



Establishment of Immune-related Gene Pair Signature to Predict Lung Adenocarcinoma Prognosis

Cell Transplantation
Volume 29: 1–14
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0963689720977131
journals.sagepub.com/home/ctj


Xueping Jiang¹, Yanping Gao¹, Nannan Zhang¹, Cheng Yuan¹,
Yuan Luo¹, Wenjie Sun¹, Jianguo Zhang¹, Jiangbo Ren²,
Yan Gong^{2,3}, and Conghua Xie^{1,4,5} 

Abstract

Tumor microenvironment (TME) has critical impacts on the pathogenesis of lung adenocarcinoma (LUAD). However, the molecular mechanism of TME effects on the prognosis of LUAD patients remains unclear. Our study aimed to establish an immune-related gene pair (IRGP) model for prognosis prediction and internal mechanism investigation. Based on 702 TME-related differentially expressed genes (DEGs) extracted from The Cancer Genome Atlas (TCGA) training cohort using the ESTIMATE algorithm, a 10-IRGP signature was established to predict LUAD patient prognosis. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses showed that DEGs were significantly associated with tumor immune response. In both TCGA training and Gene Expression Omnibus validation datasets, the risk score was an independent prognostic factor for LUAD patients using Lasso-Cox analysis, and patients in the high-risk group had poorer prognosis than those in the low-risk one. In the high-risk group, M2 macrophage and neutrophil infiltrations were higher, while the levels of T cell follicular helpers were significantly lower. The gene set enrichment analysis results showed that DNA repair signaling pathways were involved. In summary, we established an IRGP signature as a potential biomarker to predict the prognosis of LUAD patients.

Keywords

lung adenocarcinoma, tumor microenvironment, immune-related gene pair, prognosis

Introduction

Lung cancer is the most frequent malignant carcinoma, as well as the major cause of cancer death¹. Based on histopathologic characteristics, non-small cell lung cancer (NSCLC) accounts for roughly 85% of all lung carcinoma, and lung adenocarcinoma (LUAD) is the major component of NSCLC². Although LUAD patients benefited from multiple treatments, including radiotherapy, chemotherapy, and immunotherapy, the 5-year survival rate was still under 20%³. Therefore, it is urgent to identify novel biomarkers for early detection, early diagnosis, and better prediction of LUAD patient prognosis.

Tumor microenvironment (TME) composes of cells (endothelial, neuroendocrine and immune cells, etc.) and extracellular components (cytokines, chemokines and extracellular matrix, etc.). TME has crucial effects on the initiation, development, metastasis, and recurrence of carcinoma^{4,5}. As the essential ingredients of TME, immune and stromal cells play substantial roles in signal transduction, immune surveillance, immune escape, and patient prognosis^{6,7}. For instance,

the activation of CD8⁺ T cells, which are the important components of immune cells, could improve the survival in gastric cancer mouse models⁸.

¹ Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, China

² Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, China

³ Human Genetics Resource Preservation Center of Hubei Province, Human Genetics Resource Preservation Center of Wuhan University, Wuhan, Hubei, China

⁴ Hubei Key Laboratory of Tumor Biological Behaviors, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, China

⁵ Hubei Cancer Clinical Study Center, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, China

Submitted: May 30, 2020. Revised: September 28, 2020. Accepted: November 9, 2020.

Corresponding Author:

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



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

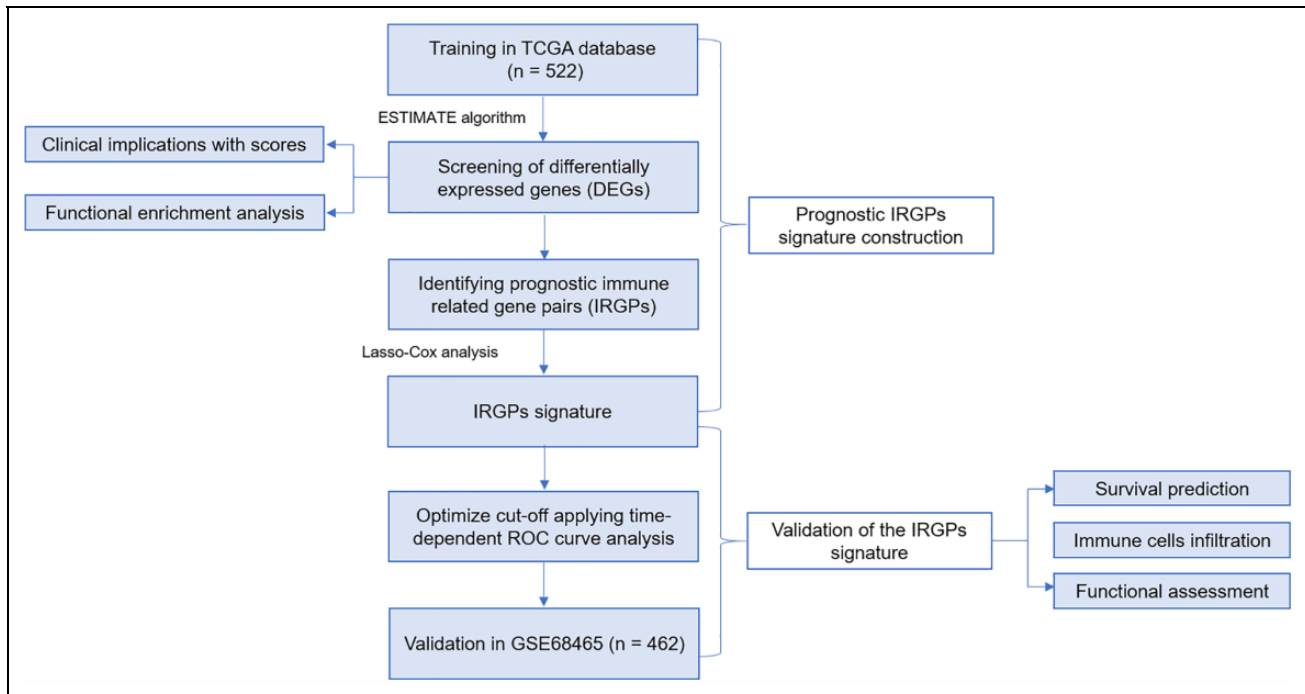


Figure 1. The workflow of current study. The TCGA-LUAD cohort was set as the training dataset, and the GSE68465 cohort was set as the validation dataset. LUAD: lung adenocarcinoma; TCGA: The Cancer Genome Atlas.

To accurately evaluate the percentage of immune and stromal cells in TME, Yoshihara et al. calculated the scores of tumor purity, as well as immune and stromal cell infiltration in tumor tissues, based on gene expression levels using the ESTIMATE algorithm⁹. Previous studies also suggested that TME-related biomarkers, which were obtained using big-data-based ESTIMATE algorithm, were helpful in histological diagnosis and prognostic prediction^{10,11}. By this algorithm, Zeng et al. identified three novel biomarkers including IL10, IGLL5, and POU2AF1 for therapy targets and prognosis evaluation in renal cell carcinoma¹¹.

Unlike traditional filtering approaches just separating high and low gene clusters according to median values, the ESTIMATE algorithm was applied in our research to screen differentially expressed genes (DEGs) associated with immune and stromal scores from TCGA database. This approach aimed to make connections between TME and the expression of genes. Increasing evidence suggested that the application of immunotherapy targeting immune checkpoints such as PD-L1 and CLTA4 were beneficial to LUAD patients in clinic¹². Previous studies reported immune-related gene signatures to estimate the prognosis of head and neck squamous cell carcinoma, breast cancer, and so on^{13,14}. Nevertheless, due to the intricate data process and technical bias of different platforms, these models were difficult to apply for clinical evaluation temporarily. In this study, we established and validated an immune-related gene pair (IRGP) signature, in which the normalization and scaling for samples were not required.

Materials and Methods

Study Design and Data Collection

The integrated research design was exhibited in Fig. 1. The LUAD patients' transcription profiles and clinical data were obtained from TCGA GDC website¹⁵. Another microarray dataset (GSE68465, $n = 462$) was downloaded from Gene Expression Omnibus (GEO) portal for further validation of the signature. The immune-related gene cohorts were obtained from ImmPort database (<https://immport.niaid.nih>)¹⁶.

Screening of DEGs

After normalization of the TCGA dataset, the immune/stromal/estimate scores were calculated using Estimate R package (<http://r-forge.rproject.org;repos=rforge;dependencies=TRUE>). Bioconductor limma package was applied to screen DEGs from the immune and stromal score groups. $|\log_2 \text{fold change (FC)}| \geq 1$ and false discovery rate (FDR) < 0.05 were set as the cut-offs to screen DEGs¹⁷. Heatmap and Venn were drawn using of heatmap and VennDiagram R packages, respectively^{18,19}.

Functional and Signal Pathways Enrichment Analysis

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to explore various functions of DEGs using clusterProfiler R package. GO analysis was divided into three parts, including

biological processes (BPs), molecular functions (MFs), and cellular components (CCs). The enrichplot and ggplot2 R packages were applied to visualize the results accurately²⁰. $P < 0.05$ was considered to be statistically significance. The Gene Set Enrichment Analysis (GSEA) was conducted with basic document named c2.cp.kegg.v7.0.symbols.gmt, which was downloaded from molecular signatures database (<https://www.gsea-msigdb.org/gsea/datasets.jsp>)²¹. The number of permutations was set as 1,000 times.

