Serican Journal of Medicine
2025, VOL. 2, NO. 1, 23130
https://doi.org/10.17161/sjm.v2i1.23130

Identification of tumor microenvironment-based gene prognostic signature with promising predictive value for chemo-immunotherapy outcomes in lung adenocarcinomas

Zhihuan Lin[1*], Haixia Xu[1*], Li Zhang[2#] and Geng Tian[1#]

[1]Department of Medical Oncology, The First Affiliated Hospital of Shenzhen University and Shenzhen Second People's Hospital, Shenzhen 518035, China; [2]Department of Medical Oncology, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China. * Zhihuan Lin and Haixia Xu contributed equally to this work. # Corresponding authors: Li Zhang, MD/PhD, Sun Yat-sen University Cancer Center, and Geng Tian, MD, Shenzhen Second People's Hospital.

ABSTRACT

Objectives: The tumor microenvironment (TME) plays a critical role in tumor progression and therapeutic response. We aimed to establish a TME-associated gene signature for lung adenocarcinoma (LUAD) patients.

Materials and Methods: We comprehensively analyzed the gene expression data of 513 LUAD patients deriving from The Cancer Genome Atlas (TCGA) database. To estimate the composition of TME, The Estimation of Stromal and Immune cells in malignant tumor tissue using the expression data (ESTIMATE) algorithm was used. We then utilized protein-protein interaction (PPI) analysis, modular analysis, least absolute shrinkage and selection operator (LASSO), and COX regression to explore the related candidate genes with survival. Eventually, a three-gene signature was constructed for risk stratification. Furthermore, we investigated its predictive value in advanced LUAD patients who received chemotherapy alone or combined with anti-PD-1 therapy.

Results: We identified three genes, namely *CCR2*, *CD40LG,* and *CCL21,* to construct a TME-associated risk stratification gene signature. This gene signature was independently linked to patients' overall survival (OS) among the TCGA dataset (HR, 1.99; 95% CI:1.36-2.93, $P < 0.001$) and GEO dataset (HR, 1.62; 95%CI:1.19-2.20, $P < 0.001$). The addition of anti-PD-1 inhibitor sintilimab to chemotherapy resulted in a significantly longer progression-free survival (PFS) (HR, 0.33;95%CI:0.20-0.56, $P < 0.001$) and OS (HR, 0.56; 95%CI:0.33-0.98, P=0.043) in low- risk patients but not in those with high- risk (PFS: HR, 0.57; 95%CI:0.24-1.31, $P = 0.173$; OS: HR, 0.66; 95% CI:0.31-1.39, P=0.273). Moreover, in patients who received sintilimab combined with chemotherapy, PFS (HR, 1.96;95%CI:1.08-3.56, P=0.024) and OS (HR, 2.31;95% CI:1.29-4.14, P=0.004) were significantly longer in high-risk group than in low-risk group, whereas no significant difference was found in chemotherapy alone.

Conclusions: This study generated a three-gene prognostic signature based on TME-associated core genes in patients with LUAD. This gene signature has a good value for prognosis prediction. In addition, this signature was correlated with treatment outcomes among patients treated with chemotherapy combined with anti-PD1 therapy.

ARTICLE HISTORY
Received: Nov. 21, 2024
Revised: Jan. 18, 2025
Accepted: Jan. 25, 2025

KEYWORDS
Lung adenocarcinoma, tumor microenvironment, ESTIMATE, overall survival, chemoimmunotherapy

Introduction

Nowadays, lung cancer remains the primary cause of cancer death, in which lung adenocarcinoma (LUAD) is the primary subtype [1]. Although the tumor, node, and metastasis (TNM) classification represents the fundamental nomenclature staging system, it is still far from satisfying to reflect cancer prognosis [2]. The tumor microenvironment (TME), which comprises several immunocytes, growth factors, and cytokines, exerts a key role in tumor growth [3-5]. It may provide important prognostic information [6]. Thus, the incorporation of TME-associated molecular biomarkers would be constructive.

Estimation of stromal and immune cells in malignant Tumor tissues using expression data (ESTIMATE) is an algorithm established to evaluate stro-

mal and immune cell composition amongst TME [7]. Previous studies have demonstrated an excellent predictive value for this big-data-based algorithm in assessing patients' prognosis with malignant tumors [8-12].

We utilized the gene expression data and corresponding clinical profiles of 513 LUAD samples derived from The Cancer Genome Atlas (TCGA) database to explore TME-associated core genes, based on the ESTIMATE algorithm. We generated a reliable TME-associated three-gene signature with prognostic value. Furthermore, this gene signature served as an indicator to predict treatment benefit from the addition of anti-PD-1 inhibitor therapy to standard chemotherapy.

Materials and methods

Datasets

TCGA-LUAD dataset, comprising gene expressions and clinical profiles, served as a training data set. Kaplan-Meier (K-M) plot, an online search tool, was used to validate the predictive value amongst each gene in our gene signature. GSE68465, which contained 442 LUAD samples was used as an external validation dataset. In addition, 171 patients with available RNA-seq profiles derived from the ORIENT-11 clinical trial (ClinicalTrials.gov: NCT03607539) were used to explore the potential predictive value of the generated gene signature in patients treated with either chemotherapy alone or in a combination of anti-PD-1 inhibitor.

TME-associated gene determination stratified by immune scores (IS) and stromal scores (SS)

The "estimate" R package was used to estimate the IS, SS, and ESTIMATE scores (ES) for each case. ES was calculated as IS plus SS. Patients were stratified into low or high-score subgroups based on median IS and SS. A subset of genes synchronously up-regulated or down-regulated in the high IS/SS subgroup were extracted, using package 'limma' with thresholds as fold change >2 and false discovery rate (FDR) < 0.05. These genes were defined as TME-associated genes. Gene biological functions were investigated by Gene ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses, via R package 'clusterProfiler,' 'enrichplot' and 'ggplot2'.

PPI network and modular analysis for core gene recognition

Candidate genes screened out above were subjected to PPI network analysis, utilizing Search Tool for the Retrieval of Interacting Genes (STRING, https://string-db.org/) with confidence >0.4 as the threshold. Isolated genes were removed. Then the remaining genes were visualized via Cytoscape software (v. 3.7.2) Molecular Complex Detection (MCODE) plug-in with the default parameters (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max depth = 100). MCODE score >10 was set as a threshold to identify the most significant modules. Genes in these models were deemed as core genes.

TME-associated gene prognostic signature construction

We screened out genes with prognostic value for overall survival (OS) using Cox regression through the 'survival' R package. For further narrowing down the number of genes for final signature construction and avoiding the overfitting problem, the least absolute shrinkage and selection operator (LASSO) regression was conducted via the 'glmnet' R package. Gene signature risk score was then calculated based on the final selected genes and their corresponding coefficients, as the following formula: Risk score=gene expression 1×coefficient 1+ gene expression 2×coefficient 2+⋯+ gene expression n×coefficient n. Receiver operating characteristic (ROC) analysis (via 'survivalROC') was performed to explore the optimal cut-off point, which is represented by the Youden index (where highest sensitivity+specificity−1) to stratify patients into high or low-risk groups. K-M plot was used to evaluate survival differences between each group.

Exploration of the potential relationship between generated gene signature and immune indicators

Enrichment levels of 29 immune signatures (including immune cells, immune-associated functions, and pathways) were calculated for each LUAD sample through the single-sample Gene Set Enrichment Analysis (ssGSEA) scores. We also extracted the immune infiltrates data (including CD4+ T cells, CD8+ T cells, B cells, neutrophils, dendritic cells, and macrophages) of TCGA LUAD samples from Tumor Immune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/) platform. Correlation

analysis for risk score and these immune indicators was generated through Spearman analysis.

Programmed death-ligand 1(PD-L1) detection and RNA sequencing of tumor samples in the ORIENT-11 study

PD-L1 tumor proportion scores (TPS) of baseline tumor samples belonging to the ORIENT-11 study were determined by immunohistochemistry using DAKO clone 22C3 pharmDx, as described before [13]. 171 qualified RNA samples were used for Illumina sequencing via the NovaSeq 6000 system. Raw counts were normalized with transcripts per million (TPM) algorithms.

