

# Development and validation of a tumor microenvironment-related prognostic signature in lung adenocarcinoma and immune infiltration analysis\*

Zhou Li<sup>1</sup>, Yanqi Feng<sup>1</sup>, Piao Li<sup>1</sup>, Shennan Wang<sup>1</sup>, Ruichao Li<sup>2</sup>, Shu Xia<sup>1</sup> (✉)

<sup>1</sup> Department of Oncology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

<sup>2</sup> Department of Geriatric, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

## Abstract

**Objective** Tumor-infiltrating immune cells and stromal cells in the tumor microenvironment (TME) significantly affect the prognosis of and immune response to lung adenocarcinoma (LUAD). In this study, we aimed to develop a novel TME-related prognostic model based on immune and stromal genes in LUAD.

**Methods** LUAD data from the TCGA database were used as the training cohort, and three Gene Expression Omnibus (GEO) datasets were used as the testing cohort. The Estimation of STromal and Immune cells in Malignant Tumor tissues using Expression data algorithm was used to analyze the immune and stromal genes involved in the TME. Kaplan-Meier and Cox regression analyses were used to identify prognostic genes and construct a TME-related prognostic model. Gene set enrichment analysis and TIMER were used to analyze the immune features and signaling pathways of the model.

**Results** A TME-related prognostic model based on six hub genes was generated that significantly stratified patients into the high- and low-risk groups in terms of overall survival. The model had strong predictive ability in both the training (TCGA) and testing (GEO) datasets and could serve as an independent prognostic factor for LUAD. Moreover, the low-risk group was characterized by greater immune cell infiltration and antitumor immune activity than the high-risk group. Importantly, the signature was closely associated with immune checkpoint molecules, which may serve as a predictor of patient response to immunotherapy. Finally, the hub genes BTK, CD28, INHA, PIK3CG, TLR4, and VEGFD were considered novel prognostic biomarkers for LUAD and were significantly correlated with immune cells.

**Conclusion** The TME-related prognostic model could effectively predict the prognosis and reflect the TME status of LUAD. These six hub genes provided novel insights into the development of new therapeutic strategies.

**Key words:** lung adenocarcinoma; tumor microenvironment; immunotherapy; immune checkpoint molecules; prognostic biomarkers

Received: 19 December 2021

Revised: 25 December 2021

Accepted: 29 December 2021

Lung cancer is the leading cause of cancer-related deaths worldwide, with a predicted 5-year survival rate of 16%<sup>[1]</sup>. More than 85% of cases are classified as non-small cell lung cancer (NSCLC), with lung adenocarcinoma (LUAD) being the most common pathological subtype<sup>[2]</sup>. In recent decades, the discovery of driver gene mutations in tumors has allowed for the introduction of personalized

molecular-targeted therapy for NSCLC<sup>[3]</sup>. However, this approach is not feasible for treating tumors that do not carry gene alterations, and the inevitable resistance to tyrosine kinase inhibitors further suggests the need for alternative therapeutic options in lung cancer patients<sup>[4]</sup>. In recent decades, immunotherapy targeting immune checkpoints has made great progress in the treatment of

✉ Correspondence to: Shu Xia. Email: xiashutj@hotmail.com

\*Supported by grants from the National Natural Science Foundation of China (No.81772471 and 82172716).

© 2021 Huazhong University of Science and Technology

NSCLC<sup>[5]</sup>. Immune checkpoint inhibitors (ICIs) enhance the antitumor effects of the immune system to obtain a potent and durable cure<sup>[6]</sup>. However, the overall response rate to ICIs is relatively low, and only one-fifth of cancer patients benefit from these agents<sup>[7]</sup>. Therefore, it is necessary to identify novel biomarkers for predicting LUAD patient survival and response to ICI therapy.

A growing body of evidence has demonstrated the importance of the tumor microenvironment (TME) in oncogenesis and tumors<sup>[8]</sup>. The TME is a complex network composed of tumor cells, immune cells, mesenchymal stem cells, fibroblast cells, endothelial cells, inflammatory mediators, and extracellular matrix<sup>[9]</sup>. The interactions between tumor cells and their surrounding supporting cells contribute to the malignant biological behaviors of cancer, such as unlimited proliferation, resistance to apoptosis, and evasion of immune surveillance<sup>[10]</sup>. Therefore, the TME significantly affects the therapeutic response to and clinical outcomes of patients with cancer. The major non-tumor components of the TME, tumor-infiltrating immune cells and stromal cells have been proposed to be valuable for the diagnostic and prognostic assessment of patients with tumors<sup>[11, 12]</sup>. The development of a comprehensive model of the TME based on immune and stromal signature genes may contribute to the prognostic evaluation of LUAD patients and predict the efficacy of immunotherapy. With advancements in sequencing techniques, bioinformatics tools such as Estimation of STromal and Immune cells in Malignant Tumor tissues using Expression data (ESTIMATE) and CIBERSORT, make it feasible to estimate the distribution of immune and stromal cells in the TME by analyzing specific gene expression signatures of immune and stromal cells<sup>[13]</sup>. This algorithm has been successfully applied to quantitative analysis of the TME in various tumors and the identification of immune and stromal genes involved in the TME, and its effectiveness has been proven<sup>[14]</sup>.

To date, several predictive models have been constructed for LUAD prognosis stratification, mainly focusing on immune-related genes or immune cells<sup>[15]</sup>. However, few studies have investigated the influence of TME on LUAD patient survival outcomes and response to immunotherapy, specifically based on immune and stromal components. To fill this knowledge gap, we aimed to develop a TME prognostic model based on immune and stromal genes to predict the survival outcomes and immune responses. In the present study, we systematically investigated the expression details and clinical significance of immune and stromal genes in the TME of LUAD and developed a novel TME-related prognostic model. In addition, we validated this model using independent datasets and analyzed its potential prognostic mechanism and association with immunotherapy responses. Our findings provide promising biomarkers for the prognostic

stratification and selection of patients responsive to ICIs, which would facilitate accurate management and appropriate personalized therapies for patients with LUAD.

## Materials and methods

### Data source and preprocessing

The gene expression profiles of 594 LUAD case were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>), along with their corresponding clinical and survival data. Datasets GSE26939, GSE37745, and GSE29016, which contained microarray expression data and clinical information of LUAD patients, were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). In our study, data from TCGA were used as the training cohort, whereas the three GEO datasets were used as validation cohorts.

### Generation of the immune score and stromal score

ESTIMATE is an algorithm for estimating the infiltration of immune and stromal cells in tumor samples by analyzing the specific gene expression signatures of immune and stromal cells. Here, we calculated the immune and stromal scores to predict the proportion of immune and stromal components in each sample using the ESTIMATE algorithm with the aid of the R software estimate package.

### Identification of differentially expressed genes related to the TME and functional enrichment analysis

All patients in the training cohort were divided into high and low immune/stromal score groups according to the median immune and stromal scores, respectively. Kaplan–Meier analysis was conducted to compare the survival difference between the two groups, and the p-value of the log-rank test was calculated. The limma package was used to identify differentially expressed genes (DEGs) between the high and low immune/stromal score groups with a fold change (FC) =1 and false discovery rate (FDR) < 0.05. DEGs between the high and low immune score groups were defined as immune DEGs, whereas the DEGs between the high and low stromal score groups were defined as stromal DEGs. Finally, the intersecting genes between the immune and stromal DEGs were considered for subsequent analysis. Heatmaps of DEGs were generated to show expression differences using the heatmap package heatmap. Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of intersecting DEGs were performed using the

clusterProfiler, enrichplot, and ggplot2 packages. Only terms with *p*- and *q*-values of < 0.05 were considered significantly enriched. Moreover, we downloaded a list of immune-related genes from the Immunology Database and Analysis Portal (Immport) to select immune-related DEGs among these DEGs using the VennDiagram package.

### **Construction and evaluation of the TME-related prognostic model in the training set**

Univariate Cox and Kaplan–Meier analyses were performed in the training cohort to identify significant prognostic genes among the immune-related DEGs. A *P* value < 0.05 in the log-rank test was considered significant. Multivariate Cox regression analysis was performed to obtain the respective coefficients ( $\beta_i$ ) of each gene. Finally, the TME prognostic model was constructed on the basis of the key prognostic genes, and the risk score of each patient was calculated on the basis of the expression level of each key prognostic gene and its regression coefficient.

Kaplan–Meier and receiver operating characteristic (ROC) analyses were used to assess the accuracy of the model in predicting clinical outcomes. Univariate and multivariate Cox regression analyses were performed to evaluate the prognostic value of the model and other common clinical factors such as age, sex, stage, and TNM stage.

### **Validation of the TME-related prognostic model in the testing set**

The feasibility and stability of the TME-related prognostic model were confirmed using the GEO database model. Patients in the three testing datasets were divided into the high- and low-risk groups according to the formula of risk score derived from the training dataset using the same methods as above. Kaplan–Meier survival analysis and ROC curve analysis were used to evaluate the performance of the six-gene prognostic model in predicting the outcomes of patients with LUAD.

