



Identification of a Novel Tumor Microenvironment-Associated Eight-Gene Signature for Prognosis Prediction in Lung Adenocarcinoma

Chao Ma^{1,2*}, Huan Luo^{1,3}, Jing Cao⁴, Xiangyu Zheng⁵, Jinjun Zhang⁵, Yanmin Zhang⁵ and Zongqiang Fu⁵

¹ Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany, ² Berlin Institute of Health Center for Regenerative Therapies and Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité – Universitätsmedizin Berlin, Berlin, Germany, ³ Klinik für Augenheilkunde, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany, ⁴ Department of Human Anatomy, College of Basic Medical Sciences, Zhengzhou University, Zhengzhou, China, ⁵ Department of Laboratory Medicine, The Second Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou, China

OPEN ACCESS

Edited by:

Anton A. Buzdin,
I.M. Sechenov First Moscow State
Medical University, Russia

Reviewed by:

Surendra Kumar Shukla,
University of Nebraska Medical
Center, United States
Ratnakar Singh,
University of Illinois
at Urbana-Champaign, United States

*Correspondence:

Chao Ma
chao.ma@charite.de
orcid.org/0000-0003-1444-4668

Specialty section:

This article was submitted to
Molecular Diagnostics
and Therapeutics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 12 June 2020

Accepted: 19 August 2020

Published: 23 September 2020

Citation:

Ma C, Luo H, Cao J, Zheng X,
Zhang J, Zhang Y and Fu Z (2020)
Identification of a Novel Tumor
Microenvironment-Associated
Eight-Gene Signature for Prognosis
Prediction in Lung Adenocarcinoma.
Front. Mol. Biosci. 7:571641.
doi: 10.3389/fmolb.2020.571641

Background: Lung cancer has become the most common cancer type and caused the most cancer deaths. Lung adenocarcinoma (LUAD) is one of the major types of lung cancer. Accumulating evidence suggests the tumor microenvironment is correlated with the tumor progress and the patient's outcome. This study aimed to establish a gene signature based on tumor microenvironment that can predict patients' outcomes for LUAD.

Methods: Dataset TCGA-LUAD, downloaded from the TCGA portal, were taken as training cohort, and dataset GSE72094, obtained from the GEO database, was set as validation cohort. In the training cohort, ESTIMATE algorithm was applied to find intersection differentially expressed genes (DEGs) among tumor microenvironment. Kaplan–Meier analysis and univariate Cox regression model were performed on intersection DEGs to preliminarily screen prognostic genes. Besides, the LASSO Cox regression model was implemented to build a multi-gene signature, which was then validated in the validation cohorts through Kaplan–Meier, Cox, and receiver operating characteristic curve (ROC) analyses. In addition, the correlation between tumor mutational burden (TMB) and risk score was evaluated by Spearman test. GSEA and immune infiltrating analyses were conducted for understanding function annotation and the role of the signature in the tumor microenvironment.

Results: An eight-gene signature was built, and it was examined by Kaplan–Meier analysis, revealing that a significant overall survival difference was seen. The eight-gene signature was further proven to be independent of other clinico-pathologic parameters via the Cox regression analyses. Moreover, the ROC analysis demonstrated that this signature owned a better predictive power of LUAD prognosis. The eight-gene signature was correlated with TMB. Furthermore, GSEA and immune infiltrating analyses showed

that the exact pathways related to the characteristics of eight-genes signature, and identified the vital roles of Mast cells resting and B cells naive in the prognosis of the eight-gene signature.

Conclusion: Identifying the eight-gene signature (INSL4, SCN7A, STAP1, P2RX1, IKZF3, MS4A1, KLRB1, and ACSM5) could accurately identify patients' prognosis and had close interactions with Mast cells resting and B cells naive, which may provide insight into personalized prognosis prediction and new therapies for LUAD patients.

Keywords: lung adenocarcinoma, tumor microenvironment, tumor immunity, gene signature, risk score, prognosis, biomarkers

INTRODUCTION

Lung cancer ranks among the top cancer-related deaths worldwide (Bray et al., 2018). Histologically, 15% of patients are classified as small cell lung cancer (SCLC), whereas the other 85% of patients are classified as non-small-cell lung cancer (NSCLC) (Gridelli et al., 2015). In NSCLC, lung adenocarcinoma (LUAD) is the most common subtype (Cancer Genome Atlas Research and Network, 2014). Over the past few decades, extensive genomic studies have identified several high-frequency genetic changes in LUAD, such as EGFR, KRAS mutations, and ALK rearrangements, which may be related to the occurrence and development of LUAD and have contributed to EGFR targeting drug development (Herbst et al., 2018). However, a large number of patients with advanced LUAD have no targeted mutations. For these patients, studies on immune checkpoints like programmed death 1 (PD-1) and cytotoxic T lymphocyte-associated antigen-4 (CTLA-4) have demonstrated the effectiveness and safety of established treatments (Hellmann et al., 2017; Xu et al., 2018), which highlights the importance of tumor microenvironment on the clinical outcomes of LUAD patients.

More and more studies have shown that the occurrence and development of tumors are the results of the dynamic interaction between tumor cells and various components (including immune cells and stromal cells) within the tumor microenvironment (Hui and Chen, 2015; Verrecchia and Redini, 2018; Shi et al., 2020). Therefore, the tumor microenvironment may become a promising target for cancer treatment (Alsaab et al., 2017). In Yoshihara et al. (2013) designed an algorithm called ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data). The algorithm analyses specific gene expression characteristics of immune and stromal cells and calculates immune and stromal scores to predict non-tumor cell infiltration (Yoshihara et al., 2013). Considering that previous research has mainly focused on one or two types of immune cells and key genes, it may create a bias against the LUAD microenvironment. Finding multiple genes from the tumor microenvironment to build a gene signature can obtain a better accuracy of the prognostic potential (Liu et al., 2019a,b; Zhou et al., 2019; Bi et al., 2020; Guo et al., 2020).

Here, we conduct comprehensive mining of The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases to determine the minimum number of potentially robust genes from tumor microenvironment that can be used

to predict the prognosis of LUAD patients. Importantly, we used the LASSO algorithm, which can effectively analyze high-dimensional sequencing data (Tibshirani, 1997). Besides, we assessed the accuracy of this eight-gene signature and validated it by comparing with tumor mutational burden (TMB) and testing in a validation cohort. Moreover, GSEA and immune infiltrating analyses were conducted to explore the role of the signature in the tumor microenvironment.

MATERIALS AND METHODS

Data Mining From TCGA and GEO

The gene expression profiles of LUAD from 515 patients, along with their clinical and survival data, were downloaded from TCGA Xena Hub¹ with cohort name: TCGA-LUAD. Besides, we selected dataset GSE72094 from the GEO database², which contains 442 LUAD cases, for the study. In our research, TCGA-LUAD was used as the training cohort, whereas GSE72094 was taken as the validation cohort.