Construction of the Prognostic Signature Based on IRGPs

The stable prognostic model was extracted from TCGA dataset. Immune-related DEGs were obtained from DEGs using ImmPort database, and 66 immune-related genes with relative high variation (evaluated by median absolute deviation > 0.5) were identified. The expression value of each one was compared with others' in the same independent sample. A total of 2,145 different pairwise combinations were obtained. If the expression level of former one was higher than the later one, this IRGP was labeled as 1. If not, it was labeled as 0. Finally, if the score of an IRGP was 0 or 1 in more than 80% of the TCGA samples, it was discarded²². The remaining 359 IRGPs were analyzed with univariate Cox regression to obtain 15 prognostic-related candidate IRGPs. Lasso-Cox analysis (iteration = 1,000) and 10-fold cross-validation were then applied to construct a more stable 10-IRGP model for prognosis evaluation by glmnet R package. The formula of the risk score model is described as follows:

$$\text{Riskscore} = \sum_{k=1}^n \beta_k * X_k$$

where β_k refers to the coefficients of each gene pair, and X_k represents 0 or 1, which is obtained from the comparison of expression level in each pair. The optimal cut-off of the IRGP was determined using time-dependent receiver operating characteristic (ROC) curve analysis ("survivalROC" R package) at 1 year in the training cohort for overall survival (OS). The cut-off value was defined as the risk score corresponding to the minimum distance between ROC curve and the point standing for 100% true positive rate and 0% false-positive rate²³.

Validation of the IRGP Signature

In GEO validation dataset, the patients were divided into the high- and low-risk groups according to the cut-off value, which obtained from the formula of risk model. The ROC curve analysis was performed and the area under the curve (AUC) was calculated. The Kaplan–Meier (K-M) survival curve analysis was used to figure out differences of OS between the high- and low-risk groups in both TCGA and GEO datasets ($\text{survdiff}(\text{formula} = \text{Surv}(\text{Time}, \text{Status}) \sim$

Table 1. Clinical Characteristics of the 522 LUAD Patients from the TCGA Cohort.

CDI9	Male (%)	Female (%)	Total (%)
Sex	242 (46.4%)	280 (53.6%)	522 (100.0%)
Age (years) ^a			
≤65	110 (21.9%)	131 (26.0%)	241 (47.9%)
>65	123 (24.5%)	139 (27.6%)	262 (52.1%)
Status			
Dead	80 (15.3%)	87 (16.7%)	167 (32.0%)
Alive	162 (31.0%)	193 (37.0%)	355 (68.0%)
TNM stage ^a			
I	163 (31.7%)	116 (22.6%)	279 (54.3%)
II	56 (10.9%)	68 (13.2%)	124 (24.1%)
III	46 (8.9%)	39 (7.6%)	85 (16.5%)
IV	12 (2.3%)	14 (2.7%)	26 (5.1%)
T ^a			
1	61 (11.7%)	111 (21.4%)	172 (33.1%)
2	143 (27.5%)	138 (26.6%)	281 (54.1%)
3	27 (5.2%)	20 (3.9%)	47 (9.1%)
4	10 (1.9%)	9 (1.8%)	19 (3.7%)
N ^a			
0	150 (29.4%)	185 (36.3%)	335 (65.7%)
1	53 (10.4%)	45 (8.8%)	98 (19.2%)
2	36 (7.1%)	39 (7.6%)	75 (14.7%)
3	0 (0%)	2 (0.4%)	2 (0.4%)
M ^a			
0	172 (45.5%)	181 (47.9%)	353 (93.4%)
1	14 (3.7%)	11 (2.9%)	25 (6.6%)

^aPartial missing of clinical characteristics.

LUAD: lung adenocarcinoma; TCGA: The Cancer Genome Atlas.

Group, data = read table))²⁴. The formula of survival curve analysis is as follows:

$$St = \frac{\text{Numbers of subjects living at the start} - \text{Numbers of subjects died}}{\text{Numbers of subjects living at the start}}$$

The concordance index (C-index) was calculated using the coxph function of "survival" R package. The uni- and multi-variate Cox regression analyses were used to validate the risk score as an independent prognostic factor.

Immune Cell Infiltration

The CIBERSORT R package (Version 1.03) was applied to quantify immune cell fractions including T cells, T cell follicular helpers (Tfh), macrophages, and neutrophils, in the high- and low-risk groups²⁵. The P -value was calculated using the deconvolution approach. $P < 0.05$ was considered as a criterion. The results were shown in radar chart by R package "fmsb."

Results

Immune and Stromal Scores Were Associated with LUAD Clinical Features

In TCGA-LUAD training dataset, three different kinds of scores were calculated: immune scores (from -936.191 to

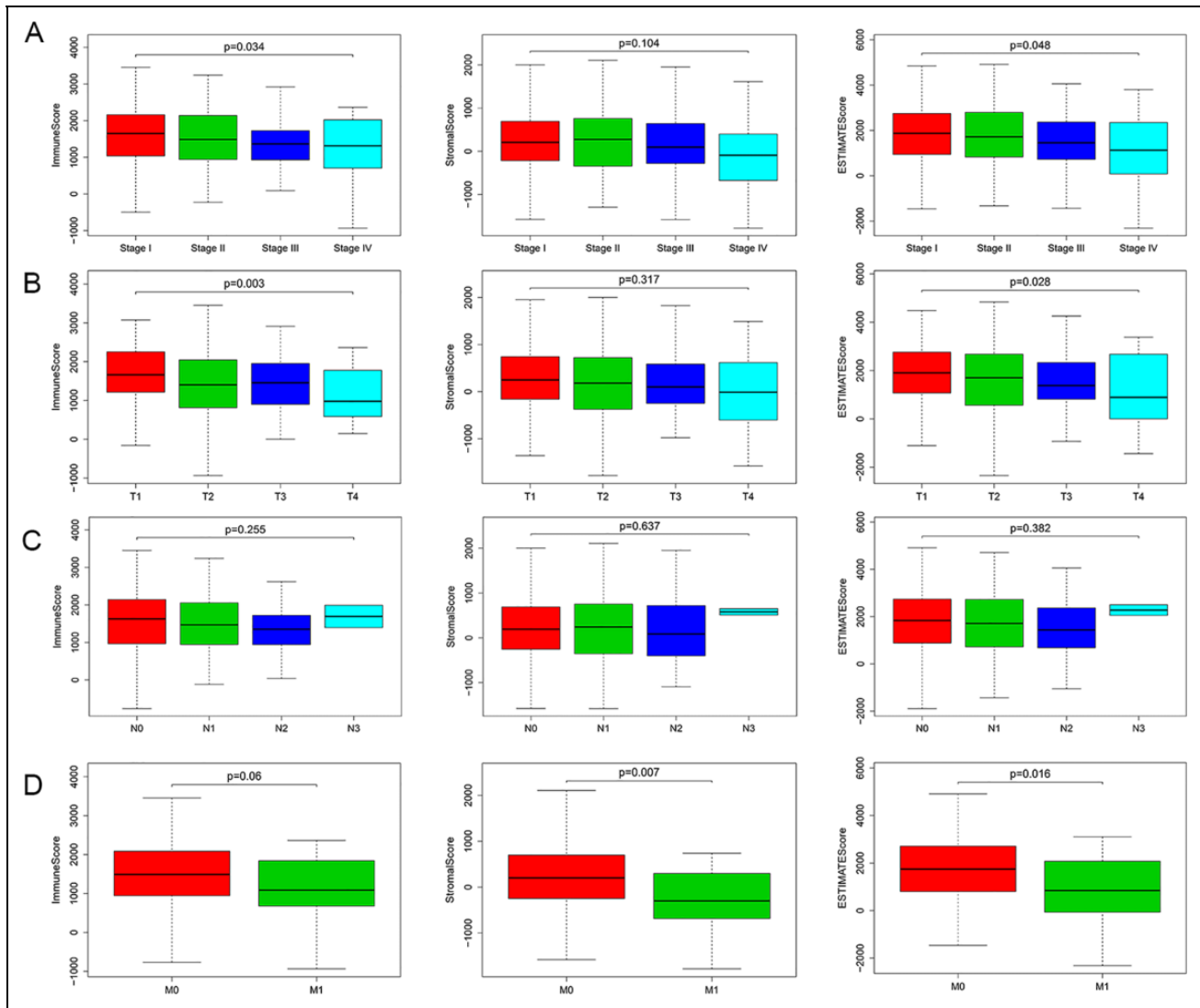


Figure 2. The correlations between immune/stromal/estimate scores and clinical features. (A) High stages were significantly associated with low immune ($P = 0.034$) and estimate ($P = 0.048$) scores. (B) Larger sizes of tumor were significantly associated with low immune ($P = 0.003$) and estimate ($P = 0.028$) scores. (C) There was no statistically significant difference between lymph node metastases and immune/stromal/estimate scores. (D) Distant metastasis was significantly associated with low stromal ($P = 0.007$) and estimate ($P = 0.016$) scores. The P -values were calculated using Wilcox test for the comparison of two groups and Kruskal test for three groups or more.

3453.015), stromal scores (from -1783.99 to 2107.561), and estimate scores (from -2344.43 to 4911.592). The clinical characteristics of LUAD patients were classified (Table 1). The tumor stages (Fig. 2A) were significantly correlated with the immune ($P = 0.034$) and estimate ($P = 0.048$) scores. The tumor sizes (Fig. 2B) were also significantly correlated with the immune ($P = 0.003$) and estimate ($P = 0.028$) scores. However, there were no statistically significant differences between these scores and lymph node metastases (Fig. 2C). The LUAD patients with distant metastases (Fig. 2D) had lower stromal ($P = 0.007$) and estimate ($P = 0.016$) scores.

