construction of chemokine family-based signature in TCGA cohort. (a) the distribution of risk score, survival plot and gene expression patterns. Patients were classified into the high and low risk groups based on the risk scores. (b) Kaplan-Meier curves of OS in total patients ($n = 497$). (c) Kaplan-Meier curves of OS at early stage (stages I&II) patients ($n = 385$). (d) Kaplan-Meier curves of OS in advanced stage (stages III&IV) patients ($n = 105$).
*OS, overall survival; HR, hazard ratio; CI, confidential interval; WT, wide-type; MUT, mutation
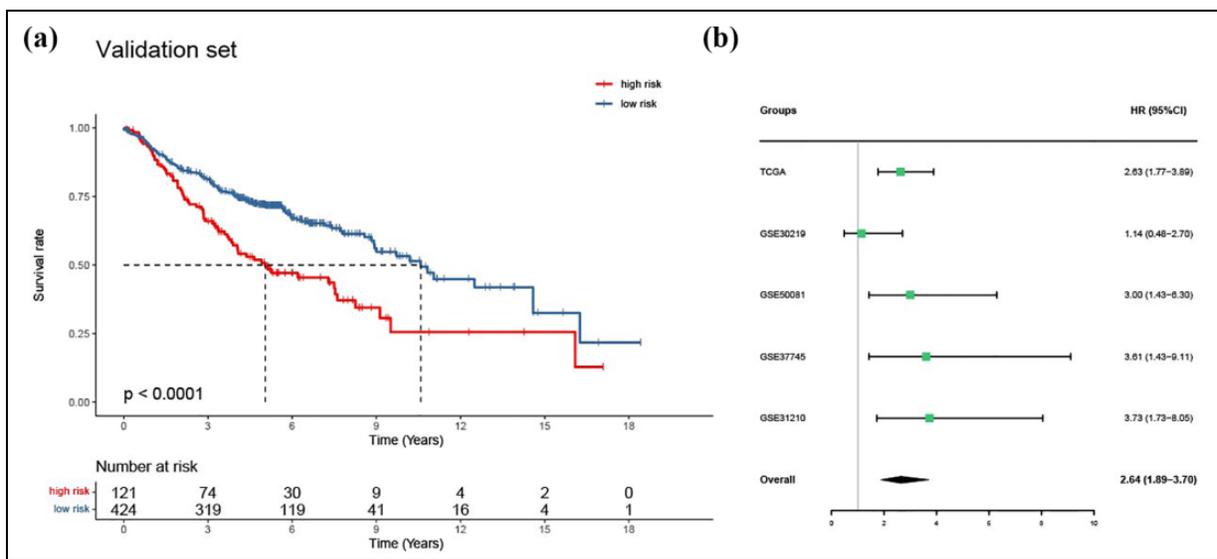

Figure 3. Validation of chemokine family-based signature in three representative GEO datasets. (a) Combination of GSE30219 ($n = 85$), GSE50081 ($n = 128$), GSE37745 ($n = 106$) and GSE31210 ($n = 226$). (b) A meta-analysis performed based on the aforementioned TCGA and GEO datasets.
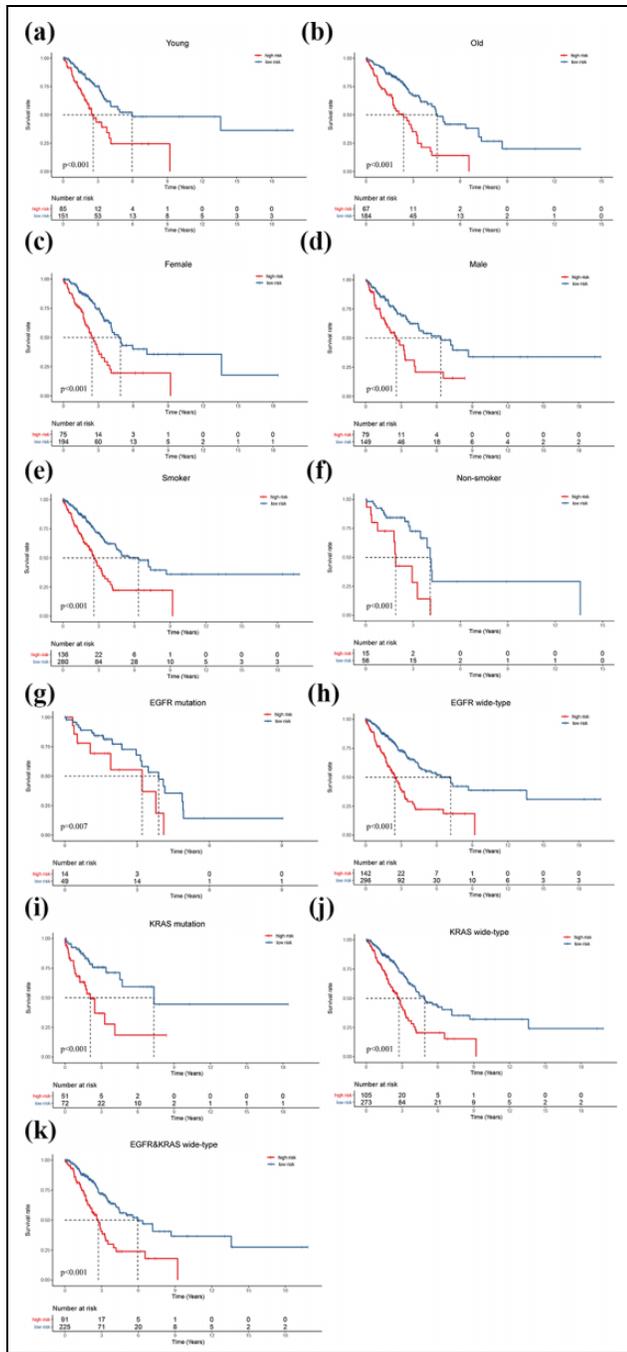
Figure 4. Validation of chemokine family-based signature in different clinical subsets, including (a) young (age ≤ 65, n = 236), (b) old (age > 65, n = 251), (c) female (n = 269), (d) male (n = 228), (e) smoker (n = 416), (f) non-smoker (n = 71), (g) EGFR mutation (n = 63), (h) EGFR wide-type (n = 438), (i) KRAS mutation (n = 123), (j) KRAS wide-type (n = 378) and (k) EGFR and KRAS wide-type (n = 316).