Immune and Stromal Scores

Immune scores and stromal scores of each case of the training cohort calculated by the ESTIMATE algorithm were downloaded from the MD Anderson Cancer Center ESTIMATE Results download page³ (Supplementary Table S1; Yoshihara et al., 2013).

Identification of the Intersection Differentially Expressed Genes (DEGs) Among Immune and Stromal Scores

All cases in the training cohort were divided into groups of high and low scores according to the median. DEGs were identified between high and low immune/stromal score groups using "limma" R package (Ritchie et al., 2015), with a cutoff of $|\log_2(\text{fold-change})| > 1$ and false discovery rate (FDR) < 0.05 . "pheatmap" R package was applied to produce heatmaps and clustering of DEGs. Genes that were upregulated in both high immune and stromal scores groups were defined as intersection-upregulated DEGs. Genes that were downregulated in both high

¹<https://tcga.xenahubs.net>

²<https://www.ncbi.nlm.nih.gov/geo/>

³<https://bioinformatics.mdanderson.org/estimate/index.html>

immune and stromal scores groups were taken as intersection-downregulated DEGs. A combination of these two intersection DEGs was the intersection DEGs.

Identification and Validation of the Prognostic Gene Signature

To continue, Kaplan–Meier analysis was performed in the training cohort, to screen for potential prognostic genes from the intersection DEGs identified in the previous step based on overall survival. Only genes with p values < 0.05 in the log-rank test were considered as significant to pass Kaplan–Meier analysis screening. Also, univariate Cox regression analysis was performed on the training cohort to look for prognostic genes from the intersection DEGs with p values < 0.05 . Same as before, only genes that showed significance in the overall survival analysis were considered to pass univariate Cox regression analysis screening.

The genes that passed both Kaplan–Meier and univariate Cox analyses were then entered into the LASSO Cox regression model analysis, which was implemented in the training cohort utilizing R software and the “glmnet” package. Ten-times cross-validations were applied to detect the best penalty parameter lambda (Tibshirani, 1997; Sauerbrei et al., 2007; Friedman et al., 2010; Goeman, 2010). According to the best lambda value, a list of prognostic genes with coefficients was obtained from the gene expression and patients’ survival data. Moreover, the risk score of each patient could be calculated based on the expression level of each prognostic gene and its corresponding coefficient.

Using the median risk score as the cutoff point, the patients in the training cohort were distributed to high-risk or low-risk groups, and Kaplan–Meier analysis was applied to evaluate the survival difference between the two groups. Besides, Cox and ROC analyses were conducted to further assess the prognostic value of the gene signature in the training cohort. Furthermore, the prognostic gene signature was validated in the validation cohort. The same formula was conducted to compute risk scores like that in the training cohort. Kaplan–Meier, Cox, and ROC analyses were implemented as described earlier.

TMB Correlation With the Gene Signature Risk Score

Tumor mutational burden is defined as the total number of non-synonymous mutations per coding area of a tumor genome. Recently, a high TMB has been identified as a genetic signature that is associated with a favorable outcome for immune checkpoint inhibitor therapy (Alexandrov et al., 2013; Gibney et al., 2016; Yuan et al., 2016). The mutation data of LUAD patients were obtained from the TCGA Xena Hub mentioned earlier. The TMB score for each LUAD patient was calculated by the formula as follows (Robinson et al., 2017; Liu J. et al., 2020): $TMB = (\text{total mutation}/\text{total covered bases}) \times 10^6$. The Spearman rank correlation coefficient was applied to assess the correlation between TMB and risk score, further evaluating the prognostic value of the gene signature identified in this study and the possibility of immune checkpoint inhibitor therapy targeting risk score. P value < 0.05 was considered statistically significant.

Gene Set Enrichment Analysis

The Hallmark (v7.1) and C7 (v7.1) gene set collections were downloaded from the Molecular Signatures Database v7.1 download page⁴. GSEA was performed based on the downloaded gene set collections using GSEA software (v4.0.3)⁵. The training cohort was taken for GSEA to reveal the functions and pathways in the DEGs between high-risk and low-risk groups. Only gene sets with $|NES| > 1$, NOM p -val < 0.05 , and FDR q -val < 0.25 were considered significant.

Correlation of Risk Score With the Proportion of 22 Kinds of Tumor-Infiltrating Immune Cells (TICs)

The CIBERSORT calculation method was used to estimate the 22 kinds of TICs abundance distribution of all tumor samples in the training cohort, and then quality filtering was performed. In total, 515 LUAD samples with p value < 0.05 were selected for the following analysis. The correlations between 22 kinds of TICs were examined by Pearson coefficient test. Spearman coefficient test was used for the correlation test between the TICs’ proportion and risk score. The Wilcoxon rank-sum test verified the differentiation of 22 kinds of immune cells between low- and high-risk groups. Besides, based on the 22 TICs’ infiltration volume and survival data of each of the 515 patients, we used the univariate Cox and Kaplan–Meier methods to screen TICs with prognostic significance. Together with the result of correlations between risk score and 22 TICs, we further assessed which kind of immune cells play a role in the prognostic zone of the gene signature identified in this study. P value < 0.05 was a statistically significant threshold.

Statistical Analysis

All statistical calculations were performed in R software. Kaplan–Meier analysis was conducted to check the prognosis difference between groups, along with the p value, which was examined in the log-rank test. Univariate and multivariate Cox proportional hazard regression analyses were conducted to assess the association between risk score and prognosis. The ROC analysis was applied to examine the sensitivity and specificity of survival prediction using the gene signature risk score. An area under the ROC curve (AUC) served as a pointer of prognostic accuracy. The R package “pROC” was used for ROC analysis, and the “delong” method is used to study the significant differences among ROC curves. In addition to noted before, all analyses p value < 0.05 was a statistically significant threshold.

