Construction and Validation of a Novel Immune-Related Gene Pairs-Based Prognostic Model in Lung Adenocarcinoma

Yafeng Liu[1,2,3,†] ⑩, Jiawei Zhou[1,2,†], Jing Wu[1,2,4], Xin Zhang[1,2], Jianqiang Guo[1,2], Yingru Xing[1,5], Jun Xie[3], Ying Bai[1,2], and Dong Hu[1,2,4] ⑩

Abstract

Object: Focus on immune-related gene pairs (IRGPs) and develop a prognostic model to predict the prognosis of patients with lung adenocarcinoma (LUAD).

Methods: First, the LUAD patient dataset was downloaded from The Cancer Genome Atlas database, and paired analysis of immune-related genes was subsequently conducted. Then, LASSO regression was used to screen prognostic IRGPs for building a risk prediction model. Meanwhile, the Gene Expression Omnibus database was used for external validation of the model. Next, the clinical predictive power of IRGPs features was assessed by uni-multivariate Cox regression analysis, the infiltration of key immune cells in high and low IRGPs risk groups was analyzed with CIBERSORT, quanTIseq, and Timer, and the key pathways enriched for IRGPs were assessed using the Kyoto Encyclopedia of Genes and Genomes. Finally, the expression and related functions of key immune cells and genes were verified by immunofluorescence and cell experiments of tissue samples.

Results: It was revealed that the risk score of 19 IRGPs could be used as accurate indicators to evaluate the prognosis of LUAD patients, and the risk score was mainly related to T cell infiltration based on CIBERSORT analysis. Two genes of IRGPs, IL6, and CCL2, were found to be closely associated with the expression of PD-1/PD-L1 and the function of T-cells. Depending on the results of tissue immunofluorescence, IL6, CCL2, and T cells were highly expressed in the LUAD tissues of patients. Furthermore, IL6 and CCL2 were positively correlated with the expression of T cells. Besides, qRT-PCR assay in four different LUAD cells proved that IL6 and CCL2 were positively correlated with the expression of PD-L1 (P < .001).

Conclusions: Based on 19 IRGPs, an effective prognosis model was established to predict the prognosis of LUAD patients. In addition, IL6 and CCL2 are closely related to the function of T-cells.

[1]School of Medicine, Anhui University of Science and Technology, Huainan, China
[2]Anhui Province Engineering Laboratory of Occupational Health and Safety, Anhui University of Science and Technology, Huainan, China
[3]Affiliated Cancer Hospital, Anhui University of Science and Technology, Huainan, China
[4]Key Laboratory of Industrial Dust Deep Reduction and Occupational Health and Safety of Anhui Higher Education Institutes, Anhui University of Science and Technology, Huainan, China
[5]Department of Clinical Laboratory, Anhui Zhongke Gengjiu Hospital, Hefei, China

[†]These authors contributed equally to this work.

Corresponding Authors:
Jing Wu, School of Medicine, Anhui University of Science and Technology, Chongren Building, No 168, Taifeng St, Huainan 232001, China.
Email: wujing8008@126.com

Ying Bai, School of Medicine, Anhui University of Science and Technology, Chongren Building, No 168, Taifeng St, Huainan 232001, China.
Email: by0319_cpu@163.com

Dong Hu, School of Medicine, Anhui University of Science and Technology, Chongren Building, No 168, Taifeng St, Huainan 232001, China.
Email: austhudong@126.com

Background

Similar to other cancers, the occurrence of lung cancer is related to the activation of proto-oncogenes, inactivation of oncogenes, and immune escape of tumor cells, which eventually leads to genetic mutations and promotes the progression of cancer.[1] Lung adenocarcinoma (LUAD) is the most frequent subtype of lung cancer worldwide, and the 5-year survival rate of LUAD patients is less than 20% and the prognosis is poor.[2,3] The classical treatments include surgical resection, postoperative chemotherapy, and postoperative radiotherapy. Since severe damage to the organism caused by radiotherapy and chemotherapy, various targeted drugs aiming at different genetic variants have emerged. Although targeted therapy and immunotherapy have improved the clinical outcome of LUAD patients, their therapeutic effects are still limited.[4] Moreover, the early diagnosis, treatment, and prognostic prediction of LUAD remain challenging in clinical, considering the limitations of traditional diagnostic markers in sensitivity and specificity.[5] Therefore, it is urgent to explore novel biomarkers and accurate molecular mechanisms to effectively predict the prognosis of LUAD patients.

In recent years, increasing evidence has shown that the immune system, including immune molecules, immune cells, and the immune microenvironment, majorly affects tumorigenesis and tumor-associated immunity exists at all stages of tumor progression.[6] Disorders of the immune system, such as changes in the number or function of immune cells, the release of chemokines and cytokines, and the expression of inhibitory receptors or their ligands, can promote tumor cell proliferation and angiogenesis, inhibit tumor cell apoptosis, and reshape the microenvironment for tumor growth, thus promoting the progress of LUAD.[7,8] Immune-related genes (IRGs) are important in regulating the development of the immune cells and the state of the immune microenvironment.[9,10] It was found that immune-related gene pairs (IRGPs) can be used as predictors of LUAD prognosis.[11,12] However, the exact IRGs-associated pathways and mechanisms of immune cells in LUAD and how these genes affect the prognosis of LUAD remain unclear, which is worth further exploration.

In this study, a prognosis model of LUAD based on IRGPs was proposed and developed, and it was verified by RNA-seq data in the GEO database. Finally, this study explored the expression of key IRGs in tissues and their relationship with immune checkpoints.

Methods

Data Download and Processing

RNA-seq data of 497 LUAD samples were downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov) as the training dataset. RNA-seq data from 442 LUAD samples (GSE68465) were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/GEO) as the testing dataset. 2498 immune-related genes were obtained from the ImmPort database (https://immport.org/shared/genelist).

This study is a retrospective study and has been approved by the Medical Ethics Committee of Anhui University of Science and Technology (No.HX-001). Verbal consent was obtained from patients for their pathological results to be used in publication and all efforts have been taken to ensure the results do not involve patients' personal information.

Construction and Verification of Immune-Related Gene Model

A pairwise comparison analysis based on the IRG expression values of the samples was used to obtain a score for each IRGPs. If the expression level of the first IRG is higher than that of the second IRG, the IRGPs were given a score of 1, otherwise, the score was 0. The scores of all immune gene pairs were calculated and analyzed in the TCGA and GEO cohorts, and the resulting immune gene pairs were then merged with the corresponding clinical data. After uni-multivariate Cox and Kaplan-Meier analysis, the IRGPs related to prognosis were screened (P < .001), which were further screened by LASSO regression to obtain more stable prognostic IRGPs for model building. The point with the best specificity and sensitivity in the ROC curve was selected as the threshold of risk score to obtain the cut-off value, and the patients were divided into high-risk and low-risk groups according to the cut-off value. Based on the screened IRGPs associated with prognosis, a prognostic feature for predicting LUAD patients was constructed. In this prognostic feature, the risk score of each sample is calculated according to the following formula (β = regression Coef, Exp = gene expression level):

$$Risk\ score = \sum_{i=1}^{n} Expi\ \beta i,$$

These results were further validated in the GEO cohort.