Validation of the Chemokine Family-Based Signature in GEO Cohorts

To validate the reliability of chemokine family-based signature in LUAD patients, 4 cohorts-GSE30219, GSE50081,

Table 4. Univariate and Multivariate Cox Regression Analyses Exploring Factors Associated with OS.

Factors	Univariable analysis			Multivariable analysis		
	HR	95%CI	P value	HR	95%CI	P value
Age						
>65 vs ≤65	1.25	0.92 -1.69	0.16			
Gender						
Male vs female	1.12	0.82 -1.52	0.47			
Stage						
III&IV vs I&II	2.46	1.78–3.40	<0.001	2.19	1.59–3.04	<0.001
Smoking history						
Yes vs no	0.90	0.59 -1.37	0.61			
EGFR status						
WT vs MUT	0.63	0.42–0.96	0.03	0.64	0.42–0.97	0.03
KRAS status						
WT vs MUT	0.96	0.67–1.39	0.84			
Risk score						
High vs low	1.92	1.59–2.32	<0.001	1.93	1.58–2.34	<0.001

GSE37745, and GSE31210 from the platform of GPL570 were downloaded and combined. Risk score for each sample was also calculated by the same formula and patients were classified into the high and low risk groups based on the optimal cut-off value (risk score = 0.1543). It is not unexpected that patients in the low risk group exhibited prolonged OS (P < 0.001, Fig. 3a).

Furthermore, we performed a meta-analysis to confirm the prognostic value of chemokine family-based signature in LUAD patients based on these 5 cohorts. Results showed that the chemokine family-based signature was indeed a risk factor for LUAD patients (pooled HR = 2.64, 95%CI [1.89–3.70], Fig. 3b).

Validation of the Chemokine Family-Based Signature in Clinical Subgroups

In order to validate the chemokine family-based signature in diverse clinical subgroups, we further divided patients from the TCGA database of different sex, age and smoking status into the high and low risk groups based on the cut-off point (risk score=1.2477). Analysis of OS time revealed that patients in the high risk group represented a significantly better prognosis across all stratified subsets (Fig. 4a-f, all P<0.05). In terms of patients with different somatic mutations, likewise, regardless of the mutation status, the chemokine family-based signature displayed robust prognostic value in each subtype, containing *EGFR* mutation, *EGFR* wide type (WT), *KRAS* mutation, *KRAS* WT, and *EGFR*/ *KRAS* WT group (Fig. 4g-k, P < 0.05).

The Chemokine Family-Based Signature Serves as an Independent risk Factor for LUAD Patients