According to the median scores, the LUAD patients were divided into corresponding high and low score groups

(immune scores^{High/Low}, stromal scores^{High/Low}, and estimate scores^{High/Low}). The patients in the immune scores^{High} or estimate scores^{High} groups had longer OS than those in the immune scores^{Low} or estimate scores^{Low} groups (Fig. 3). These results indicated that the immune and estimate scores were significantly associated with the OS of LUAD patients.

Screening DEGs Based on Immune/Stromal Scores

According to the threshold values ($|\log_2 \text{FC}| \geq 1$ and $\text{FDR} < 0.05$), the DEGs were obtained from transcription profiles of TCGA-LUAD samples using ESTIMATE algorithm. The heatmap based on the immune scores exhibited 1,092 upregulated and 302 down-regulated genes (Fig. 4A).

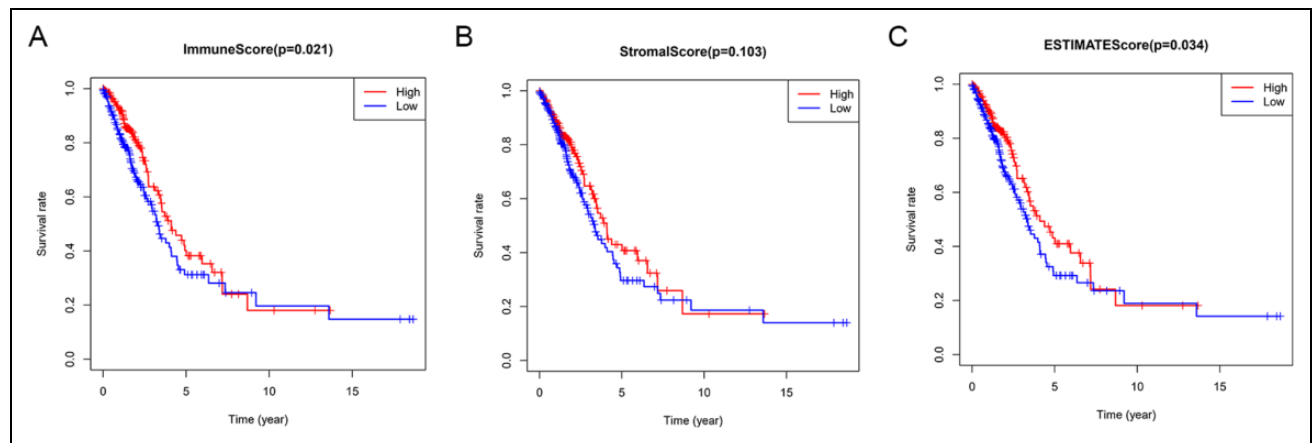


Figure 3. The correlations between the immune/stromal/estimate scores and OS. (A, C) The LUAD patients with high immune ($P = 0.021$) and estimate ($P = 0.034$) scores had better OS than those with low scores. (B) No statistically significant difference between the stromal scores and the OS of patients ($P = 0.103$). The P -value was calculated using the Log-rank test. LUAD: lung adenocarcinoma; OS: overall survival.

On the basis of the stromal scores, we identified 1,429 up-regulated genes and 211 down-regulated genes (Fig. 4B). A total of 589 up-regulated genes and 113 down-regulated genes were obtained by the intersection of corresponding groups (Fig. 4C, D). These 702 genes were defined as DEGs for subsequent analysis. Moreover, the immune/stromal scores of GSE68465 samples were also calculated (Supplemental Table 1).

Roles of DEGs During LUAD Pathogenesis

To explore the biological functions of DEGs, GO and KEGG enrichment analyses were applied. The GO terms were consisted of three parts: BPs, MFs, and CCs, and the top 10 enrichment results in each catalog were shown in Fig. 5A. The DEGs were involved in the processes of immune response, such as complement activation, antigen binding, activation of cell surface receptor, and so on. The similar results were also observed in the KEGG analysis (Fig. 5B). The consequences of KEGG functional enrichment were also mainly associated with immune response signal pathways including cytokine–cytokine receptor interaction, chemokine signaling pathway, and NF- κ B signaling pathway. These enrichment results confirmed reversely that ESTIMATE method was appropriate and accurate for DEG screening.

A 10-IRGP Risk Model was Established to Evaluate LUAD Prognosis

Based on the combined application of ImmPort database and acquired DEGs dataset, we obtained 66 immune-related DEGs, which could generate 2,145 IRGPs. After removing IRGPs with a value of 0 (or 1) in more than 80% TCGA-LUAD samples, 359 IRGPs were identified. Univariate Cox regression analysis was then applied to obtain 15 prognostic-

related candidate IRGPs ($P < 0.01$, Supplemental Table 2). Each gene pair was identified as an independent variable with coefficient trajectories (Fig. 6A). To confirm the accuracy of this risk model, 10-fold cross-validation was performed to obtain the confidence interval under each lambda (Fig. 6B). An IRGP risk model including 10 gene pairs was established (Table 2). Using time-dependent ROC curve analysis in TCGA training dataset, the optimal cut-off (value: -0.981) was applied to divide LUAD patients into the high- and low-risk groups. AUC value of 1-year survival rate was 0.702, and the C-index was 0.73 (Fig. 6C). The AUC of 2- and 5- year survival rates were 0.699 and 0.690 in the TCGA training set (Supplemental Fig. 1). Our results showed that the patients in the low-risk group had longer OS than those in the high-risk group ($P = 2.148 \times 10^{-5}$, Fig. 6D). According to the risk scores, the distribution of 10 IRGP scores and living status of LUAD patients in the training dataset were shown in Fig. 6E.

Assessment of the IRGP Model Accuracy

To better explore the accuracy of this IRGP signature, the uni- and multi-variate Cox analyses were applied in TCGA training and GEO validation datasets. Both the 10-IRGP signature and N-stage were independent prognostic elements (Table 3). In the validation dataset using the same risk model, the 1-year AUC was 0.692, the optimal cut-off score was -0.659 , and a high C-index value of 0.72 could still be achieved (Fig. 7A). The AUC of 2- and 5- year survival rates were 0.632 and 0.620 in GSE68465 validation dataset (Supplemental Fig. 1). The patients in GSE68465 were also divided into the high- and low-risk groups based on the cut-off value. The low-risk patients had longer OS than the high-risk ones in general (Fig. 7B). According to the risk scores, the distribution of the 10 IRGP scores and living status of LUAD patients in the validation dataset were shown in