RESULTS

Clinical Characteristics

The flowchart of the present research is shown in **Figure 1**. 515 LUAD cases that came from TCGA-LUAD were taken as the training cohort. The dataset GSE72094 with 442 LUAD

⁴<https://www.gsea-msigdb.org/gsea/downloads.jsp>

⁵<https://www.gsea-msigdb.org/>

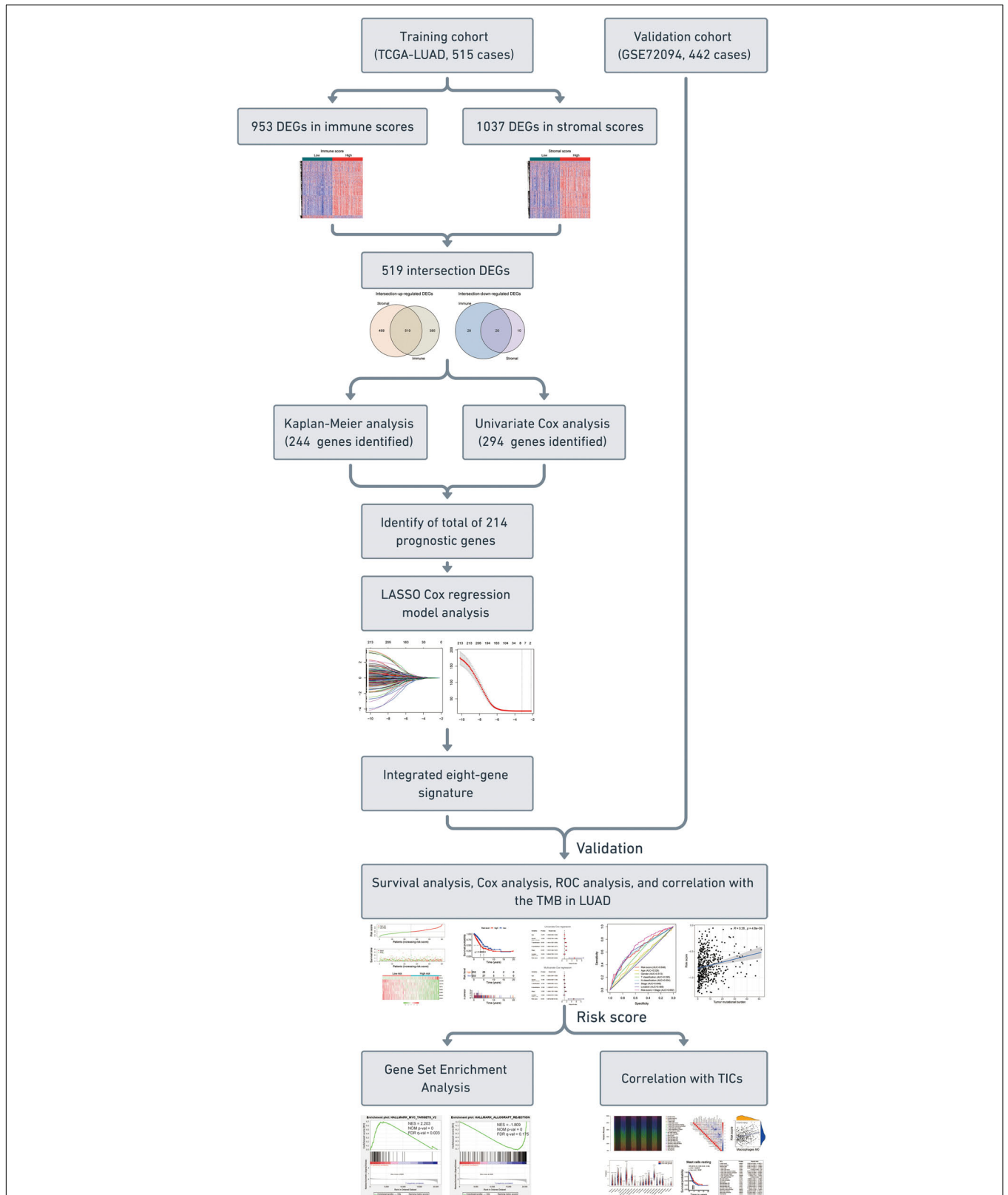


FIGURE 1 | Flow chart of the study. The study was carried out in TCGA-LUAD and GSE72094 cohorts. The training cohort was used to identify prognostic genes. The prognosis analysis was validated in the validation cohort. DEGs, differentially expressed genes; LASSO, the least absolute shrinkage and selection operator Cox regression model; ROC, receiver operating characteristic; TMB, tumor mutational burden; LUAD, lung adenocarcinoma; TICs, tumor-infiltrating immune cells.

patients was used as the validation cohort. The detailed clinical characteristics of both cohorts are summarized in **Table 1**.

Intersection DEGs Based on Immune and Stromal Scores

For identifying the DEGs among immune and stromal scores, cases in the training cohort were divided into groups of high and low scores according to their scores based on the median, and the DEG analysis was performed using the “limma” R package. **Figure 2A** shows a heatmap of 953 DEGs between immune score groups. **Figure 2B** displays a heatmap consisting of 1037 DEGs between stromal score groups. Via integrated bioinformatics analysis, we identified 519 intersection-upregulated DEGs (**Figure 2C**) and 20 intersection-downregulated DEGs (**Figure 2D**). Our subsequent analysis focused on these 539 intersection DEGs.

Construction of Prognostic Signature From the Training Cohort

Kaplan–Meier and univariate Cox regression analyses were performed on 515 LUAD patients in the training cohort to assess the prognostic relationship between the intersection DEGs and overall survival. A total of 244 genes were extracted from the Kaplan–Meier analysis (**Supplementary Table S2**), while 294 genes were identified as significant in the Cox regression analysis (**Supplementary Table S3**). In summary, the 214 genes at the intersection of the two results are defined as potential prognostic genes for the next analysis (**Supplementary Table S4**).

These genes were then subjected to LASSO Cox regression analysis, and regression coefficients were calculated. The coefficient of each gene is illustrated in **Figure 3A**. When eight genes were included, the model achieved the best performance (**Figure 3B**). These genes and corresponding coefficients are shown in **Table 2**.