Univariate and multivariate Cox regression model were performed on the TCGA dataset to explore the relationship between the signature and clinical characteristics. Advanced
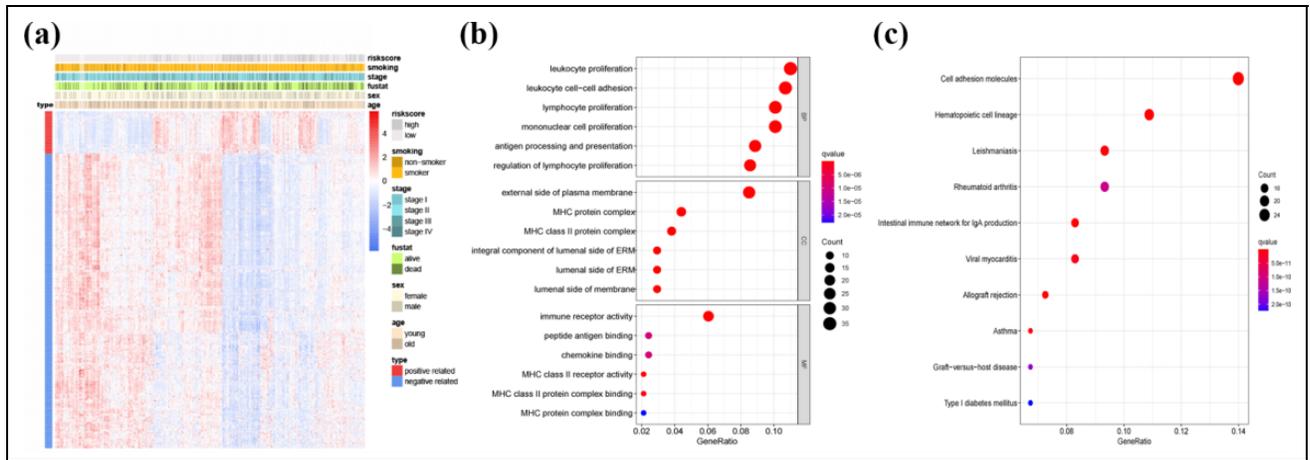
Figure 5. Function and pathway enrichment analysis of the signature related genes. (a) Heatmap of signature related genes (Pearson |r| > 0.4). (b) and (c) GO and KEGG analyses of biological processes and pathways.
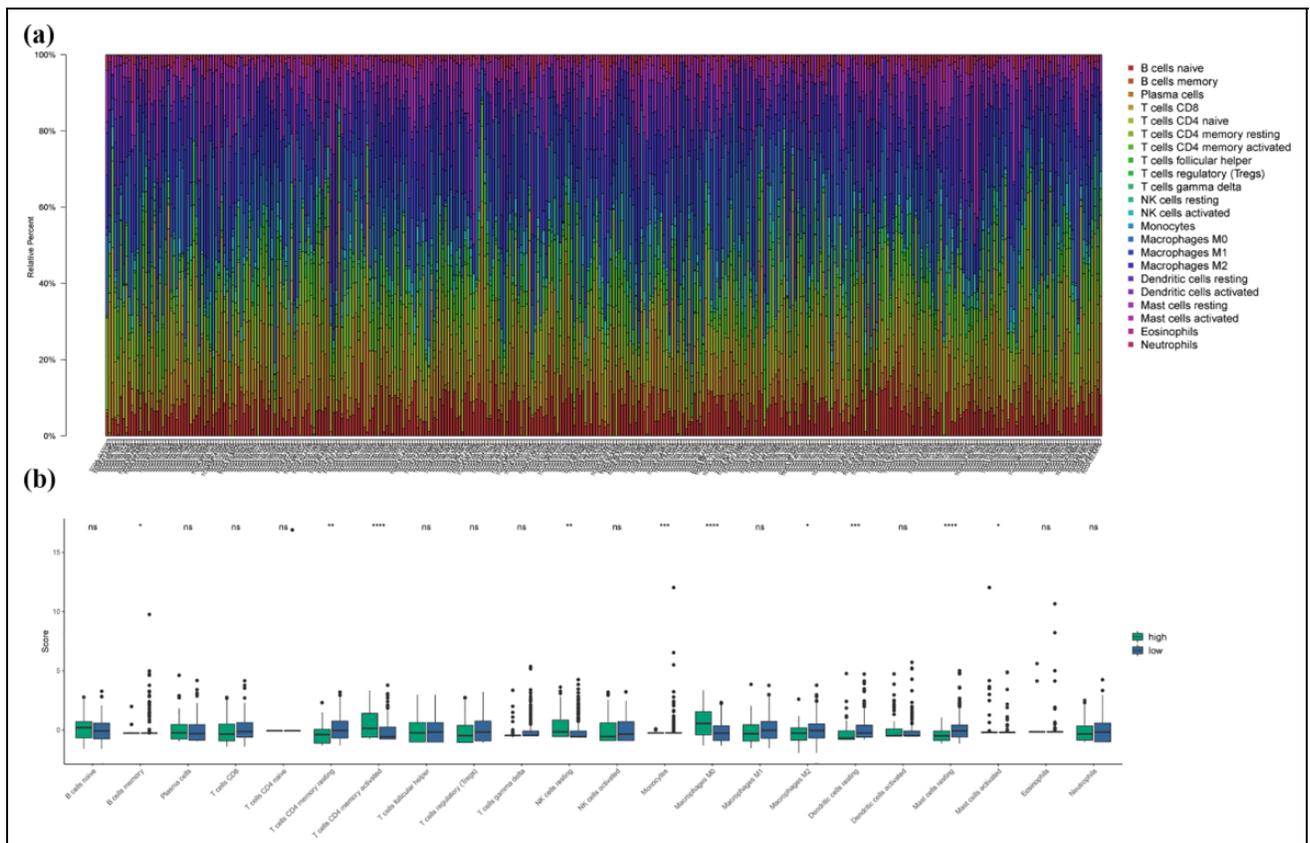

Figure 6. CIBERSORT algorithm analysis of tumor infiltrating immune cells in TCGA cohort. (a) The abundance of immune cells in each sample. (b) Comparisons of immune-cell fractions between the high and low risk groups. * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.

stages (stages III & IV) and high risk scores were correlated with unfavorable clinical outcome, while *EGFR* mutation was associated with gratifying OS period in both univariate and multivariate analysis (Table 4), indicating that the risk score was a crucial independent prognostic factor for LUAD patients.