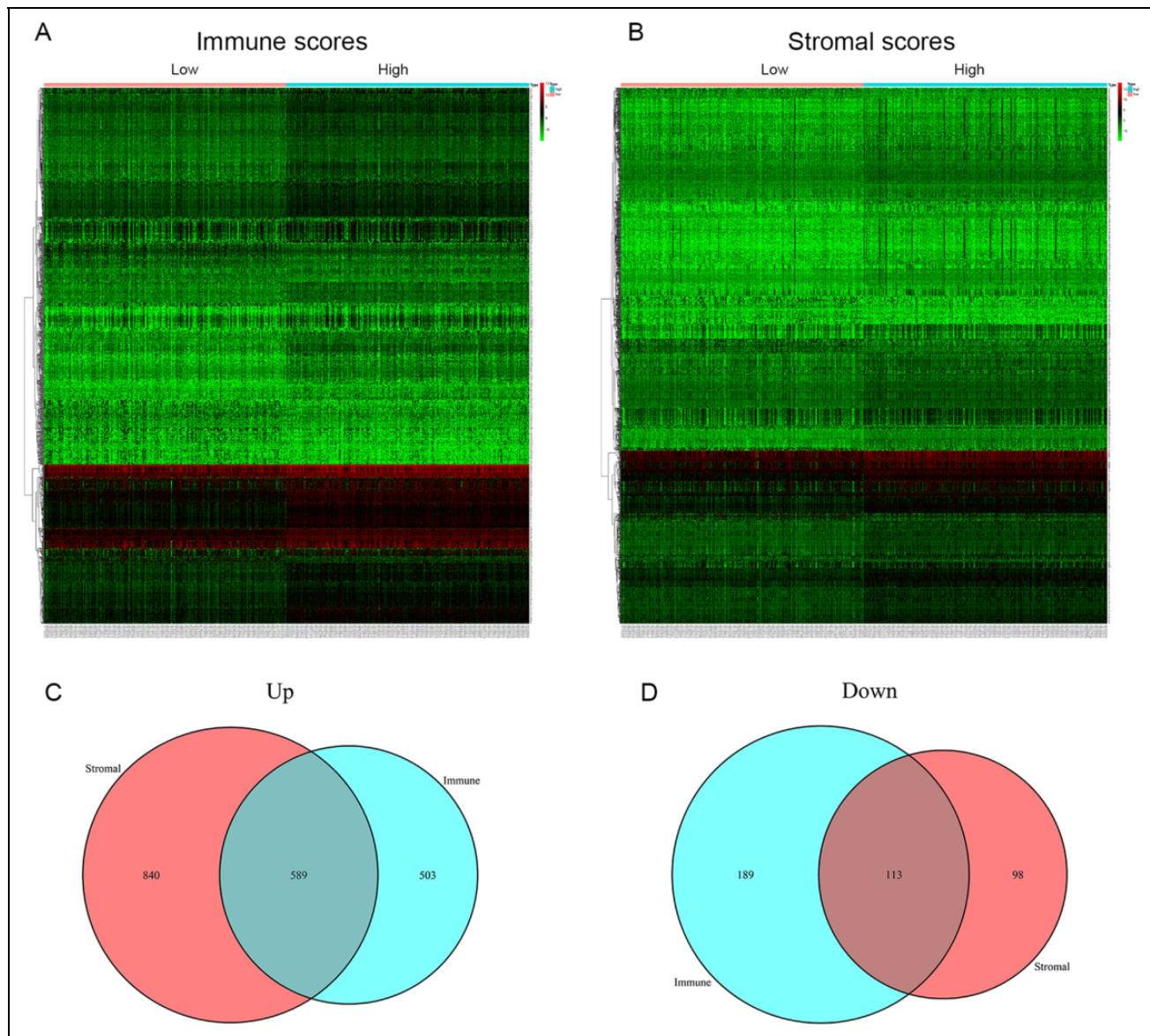


Figure 4. Gene expression profiles of immune/stromal score groups in the TCGA-LUAD training dataset. (A, B) Heatmaps of gene expressions in immune/stromal score^{High/Low} groups. $|\log_2 \text{FC}| \geq 1$ and $\text{FDR} < 0.05$ were considered as criterion. (C, D) Venn diagrams of up- and down-regulated DEGs in immune/stromal score groups. DEG: differentially expressed gene; FC: fold change; FDR: false discovery rate; LUAD: lung adenocarcinoma; TCGA: The Cancer Genome Atlas..

Fig. 7C. These results suggested that our IRGP signature was successfully established to predict the survival of LUAD patients.

Immune cell Infiltration in the High- and Low-Risk Groups

Previous studies indicated that immune cell infiltration had effects on LUAD progression, development, and prognosis. To examine the relevance of the risk scores with immune cell infiltration, CIBERSORT algorithm was used to compare the relative abundance of immune cells in different

groups. Our results suggested that regulatory T cells (Tregs), resting dendritic cells, plasma cells, CD4 memory resting T cells, and Tfh were highly expressed in the low-risk group, while M2 and M0 macrophages, activated dendritic cells, and neutrophils were lower in the low-risk group (Fig. 8 and Supplemental Fig. 2).

Functional Assessment of the IRGP Signature

To explore the signal pathways involved in this IRGP model, the functional enrichment analyses were compared in both the high- and low-risk groups using GSEA (Fig. 9 and

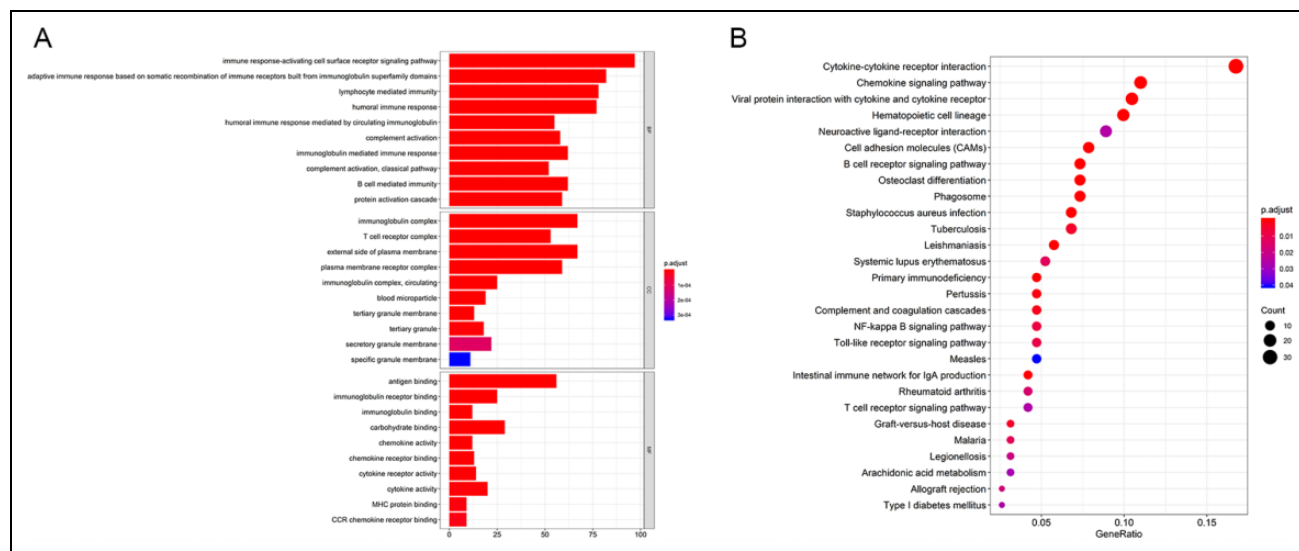


Figure 5. Functional analyses of DEGs. (A) The GO analysis of DEGs in three categories (BPs, MFs, and CCs). (B) The highly enriched signal pathways of DEGs in KEGG analysis. BPs: biological processes; CCs: cellular components; DEG: differentially expressed gene; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MFs: molecular functions.

Table 2. The Detailed Information About the I0-IRGPs Signature.

IRG 1	Full name	IRG 2	Full name	Coefficient
	CD19 molecule	CXCL6	C-X-C motif chemokine ligand 6	-0.336255502
CD19	CD19 molecule	CXCL11	C-X-C motif chemokine ligand 11	-0.121626909
CD19	CD19 molecule	TNFSF8	TNF superfamily member 8	-0.122011452
CD79B	CD79b molecule	TLR4	Toll like receptor 4	-0.1704619
IGLV1-44	Immunoglobulin lambda variable 1-44	S100P	S100 calcium binding protein P	-0.398245257
CD28	CD28 molecule	CD1E	CD1e molecule	0.304288728
CR2	Complement C3d receptor 2	PTX3	Pentraxin 3	-0.060856539
TLR7	Toll like receptor 7	PTX3	Pentraxin 3	-0.244271763
CCL23	C-C motif chemokine ligand 23	PTX3	Pentraxin 3	-0.123564823
HLA-DQAI	Major histocompatibility complex, class II, DQ alpha 1	CXCL10	C-X-C motif chemokine ligand 10	-0.230577243

IRGP: immune-related gene pair.

Supplemental Fig. 3). For the high-risk group, the enriched signal pathways were associated with DNA repair, homologous recombination, mismatch repair, and so on (Supplemental Table 3). These results provided additional evidence that immunity was associated with DNA damage-related signal pathways.

Discussion

Lung cancer draws more attention due to its substantial morbidity and mortality. The noteworthy features of lung cancer are rapid progression, high malignancy, low rates of early diagnosis, and adverse prognosis¹. LUAD is the most prevalent subtype of lung cancer. During past years, numerous studies suggested that TME played important roles in tumorigenesis, development, and metastasis^{4,26,27}. TME-related biomarkers were reported to predict prognosis for patients as novel targets for immunotherapy^{10,28}. Several immune-

related gene-related prognostic models were reported as well^{12,13}. While these researches had singular focus on TME-related or immune-related genes, the accuracy of these predictive models remained uncertain and they were difficult to be used in clinic.

Our study, for the first time, established a prognostic model for LUAD patients using ESTIMATE algorithm and Lasso-Cox analysis. Traditional IRGs were replaced by IRGPs to establish the model in our research. In the models established on IRGs, the significant genes were just extracted according to their expression levels, and the heterogeneity of patients was ignored. Applying IRGPs to construct the signature, risk scores could be calculated via comparing the relative expression of gene pairs in the same sample, and the scaling and normalization of data could be avoided.

First, we obtained 702 TME-related DEGs from the TCGA-LUAD training dataset by ESTIMATE algorithm.