### **Evaluation of immune status between the high-risk and low-risk groups stratified by prognostic model**

To explore the potential mechanism of the prognostic effects of the model, we analyzed the immune status and pathway enrichment of high-risk and low-risk samples. First, we quantified the enrichment levels of the 29 immune signatures in each LUAD sample using single-sample gene set enrichment analysis (ssGSEA) score. Based on the ssGSEA score, we performed a hierarchical clustering analysis to compare immune activities between the high-risk and low-risk samples. CIBERSORT is an algorithm used for estimating the proportion of immune

cell subsets through cell type identification by estimating the relative subsets of RNA transcripts. In this study, we used the CIBERSORT algorithm to construct 21 types of immune cell profiles in LUAD samples and compared the differences in immune cell subtypes between the high- and low-risk groups. KEGG enrichment analysis was performed to analyze the functions or pathways that were upregulated in the two groups. Finally, we compared the mRNA levels of immune checkpoints and their ligands and the expression of HLA genes in the high- and low-risk groups.

### **Comprehensive analysis of prognostic hub genes in the model**

To reveal the regulatory mechanisms of the prognostic hub genes in the TME, we systematically analyzed the genetic alterations and functional enrichment of these six genes. First, RNA expression and gene-encoding protein expression level alterations in LUAD compared with normal tissue were estimated by the Wilcoxon test and immunohistochemistry (IHC), respectively. IHC results for hub genes were obtained from the Human Protein Atlas (HPA) database. The String online database and Cytoscape software were used to construct a protein–protein interaction (PPI) network between the molecules. We then analyzed the pathways of hub genes by gene set enrichment analysis (GSEA), using the gene expression level as the phenotype. The curated KEGG gene set was downloaded from the Molecular Signature Database, and FDR < 0.05 was considered significant. Finally, we evaluated the correlation between hub gene expression and immune cell infiltration in LUAD using TIMER.

## **Results**

### **Immune scores and stromal scores were correlated with survival outcomes**

A total of 510 LUAD cases from TCGA were used as the training cohort, and three GEO datasets were used as the validation cohorts. The clinical information for all cohorts is summarized in Table 1. We calculated the immune and stromal scores of each LUAD patient in TCGA and divided them into high and low immune/stromal score groups on the basis of the median value. Kaplan–Meier survival analysis showed that patients with high immune and stromal scores showed better survival outcomes than those with low scores, with log-rank tests of *P* = 0.01 and 0.026, respectively (Fig. 1a, 1b).

### **Identification of DEGs based on immune score and stromal score**

The heatmap showed that genes in the high score group had lower expression levels than those in the low score group, both for immune and stromal scores (Figure 1C,

**Table 1** Clinical characteristics of LUAD patients included in this study

Features	Number of patients (%)			
	Training cohort TCGA ( <i>n</i> = 510)	Testing cohort1 GSE26939 ( <i>n</i> = 116)	Testing cohort2 GSE37745 ( <i>n</i> = 196)	Testing cohort3 GSE29016 ( <i>n</i> = 72)
Age (years)				
≤ 65	235 (46.08)	59 (50.86)	102 (52.04)	33 (45.83)
> 65	275 (53.92)	57 (49.14)	94 (47.96)	39 (54.17)
Gender				
Female	271 (53.14)	63 (54.31)	89 (45.41)	41 (56.94)
Male	239 (46.86)	53 (45.69)	107 (54.59)	31 (43.06)
AJCC stage				
Stage I	272 (53.33)	62 (53.45)	130 (66.33)	46 (63.89)
Stage II	124 (24.31)	19 (16.38)	35 (17.85)	15 (20.83)
Stage III	85 (16.67)	19 (16.38)	27 (13.78)	5 (6.95)
Stage IV	22 (4.31)	2 (1.72)	4 (2.04)	0 (0)
Unknown	7 (1.38)	14 (12.07)	0 (0)	6 (8.33)
T stage				
T1	168 (32.94)	—	—	25 (34.72)
T2	276 (54.12)	—	—	36 (0.50)
T3	47 (9.21)	—	—	7 (9.72)
T4	19 (3.73)	—	—	4 (5.56)
N stage				
N0	335 (65.69)	—	—	65 (90.28)
N1–3	175 (34.31)	—	—	7 (9.72)
M stage				
M0	349 (68.43)	—	—	68 (94.44)
M1	22 (4.31)	—	—	0 (0)
Unknown	139 (27.26)	—	—	4 (5.56)
Survival status				
Alive 0	184 (36.08)	50 (43.10)	51 (26.02)	22 (30.56)
Dead 1	326 (63.92)	66 (56.90)	145 (73.98)	50 (69.44)

AJCC, American Joint Committee on Cancer

D). A total of 776 immune DEGs were obtained from the comparison of immune scores (samples with high scores vs. low scores), of which 613 genes were upregulated and 163 genes were downregulated. Similarly, 792 stromal DEGs were obtained from a comparison of the stromal scores, consisting of 678 upregulated genes and 114 downregulated genes. Moreover, Venn diagrams showed that 297 DEGs were commonly upregulated in the high-score groups, and 66 DEGs were commonly downregulated (Fig. 1e, 1f). These notable DEGs were potentially determinant factors of TME status.

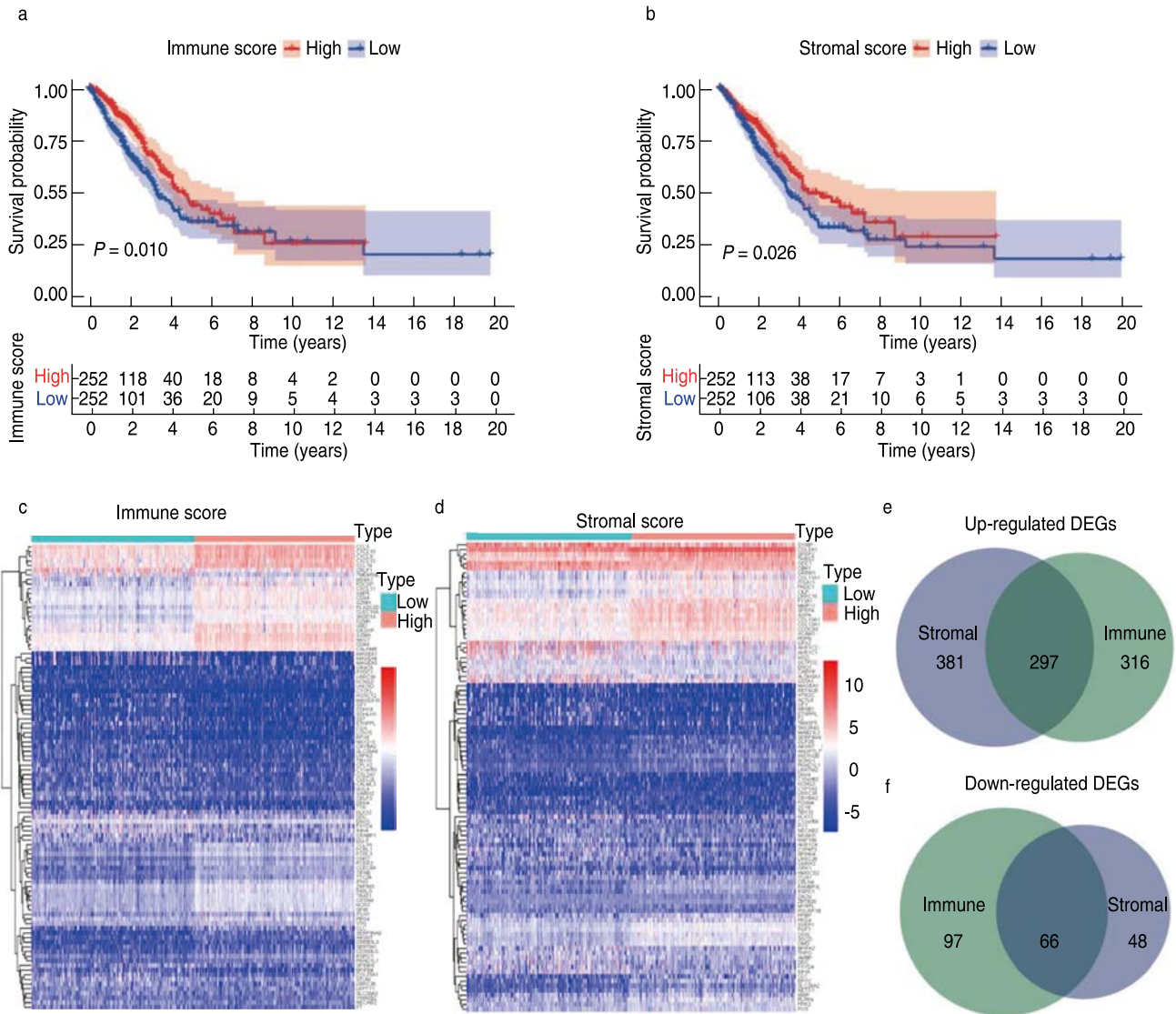
### GO and KEGG enrichment analysis

Results of the GO enrichment analysis showed that these DEGs were mainly involved in immune-related functions, such as T-cell activation and lymphocyte proliferation (Fig. 2a, 2c). KEGG analysis also revealed enrichment of the T cell receptor signaling pathway, chemokine signaling pathway, and hematopoietic cell lineage (Fig. 2b, 2d). Since the DEGs were correlated with

immune functions or pathways in LUAD, we further identified the top 89 immune-related DEGs from the Immport database for subsequent analysis (Fig. 3a).