Functional Enrichment Analysis for Chemokine Family-Based Signature Related Pathways

Genes correlated with the signature were extracted to investigate the biological functions. A total of 357 genes were finally screened out (Pearson |r| > 0.4), of which 45 were
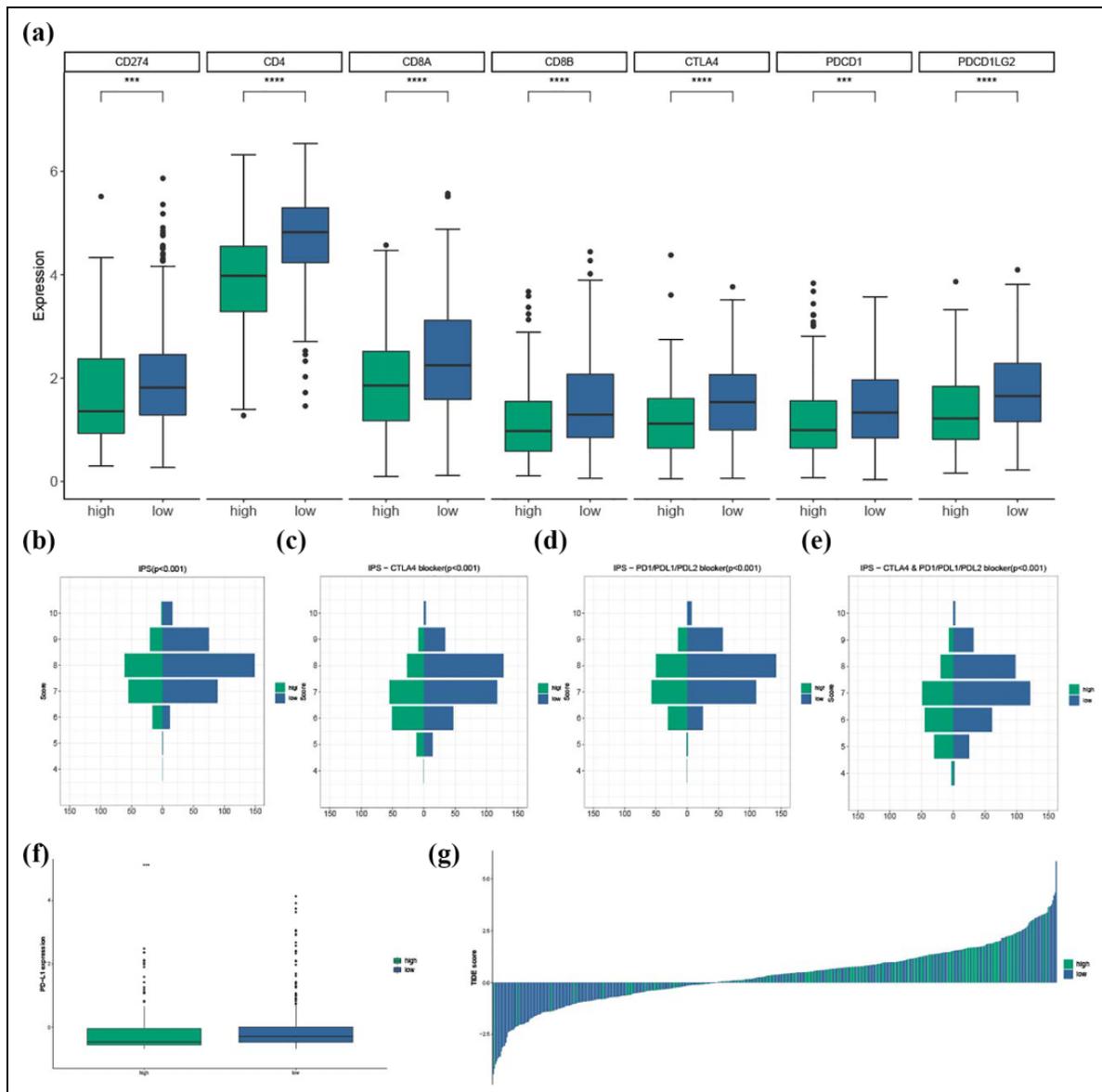
Figure 7. Expression of immune biomarkers predicting response to immunotherapy. (a) Comparisons of immune biomarker between the high and low risk groups. (b-e) Estimated IPS scores between the high and low risk groups. (f) Estimated PD-L1 expression between the high and low risk groups. (g) Waterfall of TIDE scores in both high and low risk groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

positively related and 312 were negatively related (Fig. 5a). GO and KEGG enrichment analysis were then conducted. The top six biological process, cellular component and molecular function terms were illustrated in Fig. 5b. The most significantly enriched pathways were "Cell adhesion molecules," "Hematopoietic cell lineage," "Leishmaniasis," "Rheumatoid arthritis," "Intestinal immune network for IgA production" and "Viral myocarditis" (Fig. 5c).