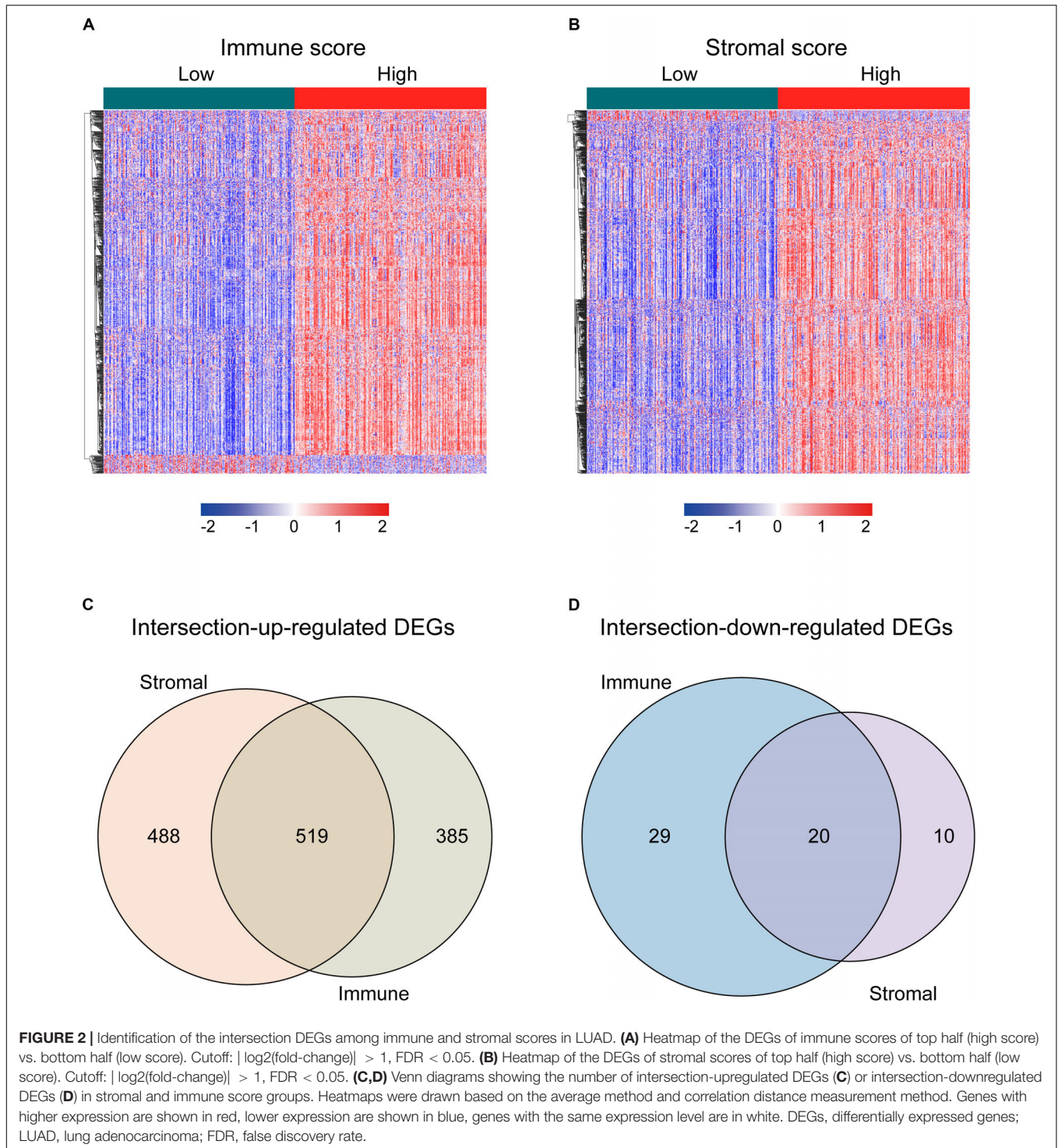
Prognostic Value of the Eight-Gene Signature in the Training and Validation Cohorts

According to the gene expression level, and the risk coefficient of each gene, the risk score of each patient was calculated, which is a linear combination of the expression level of each gene weighted by its multivariate LASSO regression coefficient. The median risk score was the cutoff value for assigning patients to high-risk or low-risk groups. The prognostic value of the risk score was evaluated by comparing the survival differences between the high-risk group and the low-risk group.

The distribution of risk scores and overall survival status and the expression profiles of the eight-gene signature of the patients in the training cohort are plotted in **Supplementary Figure S1A**. As shown in the figure, there are more deceased in high-risk patients, and the survival time is shorter than that of low-risk patients. The heat map shows that INSL4 was highly expressed in high-risk patients, whereas KLRB1, INSL4, ACSM5, SCN7A, P2RX1, MS4A1, STAP1, and IKZF3 were under-expressed in high-risk patients. Also, we examined the performance of this eight-gene signature in predicting overall

TABLE 1 | Clinical characteristics of patients involved in the study.

Characteristics	Training cohort (TCGA-LUAD, n = 515)	Validation cohort (GSE72094, n = 442)
Age at diagnosis, years		
<60	138 (26.80%)	60 (13.57%)
≥60	363 (70.49%)	361 (81.67%)
Unknown	14 (2.72%)	21 (4.75%)
Gender		
Female	269 (52.23%)	240 (54.30%)
Male	232 (45.05%)	202 (45.70%)
Unknown	14 (2.72%)	0 (0.00%)
Stage		
I	268 (52.04%)	265 (59.95%)
II	119 (23.11%)	69 (15.61%)
III	80 (15.53%)	63 (14.25%)
IV	26 (5.05%)	17 (3.85%)
Unknown	22 (4.27%)	28 (6.33%)
T classification		
T1	171 (33.20%)	NA
T2	263 (51.07%)	NA
T3	45 (8.74%)	NA
T4	19 (3.69%)	NA
Unknown	17 (3.30%)	NA
N classification		
N0	323 (62.72%)	NA
N1	94 (18.25%)	NA
N2	70 (13.59%)	NA
N3	2 (0.39%)	NA
Unknown	26 (5.05%)	NA
M classification		
M0	332 (64.47%)	NA
M1	25 (4.85%)	NA
Unknown	158 (30.68%)	NA
Tumor location		
Right	295 (57.28%)	NA
Left	194 (37.67%)	NA
Unknown	26 (5.05%)	NA
Race		
White	NA	399 (90.27%)
Other	NA	18 (4.07%)
Unknown	NA	25 (5.66%)
Kras_status		
Wild type	NA	288 (65.16%)
Mutant	NA	154 (34.84%)
Unknown	NA	0 (0.00%)
Egfr_status		
Wild type	NA	395 (89.37%)
Mutant	NA	47 (10.63%)
Unknown	NA	0 (0.00%)
Stk11_status		
Wild type	NA	374 (84.62%)
Mutant	NA	68 (15.38%)
Unknown	NA	0 (0.00%)
Tp53_status		
Wild type	NA	331 (74.89%)
Mutant	NA	111 (25.11%)
Unknown	NA	0 (0.00%)



survival in the validation cohort. As shown in **Supplementary Figure S1B**, in the high-risk group, more death happened, and shorter survival time gained. The pattern is consistent with that in the training cohort.