### Construction of the TME-related prognostic model

Univariate Cox and Kaplan–Meier analyses were conducted to determine the significant prognostic genes among the 89 immune-related DEGs. A total of 24 genes were identified as significant in the Cox regression analysis (Fig. 3b), of which 6 genes were also significant in the Kaplan–Meier analysis. Among them, higher expression levels of BTK, CD28, PIK3CG, TLR4, and VEGFD correlated positively with poor survival outcomes, whereas INHA expression correlated negatively with prognosis (Fig. 3c). Then, the six prognostic genes were subjected to multivariate Cox regression analysis, and the risk coefficient of each gene was calculated (Table 2). The TME-related prognostic model was constructed as follows: risk score =  $(-0.111958) \times \text{EXP}_{\text{BTK}} + 0.279096 \times$



**Fig. 1** Identification of differential expressed genes (DEGs). (a) Kaplan-Meier survival curve of high and low immune score groups; (b) Kaplan-Meier survival analysis for high and low stromal score groups; (c) Heatmap for DEGs generated by comparison of gene expression profiles in high and low immune score groups; (d) Heatmap for DEGs in high and low stromal score groups; (e, f) Venn diagrams showed the common up-regulated and down-regulated DEGs shared by immune and stromal score groups

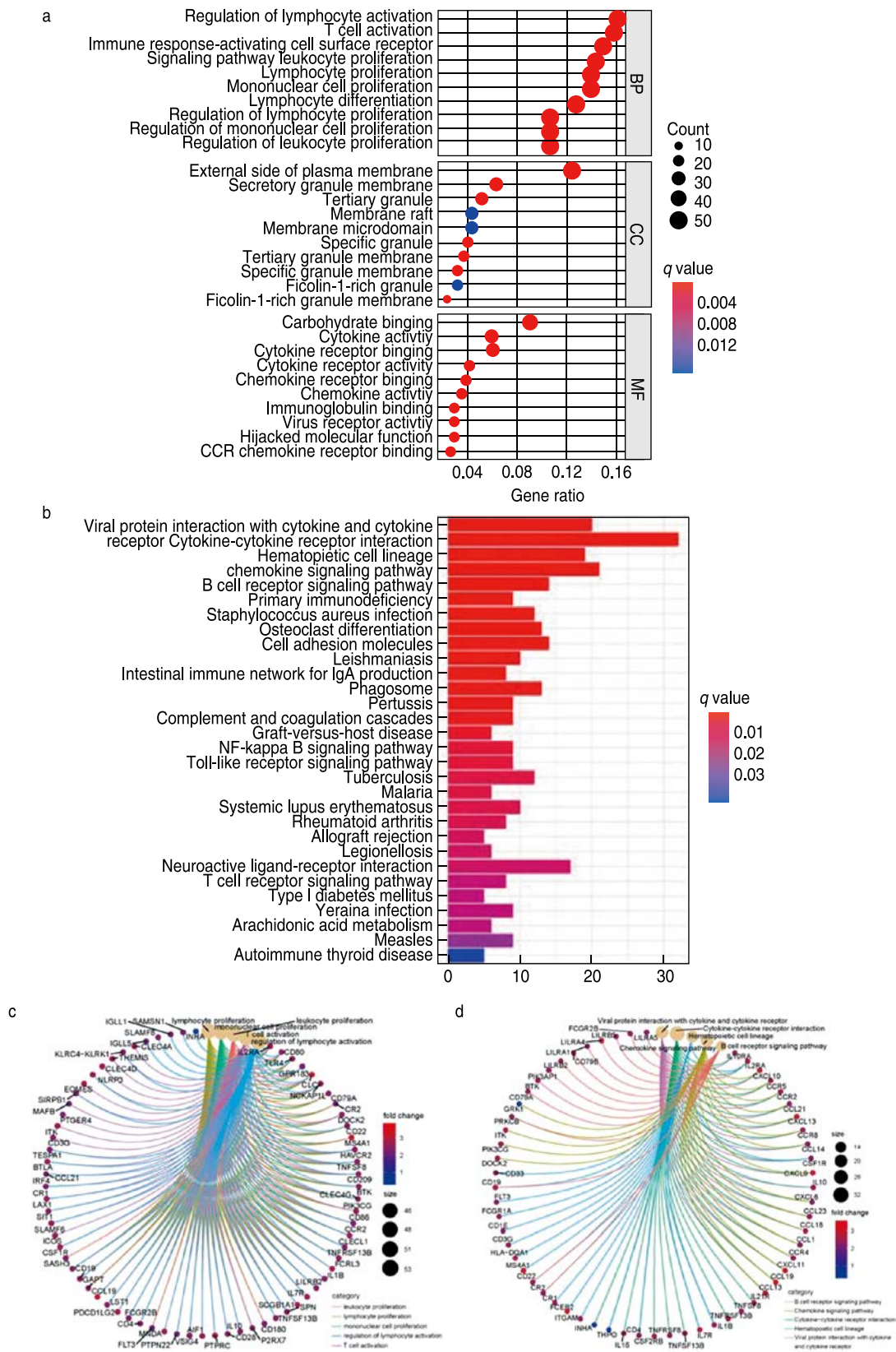
$$EXP_{CD28} + 0.008079 \times EXP_{INHA} + (-0.357674) \times EXP_{PIK3CG} + 0.099561 \times EXP_{TLR4} + (-0.102261) \times EXP_{VEGFD}$$

**Prognostic value of the TME-related model in the training and validation cohorts**

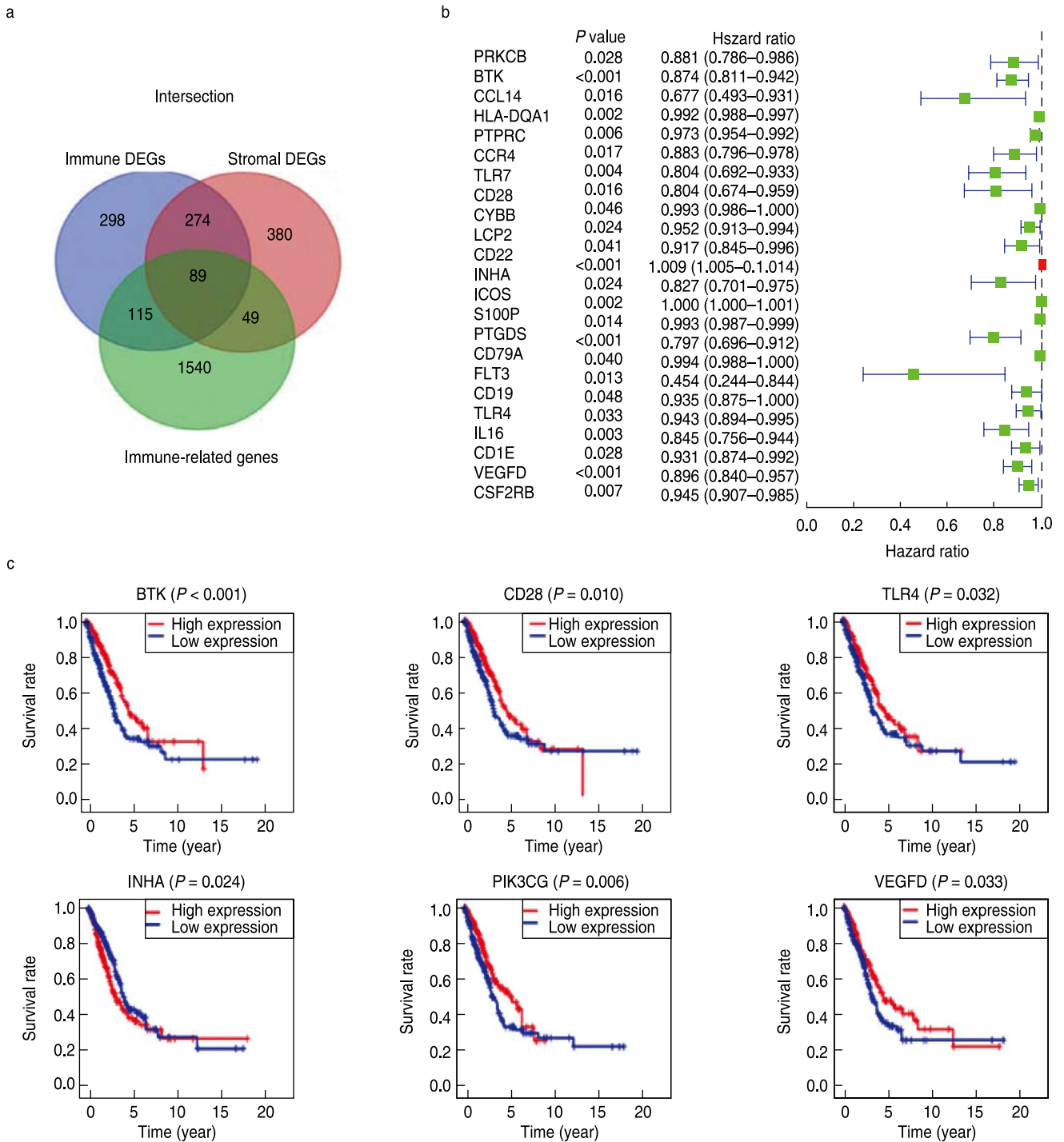
We calculated the risk score for each patient in the training cohort ( $n = 510$ ) and divided them into the high- and low-risk groups according to the median cutoff value (cutoff value:  $-0.261$ ). The Kaplan-Meier plot showed that patients in the high-risk group had worse survival outcomes than those in the low-risk group (Fig. 4a). The ROC curve of the 5-year survival prediction was drawn to assess the predictive accuracy, with an area under

the curve of 0.688 (Fig. 4b). Additionally, the risk curve indicated that the high-risk group had a higher mortality and worse prognosis than the low-risk group (Fig. 4c, 4d). The prognostic value of our model in patients with LUAD was further evaluated using other common prognostic factors. Although univariate Cox analysis indicated that the pathological stage and risk score had prognostic effects, only the risk score could be used as an independent prognostic factor ( $P < 0.001$ ; Fig. 4e, 4f).