Immune Landscape of the Chemokine Family-Based Signature

The intimate association between chemokine family and immune-related signaling pathways inspired us to conduct an intensely investigation of the signature and related TME. CIBERSORT was carried out to predict the tumor-infiltrating immune cell fractions in each sample. The abundance and different distributions of immune cells between the high and low risk groups were displayed in Fig. 6a, b. In detail, memory B cells, memory CD4 T cells, monocytes, M2 macrophages, resting dendritic cells, resting mast cells and activated mast cells were extensively enriched in the low risk group, while the high risk group had a greater population of activated memory CD4 T cells, resting NK cells and M0 macrophages. Furthermore, we explored the expression patterns of PD-L1, PD-1, CTLA4, PD-L2, CD4, CD8A, and CD8B for immunotherapy responses in these two groups and concluded that patients in the high risk group exhibited
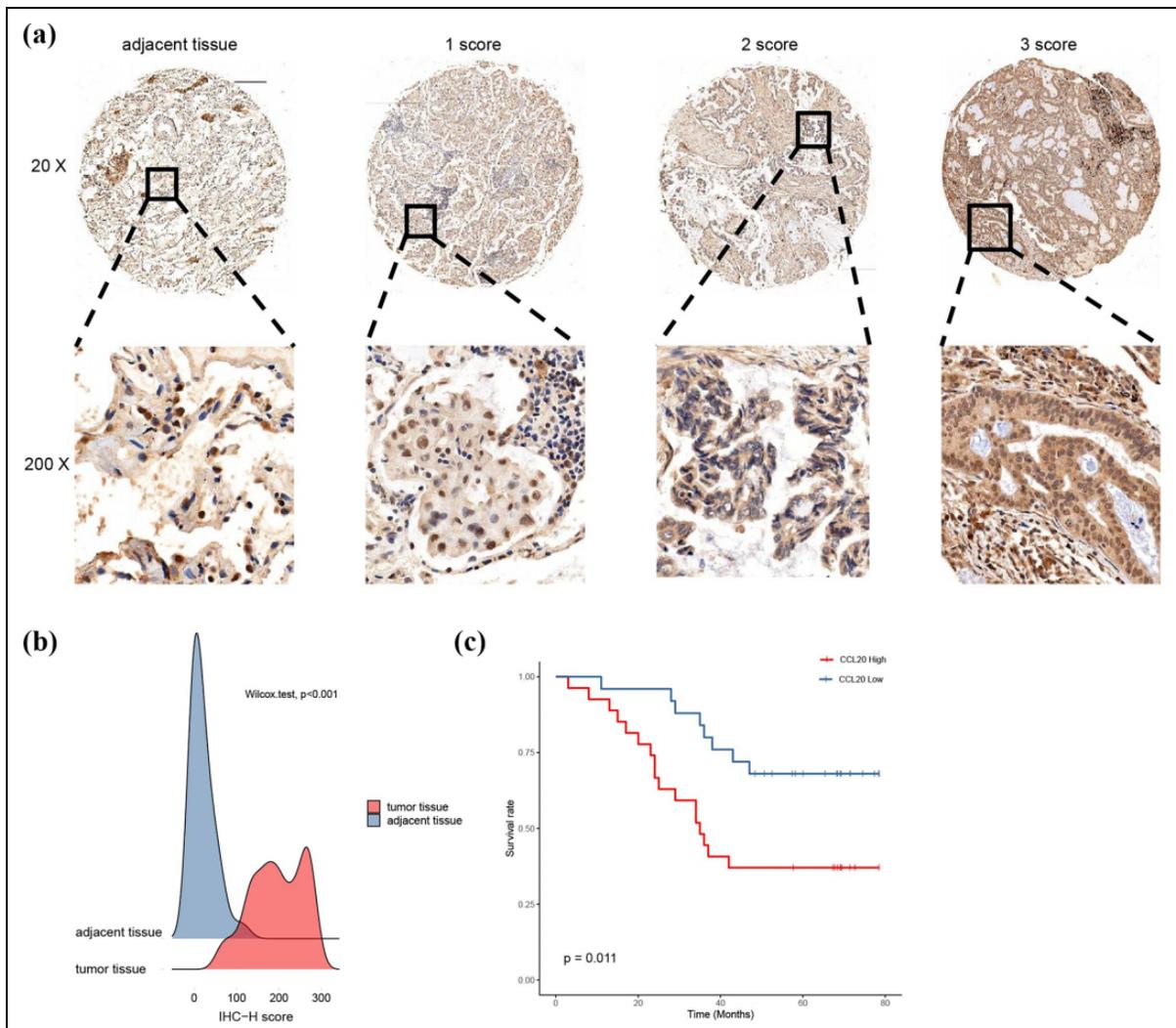
Figure 8. CCL20 was highly expressed in LUAD tissues compared to adjacent tissues and related to poor overall survival. (a) The 20 X and 200 X representative IHC staining. (b) The expression of CCL20 was significant higher in LUAD tissues compared to those in paired normal tissues ($P < 0.001$). (c) Kaplan-Meier plot of 52 patients with survival data (from tissue arrays) stratified by CCL20 expression levels. Patients possessing more CCL20 exhibited poorer overall survival ($P = 0.011$). LUAD, lung adenocarcinoma; IHC, immunohistochemical.

lower expression of immune biomarkers ($P < 0.001$, Fig. 7a). IPS scores were compared for patients in the high versus low risk group. Higher IPS scores were observed in the low risk group ($P < 0.001$) (Fig. 7b). Lower expression of immune biomarkers and IPS scores in the high risk group indicated an immunosuppressive TME and unfavorable responses to ICIs, which might be associated with the shorter OS period in high risk group. Next, TIDE method was also applied to predict response to immunotherapy. Similarly, patients in high risk group possessed lower PD-L1 expression level ($P < 0.001$) and higher TIDE scoring ($P < 0.001$), suggesting poorer response to ICIs (Fig. 7f, g).

Validation of the CCL20 Chemokine

To further valiate the results of bioinformatic analysis, CCL20, the representative risky chemokine in the signature,

was further investigated by IHC staining using a LUAD tissue microarray (Fig. S1). Considering that CCL20 was mostly expressed in cytomembrane and cytoplasm, we compared the expression levels of CCL20 in these areas. The staining patterns of CCL20 were displayed in Fig. 8a. The levels of the expression were quantitated by the IHC-H score based on tumor cell proportion and staining intensity. The IHC staining further confirmed that CCL20 was significantly elevated in LUAD tissue compared to adjacent tissue ($P < 0.001$, Fig. 8b). Samples were then split into 2 groups based on the median IHC-H score at 200, and patients possessed higher CCL20 expression displayed shorter OS period ($P = 0.011$, Fig. 8c).