Development of Nomogram

The nomogram of selected variables is analyzed based on univariate analyses in TCGA cohort. The "RMS" package was used to construct the nomogram. Calibration plots were conducted to estimate the accuracy and discriminative value of the prognostic nomogram.

Analysis of Immune Risk and Immune Cell Expression

The difference in immune cell type between the high-risk and low-risk groups was analyzed using the online tools CIBERSORT (http://cibersortx.stanford.edu/), quanTIseq (http://icbi.at/quantiseq) and Timer (cistrome.shinyapps.io/timer/).

Gene Enrichment Analysis

To clarify the potential role of risk scores signature in the treatment of immune checkpoint in LUAD, we correlated the risk scores and six immune checkpoint key genes. GO and KEGG databases were used for pathway enrichment analysis of the 19 IRGPs, and the gene sets whose P < .05 and FDR (false discovery rate) < .05 were considered significantly enriched gene sets.

Tissue Immunofluorescence Assay

Samples from LUAD patients were selected to verify the expression of key immune genes (CCL2 and IL6) and key immune cells (T-cells). 5 μm sections were taken from paraffin-embedded specimens, dewaxed, and rehydrated. The tissue sections were then placed in EDTA antigen repair buffer (pH 8.0). After natural cooling, the glass slides were placed in PBS (pH 7.4), shaken, and washed three times for 5 minutes each on a shaker. Samples were incubated with serum for 30 minutes after hydrogen peroxide blocking. Primary antibodies were added after appropriate dilution in PBS, and sections were incubated overnight in a humidified chamber at 4°C. After washing three times with PBS (pH 7.4) on a shaker, HRP-labeled secondary antibodies of the corresponding species were added to the sections to cover the tissue and incubated for 50 minutes at room temperature. Sections were washed three times with PBS (pH 7.4) for 5 minutes each time on a shaker. When the sections were slightly dry, CCL2, IL6, and CD3 primary antibodies and corresponding HRP-labeled secondary antibodies were added sequentially, and after DAPI re-staining of nuclei, the films were sealed and photographed under a microscope. All immunofluorescence (IF) sections were analyzed with ImageJ software.

Cell Experiment

Lung cancer cell lines (A549, 95D, LTEP-a-2, H1975) were obtained from the central laboratory of the School of Medicine, Anhui University of Technology. The reverse transcription kit was purchased from TaKaRa, and the quantitative real-time polymerase chain reaction (qRT-PCR) reagent, SYBR Green, was purchased from Abclonal. The RNA extraction reagent Trizol and primers were purchased from Shanghai Sangong. IL6
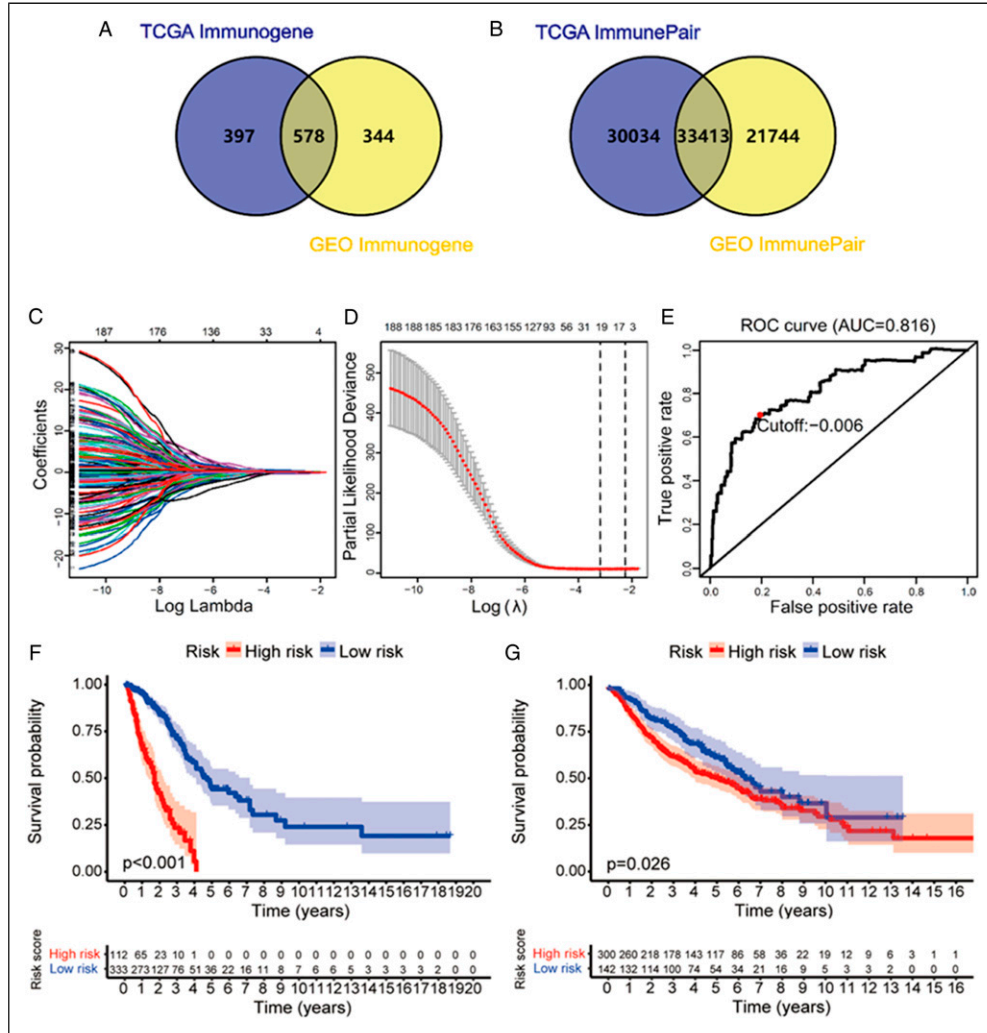
Table 1. Clinical Characteristics of LUAD Patients in TCGA.

Characteristics	n (%)
Age	
≤50	31 (33.1%)
>50	312 (66.9%)
Sex	
Male	169 (49.3%)
Female	174 (50.7%)
Stage	
I	176 (51.3%)
II	85 (24.8%)
III	61 (17.8%)
IV	21 (6.1%)
TNM-T	
T1	105 (30.6%)
T2	193 (56.3%)
T3	28 (8.1%)
T4	17 (5.0%)
TNM-N	
N0	216 (63.0%)
N1	74 (21.5%)
N2	52 (15.2%)
N3	1 (.3%)
TNM-M	
M0	322 (93.9%)
M1	21 (6.1%)

(F: 5′-ACTCACCTCTTCAGAACGAATTG-3′, R: 5′-CCATC TTTGGAAGGTTCAGGTTG-3′), CCL2 (F: 5′-CAGCCAGAT GCAATCAATGCC-3′, R: 5′-TGGAATCCTGAACCCACTTC T-3′), PD-L1 (F: 5′-TGGCATTTGCTGAACGCATTT-3′, R: 5′-TGCAGCCAGGTCTAATTGTTTT-3′), GAPDH (F: 5′-CTG GGCTACACTGAGCACC-3′, R: 5′-AAGTGGTCGTTGAGG GCAATG -3′).