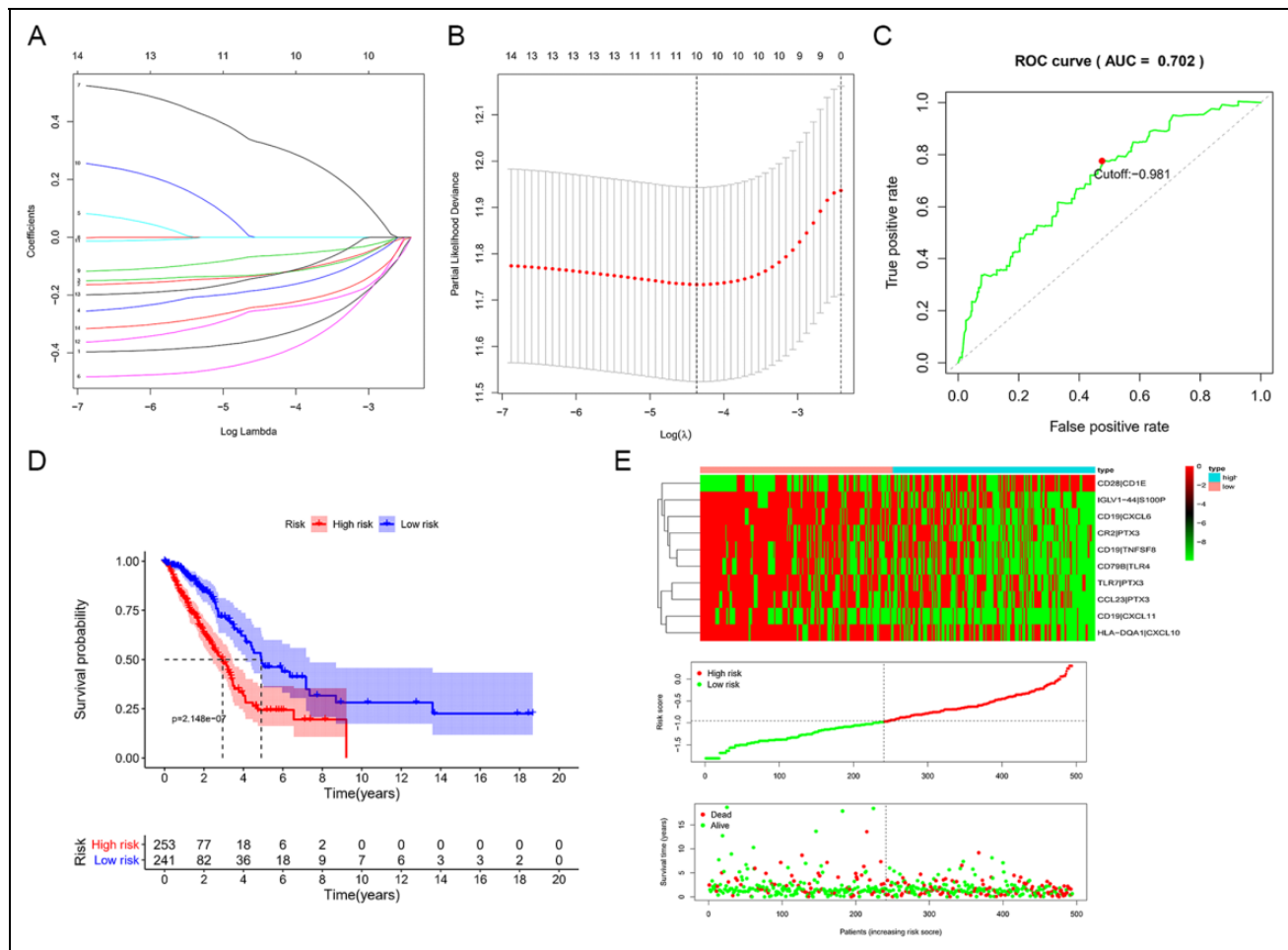


Figure 6. Construction of a 10-IRGP prognostic signature. (A) Lasso coefficient profiles of 15 IRGPs with $P < 0.01$. (B) Ten-fold cross-validation result identified optimal values of the penalty parameter λ . (C) Time-dependent ROC curve for IRGPs in the training set (1-year AUC = 0.702, cut-off value: -0.981). (D) The K-M curve in the training cohort (high- vs. low-risk, $P = 2.148 \times 10^{-7}$). (E) The heatmap showed the 10-IRGP scores based on the risk scores, the risk score distribution of TCGA-LUAD patients, and the survival state of patients according to the risk scores. AUC: area under the curve; IRGP: immune-related gene pair; K-M: Kaplan–Meier; LUAD: lung adenocarcinoma; ROC: receiver operating characteristic; TCGA: The Cancer Genome Atlas.

Table 3. Uni- and Multi-Variate Analyses of the Prognostic Factors in Both the Training and Validation Cohorts.

Variable	Single-variate factor analysis		Multivariate factor analysis	
	HR (95% CI)	P-value	HR (95% CI)	P-value
TCGA dataset				
Risk score	3.898 (2.61–5.82)	2.929E-11	3.267 (2.138–4.991)	4.392E-08
Age	1.002 (0.984–1.021)	0.784	1.021 (1.001–1.041)	0.034
Gender	1.04 (0.728–1.485)	0.827	0.906 (0.63–1.304)	0.597
T	1.622 (1.308–2.011)	1.014E-05	1.217 (0.961–1.542)	0.102
M	1.701 (0.935–3.095)	0.081	1.326 (0.706–2.491)	0.378
N	1.795 (1.467–2.197)	1.328E-08	1.447 (1.155–1.813)	0.001
GSE68465 dataset				
Risk score	1.891 (1.386–2.579)	5.666E-05	1.871 (1.361–2.572)	0.0001114
Age	1.026 (1.013–1.04)	8.123E-05	1.031 (1.017–1.045)	7.005E-06
Gender	1.426 (1.1–1.849)	0.007	1.241 (0.955–1.613)	0.105
T	1.664 (1.387–1.998)	4.466E-08	1.438 (1.193–1.734)	0.0001378
N	2.012 (1.711–2.365)	2.421E-17	2.022 (1.716–2.383)	4.04E-17

CI: confidence interval; HR: hazard ratio; TCGA: The Cancer Genome Atlas.

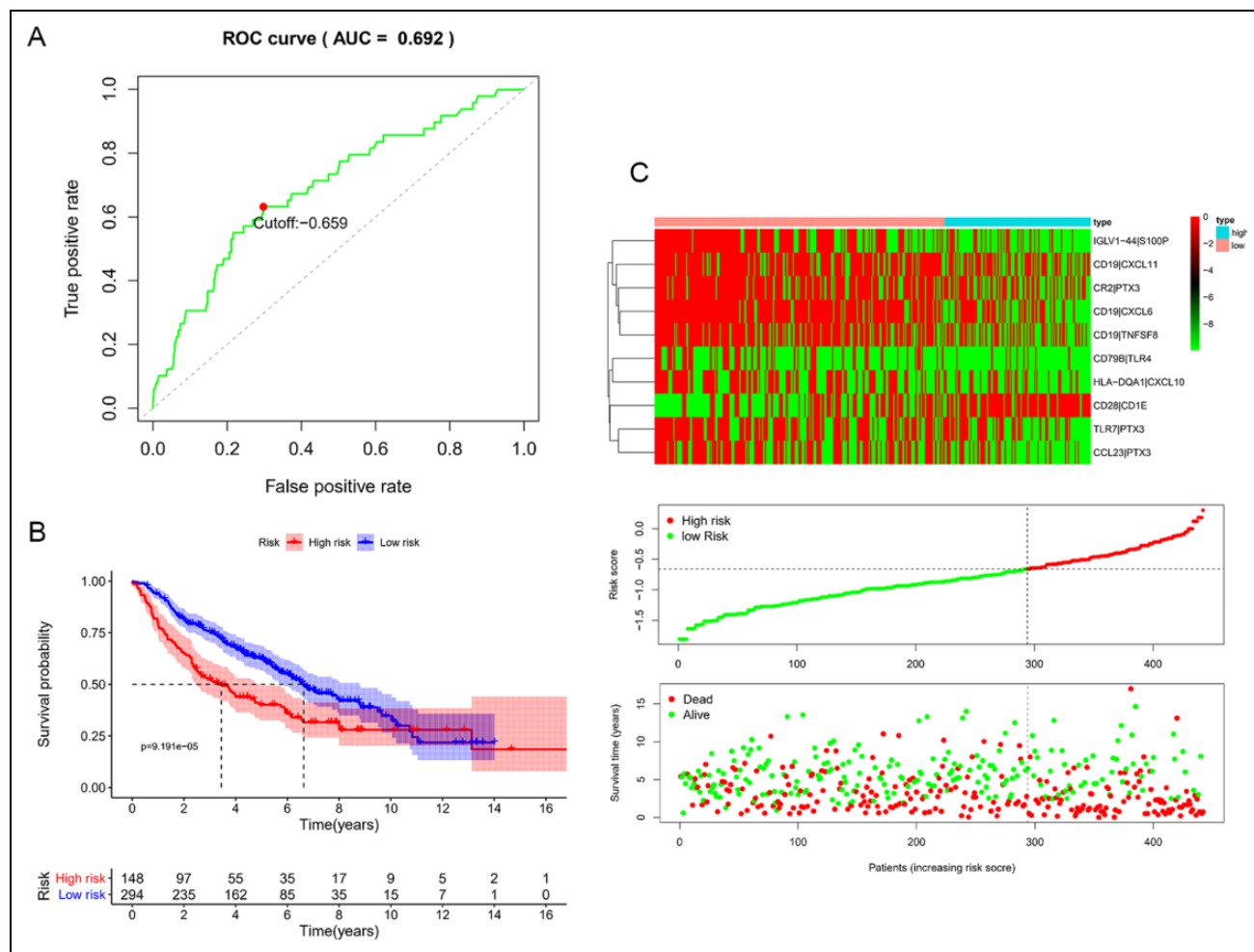


Figure 7. Validation of the 10-IRGP prognostic signature in the GEO validation dataset. (A) Time-dependent ROC curve for IRGPs (1-year AUC = 0.692, cut-off value: -0.659). (B) The K-M curve in the validation cohort (high- vs low-risk, $P = 9.191 \times 10^{-5}$). (C) The heatmap showed the 10-IRGP scores based on the risk scores, the risk score distribution of GSE68465 patients, and the survival state of patients according to the risk scores. AUC: area under the curve; GEO: Gene Expression Omnibus; IRGP: immune-related gene pair; K-M: Kaplan-Meier; ROC: receiver operating characteristic.