Consistent with the results in the training dataset, the six-gene model stratified the samples of the three GEO testing datasets into high-risk and low-risk groups. Patients with low-risk scores had better survival



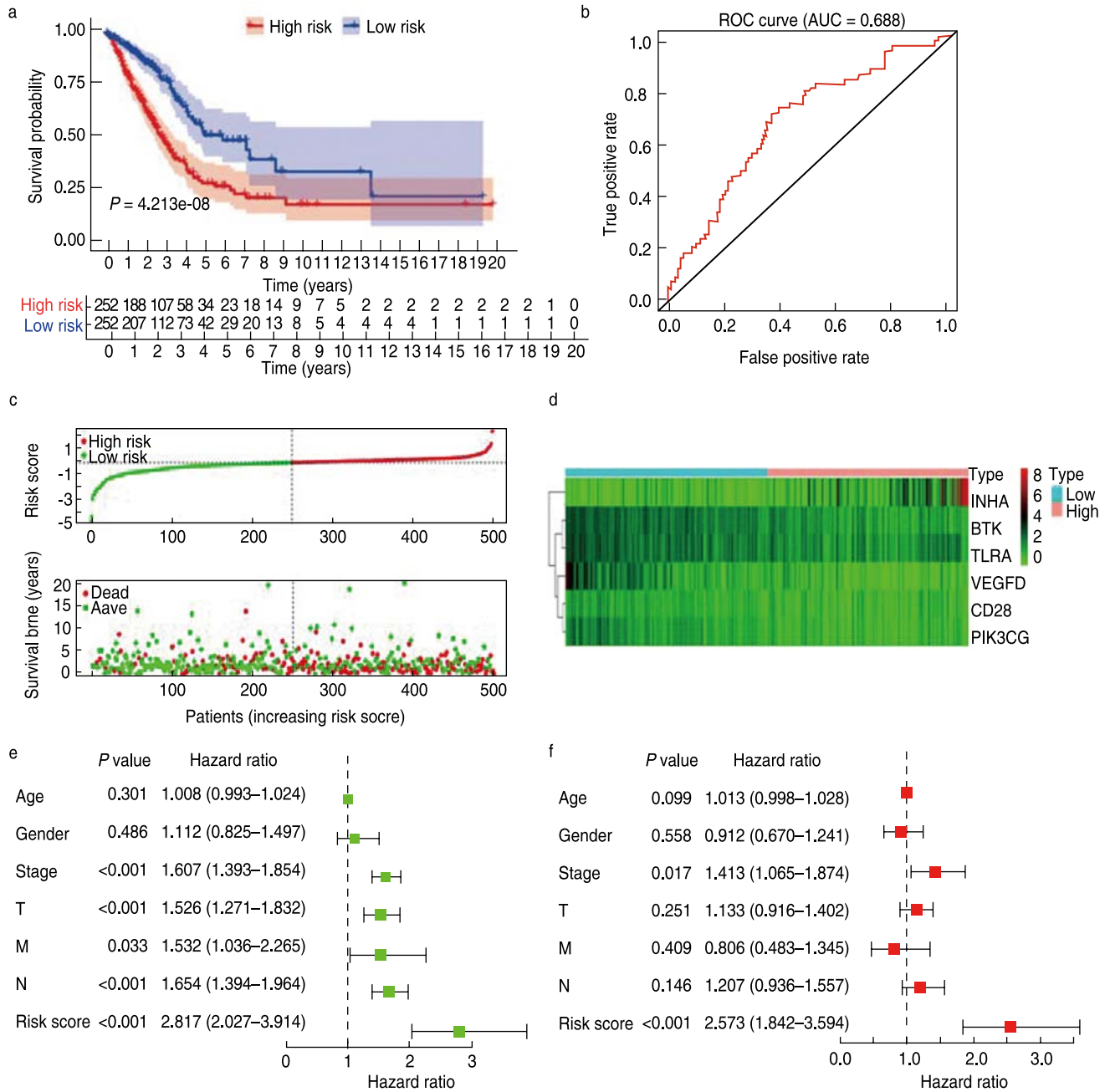




**Fig. 3** Univariate Cox and Kaplan Meier analysis for prognostic genes screening. (a) Identification of immune-related DEGs; (b) The forest plot of 24 prognostic immune-related DEGs screened out by Univariate Cox regression analysis with  $P < 0.005$ ; (c) Survival curves of the 6 prognostic genes extracted by the Kaplan-Meier analysis. Patients were labeled with high expression or low expression according to the median expression level of the 6 genes

outcomes than those in the high-risk group ( $P < 0.05$ ; Fig. 5a–5c). The areas under the ROC curves for predicting

5-year survival in the three testing datasets were 0.679, 0.666, and 0.732 (Fig. 5d–5f). These results suggest that



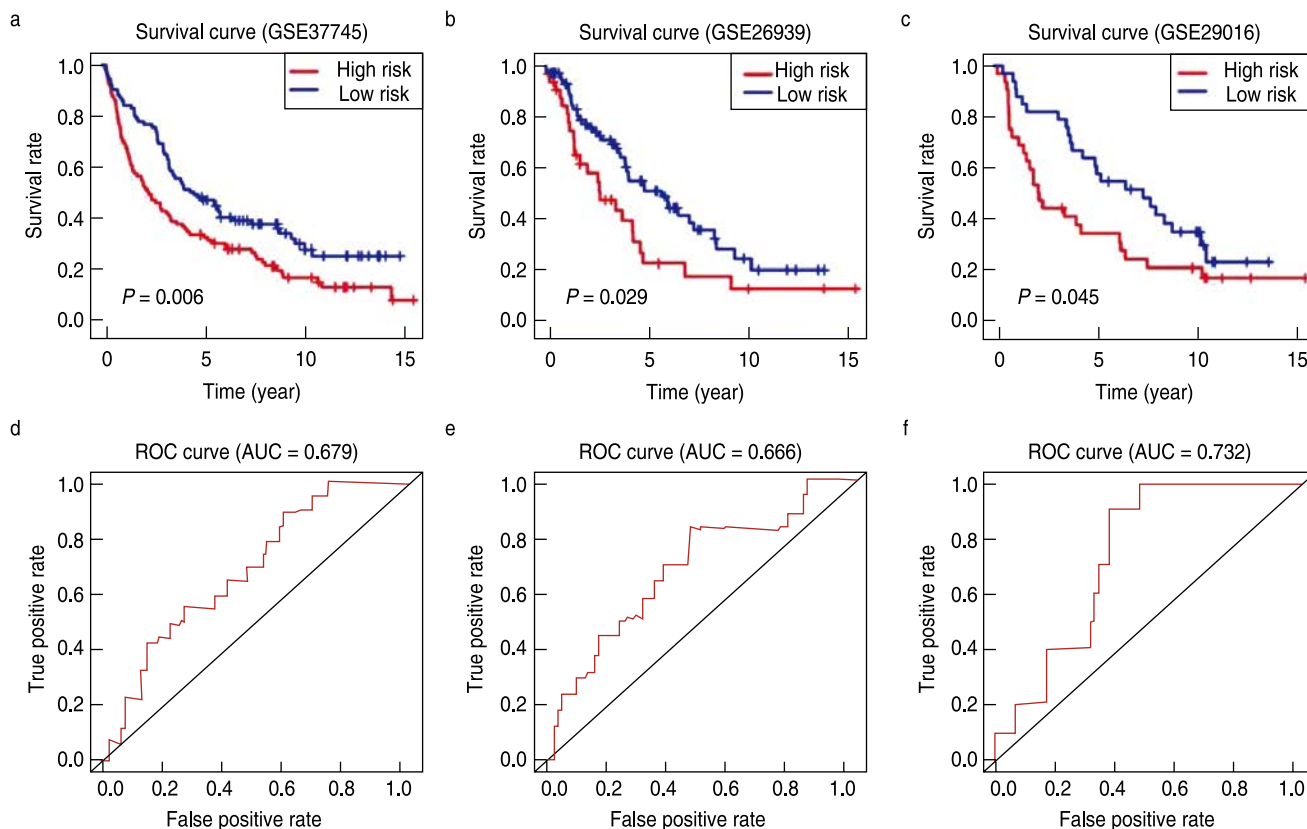
**Fig. 4** Construction and validation of the TME-related prognostic model in the training cohort. (a) Kaplan-Meier survival curve of low- and high-risk groups stratified by the TME-related prognostic model; (b) The ROC analysis of the TCGA dataset for survival prediction by the TME prognostic model; (c) The distribution of risk score and survival time in high- and low-risk groups; (d) Heatmap of the six prognostic genes; (e) The Univariate Cox analysis evaluating the prognostic effect of the model and common clinical factors; (f) Multivariate Cox analysis evaluating independent predictive ability of our model for overall survival

the TME-related prognostic model is robust in predicting the survival outcomes of patients with LUAD.