Statistical Analysis

All statistical analyses were performed using GraphPad Prism8 and R software (version 3.6.2, https://www.r-project.org/). The log-rank test was used to evaluate the relationship between IRGPs and OS in LUAD patients. Single-multifactor Cox regression analysis was used to analyze the prognostic significance of the 19 IRGPs in LUAD patients. Prognostic characteristics models were processed using the (glmnet) package. Forest plots for single-multifactor Cox analysis were processed by the R package (survival). ROC curves were processed by the R package (survival ROC). Immune cell proportion distribution plots were processed by the R package (preprocessor and limma). Gene set enrichment analysis (GSEA) was performed using the R package (fgsea and ggplot2). Quantitative experiments were

Figure 1. Screening and construction of prognosis characteristics of 19-IRGP, calculation results of optimal cut-off value and Kaplan-Meier survival analysis of LUAD patients based on 19-IRGP Risk score. (A) Venn diagram of TCGA Immunogene and GEO Immunogene (IRG). (B) Venn diagram of TCGA Immune Pair and GEO Immune Pair (IRGP). (C) Lasso regression analysis was performed using 189-IRGP. (D) Cross validation of 19-IRGP for 10 times. (E) ROC curve and optimum cut-off value based on TCGA cohort. (F) Survival analysis of high and low immune risk groups in TCGA cohort. (G) Survival analysis of high and low immune risk groups in GEO cohort.

analyzed by t-test, and P < .05 was considered a significant difference.

Results

Definition of IRGPs Characteristics and Survival Analysis

A total of 497 LUAD cases were downloaded from the TCGA queue. The samples without clinical information and patients with a survival time of 0 were removed, so 343 cases were retained. The clinical information is shown in Table 1.

Based on the immune gene list in the ImmPort database, the immune gene expression matrix of TCGA transcriptome (975 immune-associated genes in total) was obtained, and the immune gene expression matrix of GEO expression profile (922 immune-associated genes in total) was obtained in the same way. The intersection of the two sets of immune-related genes was taken to obtain 578 immune-related genes co-expressed in TCGA and GEO (Figure 1(a)). Then, these 578 IRGs were randomly paired in pairs, and the obtained data were filtered, discarding more than 80% of the IRGPs with a score of 0 or 1 in the TCGA data set, and obtained the immune gene pair expression matrix of the TCGA transcriptome (a total of 634473-IRGP) and immune gene pair expression matrix of GEO expression profile transcriptome (55157-IRGP in total). The intersection of them yielded 33413-IRGP co-expressed in TCGA and GEO (Figure 1(b)). Next, univariate

Table 2. 19-IRGP Prognostic Characteristics and Calculation Formula of Risk Score.

Gene	Coef	Risk score =
HLADQB1|PSMB8	−.211	(-.211* EXP_HLADQB1|PSMB8)+
IL6|DKK1	−.072	(-.072* EXP_IL6|DKK1)+
ISG20|SEMA4A	.13	(.130* EXP_ISG20|SEMA4A)+
CCL20|CD79 A	.445	(.445* EXP_CCL20|CD79 A)+
CYLD|VEGFC	−.26	(-.260* EXP_CYLD|VEGFC)+
GNLY|DKK1	−.036	(-.036* EXP_GNLY|DKK1)+
TLR1|DKK1	−.357	(-.357* EXP_TLR1|DKK1)+
TLR1|IL1R2	−.024	(-.024* EXP_TLR1|IL1R2)+
PDCD1|CD79 B	.342	(.342* EXP_PDCD1|CD79 B)+
CCR7|IL22RA1	−.064	(-.064* EXP_CCR7|IL22RA1)+
CCL2|SEMA3C	−.194	(-.194* EXP_CCL2|SEMA3C)+
PIK3CD|ADM	−.218	(-.218* EXP_PIK3CD|ADM)+
SEMA3C|SORT1	.108	(.108* EXP_SEMA3C|SORT1)+
DKK1|IL23 A	.003	(.003* EXP_DKK1|IL23 A)+
DKK1|OSGIN14 A	.062	(.062* EXP_DKK1|OSGIN14 A)+
DKK1|TGFA	.053	(.053* EXP_DKK1|TGFA)+
VEGFC|LCK	.045	(.045* EXP_VEGFC|LCK)+
IL1RAP|LIFR	.172	(.172* EXP_IL1RAP|LIFR)+
IL22RA1|ZAP70	.197	(.197* EXP_IL22RA1|ZAP70) +

Cox analysis (P < .001) was performed on TCGA prognostic IRGPs to obtain 189 prognostic IRGPs (Table S1).

Subsequently, LASSO regression analysis of the TCGA cohort was carried out with the variables and corresponding regression coefficients selected simultaneously (Figure 1(c)-(d)), and the prognostic features were constructed using prognostic IRGPs (Table 2). In the TCGA and GEO cohorts, the same prognostic features (19-IRGP) were used to calculate the Risk score. The accuracy of the model was verified by the ROC curve, and the 1-year area under the curve (AUC) was .816. By the risk score of the ROC curve, the cut-off value with the best specificity and sensitivity was calculated (Figure 1(e)), which was further used to divide patients into high and low-immune risk groups.

In the TCGA cohort, survival analysis was performed in these two groups (Figure 1(f)), and the results showed that the OS of the high-immune risk group was lower than that of the low-immune risk group (P < .001). In the GEO cohort (Figure 1(g)), the trend is the same (P = .026).

Validation of IRGP Risk Score and Construction of Column Line Graph Models Based on Risk Score and Different Clinical Characteristics

To further validate the correlation between the risk score of 19-IRGP and prognosis, univariate and multivariate prognostic analyses were performed in the TCGA cohort and the GEO cohort, respectively. The outcome of the univariate analysis in the TCGA cohort exhibited a significant statistical significance between stage, risk score, and prognosis in different patients (P < .001) (Figure 2(a)). While in the multi-factor analysis of the TCGA cohort, it was shown that only risk score and prognosis have a significant statistical significance (P < .001) (Figure 2(b)). The validity of this model was then verified by the GEO cohort in the same manner, there is a significant statistical significance between age, gender, T, N, stages, risk score and prognosis determined by univariate analysis (P < .05) (Figure 2(c)), and in comparison, the outcome of multi-factor analysis demonstrated that age, T, N, stage, risk score, and prognosis were significantly correlated (P < .05) (Figure 2(d)).

Given the risk score based on 19-IRGP was an independent prognostic factor for OS, as validated in TCGA and GEO cohorts, a nomogram used for predicting patients' 3-5-year OS was drawn in the TCGA cohort, which combined the risk score of 19-IRGP and other clinical characteristics (Figure 2(e)). The 3- and 5-year survival rates of LUAD patients were calculated by plotting vertical lines between the total point axis and each prognostic axis. In addition, we constructed a calibration chart and the calibration curve in the figure represented the comparison between the actual results and the predicted results. It was obvious that the model could accurately predict the 3-year and 5-year survival rate of LUAD patients, but the 5-year survival rates of LUAD patients predicted seemed more accurate (Figure 2(f)-(g)).

In addition, to know whether the prediction performance of the nomogram is better than that of a single variable, we constructed ROCs of 1 (AUC = .869), 3 (AUC = .834), and 5 (AUC = .821) years according to the predicted values of each variable and nomogram (Figure. S1).