Statistical analysis

All statistical analyses were conducted by R programming language v.3.6.1 and IBM SPSS v.23. Continuous variables were represented as mean ± SD or medians with interquartile ranges, while categorical variables were represented as frequencies. For continuous variables, the Student-t-test or Wilcoxon test was applied to test the differences. Correlation analyses were performed using the Spearman rank correlation test. Survival analysis was quantified using Kaplan-Meier and the difference was assessed by Log-rank test. Univariate and multivariate regression analyses were performed to evaluate the effects of variables on survival. Two-sided P values < 0.05 were considered statistically significant.

RESULTS

TME-associated gene identification based on the ESTIMATE algorithm

Gene expression data of 513 LUAD patients were retrieved from the TCGA database. Among these patients, 504 patients with complete profiles were selected for survival analysis. Patients' baseline characteristics and demographic data are described in Table 1. K-M curves revealed that high SS ($P = 0.021$), high IS ($P = 0.010$), and high ES ($P = 0.016$) were correlated with better OS (Supplementary Figure S1). Female patients had significantly higher IS, SS, and ES ($P < 0.001$) (Supplementary Figure S2C). Patients with advanced TNM stage possessed lower IS ($P < 0.01$) and ES ($P < 0.01$) (Supplementary Figure S2D). There were no significant differences in these three scores among patients under 60 years old and over 60 years old ($P > 0.05$) (Supplementary Figure S2B). According to median scores, we divided patients into high or low IS/SS subgroups. We then compared different expression genes between groups. As to SS, 640 genes were up-regulated and 114 down-regulated in the high-score group. In the case of IS, 690 genes were up-regulated, and 141 genes were down-regulated in the high-score group (Figure 1A). Venn diagrams (Figure 1B) demonstrated that 338 intersection genes were synchronously upregulated in the high IS/SS group, whereas 58 were synchronously down-regulated. Our subsequent analysis was focused on these 396 overlapped genes. Expression patterns were shown as a heatmap (Supplementary Figure S2A), in ascending order by ESTIMATE scores. Colors ranging from blue to red represented the low to high gene expression levels.

GO and KEGG analyses were performed to illuminate biological functions for the 396 intersection genes. Figure 1C showed that these genes were mainly enriched in immune cell activation and cytokine activation. The top KEGG enrichment pathways were "hematopoietic cell lineage", "cytokine-cytokine receptor interaction" and "viral protein interaction with cytokine and cytokine receptor" (Figure 1D).

PPI network and MCODE module analysis

396 genes screened out above were subjected to PPI analysis. 54 isolated genes without interaction were removed (Supplementary Figure S3). Next, module analysis was performed for the remaining 342 genes. After network parsing through the Cytoscape MCODE plug-in, three significant modules with MCODE score >10 were identified, as shown in Figure 2: Module 1; 28 genes with 378 edges; Module 2; 32 genes with 339 edges; Module 3; 39 genes with 276 edges. Therefore, we hypothesized that these total 99 prominent genes might exert relatively significant roles in the LUAD microenvironment.

Screen out prognostic genes and construct a three-gene prognostic signature

99 selected genes were included in the univariate Cox model, and 46 genes were identified to be associated with OS (Figure 3A). After performing LASSO regression analysis of these 46 genes, we finally identified three genes *CCR2, CD40LG, and CCL21*to establish a three-gene prognostic signature (Figure 3B-3C). The prognostic value of *CCR2, CD40LG, and CCL21* was validated individually through the K-M plotter Online Tool, which contained 719 LUAD

Table 1. Clinicopathological characteristics of LUAD training and validation sets

Characteristics	TCGA Training set (n = 504)	GSE 68465 validation set (n = 442)
Age		
≤60	157 (31.2%)	147 (33.3%)
>60	347 (68.8%)	295 (66.7%)
Gender		
Male	234 (46.4%)	223 (50.5%)
Female	270 (53.6%)	219 (49.5%)
Race		
White	388 (77.0%)	294 (66.5%)
Black	52(10.3%)	12 (2.7%)
Asia	7(1.4%)	6 (1.4%)
NA	57(11.3%)	130 (29.4%)
Smoke		
Never	71 (14.1%)	51 (11.5%)
Ever	419 (83.1%)	300 (67.9%)
NA	14 (2.8%)	91 (20.6%)
Stromal Score	151.9 (-340.1-670.4)	
Immune Score	1525.8 (937.7-2094.0)	
Estimate Score	1724.4 (741.8-2697.3)	
T stage		
T1	168 (33.3%)	150 (33.9%)
T2	272 (54.0%)	253 (57.3%)
T3	45 (8.9%)	28 (6.3%)
T4	19 (3.8%)	11 (2.5%)
N stage		
N0	337 (66.9%)	302 (68.3%)
N1	94 (18.7%)	87 (19.7%)
N2	71 (14.0%)	53 (12.0%)
N3	2 (0.4%)	0 (0%)
M stage		
M0	478 (94.8%)	434 (98.2%)
M1	26 (5.2%)	8 (1.8%)
TNM stage		
I	271 (53.8%)	115 (26.0%)
II	125 (24.8%)	257(58.2%)
III	82 (16.3%)	62(14.0%)
IV	26 (5.1%)	8 (1.8%)