### Evaluation of the immune status between low-risk and high-risk groups

The strong stratification power of the TME-related model in predicting the survival of patients with LUAD led us to explore the difference in functional characteristics between the two risk groups. The ssGSEA score of the





**Fig. 5** Validation of the TME-related prognostic model in the testing cohort. Kaplan-Meier survival curves showing overall survival outcomes of high- and low-risk groups in GSE37745 (a), GSE26939 (b) and GSE29016 (c). The ROC curves for judging the predictive accuracy of the model in GSE37745 (d), GSE26939 (e) and GSE29016 (f)

**Table 2** Genes in the TME-related prognostic model

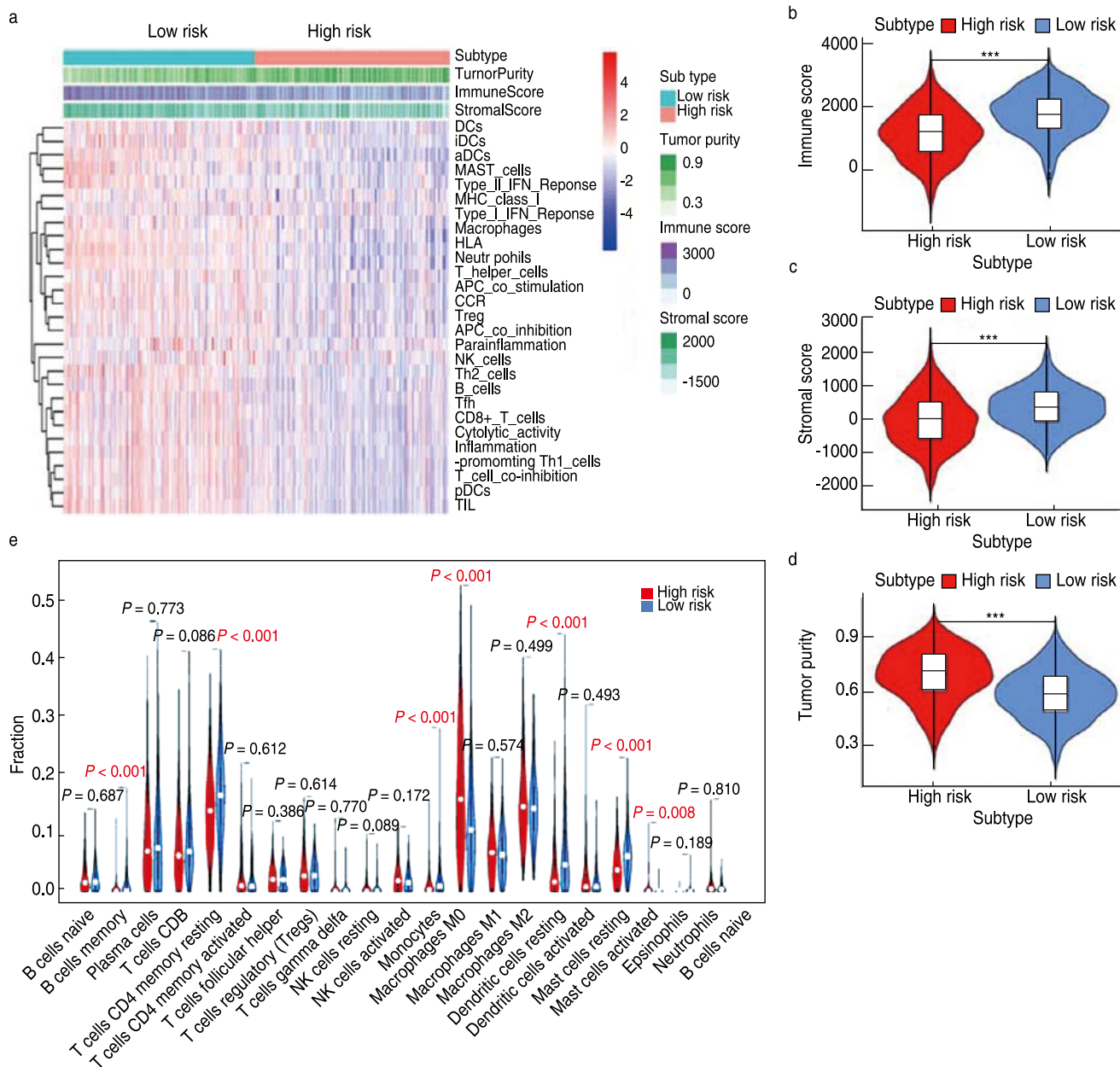
Gene symbol	Gene description	Coefficient
BTK	Bruton tyrosine kinase	-0.111958
CD28	CD28 molecule	0.279096
INHA	Inhibition alpha subunit	0.008079
PIK3CG	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma	-0.357674
TLR4	Toll like receptor 4	0.099561
VEGFD	Vascular endothelial growth factor D	-0.102261

29 immune signatures was used to evaluate the immune status of the two groups. The heatmap showed that the low-risk group had a higher immune activity than the high-risk group (Fig. 6a). Consistent with the ssGSEA results, immune and stromal scores in the low-risk group were significantly higher, and the tumor purity of the low-risk group was significantly lower (Fig. 6b–6d). This finding indicated that more immune and stromal cells infiltrated the TME of low-risk samples, whereas more tumor cells were present in high-risk samples. Moreover, we identified the immune cell subtypes in the two groups. The low-risk group had a significantly higher proportion of memory B cells, memory CD4+ T cells, monocytes,

and dendritic cells than the high-risk group, whereas the high-risk group had a markedly higher proportion of M0 macrophages (Fig. 6e). Taken together, these results suggest that patients with low-risk scores show elevated antitumor immune activity, leading to more favorable clinical outcomes.

**Potential mechanisms of the prognostic effects of the TME-related model**

GSEA was conducted to elucidate the specific regulatory mechanisms resulting in the differences in prognosis and immune status between the two risk cohorts. The results showed that the low-risk group was



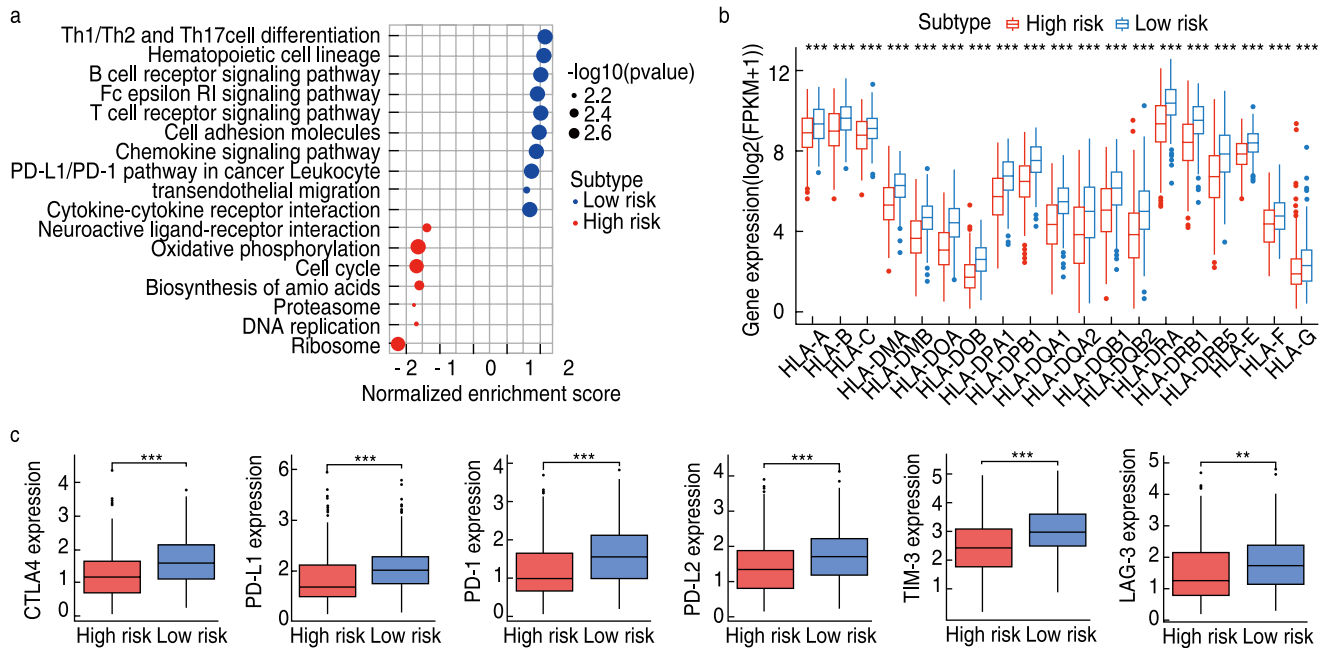
**Fig. 6** Evaluation of immune status and immune cell infiltration levels between high- and low-risk groups. (a) The heatmap of the overall immune status of high- and low-risk groups in TCGA database, showing greater heterogeneity between the two groups; (b–d) The violin plots showed the difference in immune score, stromal score and tumor purity between low- and high-risk groups. \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ ; (e) The violin plot shows the difference in the proportion of 21 kinds of immune cells between high- and low-risk groups, and the Wilcox rank-sum was used for the significance test

enriched not only in immunoregulation and immune cell activation, but also in many cancer-associated pathways, such as JAK-STAT signaling, cell adhesion molecules, and transendothelial migration (Fig. 7a). In contrast, the high-risk group was impoverished in immune signatures but enriched in metabolic signaling. Notably, most HLA genes showed significantly higher expression in the low-risk group than in the high-risk group, indicating that

local immune regulation and immunogenicity were more active in the low-risk group (Fig. 7b).