Analysis of Cell Subpopulation Associated With the Risk Score of 19-IRGP

The data from TCGA were collated using the CIBERSOR online tool to obtain the immune cell content, which was merged with the risk file to acquire a matrix of patient risk and immune cell content, which demonstrated that the expression level of eight cells was significantly different between the high and low immune risk groups (P < .05). In the high immune risk group, the infiltration of activated CD4 memory T cells and M0 macrophages was high, and in the low immune risk group, the infiltration of unactivated dendritic cells, regulatory T cells, unactivated mast cells, plasma cells, monocytes, and unactivated CD4 memory T cells was high (Figure 3(a)-(h)). To observe the proportion of 22 immune cell types in the high and low immune risk groups, the radar plots were drawn, and the results showed that M0 macrophage infiltration was highest in the high immune risk group and unactivated CD4 memory T cells infiltration was highest in the low immune risk group (Figure 3(i)), suggesting the key roles of these two cells play in LUAD.

In order to further support the difference in the proportion of immune cells between high-risk and low-risk people, we used quanTIseq and Timer algorithms for analysis. Among the
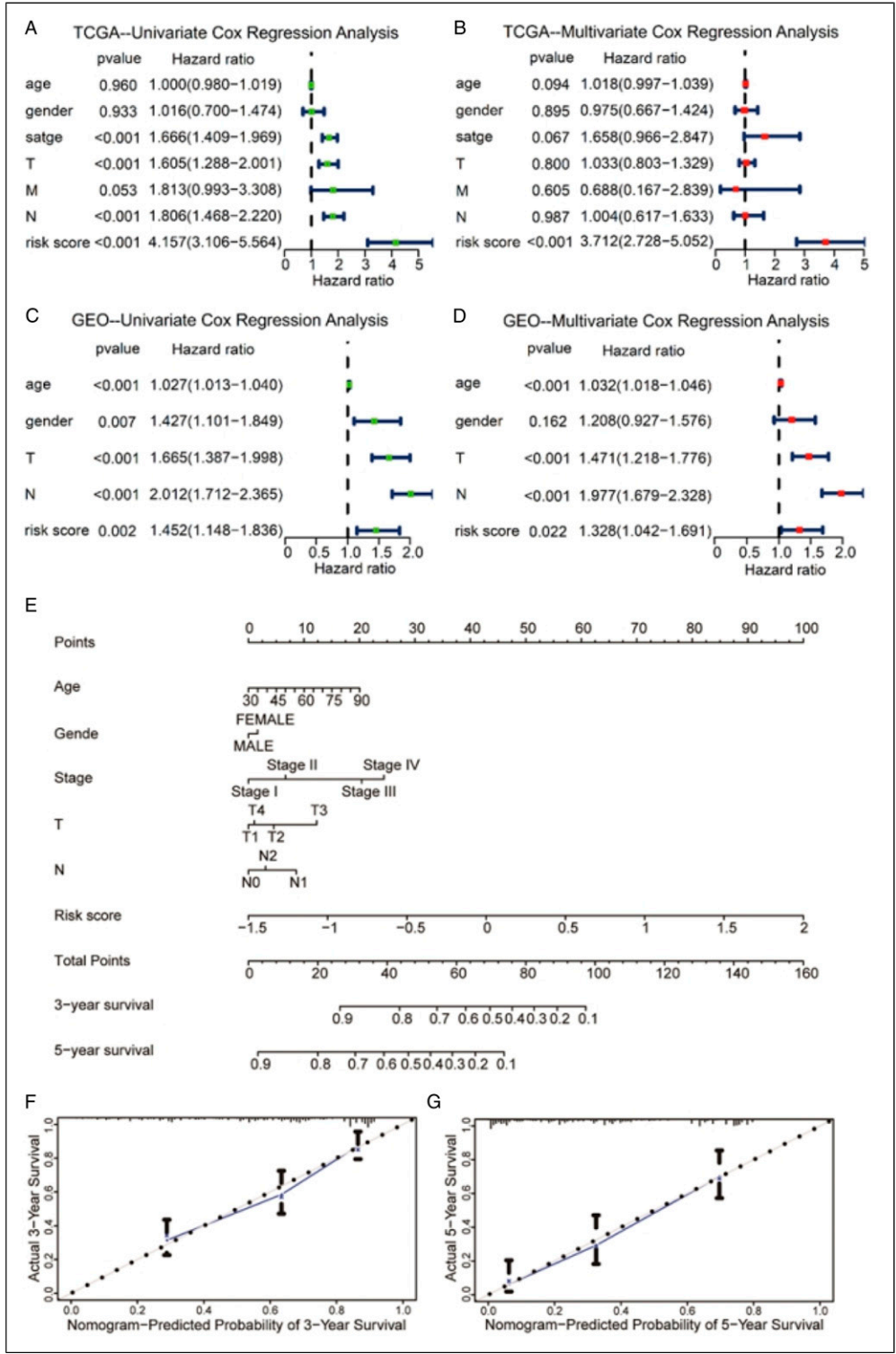
Figure 2. Univariate-multivariate Cox analysis of LUAD prognosis based on the risk score of 19-IRGP and other clinical features, nomogram and calibration chart of predicted OS of LUAD patients in 3-5 years. (A) Univariate Cox analysis was performed in the TCGA cohort. (B) Multivariate Cox analysis was performed in the TCGA cohort. (C) Univariate Cox analysis was performed in the GEO cohort. (D) Multivariate Cox analysis was performed in the GEO cohort. (E) The nomogram generated for predicting 3-year and 5-year OS of LUAD patients in TCGA cohort. (F) The calibration chart generated for predicting 3-year OS of LUAD patients in TCGA cohort. (G) The calibration chart generated for predicting 5-year OS of LUAD patients in TCGA cohort.
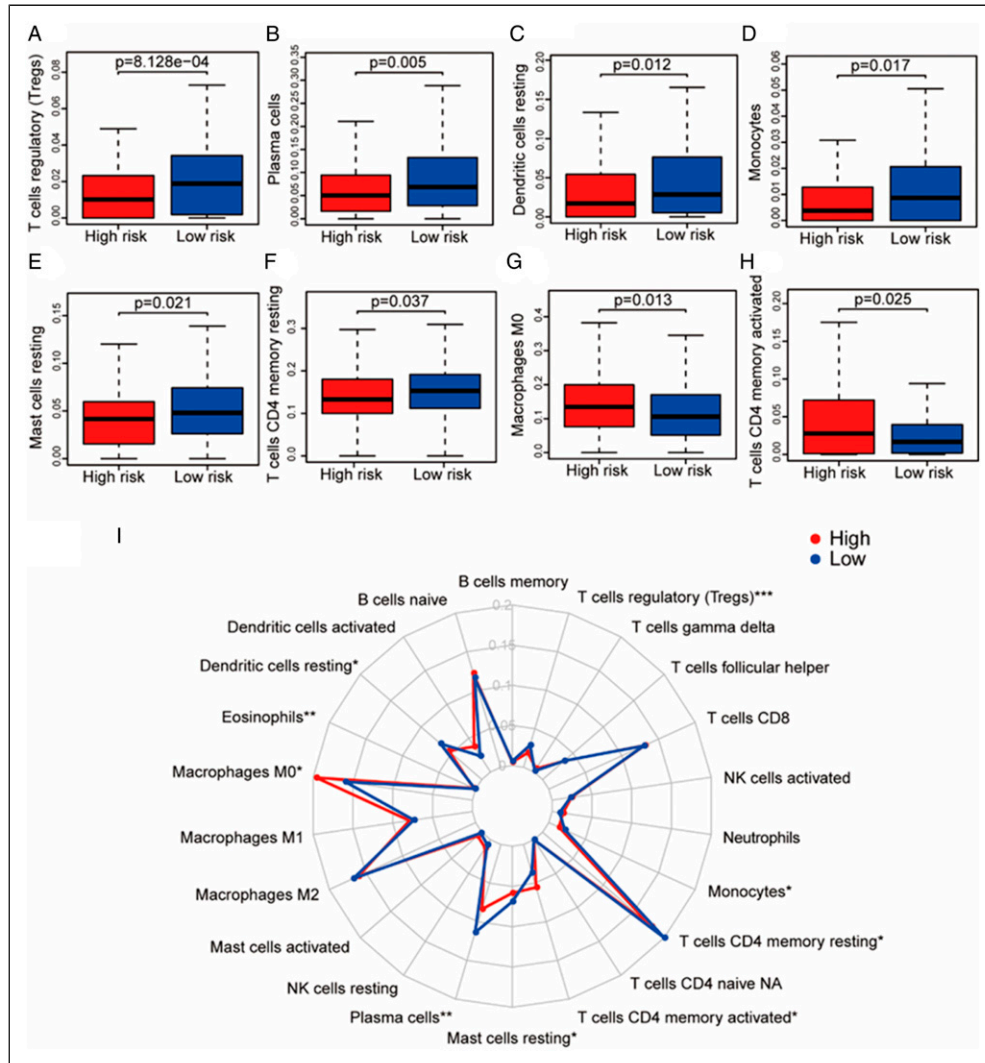
Figure 3. Box plots of infiltration ratios of 8 immune cells and radar charts of 22 immune cell infiltration distributions in high and low immune risk groups. (A) Box plot of infiltration ratio of regulatory T cell in the high and low immune risk groups. (B) Box plot of infiltration ratio of plasma cell in the high and low immune risk groups. (C) Box plot of infiltration ratio of unactivated dendritic cells in the high and low immune risk groups. (D)Box plot of infiltration ratio of monocytes in the high and low immune risk groups. (E) Box plot of infiltration ratio of unactivated mast cells in the high and low immune risk groups. (F) Box plot of infiltration ratio of unactivated CD4 memory T cells in the high and low immune risk groups. (G) Box plot of infiltration ratio of M0 macrophages in the high and low immune risk groups. (H) Box diagram of infiltration ratio of activated CD4 memory T cells in the high and low immune risk groups. (I) Radar map of infiltration distribution of 22 immune cells in the high and low immune risk groups.