TCGA, the cancer genome atlas; GEO, gene expression omnibus

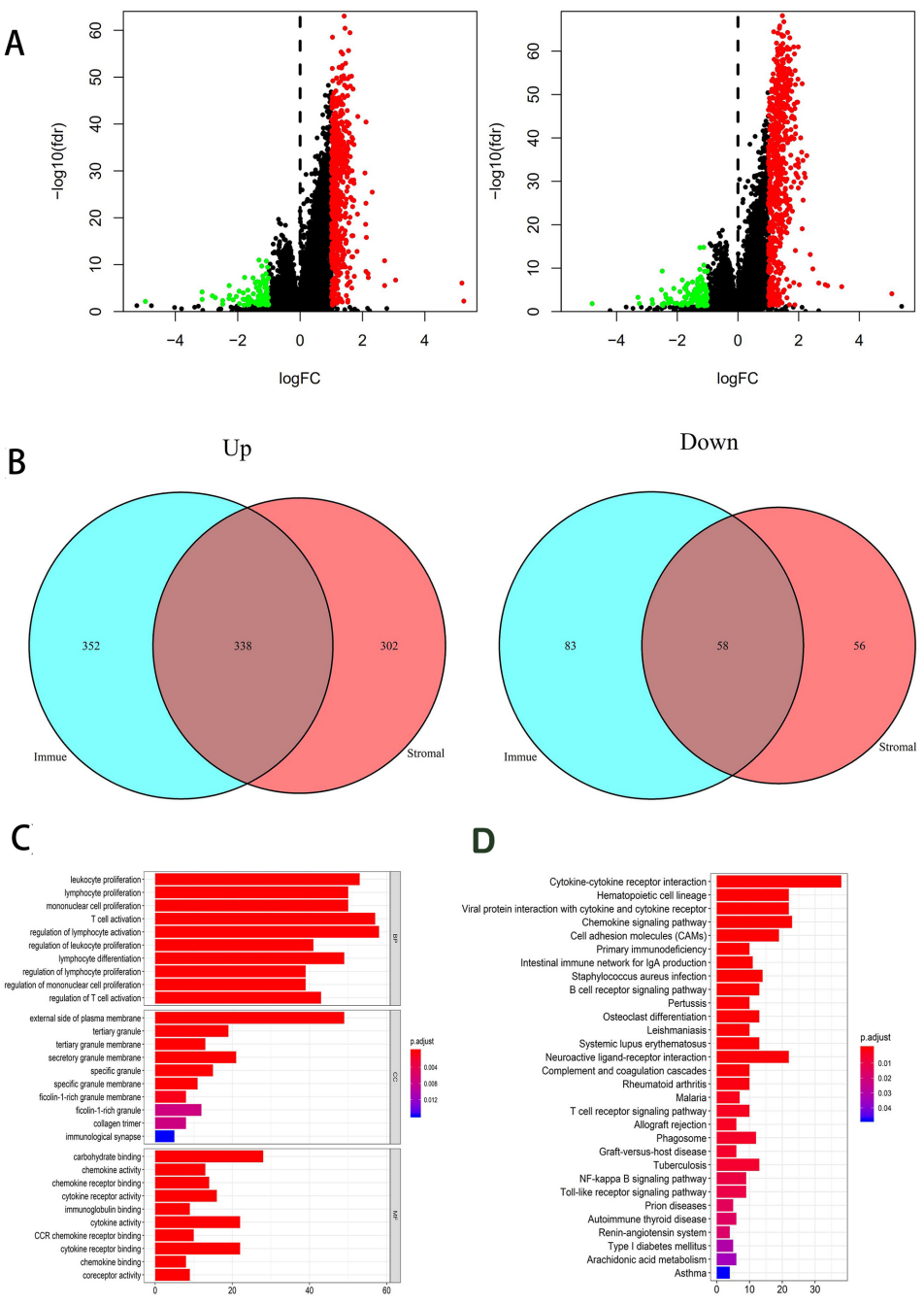

Figure 1: Identification of TME-associated genes based on immune/stromal scores and functional analysis. (A) Volcano plot of differentially expressed TME-associated genes in low versus high immune score/stromal score groups. Significantly differently expressed genes are shown in red (high expression) and green (low expression). **(B)** Venn diagram analysis of intersection genes based on immune and stromal scores (338 up- and 58 down-regulated genes). **(C)** The top 30 significantly enriched GO terms of intersection genes. **(D)** TOP KEGG pathways of intersection genes.

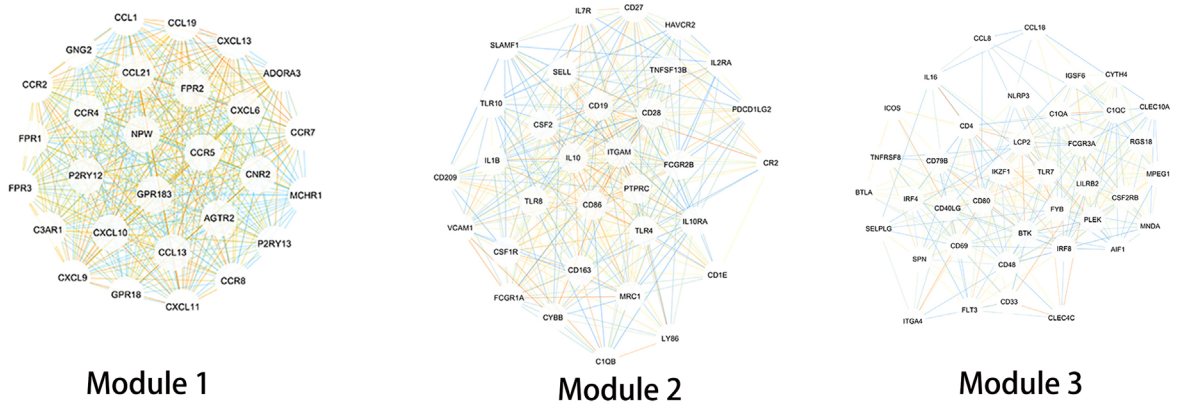

Figure 2: Three significant modules obtained with Cytoscape MCODE plug-in and rendered as separate modules (MCODE scores ≥ 10). Edges are colored gradually from blue to orange in ascending order according to the interaction score.

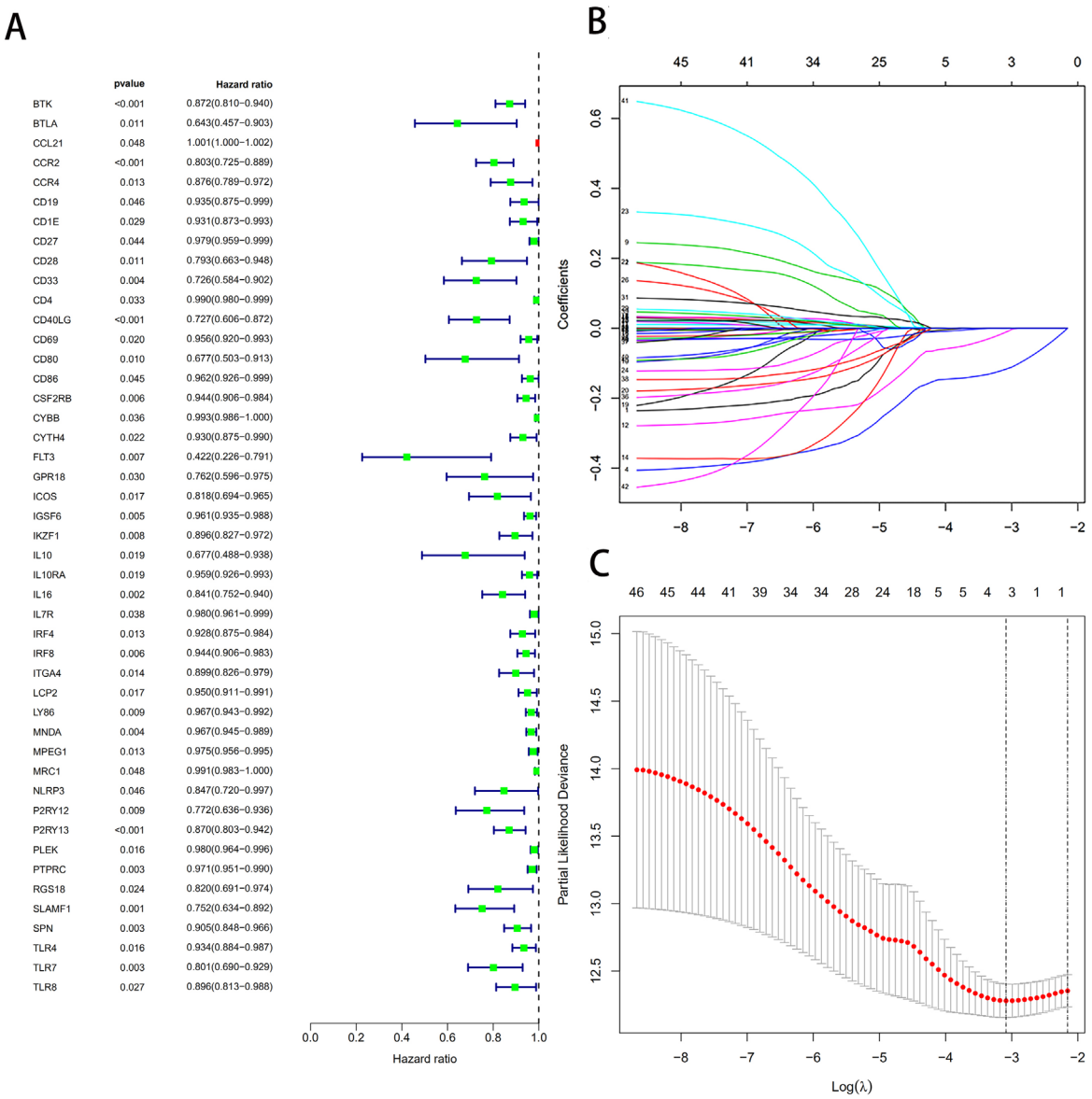

Figure 3: Screening out prognostic genes and constructing a TME-associated signature. (A) 46 prognostic genes were screened out by univariate Cox regression. **(B), (C)** LASSO regression for identifying the most discriminating subset of genes. Three optimal genes were selected.