Kaplan–Meier survival analysis showed that patients in the high-risk group were associated with a weak overall survival trend ($p = 0.00014$, **Figure 4A**) and an unfavorable 5-year prognosis

($p < 0.0001$, **Figure 4B**) in the training cohort. To confirm the efficacy of the eight-gene signature in predicting prognosis in LUAD patients, we examined the eight-gene signature in the validation cohort. Using the same classification method as before, the patients were divided into high-risk and low-risk groups according to the median risk score. Consistent with previous results, patients in the high-risk groups showed significantly

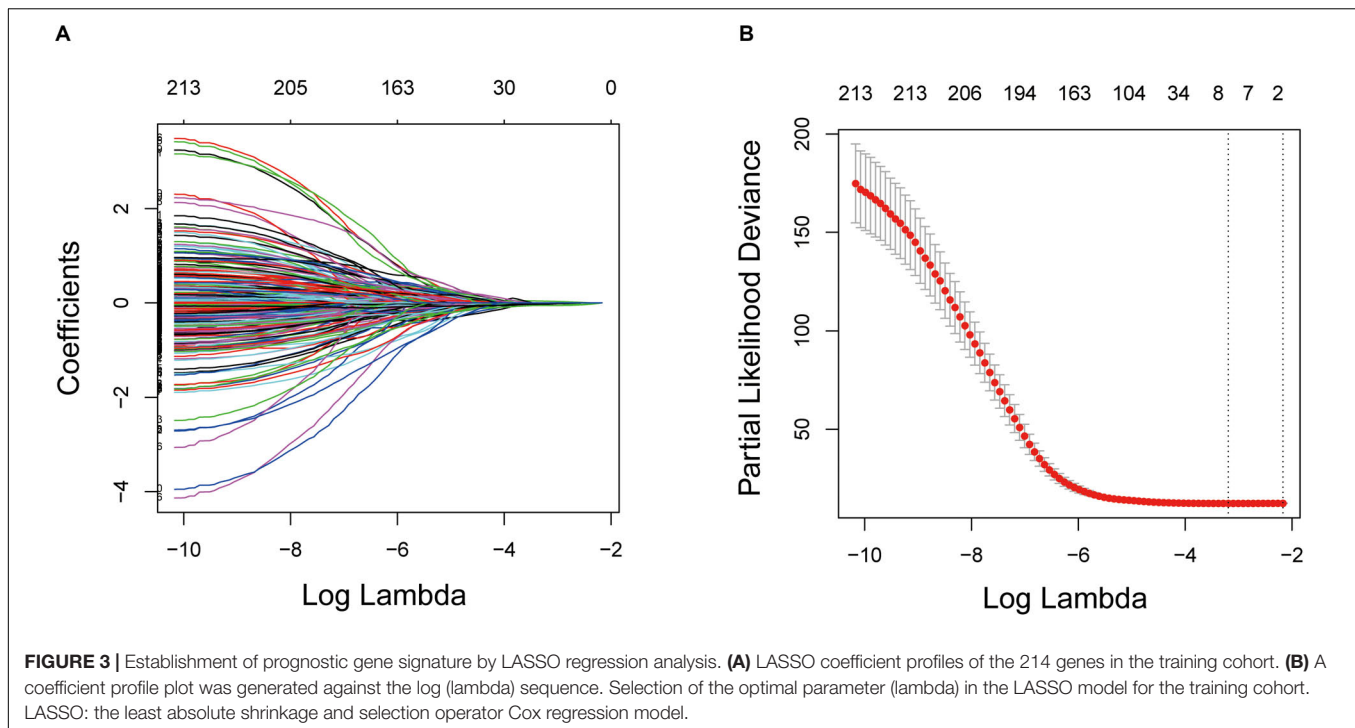


TABLE 2 | Genes in the prognostic gene signatures.

Gene symbol	Description	Risk coefficient
KLRB1	Killer Cell Lectin Like Receptor B1	-0.028631486
INSL4	Insulin Like 4	0.034674929
ACSM5	Acyl-CoA Synthetase Medium Chain Family Member 5	-0.056685884
SCN7A	Sodium Voltage-Gated Channel Alpha Subunit 7	-0.007784834
P2RX1	Purinergic Receptor P2X 1	-0.018937089
MS4A1	Membrane Spanning 4-Domains A1	-0.025293606
STAP1	Signal Transducing Adaptor Family Member 1	-0.012745735
IKZF3	IKAROS Family Zinc Finger 3	-0.020045496

worse prognosis ($p < 0.0001$, **Figure 4C**) and 5-year outcome ($p < 0.0001$, **Figure 4D**) than patients in the low-risk groups.

Univariate and multivariate Cox analyses were performed in the training and validation cohorts based overall survival, using the available co-variables including risk score, age, gender, T classification, N classification, tumor stage, tumor location, race, KRAS status, EGFR status, STK11 status, or TP53 status to confirm whether the prognostic capacity of our eight-gene signature was independent from the clinico-pathologic characteristics. In the training cohort, both univariate and multivariate Cox regression analyses indicated that the eight-gene signature was a powerful variable associated with overall survival (HR 5.002, 95% CI 2.836–8.823, $p < 0.001$, and HR 4.897, 95% CI 2.626–9.133, $p < 0.001$, respectively; **Figure 5A**). Consistent with that in the training cohort, the eight-gene signature also displayed pronounced capability in the validation cohort in predicting

overall survival (**Figure 5B**). These results proved that the eight-gene signature was to be a strong and independent variable.

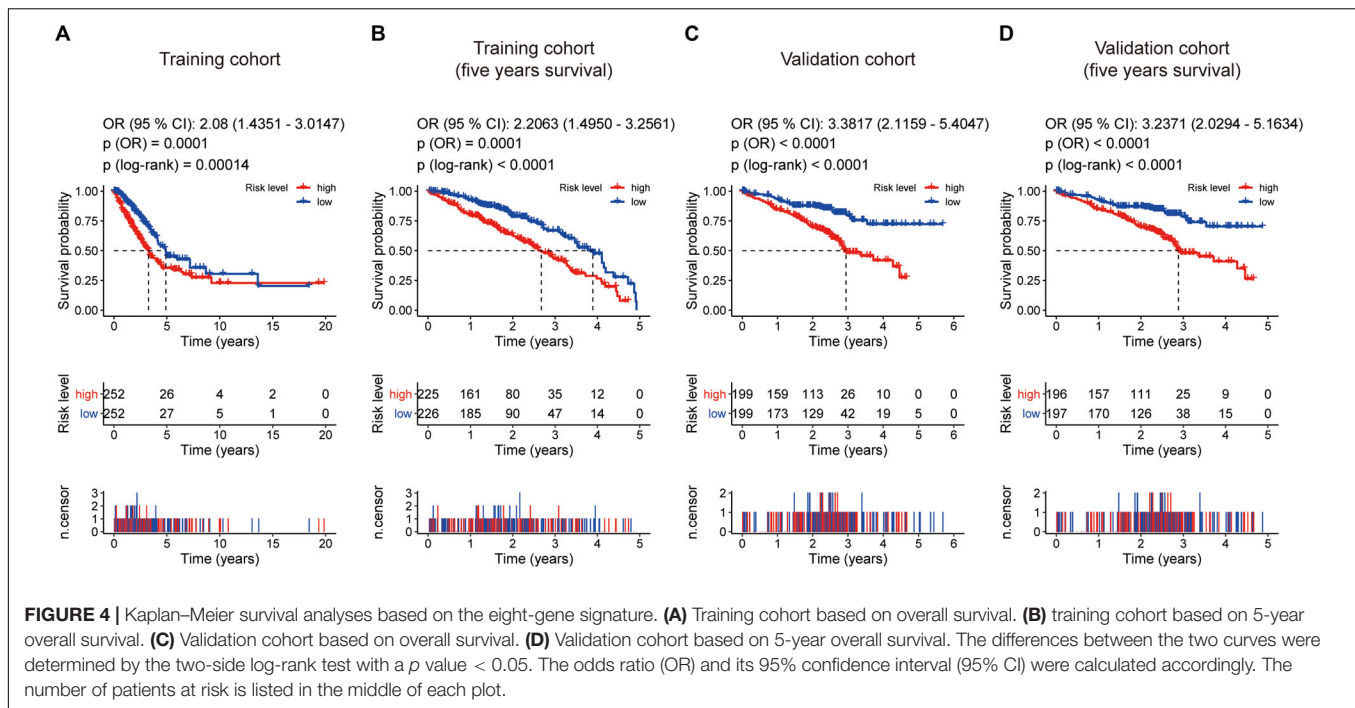
Subsequently, we conducted ROC analyses to assess how the eight-gene signature could behave in predicting prognosis. As shown in **Figure 6A**, the AUC of the eight-gene risk score model performed on overall survival in the training cohort was 0.648, which was equal to tumor stage, but superior to those of age, gender, T classification, N classification, and tumor location (0.529, 0.513, 0.595, 0.634, and 0.489, respectively). When combined risk score and tumor stage, the AUC could reach 0.692. Consistently, in the prediction model of overall survival predicted in the validation cohort, the eight-gene signature risk score also showed the best ability with AUC = 0.647, among other factors (**Figure 6B**). When putting risk score and tumor stage together for a diagnosis, the AUC was 0.680.