### Relationship between the TME-related model and immunotherapy response

In recent years, immune checkpoint proteins such as cytotoxic T lymphocyte antigen 4 (CTLA-4) or the programmed cell death ligand 1/protein 1 pathway



**Fig. 7** Functional mechanisms of the TME-related model and association with immune checkpoint molecules. (a) KEGG pathways enriched in high- and low-risk samples; (b) The expression profiles of HLA genes of low- and high-risk groups. \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ ; (c) Comparison of expression levels of CTLA-4, PD-1, PD-L1, PD-L2, TIM-3 and LAG-3 between high-risk and low-risk groups (Wilcoxon rank-sum test)

(PD-L1/PD-1) have been used as crucial targets for immunotherapy in LUAD [16]. We explored the relationship between the model and immunotherapy response by analyzing the expression of common immune checkpoints in the high- and low-risk groups and found that the expression of CTLA-4, PD-1, PD-L1, PD-L2, TIM-3, and LAG-3 in the low-risk group was significantly higher than that in the high-risk group (all  $P < 0.001$ ; Fig. 7c). This suggests that patients with low-risk scores might respond better to ICI treatment than those with high-risk scores because the expression of immune checkpoint molecules tends to be positively associated with immunotherapeutic responsiveness. Therefore, the construction of a risk cohort using our model could be a good stratification method for patients with LUAD regarding whether to conduct immunotherapy.

### The mechanism of action of prognostic hub genes in the TME

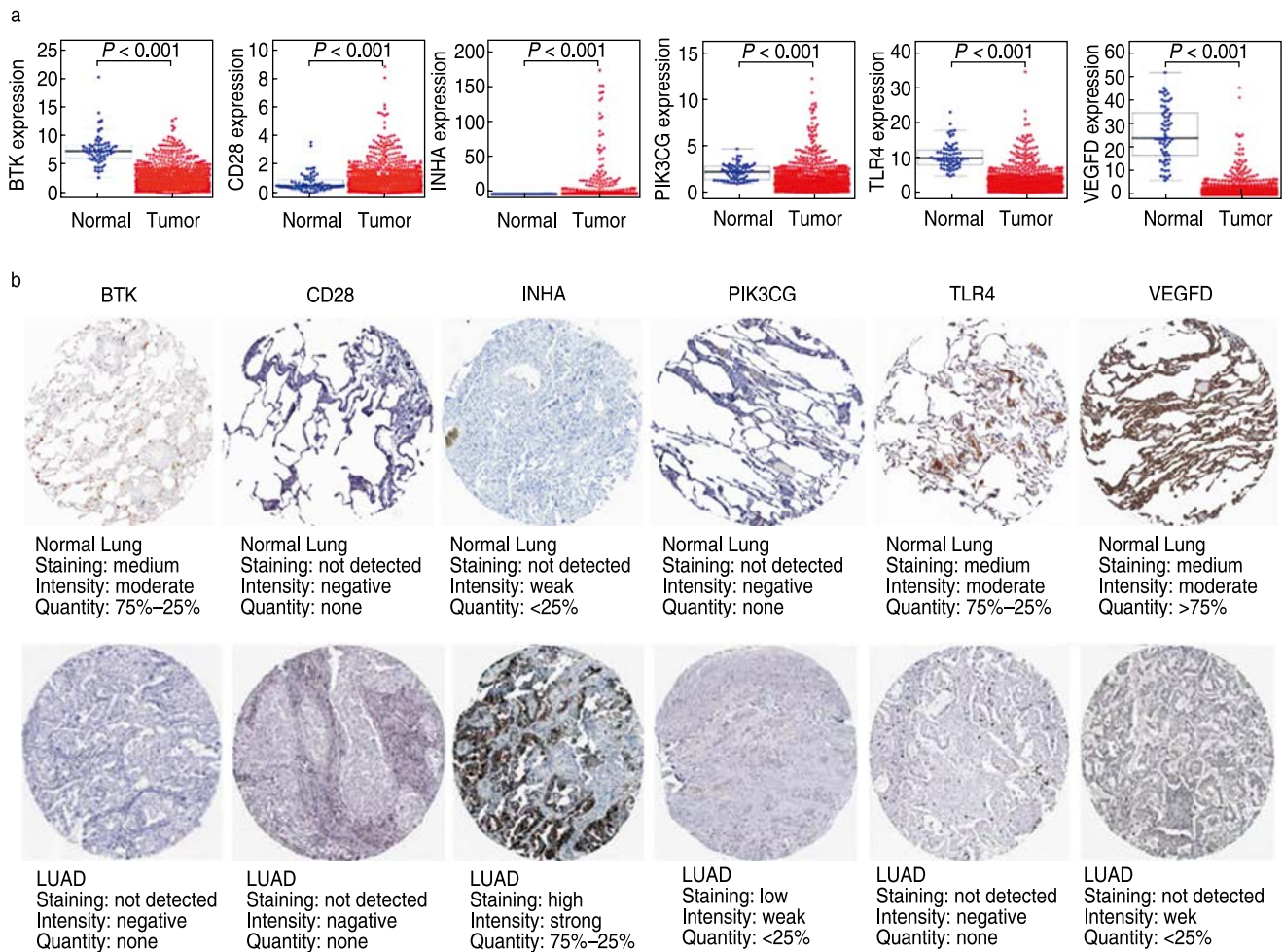
To further analyze the potential function of the hub genes, our results were verified using the HPA and TIMER databases. We found that CD28 and INHA had significantly higher expression levels in LUAD samples than in normal lung samples, whereas BTK, PIK3CG, TLR4, and VEGFD had lower expression levels in tumor tissues (Fig. 8a). In terms of protein levels, the protein expression patterns of BTK, INHA, PIK3CG, TLR4, and VEGFD were consistent with their RNA-seq

expression alterations (Fig. 8b). However, CD28 showed no significant difference. The PPI network also showed extensive interactions among BTK, CD28, PIK3CG, and TLR4.

GSEA suggested that BTK, CD28, PIK3CG, TLR4, and VEGFD were enriched in the same pathways. High expression of these five genes was mainly correlated with immune-related activities, such as antigen processing and presentation, the chemokine signaling pathway, and the JAK-STAT signaling pathway, whereas low expression of these genes was associated with metabolic pathways (Fig. 9a, 9b). In contrast, high expression of INHA was correlated with metabolic pathways, and low expression was involved in the activation of immune pathways (Fig. 9c, 9d). More importantly, the expression of BTK, CD28, PIK3CG, TLR4, and VEGFD positively correlated with the infiltration of CD4+ T cells, CD8+ T cells, B cells, neutrophils, dendritic cells, and macrophages (Fig. 10). However, INHA was negatively correlated with infiltration of the six immune cells. Collectively, these results suggest that these six hub genes affect the immune activity of the TME.

### Discussion

In this study, we aimed to identify a novel TME-related prognostic model for LUAD. We embarked on TME-related DEGs generated by comparing the immune



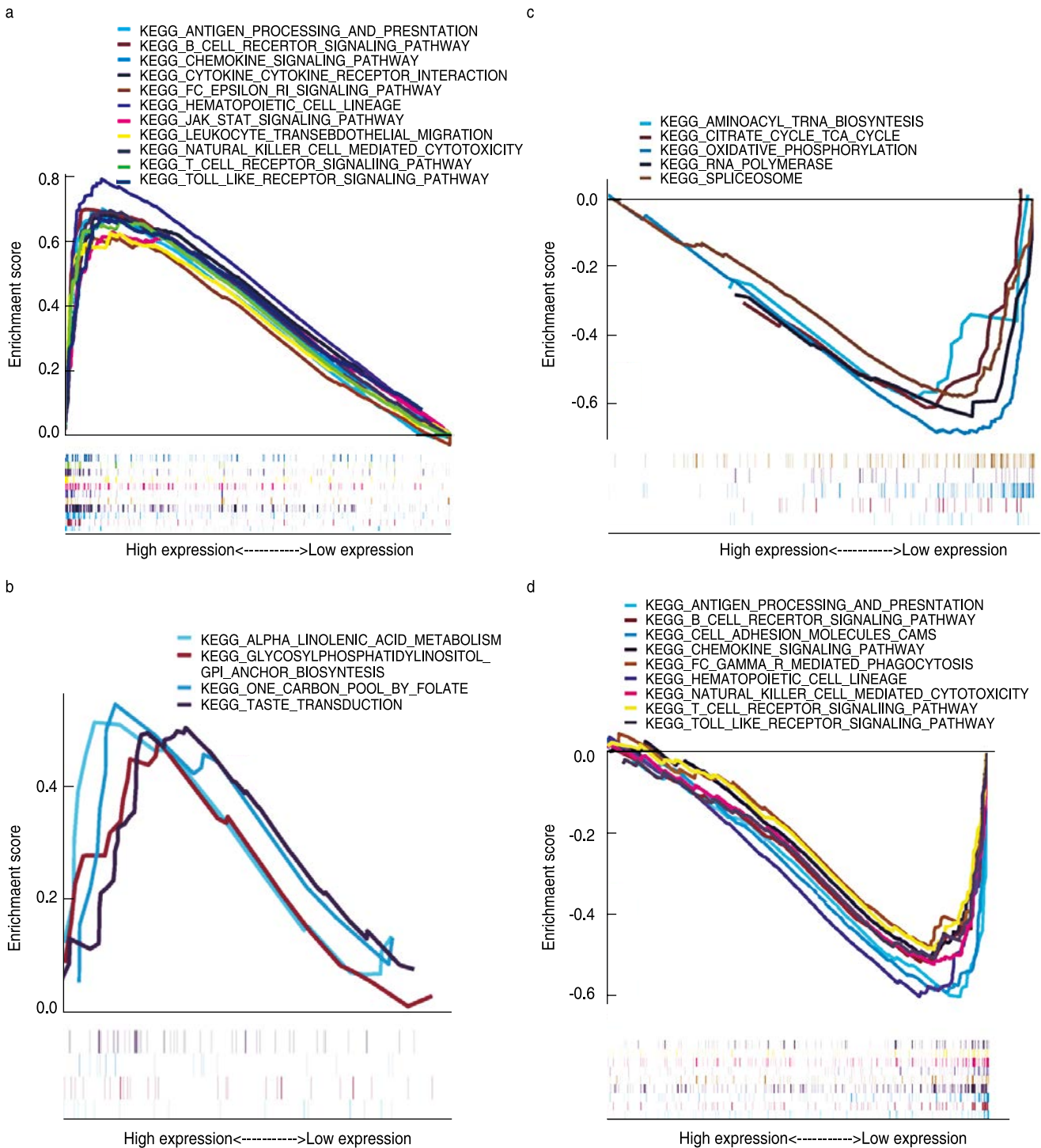
**Fig. 8** Expression profiles of the six hub genes in the model. (a) The expression levels of the six prognostic genes in LUAD samples and normal lung samples in the TCGA database (BTK, CD28, INHA, PIK3CG, TLR4 and VEGFD). Wilcox test was used to calculate the significance level between the two groups; (b) The immunohistochemistry results reflecting the gene-encoding protein levels of the six hub genes in LUAD and normal lung tissues from the HPA database