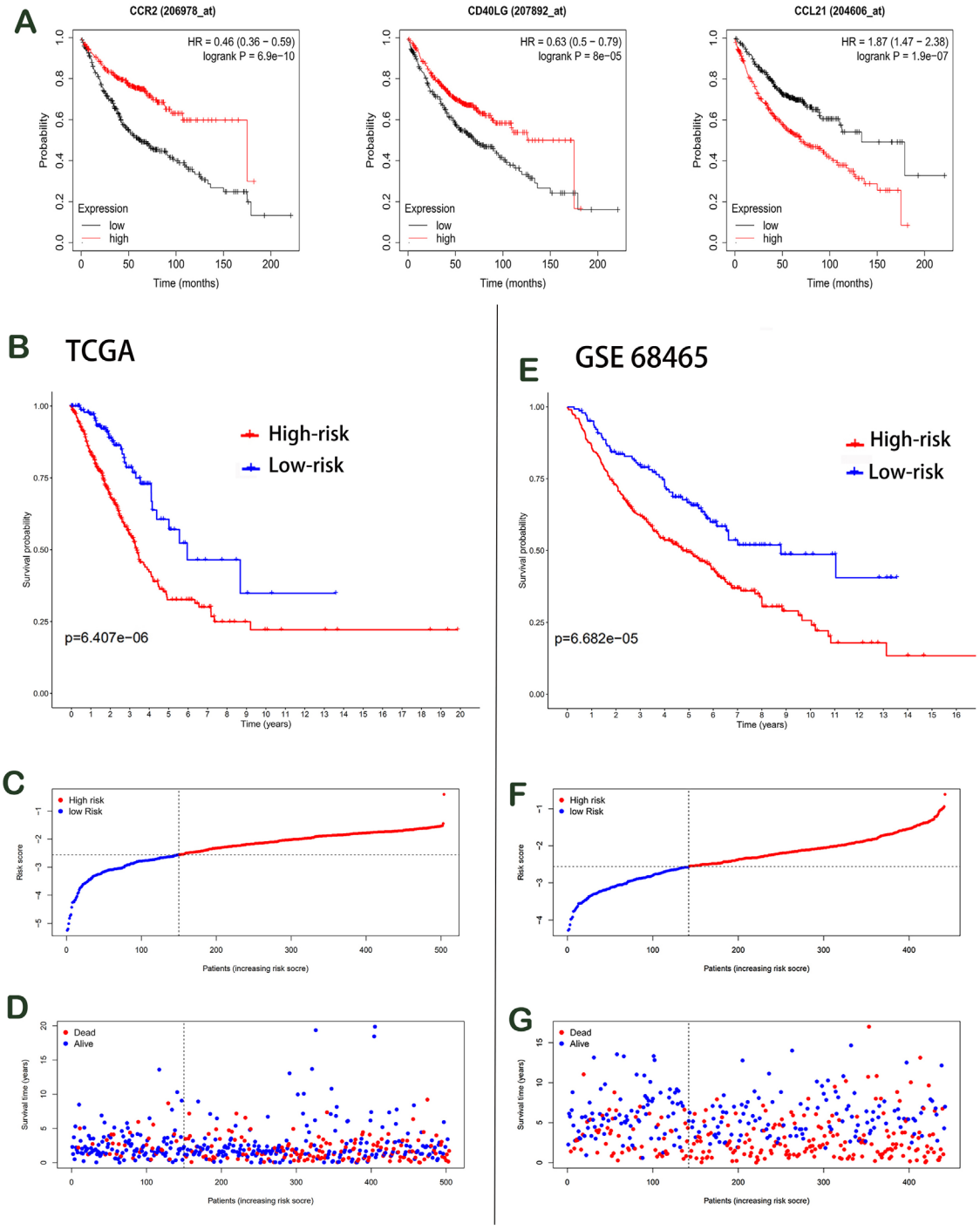

Figure 4: The predictive performance of TME-associated signature containing three genes. (A) Kaplan-Meier plots of OS for *CCR2*, *CD40LG,* and *CCL21* expression generated through an online tool: Kaplan–Meier plotter (http://kmplot.com/analysis/). **(B & E)** Kaplan–Meier curves demonstrated worse survival outcomes for LUAD patients in the high-risk group than their low-risk counterparts. **(C & F)** Survival days of LUAD patients in ascending order of risk parameters. **(D & G)** Survival status scatter plots in each patient.

samples in total with microarray gene chip expression data. As shown in Figure 4A, high expression of *CCR2* and *CD40LG* conferred an improved prognosis, while increased expression of *CCL21* conferred a poor prognosis. Risk scores for all cases were then estimated using these genes and their corresponding COX coefficients using the formula below:

Risk score= -0.289***CCR2*-0.323***CD40LG*+0.001***CCL21*.

As shown in Supplementary Figure S4, a higher risk score was demonstrated in patients with higher TNM stage ($P < 0.001$) and younger age ($P < 0.01$). Female patients tended to possess lower risk scores ($P < 0.001$).

According to ROC curves, -2.56 was determined as the optimal cut-off point for "risk score" to stratify patients into high or low-risk subgroups (Supplementary Figure S5). Patients in the high-risk group had significantly shorter OS time than those in the low-risk group, as shown in K-M curves ($P < 0.001$) (Figure 4B). We further ranked the risk scores of patients and analyzed their distributions (Figure 4C). The survival status was marked as a dot plot (Figure 4D). Furthermore, after taking clinical parameters into account, univariate and multivariate analyses were performed to explore the effect of risk score on OS. As summarized in Table 2, the risk score was significantly associated with OS (HR, 2.29;95%CI:1.56-3.36, $P < 0.001$) in the univariate analysis. After adjusting with T, N, and M tumor stage, in multivariate analysis, risk score remained as a significant prognostic factor (HR, 1.99; 95%CI:1.36-2.93, $P < 0.001$).

Correlation of the three-gene signature and immune indicators

The overall immunological status of each sample was assessed using the ssGSEA approach, through evaluating the expression profiles of 29 immunological signature gene sets provided by He and colleagues [14, 15]. Samples were ranked in ascending order according to risk score, as colors ranging from blue to red represent the low to high score levels. Heatmap showed that 29 ssGSEA scores were disparate in patients with high or low-risk scores (Figure 5A). The risk score was negatively associated with all 29 immune-associated signatures. Corresponding correlation coefficients and P-values are shown in Supplementary Table S1.

The "Diff Exp module" of the TIMER database showed that *CCR2*, *CD40LG,* and *CCL21* were differentially expressed between LUAD and normal tissues ($P < 0.001$) (Supplementary Figure S6). Then, we further confirmed the correlations between our generated signature and the infiltration of immune cells by analyzing LUAD data downloaded from this database. As shown in Supplementary Figure S7, risk score was highly negatively correlated with B-cells ($r = 0.433$, $P < 0.001$), CD8+ T cells ($r = -0.570$, $P < 0.001$), CD4+ T cells ($r = -0.631$, $P < 0.001$), dendritic cells ($r = -0.727$, $P < 0.001$) macrophages ($r = -0.458$, $P < 0.001$) and neutrophils ($r = -0.624$, $P < 0.001$).

Validation of the predictive value of the generated signature in independent cohorts

Clinical and pathological characteristics of the validation cohort GSE68465 are shown in Table 1. Consistent with the training set, high-risk patients possessed poorer prognoses compared with the low-risk ($P < 0.001$) (Figure 4E-4G). Univariate analysis showed that the high-risk score was related to poor prognosis (HR, 1.83;95%CI:1.35-2.47, $P < 0.001$) (Table 2). Multivariate COX analysis further confirmed risk score to be a significantly independent prognostic factor (HR, 1.62;95%CI:1.19-2.20, $P < 0.001$) (Table 2). Additionally, the risk score was negatively related to several infiltrated immunocytes and immune activation (Figure 5B, Supplementary Table S1). Thus, the stability of the TME -TME-associated signature was further confirmed.

The three-gene signature was applicable to predict treatment outcomes for patients receiving siltuximab combined with chemotherapy.