TMB Correlation With the Gene Signature Risk Score

Furthermore, we performed correlation analyses to assess the relationship between the eight-gene signature and LUAD TMB. Spearman test was used to assess the correlation between the TMB score and the risk score. The results showed that the eight-gene signature was significantly positively correlated with TMB ($R = 0.26$, $p = 4.8e-09$, **Supplementary Figure S2**), further, revealing that the risk score could potentially reflect the characteristics and performance of TMB in tumors.

Gene Set Enrichment Analysis With the Eight-Gene Signature

Given the negative correlation between the level of the eight-gene signature risk score and the prognosis of LUAD patients,



the GSEA was conducted between the high and the low-risk groups. As displayed in **Figure 7A** and **Supplementary Tables S5, S6**, gene sets of HALLMARK collection were mainly enriched in pathways related to MYC, E2F, G2M, and oxidative phosphorylation in the high-risk group, and KRAS and inflammatory response in the low-risk group. For C7 collection defined by the Molecular Signatures Database, gene sets related to TNF, CD8, IL6, and CD4 were significantly enriched in the high-risk score group, whereas enrichments regarding mononuclear, CD4, CD8, macrophage, and FOXP3 were mainly seen in the low-risk score group (**Figure 7B** and **Supplementary Tables S7, S8**).

Correlation of Risk Score With the Proportion of Tumor-Infiltrating Immune Cells (TICs)

To further check the correlation between the risk score and the immune microenvironment, as shown in **Supplementary Figure S3**, we used the CIBERSORT algorithm to analyze the proportion of tumor-infiltrating immune subpopulations and constructed 22 immune cell profiles in LUAD samples. Combining the results of correlation analysis (**Figure 8A** and **Supplementary Table S9**) and difference analysis (**Figure 8B**), a total of 13 TICs were associated with the eight-gene signature risk score (**Figure 8C**). Among them, NK cells resting, plasma cells, B cells naive, neutrophils, dendritic cells activated, NK cells activated, and macrophages M0 were positively correlated with risk score, whereas B cells memory, Mast cells resting, macrophages M1, dendritic cells resting, T cells gamma delta, and T cells CD8 were negatively correlated with risk score.

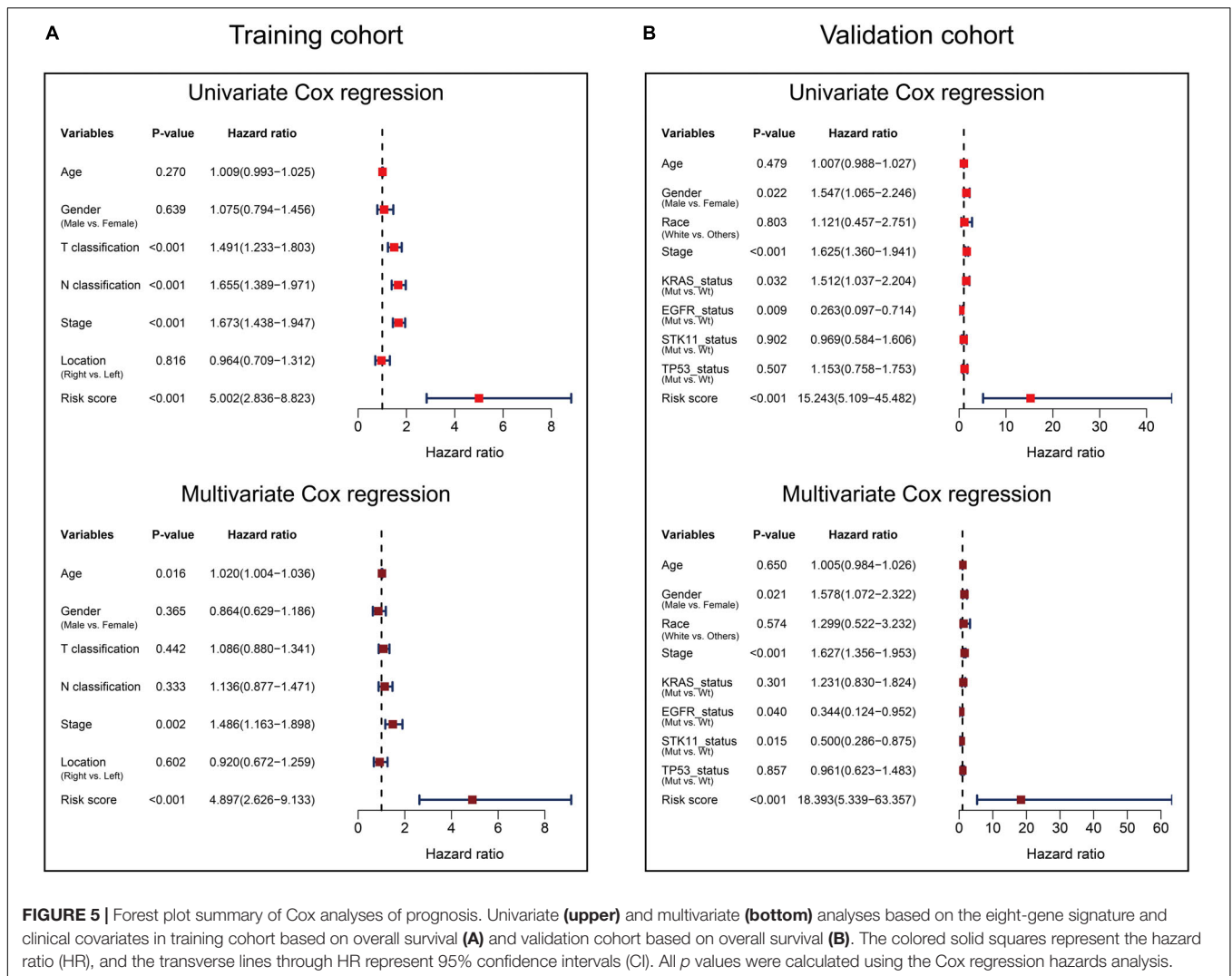
Moreover, to assess the prognostic value of each TIC, we performed univariate Cox and Kaplan–Meier analyses based