Discussion

Immunotherapy has opened a new era of anticancer treatment due to its remarkable success in solid tumors. However,

biomarkers evaluating the status of TME and predicting benefit of immunotherapy for LUAD patients are quite rare. Chemokines, on functional grounds, pleiotropically participate to inflammatory and homeostatic, with the former induced by inflammation while the latter involved in the process of immune cell homing. As such, circulating chemokines, modulated easier than tumor infiltrating lymphocytes, are closely related to the immune landscape of TME and could be a good candidate for predicting clinical response to immunotherapy. In this study, we constructed the first comprehensive analysis of the prognostic value of chemokine signature based on 71 well-specified chemokine genes from *CC*, *CXC*, *CX₃C*, *XC* subfamilies and their receptors in TCGA dataset, hoping to provide insight into the multifaceted role of chemokine families in the tumor-immune modulation.

We performed univariate Cox regression analysis followed by stepwise multivariate Cox regression analysis and screened out 3 genes that are significantly correlated with OS of LUAD. We established a risk score signature based on the expression of these 3 genes. A combination of 4 GEO datasets was applied to validate the accuracy of this signature and a meta-analysis was then performed. The chemokine-based signature, associated with adverse OS across all clinical subtypes, was regarded as an independent risk factor for LUAD patients. CIBERSORT, IPS, and TIDE scoring system were utilized to explore the immune landscape of LUAD patients. Interestingly, patients in the low risk group were more likely to have a higher IPS score and lower TIDE score, displaying increased response rate to ICIs. Thus, our model represented strong potential of novel predictors in LUAD patients.

In this analysis, we concluded that the majority of significant chemokine genes appeared as predictive factors for favorable clinical outcomes, which was in consistent with previous studies[6]. *CXCL17*, *CCL20*, and *CCR2* were extracted to establish the risk model. *CXCL17* was a mucous chemokine, with the role in cancer remaining controversial. *CXCL17* was reported to facilitate antitumor immune response at early stage intraductal papillary mucinous neoplasm by inducing dendritic cell trafficking to the tumor niche, thus enhancing the susceptibility of tumor cells exposed to cytotoxic T lymphocytes[16]. Conversely, *CXCL17* was upregulated in breast cancer and hepatocellular carcinoma, correlating with recurrence and poor prognosis[17,18], with the potential mechanisms could be attributed to increased cellular proliferation and metastasis, decreased apoptosis, and presence of immunosuppressive cells in the TME. In the setting of LUAD, previous study had confirmed an upregulation of *CXCL17* expression in LUAD patients compared with that in LUSC patients, which had an essential role in recruiting macrophages through phosphorylation of the Src/FAK pathway[19]. Our results concluded that *CXCL17* was a protective factor for LUAD patients (HR = 0.85, 95% CI [0.80–0.91], *P* < 0.001). *CCL20*, also referred to as macrophage inflammatory protein 3α or liver and activation-regulated chemokine[20,21], was considered as a

candidate target of antitumoral therapy due to its contributions to tumor progression. In lung carcinomas, investigators had revealed that *CCL20*, playing a pivotal role in tumor progression, were upregulated in patients with relapsed lung cancer and augmented cell proliferation through the ERK signaling pathway[22,23]. Downregulation of *CCL20* by docetaxel treatment was strongly implicated with prolonged OS in NSCLC patients[24], which was in consistent with our result that *CCL20* was a risky gene both in bioinformatic analysis and IHC staining, indicating that CCL20 could be a promising anti-tumor target. *CCR2* was a multifunctional and promiscuous receptor, binding not only to CCL2 but also to *CCL7*, *CCL8*, and *CCL13*[25]. The most frequent *CCR2/CCL2* axis had both pro- and anti-tumor effects in lung carcinomas. Investigations showed that the activation of *CCR2/CCL2* axis could recruit macrophages and subsequently promote cancer cell invasion[26]. In addition, the *CCR2/CCL2* axis served a pivotal role in recruiting immunosuppressive MDSC, while intervention of this signaling led to a significant reduction of tumor growth together with increased cytotoxic T cell infiltration[27]. However, in the Human Protein Atlas database, *CCR2* expression was associated with better prognosis in lung cancer[28], which was in line with ours (HR = 0.67, 95% CI [0.54–0.82], *P* < 0.001). The underlying molecular mechanisms of these genes in this signature for LUAD patients remain largely unknown and require further illumination.

The accuracy of this chemokine family-based signature was validated in four GEO cohorts and a subsequent meta-analysis was carried out to confirm the prognostic value of this signature in these datasets. Generally, this signature performed well in multiple cohorts and across different clinical subgroups. Furthermore, genes related with this signature were enriched into immune biological process and signaling pathways, indicating that the differences in OS between groups might be attributed to the heterogeneity of TME and encouraging us to further explore the potential mechanisms. Abundances of tumor-infiltrating immune cells were calculated using CIBERSORT algorithm. Our results revealed that patients in the low risk group featured a higher population of memory B cells, memory CD4 T cells and resting dendritic cells, leading to an adequate antitumor immunological reaction, which was in consistent with previous studies[29,30]. Monocytes played a duel role in different stages of tumorigenesis, functioning on the one hand to differentiate into protumoral tumor associated macrophages and on the other to recruit antitumoral NK cells in lung cancer[31,32]. In this study, monocytes were abundant in the low risk group. However, immunosuppressive M2 macrophages were also highly enriched in the low-risk group, which might indicate that patients in the low risk group might be good candidates for anti-M2 vaccination[33]. Immune biomarkers for ICIs were explored. Patients in the low risk group displayed upregulation of PD-L1, PD-1, CTLA4, PD-L2, CD4, CD8A, and CD8B as well as remarkably increased IPS scores. Besides, patients in the low risk group also presented

significantly decreased TIDE scores. Collectively, these results illustrated that this signature performed well in predicting survival and response to immunotherapy. However, further experiments are needed to validate our findings.