Following quality control, 171 samples from participants in the ORIENT-11 study were suitable for RNA sequencing. 58 patients were treated with standard chemotherapy plus placebo, while the remaining 113 patients received chemotherapy combined with PD-1 inhibitor sintilimab. We calculated the risk score for each patient and classified them into high or low-risk subgroups, according to the calculating formula and threshold described above. Consistent with the result in the TCGA and GSE68465 cohort, in the ORIENT-11 cohort, the risk score was associated with the 29 immune signatures (Figure 6C, Supplementary Table S1). In addition, low-risk group patients implied a higher PD-L1 positivity rate (≥ 1%)

Table 2. Cox regression analysis of risk score and clinical parameters for OS in training and validation cohorts

Variables	Univariate		Multivariate	
	HR (95%CI)	P	HR (95%CI)	P
Training Set				
Age (>60 vs. ≤60)	1.11 (0.81-1.53)	0.510		
Gender (female vs male)	0.94 (0.70-1.26)	0.688		
Smoke (yes vs. no)	0.89 (0.59-1.34)	0.568		
Race (white vs. others)	1.41 (0.86-2.31)	0.173		
T stage (T3-4 vs T1-2)	2.34 (1.60-3.41)	<0.001	1.74 (1.15-2.62)	0.009
N (positive vs negative)	2.55 (1.91-3.43)	<0.001	2.10 (1.49-2.97)	<0.001
M (positive vs negative)	2.27 (1.36-3.80)	0.002	1.60 (0.89-2.87)	0.114
TNM stage (per stage)	1.67 (1.45-1.91)	<0.001	1.24 (0.80-1.91)	0.337
Risk Score (high vs low)	2.29 (1.56-3.36)	<0.001	1.99 (1.36-2.93)	<0.001
Validation Set				
Age (>60 vs ≤ 60)	1.71 (1.28-2.30)	<0.001	1.03 (1.01-1.04)	<0.001
Gender (female vs male)	0.72 (0.55-0.93)	0.013	0.85 (0.65-1.10)	0.213
Smoking (yes vs no)	1.06 (0.92-1.21)	0.401		
Race (white vs others)	1.54 (0.75-3.14)	0.236		
T stage (T3-4 vs T1-2)	2.78 (1.92-4.03)	<0.001	1.69 (1.12-2.56)	0.012
Lymphatic metastasis (positive vs negative)	2.77 (2.14-3.59)	<0.001	2.05(1.48-2.84)	<0.001
Distant metastasis (positive vs negative)	2.67 (1.25-5.67)	0.011	0.65 (0.26-1.66)	0.368
TNM stage (per stage)	2.15 (1.80-2.57)	<0.001	1.61 (1.20-2.15)	0.001
Risk Score (high vs low)	1.83 (1.35-2.47)	<0.001	1.62 (1.19-2.20)	0.002

HR, hazard ratio; CI, confidence interval.

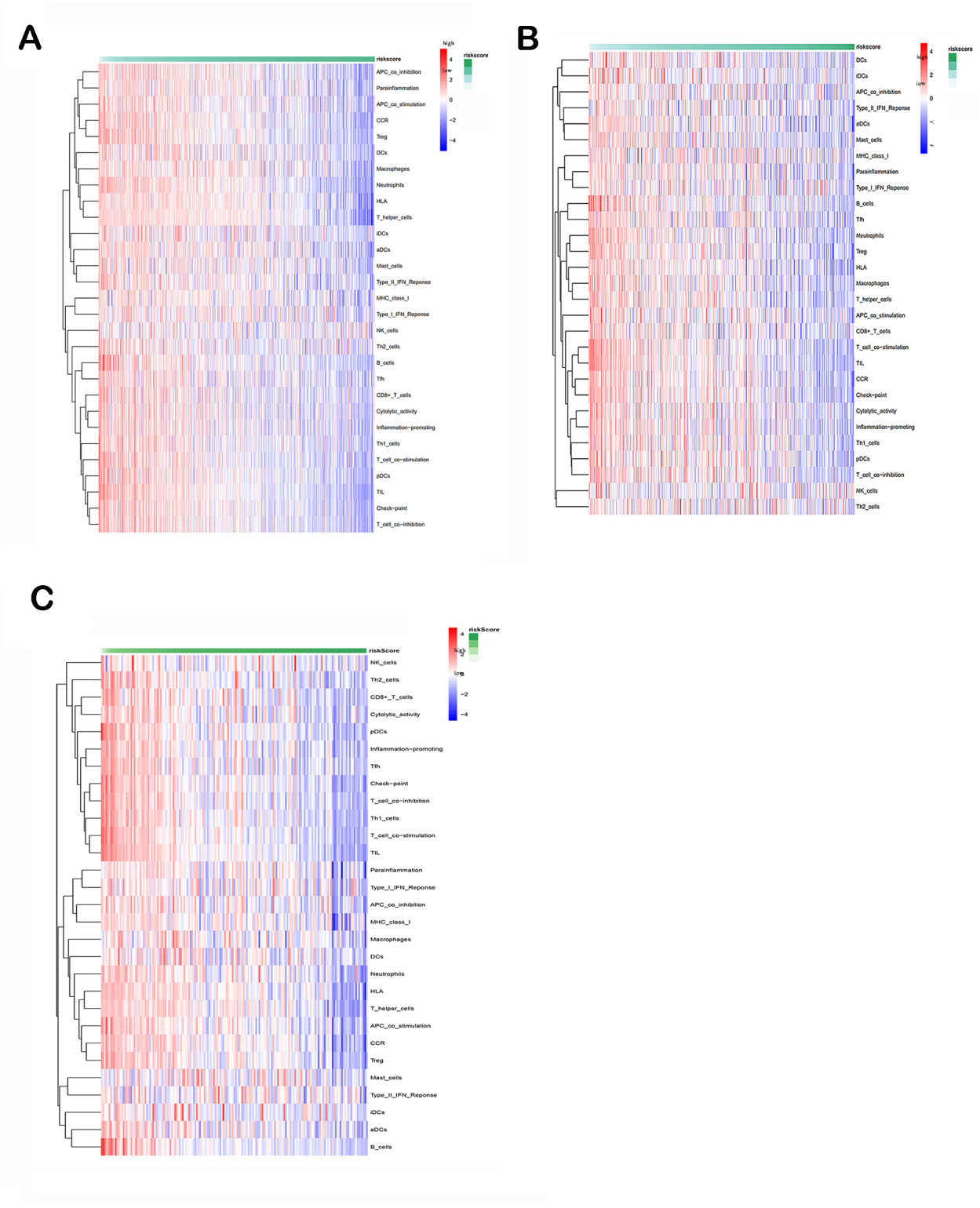

Figure 5: Correlation of the three-gene risk signature and immune status. Heatmap showed the association of risk scores, and immune status analyzed by ssGSEA in TCGA dataset (**A**), GEO validation dataset (**B**), and ORIENT-11 validation dataset (**C**). Samples are ranked in ascending order according to the risk score. Colors ranging from blue to red represented the low to high ssGSEA scores.