on the 22 TICs' infiltration volume and survival data, finding Mast cells resting, neutrophils, B cells naive, and macrophages M0 pronounced predicting the overall survival in Kaplan–Meier analyses (**Figure 9A**), whereas Mast cells activated, Mast cells resting, B cells naive, dendritic cells activated, and T cells regulatory (Tregs) have a significant prognosis value in the univariate Cox regression (**Figure 9B**). From these survival analyses results, Mast cells resting and B cells naive had potential prognostic value for LUAD patients.

The aforementioned results revealed that Mast cells resting and B cells naive show not only prognostic value in LUAD but also have significant correlations with the risk score. Thus, the significant infiltration with Mast cells resting and B cells naive potentially acted as one of the critical factors that the eight-gene signature holds to influence the outcome of LUAD pronounced.

DISCUSSION

In the present study, we built a LUAD prognostic signature by comprehensively analyzing the TCGA and GEO. By discovering the DEGs among tumor microenvironment and investigating the potential prognosis of DEGs using Kaplan–Meier, univariate Cox analyses, and the LASSO Cox regression model in the training cohort, we obtained an eight-gene signature that was pronounced related to outcome. By applying this signature in the training cohort, statistical significance was observed in univariate and multivariate Cox analysis, ROC analysis, and Kaplan–Meier curve between high-risk and low-risk groups. The prognostic ability of the eight-gene signature was also validated in the validation cohort, showing the broadness and effectiveness of the eight-gene signature in predicting LUAD prognosis.



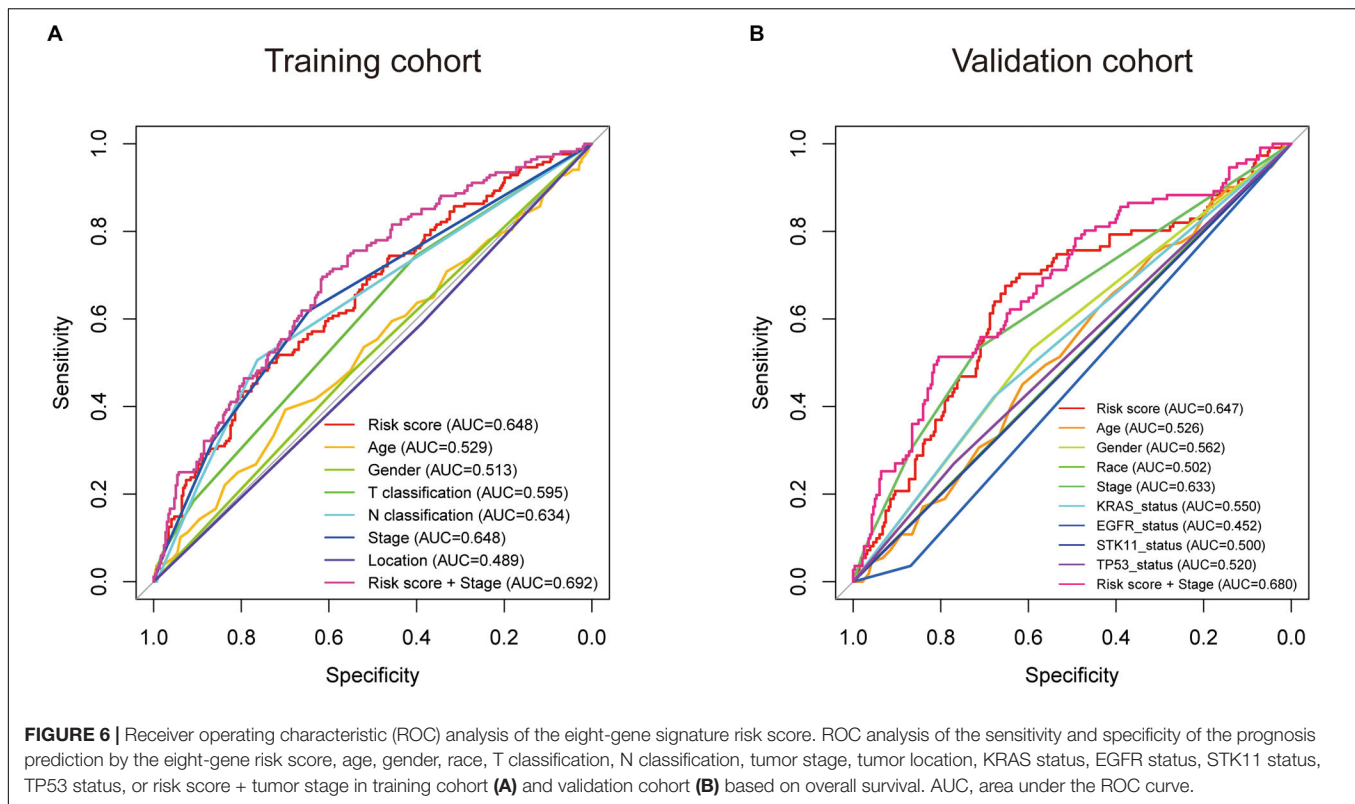
Then the GSEA and immune infiltration analyses revealed the significant pathways that linked to the eight-gene signature and the critical role that Mast cells resting and B cells naive potentially played supporting the eight-gene signature holds to influence the outcome of LUAD significantly. For research in the gene signature of LUAD, we are the first to combine tumor microenvironment scores, double screening (Kaplan–Meier and univariate Cox methods), and LASSO for training and introduce TMB for gene signature examination. What is more, we found vital immune cells that regulate the LUAD outcome among our eight-gene signature, 22 TICs, and the prognosis of 22 TICs. Such work we have done aimed to guide future research in LUAD.