10 immune cells in quanTIseq (Figure S2(a)), only Tregs and Neutrophils have differences in distribution. However, Among the 6 immune cells in the Timer algorithm (Figure S2(b)), there are distribution differences among CD4 (T-cell), CD8 (T-cell), and Macrophage.

19-IRGP Functional Enrichment Analysis

To further explore the potential biological functions of 19-IRGP, 29 immune genes from 19-IRGP were selected to perform GO and KEGG enrichment analyses. It was found that the 19-IRGP in GO were mainly enriched in pathways involved in immune cell activation including T cell differentiation and regulation (Figure 4(a)), while in KEGG, the 19-IRGP were mainly enriched in pathways involved cytokine receptors such as T cell receptor and PD-1/PD-L1 checkpoints (Figure 4(b)). It was worth noting that these pathways are closely related to the regulation of IL6 and CCL2. Therefore, the correlation of IL6/CCL2 and PD-1/PD-L1 expression in LUAD tissues was detected using the GEPIA network database, which suggested that the expression of IL6/CCL2 and PD-L1 was positively correlated
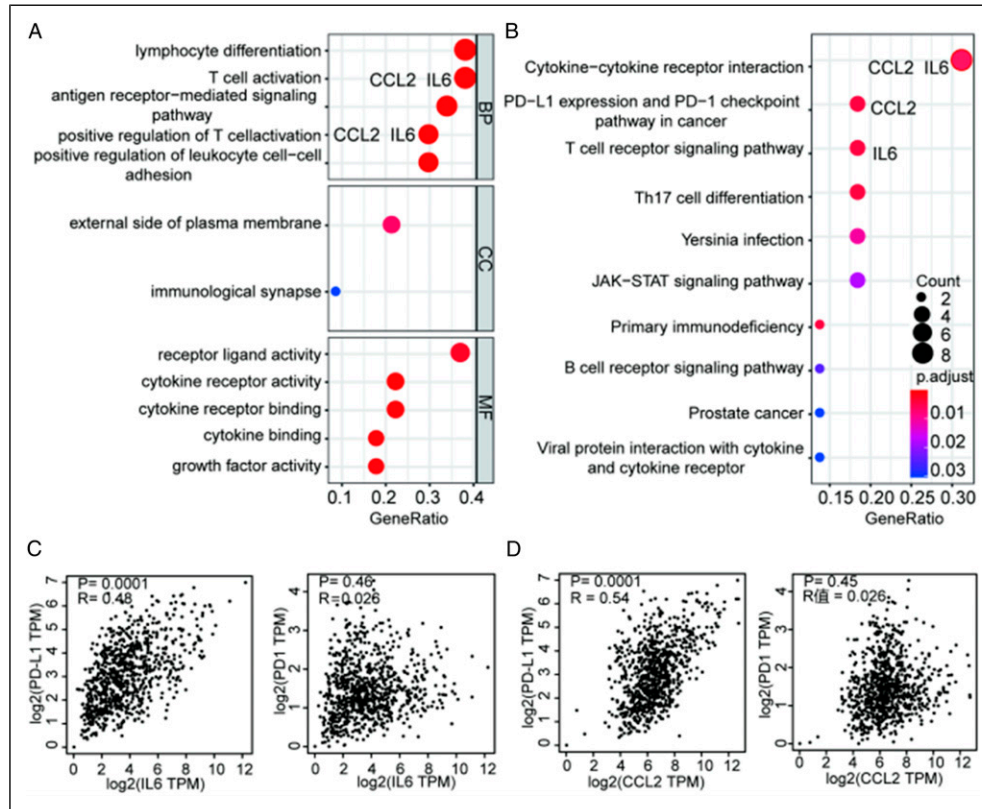
Figure 4. Functional enrichment analysis of 19-IRGP; (A) Results of GO enrichment analysis of 19-IRGP. (B) Results of KEGG enrichment analysis of 19-IRGP. (C) Correlation analysis of IL6 with PD1/PD-L1. (D) Correlation analysis of CCL2 with PD1/PD-L1.

and no significant correlation was found between IL6/CCL2 and PD1 (Figure 4(c)-(d)).