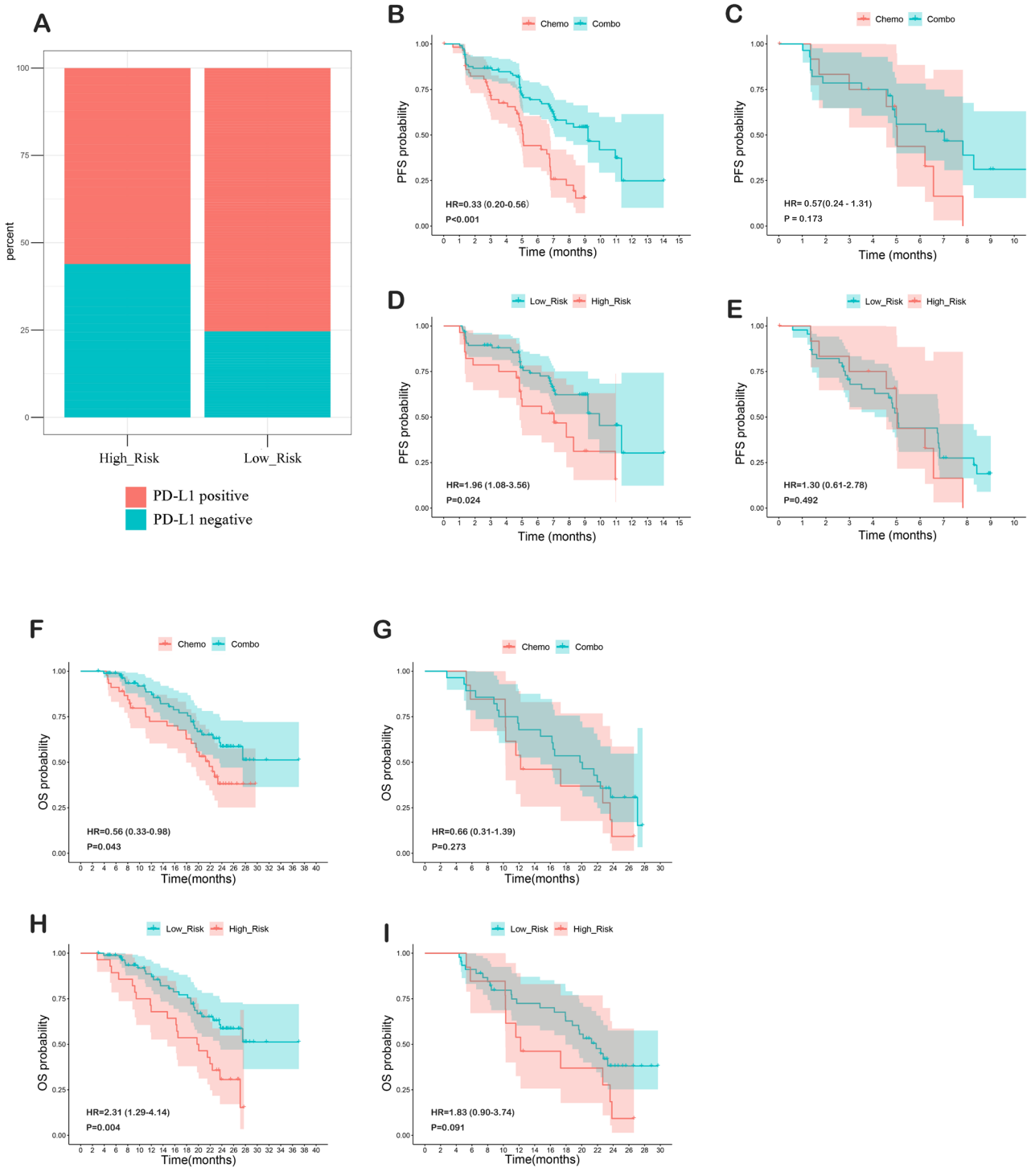

Figure 6: The three-gene signature displayed predictive value for chemo-immunotherapy outcome in patients with LUAD. (A) Low-risk group implied a higher PD-L1 positivity rate (>=1%) than the high-risk group. Patients belonging to a low-risk group (**B & F**) but not those belonging to high-risk groups (**C & G**), benefited from additional sintilimab to chemotherapy. **(D & H)** In the sintilimab plus chemotherapy subset, the low-risk group possessed longer PFS and OS time. **(E & I)** In the chemotherapy subset, patients assigned to the high- or low-risk group displayed similar PFS and OS time. Chemo: chemotherapy group; Combo: combination therapy

compared with those in the high-risk group (75.4% vs. 56.1%, $P = 0.029$) (Figure 6A). As shown in Figure 6B-6F, low-risk group patients showed PFS (HR = 0.33; 95% CI:0.20-0.56, $P < 0.001$) and OS (HR, 0.56;95% CI:0.33-0.98, $P = 0.043$) benefit from combination therapy compared to chemotherapy alone, while in patients with high risk, no benefit was observed (PFS: HR, 0.57; 95%CI:0.24-1.31, $P = 0.173$; OS: HR, 0.66; 95%CI:0.31-1.39, $P = 0.273$) (Figure 6C-6G). Furthermore, in patients who received combination therapy, the high-risk group possessed shorter PFS (HR, 1.96; 95% CI:1.08-3.55, $P = 0.024$) and OS (HR, 2.31;95%CI:1.29-4.14, $P = 0.004$) compared to the low-risk group (Figure 6D-6H). However, in patients treated with chemotherapy alone, no significant difference was seen between high and low-risk groups (PFS: HR, 1.30; 95% CI:0.61-2.78, $P = 0.492$; OS: HR, 1.83; 95%CI:0.90-3.74, $P = 0.091$) (Figure 6E-6I).

Discussion

TME is involved in the development, progression, and relapse of several malignant tumors [5]. As significant components of TME, immune cells, and stromal cells are commonly detected by immunohistochemistry and flow cytometry [16]. However, the utility of both immunohistochemistry and flow cytometry is restricted by the limited number of markers for evaluating diverse immune cells simultaneously [17]. Alternatively, the estimation of immune or stromal cells in the TME using gene expression represents a rational option [7, 18, 19]. Accumulating evidence has demonstrated the prognosis value of ESTIMATE algorithm-based scores in malignancies, although contradicting conclusions have been drawn into consideration as favorable or unfavorable factors [8, 20].

The present study aimed to construct a prognostic gene signature for patients with LUAD. First, we used the ESTIMATE algorithm to assess the composition of TME in LUAD tumor tissue. Results indicated that high SS, IS, and ES scores conferred longer OS. In addition, these scores were associated with multiple clinicopathological features such as gender and TNM stage. Second, we identified intersected genes that were expressed differentially between high and low IS/SS groups, considering these genes were most highly conserved. Functional analysis indicated that these genes were significantly enriched in immune-related processes. Through PPI and module analysis, we further selected the core candidate genes that could better represent TME. After performing COX regression and LASSO regression, a TME-associated signature containing three genes was finally generated, which showed a good prognostic value for survival in patients with LUAD. Furthermore, this gene signature displayed predictive value for treatment outcomes in advanced LUAD patients who received a combination of chemotherapy with an anti-PD-1 inhibitor.

Our three-gene signature included three genes, *CCR2,* the primary cytokine receptor responsible for monocyte trafficking expressed in monocytes and liver macrophages. A recent report showed that this gene is associated with inflammatory activation and tumor environment. However whether it has a pro-tumoral or antitumoral function remained unclear [21, 22]. *CD40LG* encoded CD40 ligand, found in the surface of immune cells, is primarily expressed by activated CD4+ T cells [23]. In addition, *CD40LG* mediates several essential functions in the regulation of cellular and humoral immunity [24]. Much of the works of literature on *CD40LG* were focused on its association with several autoimmune diseases including systemic lupus erythematosus and rheumatoid arthritis [23, 25, 26]. However, the relationship between *CD40LG* and tumors has been rarely reported. *CCL21* seemed to assume double roles in tumorigenesis and development. Vitro experiments found that overexpression of *CCL21* suppressed tumorigenesis [27, 28]. It also plays an important role in immune tolerance and downregulation of inflammatory responses [29, 30]. This explained why *CCL21* exhibited low expression in tumor tissue compared with normal tissue. Nevertheless, our study showed that high *CCL21* was associated with a worse prognosis in LUAD patients. This may be because immune tolerance generally means immune response downregulation. Higher *CCL21* signifies a relatively low anti-tumor immune response. Hence, LUAD patients with high *CCL21* may be more likely to develop recurrence after operation and less likely to respond to immunotherapy, thus worse prognosis would be obtained.

Immunotherapy, especially immune checkpoint inhibitors (ICIs) therapy, has marked a revolution in lung cancer treatment [31, 32]. The so-called immune phenotypes, which were classified based on TME, have attracted wide attention for their association with treatment response to ICIs [33, 34]. Tumors with a higher degree of immune infiltrate, particularly tumor-infiltrating lymphocytes (TILs), are known

as the "hot tumors" which are thought to respond better to ICIs therapy, whereas tumors with a lower level of immune infiltrate are referred to as "non-inflamed" or "cool tumors", which are characterized to be less efficacious to ICIs therapy. As indicated in our study, our constructed three-gene signature was associated with several immune indicators. Patients in the high-risk group tended to have less tumor-infiltrating immune cells and less active inflammatory response. Thus, we considered this signature might serve as a promising biomarker for guiding the application of immunotherapy. Hence, we made preliminary explorations of its potential in predicting the outcomes of LUAD checkpoint inhibitor therapy, based on 171 available RNA-seq profiles derived from the ORIENT-11 study cohort. Results further validated its association with immune indicators, such as 29 immune signatures calculated by ssGSEA as well as PD-L1. As we know, effective biomarkers for identifying the most beneficiary population to receive immune checkpoint inhibitor therapy combined with chemotherapy are still lacking [35]. Our current study indicated that only patients in the low-risk group, but not those in the high-risk group, gained benefit from the addition of anti-PD-1 to chemotherapy. Moreover, this signature was able to predict PFS and OS in the combination therapy subgroup population, but not in the chemotherapy alone subgroup population. These results suggested the potential of this three-gene signature in distinguishing appropriate candidates to adopt additional immune checkpoint inhibitors to standard chemotherapy.