The GO and KEGG analysis showed that these DEGs participated in inflammatory response and innate/adaptive immune response. These results accorded with previous reports on the functions of immune and stromal cells in the TME of LUAD patients^{29–31}. We extracted 16 immune genes from these DEGs to build a 10-IRGP prognostic signature via Lasso-Cox analysis. Most of these genes were cytokines and chemokine ligands. These genes participated in the progression of tumors and affected the prognosis of patients by identifying and presenting antigen in immune system. Previous studies analyzed the roles of PTX3 in cancer progression. It has been proved that PTX3 is produced by lung cancer cells, which consists of higher PTX3 plasma levels in lung cancer patients than healthy people³². Furthermore, the higher expression of PTX3, the worse progression-free survival for lung cancer patients³³. The deficiency of PTX3 promoted cancer-associated inflammatory response

and activated complement pathways, such as upregulation of C3 and C5a^{34,35}. In lung cancer, CXCL6 promotes tumor metastasis via targeting miR-515-5p³⁶. In colorectal cancer, the poor prognosis was attributed to CXCL10/11 secretion, enhancing the infiltration of tumor-associated macrophages into tissues and promoting cell invasion^{37,38}. CXCL10 also induces chemotaxis of CXCR3⁺ T lymphocytes and CD11c⁺ DCs into the TME to eliminate tumor cells after IFN- γ treatment³⁹. CCL23 promotes the migration of ovarian cancer through ERK1/2 and PI3 K pathway activation⁴⁰. Moreover, CCL23 binds to CCR1 and thus exerts chemotactic activities on monocytes, DCs, and resting T lymphocytes⁴¹. TLR4 was demonstrated to promote the progression of tumor growth and apoptosis resistance via secreting various factors, such as antiapoptotic protein. However, activated TLR4 also has the antitumor effects. Activated TLR4 suppressed tumor progression via

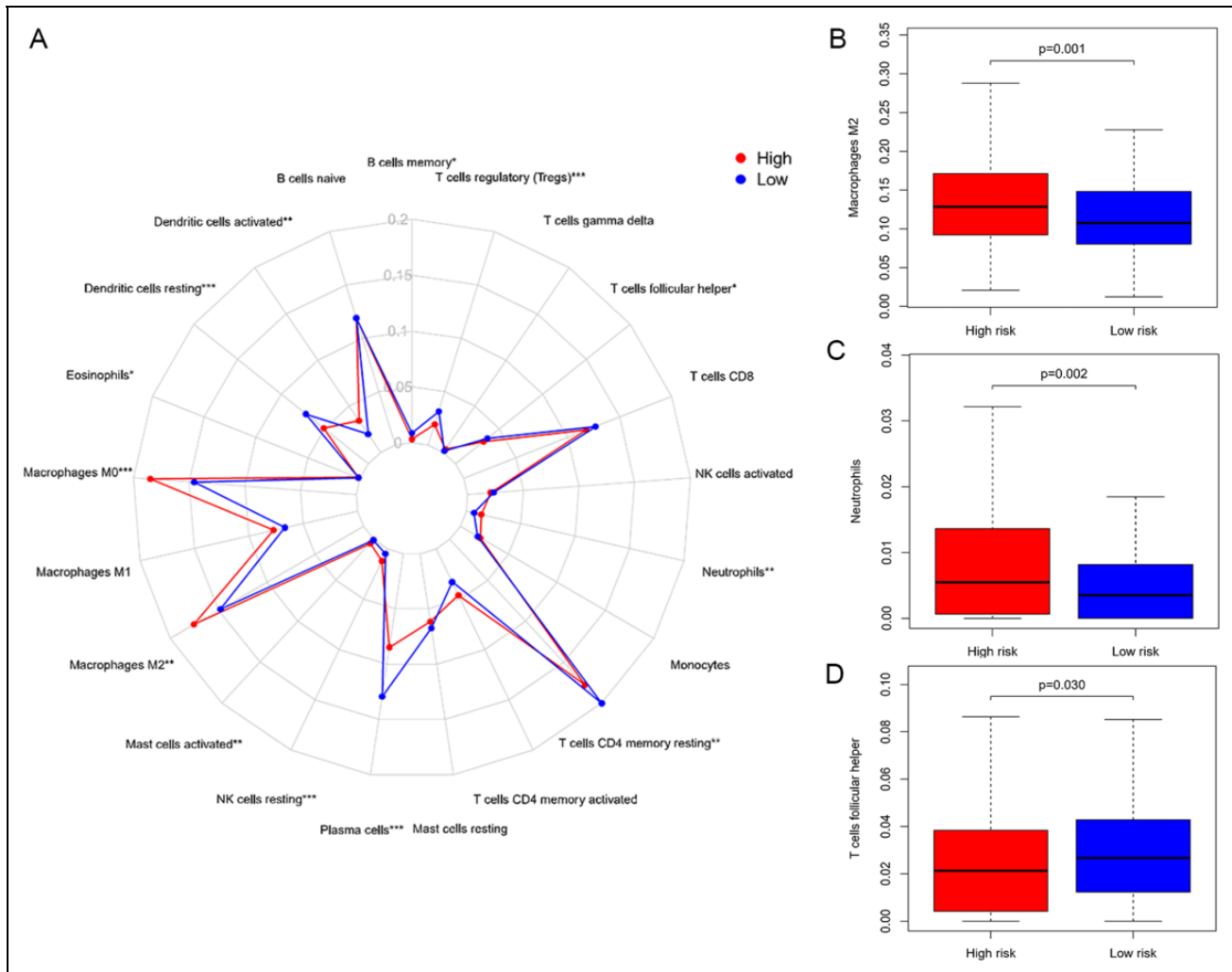


Figure 8. Immune cell infiltration in both the high- and low-risk groups. (A) The relative abundance of immune cells in different risk groups. (B, C) M2 macrophage ($P = 0.001$) and neutrophil ($P = 0.002$) infiltrations in the low-risk group were significantly lower than those in the high-risk groups. (D) Tfh was significantly higher in the low-risk group ($P = 0.03$). Tfh: T cell follicular helpers.

increasing CD8⁺ T cell infiltration in osteosarcoma⁴² and developed resistance to anti-epidermal growth factor receptor (anti-EGFR) therapy in head and neck squamous cell carcinoma⁴³. It had been proved that tumor-derived autophagosomes induce B cell activation and T cell immune responses. Macrophages enhance this process through the TLR4 and MyD88 signaling pathways⁴⁴. The expression of TLR7 had different predictive values for different cancers. Its higher levels indicated lower OS in breast cancer and lung cancer⁴⁵, but the higher OS in melanoma^{46,47}. Costimulation of CD9-specific chimeric antigen receptors and CD28 was reported to have an improved antitumor effect in adoptively transferred T cells⁴⁸. In the upper dermis of the basal cell carcinoma lesion, the percentage of CD30 ligand (CD30 L, TNFSF8) positive mast cells were increased significantly^{49,50}. Researches showed that high expression of S100P was correlated with poor clinical outcomes. Moreover, S100P was specifically overexpressed in lung metastatic tissues⁵¹.

Based on this IRGP model, the patients were divided into the high- and low-risk groups. Overall, the K-M analysis showed that low-risk patients had longer OS than high-risk ones. However, two patients in the high-risk group and one patient in the low-risk group were survived after 14 years in validation dataset. This phenomenon probably resulted from the heterogeneity of the patients, the limited number of samples and datasets included in this study, and the different efficiency of clinical treatments such as radio- or chemotherapy in patients. The details of immunotherapy were lacked in the GSE68465 dataset, which might contribute to the survival of these two high-risk patients. More importantly, for the majority (439/442) of patients and for less than 14 years, our model presented excellent predictive functions. The high AUC and C-index values indicated the good performance of our model.

To understand the immune mechanisms in different groups, the immune cell infiltration analyses were explored