Despite of the aforementioned compelling results, there still exist some limitations that ought to be acknowledged. First, most cases enrolled in the study were obtained from public databases and estimated by bioinformatic methods. Only the role of CCL20 chemokine was validated in tissue microarray by IHC staining. Independent clinical samples and further experiment are still needed to validate other chemokines in this signature. Second, we only focused on the chemokine members, with some crucial genes failing to be investigated. Further researches should certainly take a wider scope. Third, response to immunotherapy were validated indirectly as patients treated with immunotherapy were not included in this study. However, the compelling results outweighed these limitations.

Conclusion

This study investigated the role of chemokine families in LUAD patients, and 13 genes were perceived to be predictors of OS period. A three gene-based signature was subsequently constructed, and the predictive accuracy, related TME and ICI response were then estimated. These findings may serve as a reliable clinical prediction tool and optimize the therapeutic potential of ICIs.

Author Contributions

JC, QW, and CX designed the study. YG, XL, and XJ collected the mRNA transcriptome data and clinical information from public databases. JC and XL performed the immunohistochemical staining. YG, ZZ, JL, and XL performed statistical analyses. JC and XL wrote the manuscript. QW and CX improved and revised the manuscript. All authors read and approved the final manuscript.

Jiarui Chen, and Xingyu Liu are authors have contributed equally to this work.

Availability of Data and Materials

The datasets analyzed in this study can be found in TCGA portal (https://portal.gdc.cancer.gov/) and GEO portal (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30219& https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50081& https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37745& https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31210). Further inquiries can be directed to the corresponding author.

Ethical Approval

The human lung adenocarcinoma tissue arrays were purchased from the the tissue array company Zhuoli Biotechnology Co, Shanghai, China. The samples come from the National Human Genetic Resources Sharing Service Platform (2005DKA21300).

Statement of Human and Animal Rights

All following experiments and human rights were complied with relevant guidelines and regulations.

Statement of Informed Consent

There are no subject in this article and informed consent is not applicable.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iDs

Jiarui Chen https://orcid.org/0000-0002-2934-016X
Conghua Xie https://orcid.org/0000-0001-6623-9864

Supplemental Material

Supplemental material for this article is available online.

Reference

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin. 2021;71(1):7–33.
2. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. Lancet Oncol. 2011;12(2):175–180.
3. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med. 2016; 375(19):1823–1833.
4. Yang K, Li J, Sun Z, Zhao L, Bai C. Retreatment with immune checkpoint inhibitors in solid tumors: a systematic review. Ther Adv Med Oncol. 2020;12:1758835920975353.
5. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol. 2016;17(12):e542–e551.
6. Letourneur D, Danlos FX, Marabelle A. Chemokine biology on immune checkpoint–targeted therapies. Eur J Cancer. 2020; 137:260–271.
7. Low-Marchelli JM, Ardi VC, Vizcarra EA, van Rooijen N, Quigley JP, Yang J. Twist1 induces CCL2 and recruits macrophages to promote angiogenesis. Cancer Res. 2013;73(2):662–671.
8. Mou T, Xie F, Zhong P, Hua H, Lai L, Yang Q, Wang J. MiR-345-5p functions as a tumor suppressor in pancreatic cancer by directly targeting CCL8. Biomed Pharmacother. 2019;111: 891–900.

9. Dagouassat M, Suffee N, Hlawaty H, Haddad O, Charni F, Laguillier C, Vassy R, Martin L, Schischmanoff P-O, Gattegno L, Oudar O, et al. Monocyte chemoattractant protein-1 (MCP-1)/CCL2 secreted by hepatic myofibroblasts promotes migration and invasion of human hepatoma cells. Int J cancer. 2010; 126(5):1095–1108.

10. Bockorny B, Semenisty V, Macarulla T, Borazanci E, Wolpin BM, Stemmer SM, Golan T, Geva R, Borad MJ, Pedersen KS, Park JO, et al. BL-8040, a CXCR4 antagonist, in combination with pembrolizumab and chemotherapy for pancreatic cancer: the COMBAT trial. Nat Med. 2020;26(6):878–885.

11. Doi T, Muro K, Ishii H, Kato T, Tsushima T, Takenoyama M, Oizumi S, Gemmoto K, Suna H, Enokitani K, Kawakami T, et al. A phase I study of the anti-CC chemokine receptor 4 antibody, mogamulizumab, in combination with nivolumab in patients with advanced or metastatic solid tumors. Clin Cancer Res. 2019;25(22):6614–6622.

12. Pelpola S, Ball R, Yang J, Sundaram V, Lavezo J. Macrophage exclusion after radiation therapy (MERT): a first in human phase I/II trial using a CXCR4 inhibitor in Glioblastoma. 2020;25(23):6948–6957.