Meanwhile, gene set enrichment analysis (GSEA) was performed in the TCGA dataset using the mean value of the high immune risk group compared to that of the low immune risk group, leading to the significantly correlated GO pathway enrichment plots (Figure S3) and bubble plots (Figure S4). According to the results of GSEA enrichment analysis, high-risk represented a high 19-IRGP enriched in pathways including biological process (BP) and hormone metabolism, steroid metabolism, chromosome agglutination, adaptive immune response, adaptive immune response based on somatic cell reorganization, B cell-mediated immune response, lymphocyte-mediated immune response, regulation of B cell activation, regulation of lymphocyte activation, antigen receptor-mediated pathways, FC and FCε receptor pathways, and pathways that regulate immune responses from cell surface receptors. From the analysis of the cellular component (CC, cellular component), it is mainly enriched outside the plasma membrane. From the perspective of molecular function (MF, molecular function), it is mainly enriched in the activation of complement, activation of T lymphocytes, and phagocytosis. We found IRGPs were mainly enriched outside the plasma membrane. While in terms of molecular function (MF), IRGPs were mainly enriched in the activation of complement, activation of T lymphocytes, and phagocytosis.
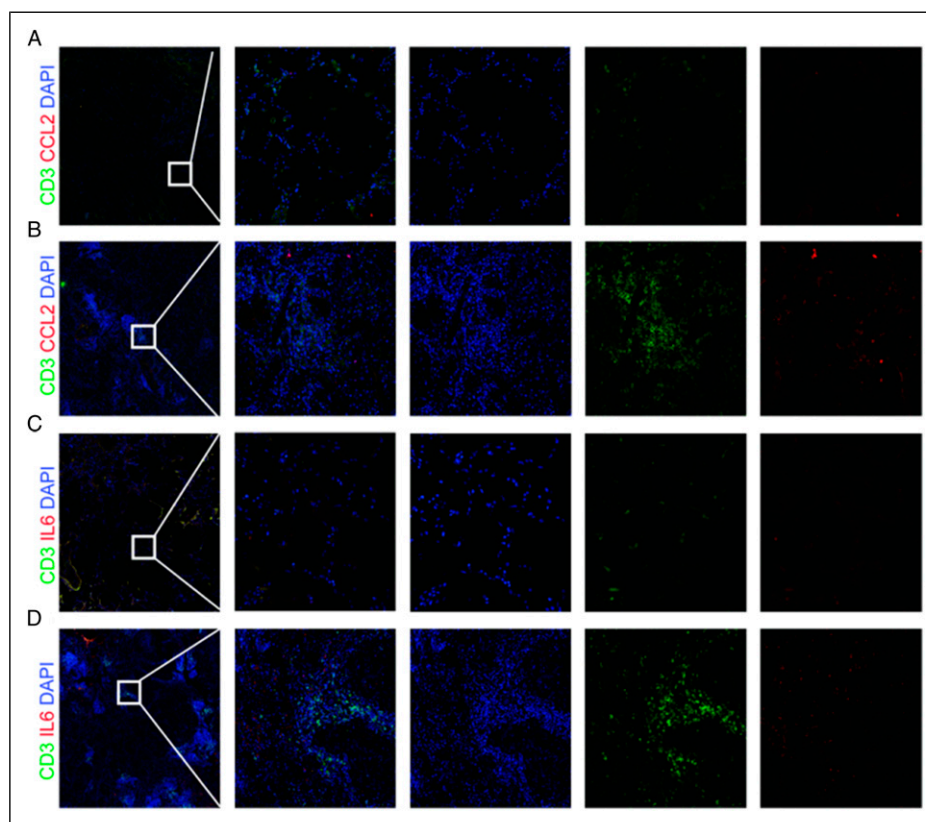
Immunofluorescence detection of IL6, CCL2, and T cell expression in tissues of LUAD patients

The expression levels of CCL2, IL6, and T cells in cancer and adjacent tissues of LUAD patients were detected by immunofluorescence, and the results showed that the expression levels of CCL2, IL6, and T cells (CD3) were significantly upregulated in cancer tissues (Figure 5(a)-(d)), which was consistent with the previous results of CIBERSORT and pathway enrichment analysis. Besides, there was a trend in the enrichment of T cells where CCL2 and IL6 were highly expressed (Figure 5(a)-(d)). Together, these results suggested that CCL2 and IL6 might be closely related to the activation and activity of T cells.

Analysis of IL6, CCL2 and PD-L1 Expression in LUAD Cells

To investigate the correlation of IL6, CCL2, and PD-L1 expression, four LUAD cells (A549, 95D. LTEP-a-2, H1975) (Figure 6(a)) were selected and detected by qRT-PCR. It was shown that the expression of IL6, CCL2, and PD-L1 was the lowest in A549 and the highest in H1975 (Figure 6(b)-(d)), and the expression of IL6, CCL2, and PD-L1 were significantly and positively correlated (Figure 6(e)).

Figure 5. Immunofluorescence detection of IL6, CCL2 and T cell expression. (A-B) Expression levels of CCL2 and T cells in cancer and adjacent tissues of LUAD patients. (C-D) Expression levels of IL6 and T cells in cancer and adjacent tissues of LUAD patients.

Discussion

In this study, bioinformatics methods were used to establish a model that could accurately predict the prognosis of LUAD patients. Meanwhile, immune-related genes and cells that affect the survival and prognosis of LUAD patients were explored, and their potential pathways and mechanisms of regulating immune cells were further investigated.

The risk score was an independent prognostic factor and was superior to stage staging, indicating that the risk score could be used as an independent prognostic factor to reflect the prognosis of patients, and the prognosis of patients became worse as the risk score increased. Subsequently, we constructed a nomogram and calibration plot of risk score combined with other clinical traits to predict OS. However, the weights of T4 and N2 in the nomogram are inconsistent with the cognition in clinical practice, which may be due to the defects caused by the small number of late populations in the TCGA cohort.