Currently, Cox regression remains the most popular approach in establishing prognostic models, which plays an important role in guiding clinical decision-making [36, 37]. The preliminary attempts at artificial intelligence like random forest, neural networks have gradually been made [38, 39]. However, they act like black boxes with inscrutable inner structures, which lead to obscure clinical interpretability [40, 41]. In our current study, COX regression acted as a basis in model construction, which was easy to interpret and apply to clinical scenarios. To avoid overfitting, LASSO regression was utilized [42]. This model performed well in both the training set and the validation set. Hence, we consider that our model is rational.

Our present study has some limitations. Firstly, this study was retrospective, clinical characteristics of the included samples, especially those derived from publicly available datasets were not comprehensive, which might cause some bias. Secondly, various TME-associated signatures have been reported in published literature [43-45]. The model discrepancy may mainly be due to different approaches in constructing models. Due to the lack of space, it is unrealistic to comprehensively assess them. Large real-world clinical datasets or system assessments like meta-analysis may be feasible ways to address it.

Conclusions

The present study performed a comprehensive analysis of gene expression profiles in patients with LUAD, based on ESTIMATE algorithm-based scores. We established a TME-associated signature containing three genes that could stratify patients with different survival outcomes. This signature demonstrated a close association with multiple immune indicators. It also displayed promising potential to be a biomarker for predicting anti-PD1 therapy benefits. The underlying molecular mechanism warranted further investigation to gain additional insight.

References

1. Dela Cruz CS, Tanoue LT, Matthay RA: **Lung cancer: epidemiology, etiology, and prevention**. *Clin Chest Med* 2011, **32**(4):605-644; doi:10.1016/j.ccm.2011.09.001; PMC3864624.
2. Detterbeck FC, Chansky K, Groome P, Bolejack V, Crowley J, Shemanski L, Kennedy C, Krasnik M, Peake M, Rami-Porta R *et al*: **The IASLC Lung Cancer Staging Project: Methodology and Validation Used in the Development of Proposals for Revision of the Stage Classification of NSCLC in the Forthcoming (Eighth) Edition of the TNM Classification of Lung Cancer**. *J Thorac Oncol* 2016, **11**(9):1433-1446; doi:10.1016/j.jtho.2016.06.028.
3. Hanahan D, Coussens LM: **Accessories to the crime: functions of cells recruited to the tumor microenvironment**. *Cancer Cell* 2012, **21**(3):309-322; doi:10.1016/j.ccr.2012.02.022.
4. Barnes TA, Amir E: **HYPE or HOPE: the prognostic value of infiltrating immune cells in cancer**. *Br J Cancer* 2017, **117**(4):451-460; doi:10.1038/bjc.2017.220; PMC5558691.
5. Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, Gong Z, Zhang S, Zhou J, Cao K *et al*: **Role of tumor microenvironment in tumorigenesis**. *J Cancer* 2017, **8**(5):761-773; doi:10.7150/jca.17648; PMC5381164.

6. Li HY, McSharry M, Bullock B, Nguyen TT, Kwak J, Poczobutt JM, Sippel TR, Heasley LE, Weiser-Evans MC, Clambey ET *et al*: **The Tumor Microenvironment Regulates Sensitivity of Murine Lung Tumors to PD-1/PD-L1 Antibody Blockade**. *Cancer Immunol Res* 2017, **5**(9):767-777; doi:10.1158/2326-6066.CIR-16-0365; PMC5787226.
7. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, *et al*: **Inferring tumor purity and stromal and immune cell admixture from expression data**. *Nat Commun* 2013, **4**:2612; doi:10.1038/ncomms3612; PMC3826632.
8. Pan XB, Lu Y, Huang JL, Long Y, Yao DS: **Prognostic genes in the tumor microenvironment in cervical squamous cell carcinoma**. *Aging (Albany NY)* 2019, **11**doi:10.18632/aging.102429.
9. Yang S, Liu T, Nan H, Wang Y, Chen H, Zhang X, Zhang Y, Shen B, Qian P, Xu S *et al*: **Comprehensive analysis of prognostic immune-related genes in the tumor microenvironment of cutaneous melanoma**. *J Cell Physiol* 2020, **235**(2):1025-1035; doi:10.1002/jcp.29018.
10. Li S, Xu W: **Mining TCGA database for screening and identification of hub genes in kidney renal clear cell carcinoma microenvironment**. *J Cell Biochem* 2019doi:10.1002/jcb.29511.
11. Chen J, Hou C, Wang P, Yang Y, Zhou D: **Grade II/III Glioma Microenvironment Mining and Its Prognostic Merit**. *World Neurosurg* 2019doi:10.1016/j.wneu.2019.08.253.
12. Yan H, Qu J, Cao W, Liu Y, Zheng G, Zhang E, Cai Z: **Identification of prognostic genes in the acute myeloid leukemia immune microenvironment based on TCGA data analysis**. *Cancer Immunol Immunother* 2019, **68**(12):1971-1978; doi:10.1007/s00262-019-02408-7.
13. Yang Y, Wang Z, Fang J, Yu Q, Han B, Cang S, Chen G, Mei X, Yang Z, Ma R *et al*: **Efficacy and Safety of Sintilimab Plus Pemetrexed and Platinum as First-Line Treatment for Locally Advanced or Metastatic Nonsquamous NSCLC: a Randomized, Double-Blind, Phase 3 Study (Oncology pRogram by InnovENT anti-PD-1-11)**. *J Thorac Oncol* 2020, **15**(10):1636-1646; doi:10.1016/j.jtho.2020.07.014.
14. Hanzelmann S, Castelo R, Guinney J: **GSVA: gene set variation analysis for microarray and RNA-seq data**. *BMC Bioinformatics* 2013, **14**:7; doi:10.1186/1471-2105-14-7; PMC3618321.
15. He Y, Jiang Z, Chen C, Wang X: **Classification of triple-negative breast cancers based on Immunogenomic profiling**. *J Exp Clin Cancer Res* 2018, **37**(1):327; doi:10.1186/s13046-018-1002-1; PMC6310928.
16. Rohr-Udilova N, Klinglmuller F, Schulte-Hermann R, Stift J, Herac M, Salzmann M, Finotello F, Timelthaler G, Oberhuber G, Pinter M *et al*: **Deviations of the immune cell landscape between healthy liver and hepatocellular carcinoma**. *Sci Rep* 2018, **8**(1):6220; doi:10.1038/s41598-018-24437-5; PMC5906687.
17. Zhou R, Zhang J, Zeng D, Sun H, Rong X, Shi M, Bin J, Liao Y, Liao W: **Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer**. *Cancer Immunol Immunother* 2019, **68**(3):433-442; doi:10.1007/s00262-018-2289-7; PMC6426802.
18. 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; doi:10.1038/nmeth.3337; PMC4739640.
19. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S *et al*: **Comprehensive analyses of tumor immunity: implications for cancer immunotherapy**. *Genome Biol* 2016, **17**(1):174; doi:10.1186/s13059-016-1028-7; PMC4993001.
20. 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; doi:10.18632/aging.101415; PMC5940130.
21. Maximov V, Chen Z, Wei Y, Robinson MH, Herting CJ, Shanmugam NS, Rudneva VA, Goldsmith KC, MacDonald TJ, Northcott PA *et al*: **Tumour-associated macrophages exhibit anti-tumoural properties in Sonic Hedgehog medulloblastoma**. *Nat Commun* 2019, **10**(1):2410; doi:10.1038/s41467-019-10458-9; PMC6546707.
22. Flores CT, Wildes TJ, Drake JA, Moore GL, Dean BD, Abraham RS, Mitchell DA: **Lin(-)CCR2(+) hematopoietic stem and progenitor cells overcome resistance to PD-1 blockade**. *Nat Commun* 2018, **9**(1):4313; doi:10.1038/s41467-018-06182-5; PMC6192988.
23. Roman-Fernandez IV, Sanchez-Zuno GA, Padilla-Gutierrez JR, Cerpa-Cruz S, Hernandez-Bel-

lo J, Valle Y, Ramirez-Duenas MG, Carrillo C, Munoz-Valle JF: **The 3'-UTR (CA)n microsatellite on CD40LG gene as a possible genetic marker for rheumatoid arthritis in Mexican population: impact on CD40LG mRNA expression**. *Clin Rheumatol* 2018, **37**(2):345-353; doi:10.1007/s10067-017-3853-9.