13. Zhang C, Zhang G, Sun N, Zhang Z, Zhang Z, Luo Y, Che Y, Xue Q, He J. Comprehensive molecular analyses of a TNF family-based signature with regard to prognosis, immune features, and biomarkers for immunotherapy in lung adenocarcinoma. EBioMedicine. 2020;59:102959.

14. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal Genotype-Immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017;18(1):248–262.

15. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24(10):1550–1558.

16. Hiraoka N, Yamazakiitoh R, Ino Y, Mizuguchi Y, Yamada T, Hirohashi S, Kanai Y. CXCL17 and ICAM2 are associated with a potential anti-tumor immune response in early intraepithelial stages of human pancreatic carcinogenesis. Gastroenterology. 2011;140(1):310–321.

17. Li L, Yan J, Xu J, Liu C-Q, Zhen Z-J, Chen H-W, Ji Y, Wu Z-P, Hu J-Y, Zheng L, Lau WY. CXCL17 expression predicts poor prognosis and correlates with adverse immune infiltration in hepatocellular carcinoma. PLoS One. 2014;9(10):e110064.

18. Guo YJ, Zhou YJ, Yang XL, Shao ZM, Ou ZL. The role and clinical significance of the CXCL17-CXCR8 (GPR35) axis in breast cancer. Biochem Biophys Res Commun. 2017;493(3): 1159–1167.

19. Choreño-Parra JA, Thirunavukkarasu S, Zúñiga J, Khader SA. The protective and pathogenic roles of CXCL17 in human health and disease: potential in respiratory medicine. Cytokine Growth Factor Rev. 2020;53:53–62.

20. Kadomoto S, Izumi K, Mizokami A. The CCL20-CCR6 axis in cancer progression. Int J Mol Sci. 2020;21(15):5186.

21. Hromas R, Gray PW, Chantry D, Godiska R, Krathwohl M, Fife K, Bell GI, Takeda J, Aronica S, Gordon M, Cooper S, et al. Cloning and characterization of exodus, a novel beta-chemokine. Blood. 1997;89(9):3315–3322.

22. Zhang XP, Hu ZJ, Meng AH, Duan GC, Zhao QT, Yang J. Role of CCL20/CCR6 and the ERK signaling pathway in lung adenocarcinoma. Oncol Lett. 2017;14(6):8183–8189.

23. Wang B, Shi L, Sun X, Wang L, Wang X, Chen C. Production of CCL20 from lung cancer cells induces the cell migration and proliferation through PI3 K pathway. J Cell Mol Med. 2016; 20(5):920–929.

24. Zhang C-Y, Qi Y, Li X-N, Yang Y, Liu D-L, Zhao J, Zhu D-Y, Wu K, Zhou X-D, Zhao S. The role of CCL20/CCR6 axis in recruiting Treg cells to tumor sites of NSCLC patients. Biomed Pharmacother. 2015;69:242–248.

25. Lim SY, Yuzhalin AE, Gordon-Weeks AN, Muschel RJ. Targeting the CCL2-CCR2 signaling axis in cancer metastasis. Oncotarget. 2016;7(19):28697–28710.

26. Li X, Tai HH. Activation of thromboxane A2 receptor (TP) increases the expression of monocyte chemoattractant protein-1 (MCP-1)/chemokine (C-C motif) ligand 2 (CCL2) and recruits macrophages to promote invasion of lung cancer cells. PLoS One. 2013;8(1): e54073.

27. Mittal P, Wang L, Akimova T, Leach CA, Clemente JC, Sender MR, Chen Y, Turunen BJ, Hancock WW. The CCR2/MCP-1 Chemokine Pathway and Lung Adenocarcinoma. Cancers (Basel). 2020;12(12):3723.

28. Korbecki J, Kojder K, Simińska D, Bohatyrewicz R, Gutowska I, Chlubek D, Baranowska-Bosiacka I. CC Chemokines in a tumor: a review of pro-cancer and anti-cancer properties of the ligands of receptors CCR1, CCR2, CCR3, and CCR4. Int J Mol Sci. 2020;21(21):8412.

29. Liao Y, He D, Wen F. Analyzing the characteristics of immune cell infiltration in lung adenocarcinoma via bioinformatics to predict the effect of immunotherapy. Immunogenetics. 2021; 73(5):369–380.

30. Bremnes RM, Busund L-T, Kilvær TL, Andersen S, Richardsen E, Paulsen EE, Hald S, Khanehkenari MR, Cooper WA, Kao SC, Dønnem T. The role of tumor-infiltrating lymphocytes in development, progression, and prognosis of non-small cell lung cancer. J Thorac Oncol. 2016;11(6):789–800.

31. Olingy CE, Dinh HQ, Hedrick CC. Monocyte heterogeneity and functions in cancer. J Leukoc Biol. 2019;106(2):309–322.

32. Lavin Y, Kobayashi S, Leader A, Amir E-AD, Elefant N, Bigenwald C, Remark R, Sweeney R, Becker CD, Levine JH, Meinhof K, et al. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. Cell. 2017; 169(4):750–765.

33. Carrato C, Alameda F, Esteve-Codina A, Pineda E, Arpí O, Martinez-García M, Mallo M, Gut M, Lopez-Martos R, Del Barco S, Ribalta T. Glioblastoma TCGA mesenchymal and IGS 23 tumors are identifiable by IHC and have an immune-phenotype indicating a potential benefit from immunotherapy. Clin Cancer Res. 2020;26(24):6600–6609.