The key IRGPs screened are mainly functionally related to the recognition and expression of immune cells and antigens, involved in the composition of the immune microenvironment and mediating the immune response, many of which have been proven to play an important role in the occurrence and development of tumors. For example, the deubiquitinating enzyme CYLD can promote cell survival, proliferation, and differentiation, and plays key roles in regulating inflammatory and natural immune responses.[13] CYLD also regulates DC function through the NF-κB signaling pathway and suppresses the progression of LUAD.[14] Interleukin 6 (IL-6) can promote the metastatic colonization of colon cancer cells and has an abnormal anti-tumor immune function, it is identified as a potential target for the treatment of colorectal cancer by improving host immunity.[15] Programmed cell death 1 (PDCD1 or PD-1) mainly functions by inducing and maintaining autoimmune tolerance, after it binds to CD3-TCR, it can directly inhibit T cell activation, thereby leading to anti-tumor immunity and evasion of the immune system attacks. Evidence has shown that PDCD1 is a potential anticancer target in a variety of tumors including lung cancer.[16] CCL2 (monocyte chemotactic and activating factor, chemokine ligand (2) is associated with immune regulation and inflammatory processes, has chemotactic effects on monocytes and basophils, is involved in macrophage CCR5 pathway and ERK signaling, and serum CCL2 levels can be used to monitor and predict clinical outcomes in patients with refractory advanced NSCLC treated with drugs.[17] It has been proved that ERα activates the CCL2/CCR2 axis and promotes macrophage infiltration, M2 polarization, and MMP
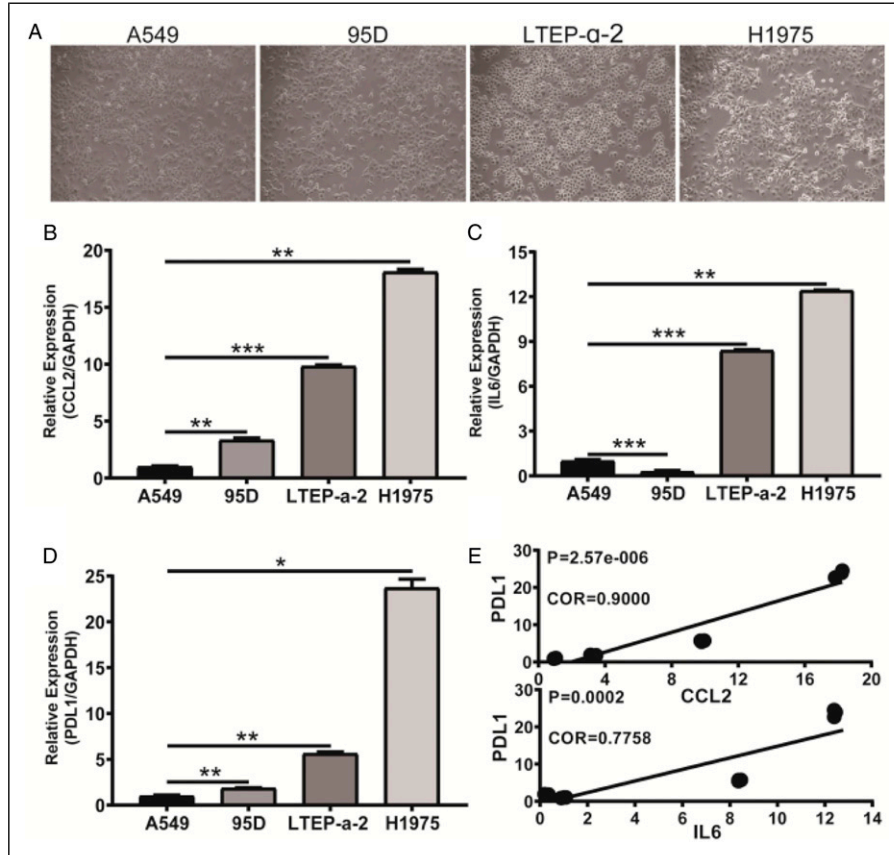
Figure 6. The expression of IL6, CCL2 and PD-L1 in four LUAD cells was detected by qRT-PCR. (A) Four different LUAD cells. (B) Expression of CCL2 in four LUAD cells. (C) Expression of IL6 in four LUAD cells. (D) Expression of PD-L1 in four LUAD cells. (E) Correlation analysis of IL6, CCL2 and PD-L1.

9 production, thereby increasing the invasion of non-small cell lung cancer cells.[18] High expression of SEMA3C (signaling protein 3C) is associated with increased cancer cell invasion and adhesion.[19] DKK1 (WNT signaling pathway inhibitor (1) is highly expressed in a variety of cancers and may promote cancer cell proliferation, invasion, and growth, and DDK1 is highly expressed in lung cancer cells and tissues.[20] TGFA (transforming growth factor α), which binds to EGFR, activates the signal pathway of cell proliferation, differentiation, and development, which is related to the progress of various cancers.[21] Cell subset analysis revealed a high proportion of activated CD4 memory T cells and M0 macrophage immune infiltration in the high immune risk group. Studies have shown that in damaged lung tissue, M0 macrophages with pro-inflammatory and pro-fibrotic phenotypes are readily differentiated under various stimuli and play a crucial role in the progression of lung cancer,[22] whereas activated CD4 memory T cells can stimulate the body to produce an adaptive immune response by clearing antigens more quickly and effectively through immune memory,[23,24] which is consistent with our findings. In LUAD patients, high expression of M0 macrophages predicts a poor

prognosis, and M0 macrophages are closely related to anti-tumor immunosuppression,[25] therefore, antagonists or other targeted drugs that block M0 macrophages may contribute to LUAD therapy.

GO and KEGG enrichment analysis showed that the 19 IRGPs were mainly closely related to the expression of T cells and various immune responses and the role of cytokine receptors (such as PD-1/PD-L1). For example, activated regulatory T cells are highly expressed in high-risk patients, by the finding that CD4~+CD25~+Foxp3~+Treg regulatory T cells are highly expressed in lung cancer tissue and peripheral blood, promoting the further deterioration of the tumor.[26] In the pathway analysis, it was found that the enrichment of key immune genes such as CCL2 (chemokine ligand (2) and IL6 (interleukin (6) is mainly determined by the function of T cells, and CCL2 is also closely related to the PD-1/PD-L1 pathway.[27-29] It has been reported that the PD-1/PD-L1-mediated immunosuppressive pathway can lead to anti-tumor immunity and evade the attack of the immune system, thus PD-1/PD-L1 has become a potential therapeutic target for various cancers including lung cancer.[30] In addition, studies have proved that blocking the expression of IL6 and CCL2 can significantly improve the immunotherapy effect of

PD-1/PD-L1,[31,32] but the mechanism of how IL6 and CCL2 improve immunotherapy is still unclear. Our research has found that IL6 and CCL2 might regulate the proliferation of T cells by affecting their activity, and CCL2 might increase the expression of PD-1/PD-L1 after binding to the relevant receptors on T cells, thereby affecting the therapeutic effect. Further, the results of human tissue immunofluorescence confirmed that the expression of IL6 and CCL2 was positively correlated with the expression of T cells. In the four LUAD cells, IL6 and CCL2 were proved to be positively correlated with PD-L1, which was consistent with the results of previous studies.[33,34] Therefore, we speculated that IL6 and CCL2 might affect the survival and prognosis of patients by regulating the proliferation of T cells and the expression of PD-1/PD-L1. While the detailed mechanism needs to be further studied.

IRGs play an important role in the occurrence and development of tumors. Previous studies only focused on the functions of single or multiple genes, but seldom discussed the research value of highly correlated gene combinations. This study starts with the correlation between immune genes, and builds a survival prediction model by screening gene pairs with strong correlation. In addition, the prediction effect of the model is better. This shows that there is still a great research prospect for the diagnosis, treatment and prognosis of lung cancer by exploring the interaction between genes from gene pairs. However, this research still has some limitations. (1) Due to the lack of verification of immunotherapy samples, the application scenarios of risk scoring can't be further expanded. (2) The interaction between genes also needs further study and explanation. (3) As all samples are collected from public databases. Some samples lack some clinical information, such as treatment, molecular subtypes, etc., which may lead to information bias.

Conclusion

This study constructed a prognostic model of LUAD patients based on immune gene pairs. Hub genes CCL2 and IL-6 have been proven to affect the prognosis and immune checkpoint of patients by affecting the activation of T cells, which provides new insights and ideas for individualized treatment of LUAD.

Acknowledgments

BY, HD and WJ: conception and design, and study supervision. ZJ, LY, ZX, GJ, XY and XJ: development of methodology, analysis and interpretation of data, and writing of the manuscript. BY, WJ and HD: review of the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: No conflict of interest exits in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

Ethics Statement

This research has been approved by the Medical Ethics Committee of Anhui University of Science and Technology (No, HX-001).