and stromal scores in LUAD samples. Subsequently, a list of TME-related genes that contribute to the survival outcomes of patients with LUAD was extracted. Finally, a six-gene prognostic model, based on prognostic TME-related genes, was constructed. Both immune and stromal genes in LUAD samples were analyzed to better reflect the complete TME status. Furthermore, we validated its prognostic value in three testing sets from the GEO datasets. Kaplan-Meier and ROC analyses revealed the strong predictive ability of our model for LUAD prognosis in both the training and testing sets. Univariate and multivariate Cox analyses confirmed the independent prognostic value of the six-gene model. Accordingly, unlike those developed previously, the TME-related prognostic model developed herein could reflect the tumor immune microenvironment status and predict the prognosis of LAUD more accurately. Moreover, an enhanced understanding of the model and related hub

genes would help to elucidate regulatory mechanisms of the TME and develop new treatment strategies.

Experimental and clinical studies have demonstrated that the immune and stromal components of the TME play significant roles in lung cancer development and progression<sup>[17]</sup>. The immune and stromal cells infiltrating the TME are composed of different cell types. As the most important immune cells in the TME, tumor-infiltrating T lymphocytes execute key effector cytotoxic functions and mediate responses to ICIs<sup>[18]</sup>. Tumor-associated macrophages are another class of immune cells that interact with lung cancer cells. Macrophage-tumor cell interactions lead to the release of pro-inflammatory cytokines, chemokines, and growth factors, which in turn recruit additional inflammatory cells to the microenvironment<sup>[19]</sup>. Other immune cells in the TME include B cells, NK cells, dendritic cells (DCs), T regulatory cells (Tregs), and B regulatory cells (Bregs). Cancer-associated fibroblasts

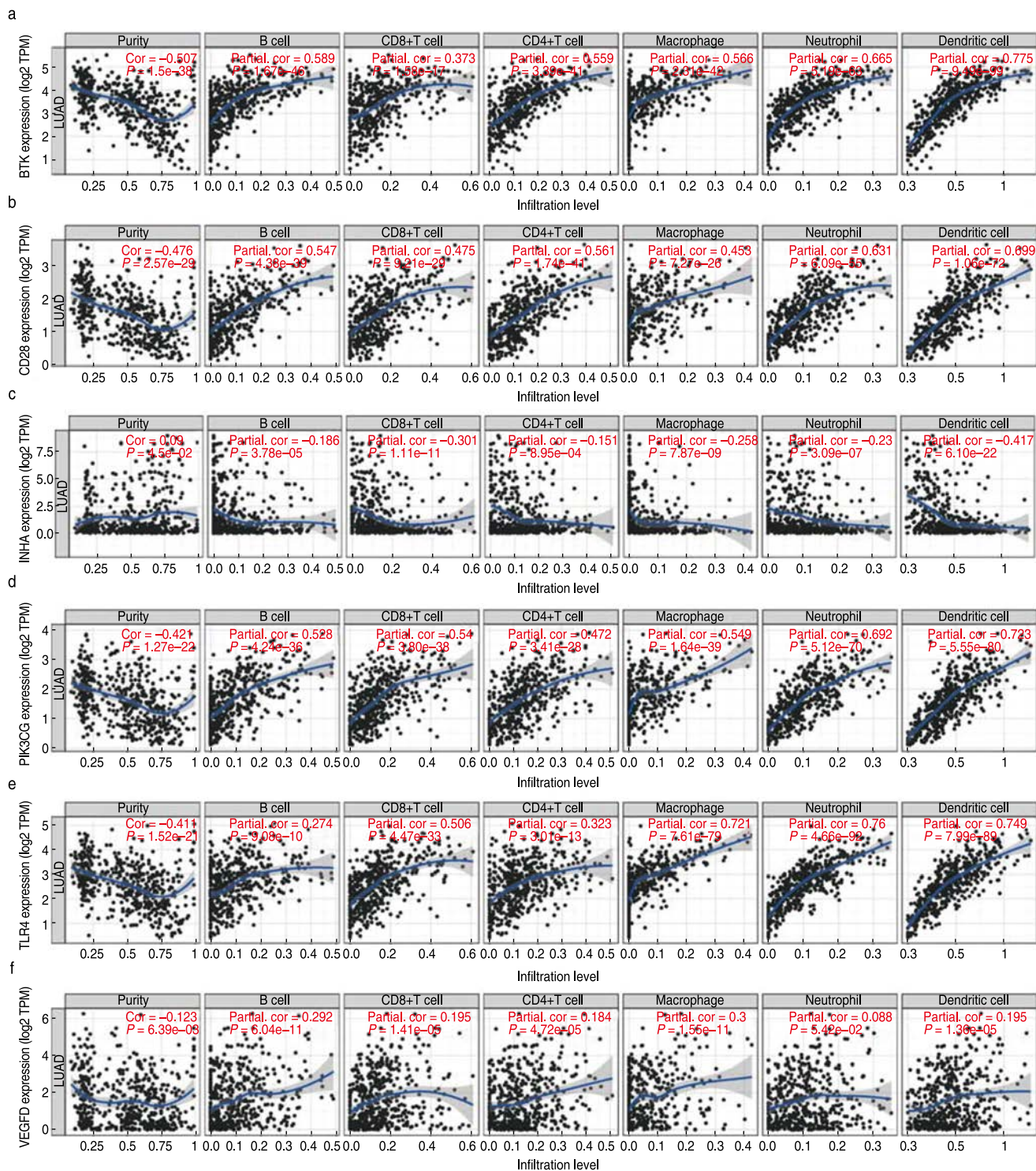




**Fig. 9** The GSEA enrichment analysis of the hub prognostic genes. (a) The enrichment pathways of high expression of BTK, CD28, PIK3CG, TLR4 and VEGFD. Each line representing one particular pathway with unique color, only pathways with  $p$  and  $q < 0.05$  were considered significant. And only several leading gene sets were displayed in the plot; (b) The enrichment pathways of low expression of the five genes; (c) Enrichment pathways of high INHA expression; (d) Enrichment pathways of the low INHA expression

are the most abundant stromal cells in the TME and play critical roles in the inflammatory response and immune suppression of tumors [20]. Fibroblasts promote tumor

progression via multiple pathways, including regulation of the extracellular matrix, production of growth factors or cytokines, and promotion of angiogenesis, whereas



**Fig. 10** Correlation between hub prognostic genes and immune cell infiltration. (a–f) The gene expression levels against tumor purity are displayed on the left-most panel

some fibroblast subtypes also show antitumor activities by secreting immunosuppressive cytokines<sup>[21]</sup>.

To determine the distinct gene expression profiles

in the TME with respect to immune and stromal components, DEGs based on immune and stromal scores were screened. Six prognostic hub genes among these



DEGs were identified (survival positive correlation: INHA; negative correlation: BTK, CD28, PIK3CG, TLR4, and VEGFD). Interestingly, functional analysis showed that BTK, CD28, PIK3CG, TLR4, and VEGFD could promote immune infiltration, while INHA inhibited immune cell infiltration. These results suggest that the six hub genes participate in the immune regulation of the TME and affect the prognosis of patients with LUAD, which might be potential immune prognostic markers and therapeutic targets for LUAD.