24. Elgueta R, Benson MJ, de Vries VC, Wasiuk A, Guo Y, Noelle RJ: **Molecular mechanism and function of CD40/CD40L engagement in the immune system**. *Immunol Rev* 2009, **229**(1):152-172; doi:10.1111/j.1600-065X.2009.00782.x; PMC3826168.
25. Lee SH, Lee EB, Shin ES, Lee JE, Cho SH, Min KU, Park HW: **The Interaction Between Allelic Variants of CD86 and CD40LG: A Common Risk Factor of Allergic Asthma and Rheumatoid Arthritis**. *Allergy Asthma Immunol Res* 2014, **6**(2):137-141; doi:10.4168/aair.2014.6.2.137; PMC3936042.
26. Citores MJ, Rua-Figueroa I, Rodriguez-Gallego C, Durantez A, Garcia-Laorden MI, Rodriguez-Lozano C, Rodriguez-Perez JC, Vargas JA, Perez-Aciego P: **The dinucleotide repeat polymorphism in the 3'UTR of the CD154 gene has a functional role on protein expression and is associated with systemic lupus erythematosus**. *Ann Rheum Dis* 2004, **63**(3):310-317; doi:10.1136/ard.2003.006148; PMC1754911.
27. Vicari AP, Ait-Yahia S, Chemin K, Mueller A, Zlotnik A, Caux C: **Antitumor effects of the mouse chemokine 6Ckine/SLC through angiostatic and immunological mechanisms**. *J Immunol* 2000, **165**(4):1992-2000; doi:10.4049/jimmunol.165.4.1992.
28. Nakayama A, Roquid KA, Iring A, Strilic B, Gunther S, Chen M, Weinstein LS, Offermanns S: **Suppression of CCL2 angiocrine function by adrenomedullin promotes tumor growth**. *J Exp Med* 2023, **220**(1)doi:10.1084/jem.20211628; PMC9665902.
29. Shields JD, Kourtis IC, Tomei AA, Roberts JM, Swartz MA: **Induction of lymphoidlike stroma and immune escape by tumors that express the chemokine CCL21**. *Science* 2010, **328**(5979):749-752; doi:10.1126/science.1185837.
30. Forster R, Davalos-Misslitz AC, Rot A: **CCR7 and its ligands: balancing immunity and tolerance**. *Nat Rev Immunol* 2008, **8**(5):362-371; doi:10.1038/nri2297.
31. Gettinger SN, Horn L, Gandhi L, Spigel DR, Antonia SJ, Rizvi NA, Powderly JD, Heist RS, Carvajal RD, Jackman DM, *et al*: **Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer**. *J Clin Oncol* 2015, **33**(18):2004-2012; doi:10.1200/JCO.2014.58.3708; PMC4672027.
32. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, Patnaik A, Aggarwal C, Gubens M, Horn L *et al*: **Pembrolizumab for the treatment of non-small-cell lung cancer**. *N Engl J Med* 2015, **372**(21):2018-2028; doi:10.1056/NEJMoa1501824.
33. Maleki Vareki S: **High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors**. *J Immunother Cancer* 2018, **6**(1):157; doi:10.1186/s40425-018-0479-7; PMC6307306.
34. Bonaventura P, Shekarian T, Alcazer V, Valladeau-Guilemond J, Valsesia-Wittmann S, Amigorena S, Caux C, Depil S: **Cold Tumors: A Therapeutic Challenge for Immunotherapy**. *Front Immunol* 2019, **10**:168; doi:10.3389/fimmu.2019.00168; PMC6376112.
35. Yan Y, Kumar AB, Finnes H, Markovic SN, Park S, Dronca RS, Dong H: **Combining Immune Checkpoint Inhibitors With Conventional Cancer Therapy**. *Front Immunol* 2018, **9**:1739; doi:10.3389/fimmu.2018.01739; PMC6072836.
36. Guinney J, Wang T, Laajala TD, Winner KK, Bare JC, Neto EC, Khan SA, Peddinti G, Airola A, Pahikkala T *et al*: **Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: development of a prognostic model through a crowdsourced challenge with open clinical trial data**. *Lancet Oncol* 2017, **18**(1):132-142; doi:10.1016/S1470-2045(16)30560-5; PMC5217180.
37. Halabi S, Lin CY, Kelly WK, Fizazi KS, Moul JW, Kaplan EB, Morris MJ, Small EJ: **Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer**. *J Clin Oncol* 2014, **32**(7):671-677; doi:10.1200/JCO.2013.52.3696; PMC3927736 are found at the end of this article.

38. Giner-Galvan V, Pomares-Gomez FJ, Quesada JA, Rubio-Rivas M, Tejada-Montes J, Baltasar-Corral J, Taboada-Martinez ML, Sanchez-Mesa B, Arnalich-Fernandez F, Del Corral-Beamonte E *et al*: **C-Reactive Protein and Serum Albumin Ratio: A Feasible Prognostic Marker in Hospitalized Patients with COVID-19**. *Biomedicines* 2022, **10**(6)doi:10.3390/biomedicines10061393; PMC9219981.
39. Gammall J, Lai AG: **Pan-cancer prognostic genetic mutations and clinicopathological factors associated with survival outcomes: a systematic review**. *NPJ Precis Oncol* 2022, **6**(1):27; doi:10.1038/s41698-022-00269-5; PMC9021198.
40. Morse S, Talty K, Kuiper P, Scioletti M, Heymsfield SB, Atkinson RL, Thomas DM: **Machine learning prediction of combat basic training injury from 3D body shape images**. *PLoS One* 2020, **15**(6):e0235017; doi:10.1371/journal.pone.0235017; PMC7326186.
41. Wang KYX, Pupo GM, Tembe V, Patrick E, Strbenac D, Schramm SJ, Thompson JF, Scolyer RA, Muller S, Tarr G, *et al*: **Cross-Platform Omics Prediction procedure: a statistical machine learning framework for wider implementation of precision medicine**. *NPJ Digit Med* 2022, **5**(1):85; doi:10.1038/s41746-022-00618-5; PMC9253123 Evaxion, Provectus Biopharmaceuticals Australia, Qbiotics, Novartis, MSD Sharp & Dohme, NeraCare, AMGEN., Bristol-Myers Squibb, Myriad Genetics, GlaxoSmithKline. J.F.T. has received honoraria for advisory board participation from BMS Australia, MSD Australia, GlaxoSmithKline, and Provectus Inc., and travel support from GlaxoSmithKline and Provectus Inc. All other authors declare no competing interests.
42. Tibshirani R: **The lasso method for variable selection in the Cox model**. *Stat Med* 1997, **16**(4):385-395; doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3.
43. Xu F, Chen JX, Yang XB, Hong XB, Li ZX, Lin L, Chen YS: **Analysis of Lung Adenocarcinoma Subtypes Based on Immune Signatures Identifies Clinical Implications for Cancer Therapy**. *Mol Ther Oncolytics* 2020, **17**:241-249; doi:10.1016/j.omto.2020.03.021; PMC7183104.
44. Li N, Wang J, Zhan X: **Identification of Immune-Related Gene Signatures in Lung Adenocarcinoma and Lung Squamous Cell Carcinoma**. *Front Immunol* 2021, **12**:752643; doi:10.3389/fimmu.2021.752643; PMC8649721.
45. Zhang P, Pei S, Liu J, Zhang X, Feng Y, Gong Z, Zeng T, Li J, Wang W: **Cuproptosis-related lncRNA signatures: Predicting prognosis and evaluating the tumor immune microenvironment in lung adenocarcinoma**. *Front Oncol* 2022, **12**:1088931; doi:10.3389/fonc.2022.1088931; PMC9887198.