ORCID iDs

Yafeng Liu https://orcid.org/0000-0003-3930-4872
Dong Hu https://orcid.org/0000-0002-8995-4729

Supplemental Material

Supplemental material for this article is available online.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin*. 2019;69(1):7-34.
2. Behera M, Owonikoko TK, Gal AA, et al. Lung adenocarcinoma staging using the 2011 IASLC/ATS/ERS classification: A pooled analysis of adenocarcinoma in situ and minimally invasive adenocarcinoma. *Clin Lung Cancer*. 2016;17(5): e57-e64.
3. Jiawei Z, Min M, Yingru X, et al. Identification of key genes in lung adenocarcinoma and establishment of prognostic mode. *Front Mol Biosci*. 2020;7:561456.
4. Liu Y, Yang S, Wang K, et al. Cellular senescence and cancer: Focusing on traditional Chinese medicine and natural products. *Cell Prolif*. 2020;53(10):e12894.
5. Wu K, Yu S, Liu Q, Bai X, Zheng X. The clinical significance of CXCL5 in non-small cell lung cancer. *OncoTargets Ther*. 2017; 10:5561-5573.
6. Chen W, Ou M, Tang D, Dai Y, Du W. Identification and validation of immune-related gene prognostic signature for hepatocellular carcinoma. *J Immunol Res*. 2020;2020:5494858-5494914.
7. Sun XY, Yu SZ, Zhang HP, Li J, Guo W. A signature of 33 immune-related gene pairs predicts clinical outcome in hepatocellular carcinoma. *Cancer Med*. 2020;9(8):2868-2878.

8. Guan X. Cancer metastases: Challenges and opportunities. *Acta Pharm Sin B*. 2015;5(5):402-418.

9. Wang Q, Li M, Yang M, et al. Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: Implications for immune checkpoint blockade therapy. *Aging (Albany NY)*. 2020;12(4):3312-3339.

10. Xu F, Zhan X, Zheng X, et al. A signature of immune-related gene pairs predicts oncologic outcomes and response to immunotherapy in lung adenocarcinoma. *Genomics*. 2020;112(6): 4675-4683.

11. Wu C, Hu Q, Ma D. Development of an immune-related gene pairs signature for predicting clinical outcome in lung adenocarcinoma. *Sci Rep*. 2021;11(1):3611.

12. Cai HY, Yang HS, Shan SC, et al. A novel signature based on immune-related gene pairs and clinical features to predict prognosis and treatment effect in "driver gene negative" lung adenocarcinoma. *Cancer Med*. 2022;11(11): 2259-2270.

13. Lork M, Kreike M, Staal J, Beyaert R. Importance of validating antibodies and small compound inhibitors using genetic knockout studies-T cell receptor-induced CYLD phosphorylation by IKKε/ TBK1 as a case study. *Front Cell Dev Biol*. 2018;6:40.

14. Zhao M, Xin XF, Zhang JY, Dai W, Lv T, Song Y. LncRNA GMDS-AS1 inhibits lung adenocarcinoma development by regulating miR-96-5p/CYLD signaling. *Cancer Med*. 2020; 9(3):1196-1208.

15. Toyoshima Y, Kitamura H, Xiang H, et al. IL6 modulates the immune status of the tumor microenvironment to facilitate metastatic colonization of colorectal cancer cells. *Cancer Immunol Res*. 2019;7(12):1944-1957.

16. Liebler DC, Holzer TR, Haragan A, et al. Analysis of immune checkpoint drug targets and tumor proteotypes in non-small cell lung cancer. *Sci Rep*. 2020;10(1):9805.

17. Lu J, Zhong H, Chu T, et al. Role of anlotinib-induced CCL2 decrease in anti-angiogenesis and response prediction for nonsmall cell lung cancer therapy. *Eur Respir J*. 2019;53(3): 1801562.

18. He M, Yu W, Chang C, et al. Estrogen receptor α promotes lung cancer cell invasion via increase of and cross-talk with infiltrated macrophages through the CCL2/CCR2/MMP9 and CXCL12/ CXCR4 signaling pathways. *Mol Oncol*. 2020;14(8): 1779-1799.

19. Vaitkienė P, Skiriutė D, Steponaitis G, Skauminas K, Tamasauskas A, Kazlauskas A. High level of Sema3C is associated with glioma malignancy. *Diagn Pathol*. 2015;10:58.

20. Li YF, Zhang J, Yu L. Circular RNAs regulate cancer onset and progression via Wnt/β-catenin signaling pathway. *Yonsei Med J*. 2019;60(12):1117-1128.

21. Liang H, Liu X, Wang M. Immunotherapy combined with epidermal growth factor receptor-tyrosine kinase inhibitors in non-small-cell lung cancer treatment. *OncoTargets Ther*. 2018; 11:6189-6196.

22. Liu SS, Lv XX, Liu C, et al. Targeting degradation of the transcription factor C/EBPβ reduces lung fibrosis by restoring activity of the ubiquitin-editing enzyme A20 in macrophages. *Immunity*. 2019;51(3):522-534.

23. Gong J, Chehrazi-Raffle A, Placencio-Hickok V, Guan M, Hendifar A, Salgia R. The gut microbiome and response to immune checkpoint inhibitors: Preclinical and clinical strategies. *Clin Transl Med*. 2019;8(1):9.

24. Silva TF, Tomiotto-Pellissier F, Sanfelice RA, et al. A 21st century evil: Immunopathology and new therapies of COVID-19. *Front Immunol*. 2020;11:562264.

25. Meng L, He X, Zhang X, et al. Predicting the clinical outcome of melanoma using an immune-related gene pairs signature. *PLoS One*. 2020;15(10):e0240331.

26. Guo X, Zhang Y, Zheng L, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med*. 2018;24(7):978-985.

27. Wang Y, Zhang X, Yang L, Xue J, Hu G. Blockade of CCL2 enhances immunotherapeutic effect of anti-PD1 in lung cancer. *J Bone Oncol*. 2018;11:27-32.

28. Kuo IY, Yang YE, Yang PS, et al. Converged Rab37/IL-6 trafficking and STAT3/PD-1 transcription axes elicit an immunosuppressive lung tumor microenvironment. *Theranostics*. 2021;11(14):7029-7044.

29. Ohno Y, Toyoshima Y, Yurino H, et al. Lack of interleukin-6 in the tumor microenvironment augments type-1 immunity and increases the efficacy of cancer immunotherapy. *Cancer Sci*. 2017;108(10):1959-1966.

30. Zhao Y, Lee CK, Lin CH, et al. PD-L1: CD80 Cis-heterodimer triggers the co-stimulatory receptor CD28 while repressing the inhibitory PD-1 and CTLA-4 pathways. *Immunity*. 2019;51(6): 1059-1073.

31. Tsukamoto H, Fujieda K, Miyashita A, et al. Combined blockade of IL6 and PD-1/PD-L1 signaling abrogates mutual regulation of their immunosuppressive effects in the tumor microenvironment. *Cancer Res*. 2018;78(17): 5011-5022.

32. Choi J, Lee HJ, Yoon S, et al. Blockade of CCL2 expression overcomes intrinsic PD-1/PD-L1 inhibitor-resistance in transglutaminase 2-induced PD-L1 positive triple negative breast cancer. *Am J Cancer Res*. 2020;10(9):2878-2894.

33. Liu C, Yang Y, Chen C, et al. Environmental eustress modulates β-ARs/CCL2 axis to induce anti-tumor immunity and sensitize immunotherapy against liver cancer in mice. *Nat Commun*. 2021;12(1):5725.

34. Tanaka I, Dayde D, Tai MC, et al. SRGN-triggered aggressive and immunosuppressive phenotype in a subset of TTF-1-negative lung adenocarcinomas. *J Natl Cancer Inst*. 2021;114:djab183.