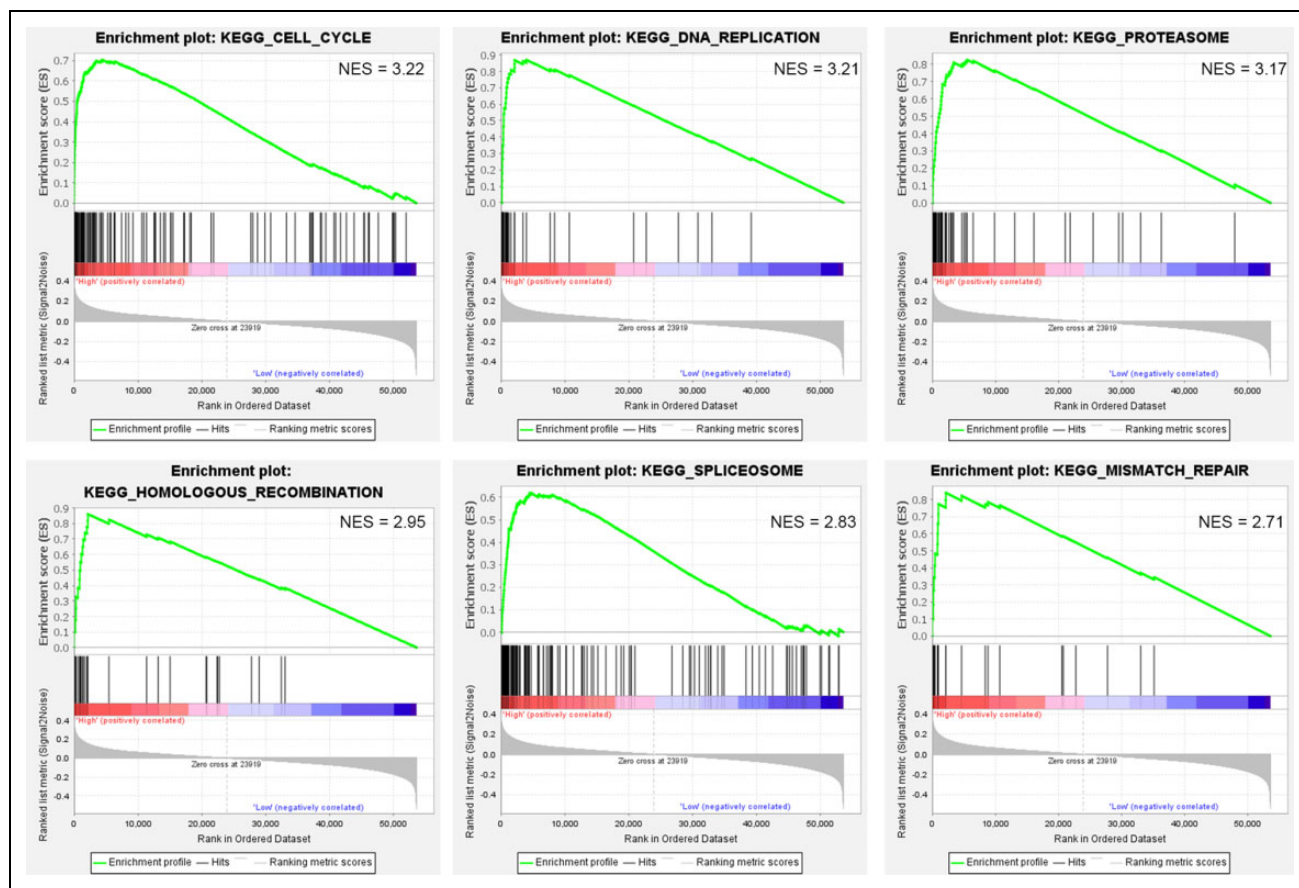


Figure 9. Functional assessment of the IRGP signature using GSEA. The results showed that the top six vital KEGG pathways in the high-risk group. GSEA: Gene Set Enrichment Analysis; IRGP: immune-related gene pair; KEGG: Kyoto Encyclopedia of Genes and Genomes.

with CIBERSORT. We founded that M2 macrophage and neutrophil infiltration in the low-risk group were significantly lower than those in the high-risk group. Previous studies proved that high levels of M2 macrophages were associated with poor OS in diffused large B-cell lymphoma and gastric cancer^{52,53}. The number of neutrophils could be a negative prognostic factor in many types of carcinoma, such as lung and ovarian cancers⁵⁴. In addition, Tfh was significantly higher in the low-risk group. The density of Tfh cells was positively associated with patients' survival in colorectal and breast cancer⁵⁵. Our research showed that Tregs were highly expressed in the low-risk group, and they were partially related to the clinical response for treatments in patients. It was proved that Tregs were activated in response to radiation damage, but there was an important homeostatic mechanism between radiotherapeutic benefit and damage. Because of the balance between increased Tregs and radiotherapy effects, patients would get more benefits from radiotherapy combined with immune modulation about Tregs⁵⁶. It also indicated better survival for patients. This was consistent with better OS in the low-risk group.

Moreover, the GSEA was used to investigate signal pathways involved with IRGP-related immune process. In the

high-risk group, most pathways were involved in DNA repair to ensure the genome integrality. It was worth to note that radiotherapy, as the traditional therapy, enhanced immune response through radio-related DNA damage. The DNA repair capacity of damaged cells was associated with resistance to radiation⁵⁷. The poor survival in the high-risk group might partly resulted from the strong capacity for DNA repair.

In clinical diagnosis and treatment, as the genetic heterogeneity, patients showed different survival outcomes, even though they had same clinical characteristics, pathologic conditions, and treatments⁵⁸. So, our prognosis model can serve as an individualized, single-sample estimate for LUAD patients' survival. Meanwhile, the IRGP signature can be used to predict the infiltration of immune cells in patients, which is conducive to the prediction of patient sensitivity to immunotherapy. Our study holds great promise for the clinical application. In the same patient, we can analyze the expression of 16 genes in the IRGP signature and calculated the scores by particular pairwise comparison. The patients could be divided into the high- or low- risk group according to the risk scores and then predicted with different survival outcomes. Meanwhile, the sensitivity to immunotherapy for

patients can be predicted according to the infiltration of immune cells in different risk groups.

There were several limitations in our current study. First, there were only one TCGA-LUAD training and one GSE68465 validation datasets. Second, our study was based on publicly accessible database using the computational biology algorithm and these results should be further validated with more experiments *in vivo* and *in vitro*. Third, the heterogeneity of tumors might bring unavoidable sampling bias. In future, these bias could be reduced with the expanding of datasets and increasing the sample counts.

Conclusions

An IRGP signature was established to predict LUAD prognosis. Based on this model, a novel classification method for LUAD patients was developed mainly based on the risk score, which was an independent predictive factor. Moreover, our studies provided promising perspectives about therapeutical targets for LUAD immunotherapy.

Author Contributions

XJ, YG, and CX designed the study. XJ, YG, NZ, and CY collected the mRNA transcriptome data and clinical information from public databases. XJ, YL, WS, and JZ performed statistical analyses. XJ and JR wrote the manuscript. YG and CX improved and revised the manuscript. All authors read and approved the final manuscript.

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=GSE68465>).

Ethical Approval

As the work is a bioinformatics analysis article, ethical approval was not necessary and all the data were retrieved from the free online databases.

Statement of Human and Animal Rights

This article does not contain studies with human or animal subjects.

Statement of Informed Consent

There are no human subjects 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

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by National Natural Science Foundation of China (81773236, 81800429, and 81972852), Key Research & Development Project of Hubei Province (2020BCA069), Nature Science Foundation of Hubei Province (2020CFB612), Health Commission of Hubei Province Medical Leading Talent Project, Health Commission of Hubei Province Scientific Research Project

(WJ2019H002 and WJ2019Q047), Young & Middle-Aged Medical Backbone Talents of Wuhan, and Zhongnan Hospital of Wuhan University Science, Technology and Innovation Seed Fund (znp2018028, znp2018070, znp2019001, znp2019048, and ZNJC201922).

ORCID iD

Conghua Xie  <https://orcid.org/0000-0001-6623-9864>

Supplemental Material

Supplemental material for this article is available online.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature.* 2018; 553(7689):446–454.
3. Osmani L, Askin F, Gabrielson E, Li QK. Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): Moving from targeted therapy to immunotherapy. *Semin Cancer Biol.* 2018;52(Pt 1):103–109.
4. Wu T, Dai Y. Tumor microenvironment and therapeutic response. *Cancer Lett.* 2017;387:61–68.
5. Gao Y, Khan GJ, Wei X, Zhai KF, Sun L, Yuan S. DT-13 inhibits breast cancer cell migration via non-muscle myosin II-A regulation in tumor microenvironment synchronized adaptations. *Clin Transl Oncol.* 2020.22(9):1591–1602
6. Zhang W, Hu X, Liang J, Zhu Y, Zeng B, Feng L, Zhao C, Liu S, Liu B, Zhang K. oHSV2 can target murine colon carcinoma by altering the immune status of the tumor microenvironment and inducing antitumor immunity. *Mol Ther Oncolytics.* 2020; 16:158–171.
7. Zhou Q, Bauden M, Andersson R, Hu D, Marko-Varga G, Xu J, Sasor A, Dai H, Pawlowski K, Said Hilmersson K, Chen X, et al. YAP1 is an independent prognostic marker in pancreatic cancer and associated with extracellular matrix remodeling. *J Transl Med.* 2020;18(1):77.
8. He W, Zhang H, Han F, Chen X, Lin R, Wang W, Qiu H, Zhuang Z, Liao Q, Zhang W, Zhang W, et al. CD155T/TIGIT signaling regulates CD8(+) T-cell metabolism and promotes tumor progression in human gastric cancer. *Cancer Res.* 2017; 77(22):6375–6388.
9. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4(1):1–1.
10. Jia D, Li S, Li D, Xue H, Yang D, Liu Y. Mining TCGA database for genes of prognostic value in glioblastoma microenvironment. *Aging (Albany NY).* 2018;10(4):592–605.
11. Zeng Q, Zhang W, Li X, Lai J, Li Z. Bioinformatic identification of renal cell carcinoma microenvironment-associated