Among these six genes, Bruton tyrosine kinase (BTK) is a member of the Tec kinase family. As a key component of the B-cell antigen receptor signaling pathway, BTK plays a vital role in B lymphocyte development, differentiation, and signaling<sup>[22]</sup>. Ibrutinib is a small-molecule irreversible inhibitor of BTK that has been approved for the treatment of hematological malignancies and some solid tumors owing to its ability to inhibit tumor growth by modifying the tumor microenvironment and its potential synergistic activity with ICIs<sup>[23,24]</sup>. CD28 is a key T-cell costimulatory molecule that binds to B7 molecules, which are involved in the regulation of T cell proliferation and activation, along with cytokine production<sup>[25]</sup>. Previous studies have demonstrated that CD28 can predict the response to anti-PD-1 therapy in patients with lung cancer<sup>[26]</sup>. PIK3CG encodes the PI3Ky enzyme, which can activate the signaling molecule Akt and modulate various cell functions such as cell proliferation, migration, and adhesion<sup>[27]</sup>. Novel PI3K inhibitors are important for the treatment of hematologic malignancies<sup>[28]</sup>. The protein encoded by TLR4 is a member of the toll-like receptor family. Studies have shown that TLR4 is highly expressed in immune cells, such as monocytes and lymphocytes, but is expressed at low levels in epithelial, endothelial, and cancer cells. Thus, TLR4 agonists have been widely explored as potential immunotherapeutic agents for the treatment of cancer<sup>[29]</sup>. VEGFD belongs to the vascular endothelial growth factor family and can induce both angiogenesis and lymphangiogenesis<sup>[30]</sup>. Clinical studies have shown that low expression of VEGFD is a predictor of greater survival benefits from bevacizumab treatment in patients with CRC<sup>[31]</sup>. INHA encodes a member of the transforming growth factor-beta (TGF-beta) superfamily of proteins. However, its function in lung cancer remains unknown. Our results suggest that high expression of INHA is associated with a poor prognosis of LUAD. A possible mechanism might be that INHA is involved in vascularization and tumor metastasis, leading to a poor prognosis<sup>[32]</sup>. Further studies are required to clarify the role of these hub genes in the TME in the initiation and development of LUAD.

Finally, a TME-related prognostic model was developed using the six hub immune genes for survival prediction. The low-risk group showed higher expression

of HLA genes. HLA-related genes play a significant role in immune regulation, and their expression is advantageous for immunotherapy efficacy<sup>[33]</sup>. Our results showed elevated antitumor immune activity in the low-risk group, which could explain why the low-risk group had more favorable clinical outcomes than the high-risk group. In addition to survival prediction, this TME-related signature was also a predictor of patient response to ICI treatment. To date, many biomarkers have been verified to indicate the efficacy of ICI treatment, including TMB, PD-L1 expression level, neoantigens, and gut microbiota<sup>[34]</sup>. Generally, most biomarkers reflect the status of the tumor immune microenvironment in a certain aspect. Thus, a prognostic model based on the TME may aid in the stratification of patients with LUAD to identify those responsive to immunotherapy. It is possible that patients with low-risk scores are more sensitive to immunotherapy than those with high-risk scores, since immune checkpoint molecules are more highly expressed in low-risk groups, and the increased levels of immune checkpoints indirectly indicate pre-existing T cell activation in the low-risk group.

This study has some limitations. The six-gene model was derived from retrospective data, and more prospective data are needed to validate our results. Second, this study lacked basic experiments to validate the function of the six hub genes and their association with immune cell infiltration. Third, patients receiving immunotherapy were not included in this study; therefore, the predictive ability of the model for immunotherapy response was evaluated indirectly.

## Conclusions

In conclusion, we constructed a TME-related prognostic model to predict LUAD patient survival outcomes and responses to immunotherapy. Patients with low risk scores had better prognoses and were expected to benefit from ICI treatment. This model might be valuable for prognostic management and patient selection before immunotherapy and deserves further validation. A significant association was observed between the hub genes and patient prognosis and immune infiltration, providing novel insights for the development of new treatment strategies.

## Acknowledgments

We are grateful to the TCGA database and GEO database for the source of data used in our study.

## Conflicts of interest

The authors indicated no potential conflicts of interest.

## References

1. Bray F, J Ferlay, I Soerjomataram, *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 2018, 68: 394–424.
2. Schneider BJ, N Ismaila, N Altorki. Lung cancer surveillance after definitive curative-intent therapy: ASCO Guideline Summary. *JCO Oncol Pract*, 2020, 16: 83–86.
3. Herbst RS, D Morgensztern, C Boshoff. The biology and management of non-small cell lung cancer. *Nature*, 2018, 553: 446–454.
4. Fenizia F, R Pasquale, C Roma, *et al.* Measuring tumor mutation burden in non-small cell lung cancer: tissue versus liquid biopsy. *Transl Lung Cancer Res*, 2018, 7: 668–677.
5. Zou W, JD Wolchok, L Chen. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. *Sci Transl Med*, 2016, 8: 328rv4.
6. Rotte A, JY Jin, V Lemaire. Mechanistic overview of immune checkpoints to support the rational design of their combinations in cancer immunotherapy. *Ann Oncol*, 2018, 29: 71–83.
7. Braun DA, KP Burke, EM Van Allen. Genomic Approaches to Understanding Response and Resistance to Immunotherapy. *Clin Cancer Res*, 2016, 22: 5642–5650.
8. Wood SL, M Pernemalm, PA Crosbie, *et al.* The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets. *Cancer Treat Rev*, 2014, 40: 558–566.
9. Hanahan D, LM Coussens. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell*, 2012, 21: 309–322.
10. Quail DF, JA Joyce. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*, 2013, 19: 1423–1437.
11. Guo X, Y Zhang, L Zheng, *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med*, 2018, 24: 978–985.
12. Bussard KM, L Mutkus, K Stumpf, *et al.* Tumor-associated stromal cells as key contributors to the tumor microenvironment. *Breast Cancer Res*, 2016, 18: 84.
13. Carter SL, K Cibulskis, E Helman, *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*, 2012, 30: 413–421.
14. Alonso MH, S Ausso, A Lopez-Doriga, *et al.* Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *Br J Cancer*, 2017, 117: 421–431.
15. Luo C, M Lei, Y Zhang, *et al.* Systematic construction and validation of an immune prognostic model for lung adenocarcinoma. *J Cell Mol Med*, 2020, 24: 123–144.
16. Nishino M, NH Ramaiya, H Hatabu, *et al.* Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*, 2017, 14: 655–668.
17. Noonan DM, A De Lerma Barbaro, N Vannini, *et al.* Inflammation, inflammatory cells and angiogenesis: decisions and indecisions. *Cancer Metastasis Rev*, 2008, 27: 31–40.
18. Chen DS, I Mellman. Elements of cancer immunity and the cancer-immune set point. *Nature*, 2017, 541: 321–330.
19. Chanmee T, P Ontong, K Konno, *et al.* Tumor-associated macrophages as major players in the tumor microenvironment. *Cancers (Basel)*, 2014, 6: 1670–1690.
20. Chen X, E Song. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Discov*, 2019, 18: 99–115.
21. Orimo A, PB Gupta, DC Sgroi, *et al.* Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell*, 2005, 121: 335–348.
22. Mohamed AJ, L Yu, CM Backesjo, *et al.* Bruton's tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain. *Immunol Rev*, 2009, 228: 58–73.
23. Gayko U, M Fung, F Clow, *et al.* Development of the Bruton's tyrosine kinase inhibitor ibrutinib for B cell malignancies. *Ann N Y Acad Sci*, 2015, 1358: 82–94.
24. Molina-Cerrillo J, T Alonso-Gordoa, P Gajate, *et al.* Bruton's tyrosine kinase (BTK) as a promising target in solid tumors. *Cancer Treat Rev*, 2017, 58: 41–50.
25. Esensten JH, YA Helou, G Chopra, *et al.* CD28 costimulation: from mechanism to therapy. *Immunity*, 2016, 44: 973–988.
26. Kamphorst AO, A Wieland, T Nasti, *et al.* Rescue of exhausted CD8 T cells by PD-1-targeted therapies is CD28-dependent. *Science*, 2017, 355: 1423–1427.
27. Lien EC, CC Dibble, A Toker. PI3K signaling in cancer: beyond AKT. *Curr Opin Cell Biol*, 2017, 45: 62–71.
28. Pei Y, N Hwang, F Lang, *et al.* Quassinoid analogs with enhanced efficacy for treatment of hematologic malignancies target the PI3Kgamma isoform. *Commun Biol*, 2020, 3: 267.
29. Maryam A Shetab Boushehri, Alf Lamprecht. TLR4-based immunotherapeutics in cancer: a review of the achievements and shortcomings. *Mol Pharm*, 2018, 15: 4777–4800.
30. Janiszewska M, DP Tabassum, Z Castano, *et al.* Subclonal cooperation drives metastasis by modulating local and systemic immune microenvironments. *Nat Cell Biol*, 2019, 21: 879–888.
31. Weickhardt AJ, DS Williams, CK Lee, *et al.* Vascular endothelial growth factor D expression is a potential biomarker of bevacizumab benefit in colorectal cancer. *Br J Cancer*, 2015, 113: 37–45.
32. Singh P, LM Jenkins, B Horst, *et al.* Inhibin Is a Novel Paracrine Factor for Tumor Angiogenesis and Metastasis. *Cancer Res*, 2018, 78: 2978–2989.
33. Chowell D, C Krishna, F Pierini, *et al.* Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat Med*, 2019, 25: 1715–1720.
34. Patel SP, R Kurzrock. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. *Mol Cancer Ther*, 2015, 14: 847–856.

DOI 10.1007/s10330-021-0545-5

Cite this article as: Li Z, Feng YQ, Li P, *et al.* Development and validation of a tumor microenvironment-related prognostic signature in lung adenocarcinoma and immune infiltration analysis. *Oncol Transl Med*, 2021, 7: 253–268.