- biomarkers with therapeutic and prognostic value. *Life Sci.* 2020;243:117273.
12. Ogino S, Nowak JA, Hamada T, Phipps AI, Peters U, Milner DA Jr., Giovannucci EL, Nishihara R, Giannakis M, Garrett WS, Song M. Integrative analysis of exogenous, endogenous, tumour and immune factors for precision medicine. *Gut.* 2018; 67(6):1168–1180.
 13. She Y, Kong X, Ge Y, Yin P, Liu Z, Chen J, Gao F, Fang S. Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma. *Cancer Cell Int.* 2020; 20(1):22.
 14. Xie P, Ma Y, Yu S, An R, He J, Zhang H. Development of an immune-related prognostic signature in breast cancer. *Front Genet.* 2019;10:1390.
 15. Tomczak K, Czerwinka P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn).* 2015;19(1A):A68–77.
 16. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J, Thomson E, et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol Res.* 2014;58(2-3):234–239.
 17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43(7):e47.
 18. Galili T, O'Callaghan A, Sidi J, Sievert C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics.* 2018;34(9):1600–1602.
 19. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics.* 2011;12:35.
 20. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–287.
 21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–15550.
 22. Nie H, Bu F, Xu J, Li T, Huang J. 29 immune-related genes pairs signature predict the prognosis of cervical cancer patients. *Sci Rep.* 2020;10(1):14152.
 23. Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncol.* 2017;3(11):1529–1537.
 24. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res.* 2010;1(4): 274–278.
 25. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–457.
 26. Curtis M, Mukherjee A, Lengyel E. The tumor microenvironment takes center stage in ovarian cancer metastasis. *Trends Cancer.* 2018;4(8):517–519.
 27. Mittal S, Brown NJ, Hoken I. The breast tumor microenvironment: role in cancer development, progression and response to therapy. *Expert Rev Mol Diagn.* 2018;18(3):227–243.
 28. Zhang Z, Li Z, Liu Z, Zhang X, Yu N, Xu Z. Identification of microenvironment-related genes with prognostic value in clear cell renal cell carcinoma. *J Cell Biochem.* 2020.121(7): 3606–3615.
 29. Tan Q, Huang Y, Deng K, Lu M, Wang L, Rong Z, Zhao W, Li S, Xu Z, Fan L, Li K, et al. Identification immunophenotyping of lung adenocarcinomas based on the tumor microenvironment. *J Cell Biochem.* 2020.121(11):4569–4579
 30. Angel PM, Bruner E, Bethard J, Clift CL, Ball L, Drake RR, Feghali-Bostwick C. Extracellular matrix alterations in low-grade lung adenocarcinoma compared with normal lung tissue by imaging mass spectrometry. *J Mass Spectrom.* 2020;55(4): e4450.
 31. Wang W, Ren S, Wang Z, Zhang C, Huang J. Increased expression of TTC21A in lung adenocarcinoma infers favorable prognosis and high immune infiltrating level. *Int Immunopharmacol.* 2020;78:106077.
 32. Planque C, Kulasingam V, Smith CR, Reckamp K, Goodglick L, Diamandis EP. Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. *Mol Cell Proteomics.* 2009;8(12): 2746–2758.
 33. Berg J, Halvorsen AR, Bengtson MB, Tasken KA, Maelandsmo GM, Yndestad A, Halvorsen B, Brustugun OT, Aukrust P, Ueland T, Helland A. Levels and prognostic impact of circulating markers of inflammation, endothelial activation and extracellular matrix remodelling in patients with lung cancer and chronic obstructive pulmonary disease. *BMC Cancer.* 2018;18(1):739.
 34. Garlanda C, Bottazzi B, Magrini E, Inforzato A, Mantovani A. PTX3, a Humoral pattern recognition molecule, in innate immunity, tissue repair, and cancer. *Physiol Rev.* 2018;98(2): 623–639.
 35. Doni A, Stravalaci M, Inforzato A, Magrini E, Mantovani A, Garlanda C, Bottazzi B. The Long Pentraxin PTX3 as a link between innate immunity, tissue remodeling, and cancer. *Front Immunol.* 2019;10:712.
 36. Li J, Tang Z, Wang H, Wu W, Zhou F, Ke H, Lu W, Zhang S, Zhang Y, Yang S, Ni S, et al. CXCL6 promotes non-small cell lung cancer cell survival and metastasis via down-regulation of miR-515-5p. *Biomed Pharmacother.* 2018;97:1182–1188.
 37. Chen J, Chen QL, Wang WH, Chen XL, Hu XQ, Liang ZQ, Cao YB, Cao YM, Su SB. Prognostic and predictive values of CXCL10 in colorectal cancer. *Clin Transl Oncol.* 2020.22(9): 1548–1564.
 38. Zeng YJ, Lai W, Wu H, Liu L, Xu HY, Wang J, Chu ZH. Neuroendocrine-like cells -derived CXCL10 and CXCL11 induce the infiltration of tumor-associated macrophage leading to the poor prognosis of colorectal cancer. *Oncotarget.* 2016; 7(19):27394–27407.
 39. Sharma S, Yang SC, Hillinger S, Zhu LX, Huang M, Batra RK, Lin JF, Burdick MD, Strieter RM, Dubinett SM. SLC/CCL21-

- mediated anti-tumor responses require IFN γ , MIG/CXCL9 and IP-10/CXCL10. *Mol Cancer*. 2003;2:22.
40. Krishnan V, Tallapragada S, Schaar B, Kamat K, Chanana AM, Zhang Y, Patel S, Parkash V, Rinker-Schaeffer C, Folkens AK, Rankin EB, et al. Omental macrophages secrete chemokine ligands that promote ovarian cancer colonization of the omentum via CCR1. *Commun Biol*. 2020;3(1):524.
 41. Tamassia N, Bianchetto-Aguilera F, Arruda-Silva F, Gardiman E, Gasperini S, Calzetti F, Cassatella MA. Cytokine production by human neutrophils: revisiting the “dark side of the moon”. *Eur J Clin Invest*. 2018;48(Suppl 2):e12952.
 42. Yahiro K, Matsumoto Y, Yamada H, Endo M, Setsu N, Fujiwara T, Nakagawa M, Kimura A, Shimada E, Okada S, Oda Y, et al. Activation of TLR4 signaling inhibits progression of osteosarcoma by stimulating CD8-positive cytotoxic lymphocytes. *Cancer Immunol Immunother*. 2020;69(5):745–758.
 43. Ju H, Hu Z, Lu Y, Wu Y, Zhang L, Wei D, Guo W, Xia W, Liu S, Ren G, Hu J. TLR4 activation leads to anti-EGFR therapy resistance in head and neck squamous cell carcinoma. *Am J Cancer Res*. 2020;10(2):454–472.
 44. Yang Y, Feng R, Wang YZ, Sun HW, Zou QM, Li HB. Toll-like receptors: triggers of regulated cell death and promising targets for cancer therapy. *Immunol Lett*. 2020;223:1–9.
 45. Urban-Wojciuk Z, Khan MM, Oyler BL, Fahraeus R, Marek-Trzonkowska N, Nita-Lazar A, Hupp TR, Goodlett DR. The Role of TLRs in anti-cancer immunity and tumor rejection. *Front Immunol*. 2019;10:2388.
 46. Shi S, Xu C, Fang X, Zhang Y, Li H, Wen W, Yang G. Expression profile of Tolllike receptors in human breast cancer. *Mol Med Rep*. 2020;21(2):786–794.
 47. Zhang M, Yan Z, Wang J, Yao X. Toll-like receptors 7 and 8 expression correlates with the expression of immune biomarkers and positively predicts the clinical outcome of patients with melanoma. *Oncotargets Ther*. 2017;10:4339–4346.
 48. Kowolik CM, Topp MS, Gonzalez S, Pfeiffer T, Olivares S, Gonzalez N, Smith DD, Forman SJ, Jensen MC, Cooper LJ. CD28 costimulation provided through a CD19-specific chimeric antigen receptor enhances in vivo persistence and antitumor efficacy of adoptively transferred T cells. *Cancer Res*. 2006;66(22):10995–11004.
 49. Kennedy MK, Willis CR, Armitage RJ. Deciphering CD30 ligand biology and its role in humoral immunity. *Immunology*. 2006;118(2):143–152.
 50. Diaconu NC, Kaminska R, Naukkarinen A, Harvima RJ, Nilsson G, Harvima IT. Increase in CD30 ligand/CD153 and TNF- α expressing mast cells in basal cell carcinoma. *Cancer Immunol Immunother*. 2007;56(9):1407–1415.
 51. Arumugam T, Logsdon CD. S100P: a novel therapeutic target for cancer. *Amino Acids*. 2011;41(4):893–899.
 52. Gambardella V, Castillo J, Tarazona N, Gimeno-Valiente F, Martinez-Ciarpaglini C, Cabeza-Segura M, Rosello S, Roda D, Huerta M, Cervantes A, Fleitas T. The role of tumor-associated macrophages in gastric cancer development and their potential as a therapeutic target. *Cancer Treat Rev*. 2020;86:102015.
 53. Nam SJ, Go H, Paik JH, Kim TM, Heo DS, Kim CW, Jeon YK. An increase of M2 macrophages predicts poor prognosis in patients with diffuse large B-cell lymphoma treated with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone. *Leuk Lymphoma*. 2014;55(11):2466–2476.
 54. Futosi K, Fodor S, Mocsai A. Neutrophil cell surface receptors and their intracellular signal transduction pathways. *Int Immunopharmacol*. 2013;17(3):638–650.
 55. Vinuesa CG, Linterman MA, Yu D, MacLennan IC. Follicular Helper T cells. *Annu Rev Immunol*. 2016;34:335–368.
 56. Kachikwu EL, Iwamoto KS, Liao YP, DeMarco JJ, Agazaryan N, Economou JS, McBride WH, Schaeue D. Radiation enhances regulatory T cell representation. *Int J Radiat Oncol Biol Phys*. 2011;81(4):1128–1135.
 57. Li L, Zhu T, Gao YF, Zheng W, Wang CJ, Xiao L, Huang MS, Yin JY, Zhou HH, Liu ZQ. Targeting DNA Damage Response in the Radio(Chemo)therapy of Non-Small Cell Lung Cancer. *Int J Mol Sci*. 2016;17(6):839.
 58. Wu J, Zhao Y, Zhang J, Wu Q, Wang W. Development and validation of an immune-related gene pairs signature in colorectal cancer. *Oncimmunology*. 2019;8(7):1596715.