After we constructed the eight-gene signature, we first confirmed its capacity to distinguish the prognosis of patients effectively. As shown in **Supplementary Figure S1A**, the high-risk zone not only counted more deaths but also the patients in it presented a shorter survival time than that in the low-risk zone. Moreover, the heatmap indicated that each of these eight genes had a differential expression pattern between the low- and high-risk groups. Importantly, this eight-gene signature also owned

pronounced performance in the validation cohort for predicting overall survival (**Supplementary Figure S1B**).

In addition, we examined the prognostic value of the eight-gene signature by Kaplan–Meier analysis in the training cohort and the validation cohort based on overall survival and 5-year outcome, finding all of its predicting ability to be significant in LUAD patients (**Figure 4**). Furthermore, univariate and multivariate Cox analyses were performed in the two cohorts to confirm whether our eight-gene signature can be independent from other variables in predicting LUAD outcome. As plotted in **Figure 5**, regardless of the training cohort or validation cohort, and whether it is univariate or multivariate Cox regression analysis, the variable of risk score was always statistically significant. The results verified the predictive ability of the risk score and its independence.

To further assess the predictive power of this eight-gene signature, ROC analysis was conducted. In diagnostic tests, we can use AUC to check the accuracy and determine the predictive capacity of biomarkers (Hanley and McNeil, 1982). ROC analysis indicated that the AUC of the eight-gene signature



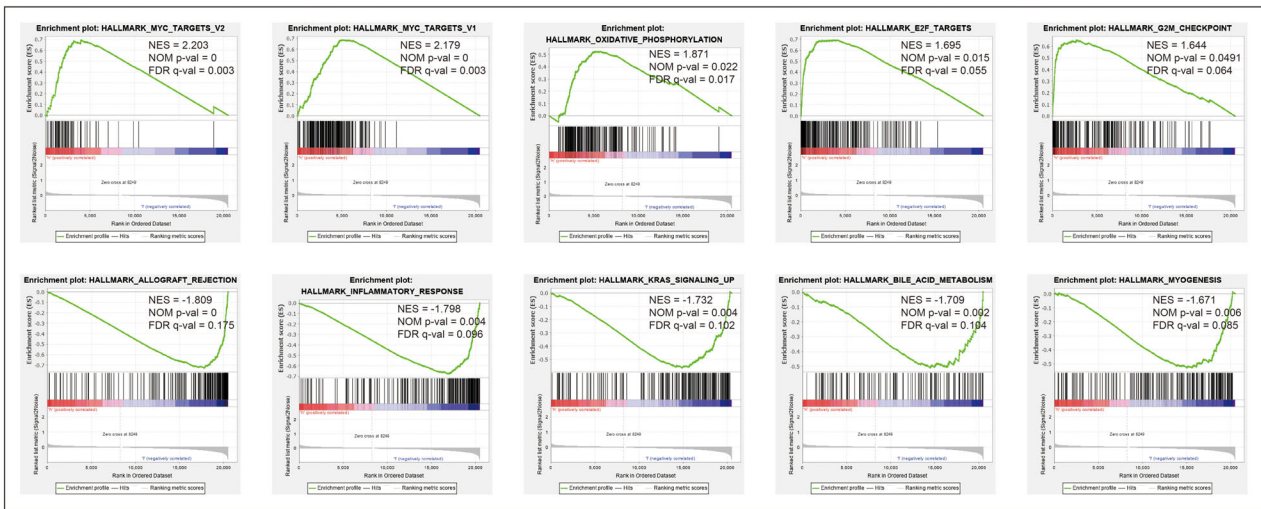
stayed superior to other factors and could enhance the diagnostic ability when combined with the tumor stage (Figure 6). These ROC results again suggested that our signature might strengthen the predictive accuracy of prognosis in LUAD.

Our signature was composed of eight genes, which were INSL4, SCN7A, STAP1, P2RX1, IKZF3, MS4A1, KLRB1, and ACSM5, respectively. In the signature model, INSL4 was the unfavorable genes for the outcome, whereas other genes presented protective function on the prognosis of LUAD patients. INSL4 is a member of the insulin/IGF/relaxin superfamily that is restrictively expressed in the placenta (Laurent et al., 1998; Lu et al., 2005; Pollak, 2012). INSL4 upregulation was previously identified in a breast cancer cell subclone with increased invasiveness through *in vitro* selection (Brandt et al., 2002, 2005). INSL4 was also discovered as a novel target downstream of LKB1 deficiency (Yang et al., 2018). INSL4 signaling is a potential vital for LKB1-inactivated non-small-cell lung carcinoma, and its introduction helps to develop novel and effective anti-tumor strategies (Yang et al., 2018). SCN7A encodes a voltage-dependent sodium channel of the excitable membrane, and it is reported to be downregulated in colorectal carcinoma at the mRNA level (Ostasiewicz et al., 2016). Recently, a new comprehensive bioinformatics study showed that low expression of SCN7A in non-small-cell lung cancer patients was associated with poor survival (Liu Y. et al., 2020). STAP1 is a relatively unknown protein and recruits signaling proteins to receptor tyrosine kinases (Steeghs et al., 2018). STAP1 is reported to be a docking protein downstream of Tec protein tyrosine kinase, which is involved in the B-cell receptor

signaling (Steeghs et al., 2018). However, the involvement of the STAP1 protein in cancer thus far remains elusive. The P7 receptor family consists of seven ionotropic P2X receptors (Adriaenssen and Timmermans, 2004). P2X receptors have been specifically described in many immune cells (including platelets, lymphocytes, and macrophages), where P2X receptors are involved in the regulation of multiple functions, including platelet aggregation, apoptosis, migration, and cytokine release (Wareham et al., 2009). Wareham et al. (2009) confirmed the presence of functional P2RX1 in LAD2 cells and human lung mast cells. With the development of leukemia, the expression of P2RX1 in splenic macrophages increases. However, how P2RX1 affects lung cancer remains unclear (Chen et al., 2014). IKZF3 is a member of the Ikaros family of transcription factors, which are the crucial regulators of lymphoid differentiation (Menezes et al., 2013). IKZF3 expressions were associated with longer median progression-free survival and overall survival in multiple myeloma patients (Pourabdollah et al., 2016). IKZF3 expression also was a favorable indicator of multiple myeloma patients who received lenalidomide-based therapy (Pourabdollah et al., 2016). The IKZF3 mutation has been shown to exist in the chronic phase and blast crisis of chronic myeloid leukemia, indicating the potential role of this gene in myeloid leukemia (Menezes et al., 2013). Early lung cancer was found to have high IKZF3 expression, which indicates that IKZF3 induction often occurs before the clinical detection of lung cancer (Li et al., 2014; Terada and Liu, 2014). The MS4A1 gene located on 11q12 encodes a member of the membrane-spanning 4 domains, subfamily A, B lymphocyte antigen CD20, which plays a role

A

Top enriched gene sets in HALLMARK collection



B

Top enriched gene sets in C7 collection

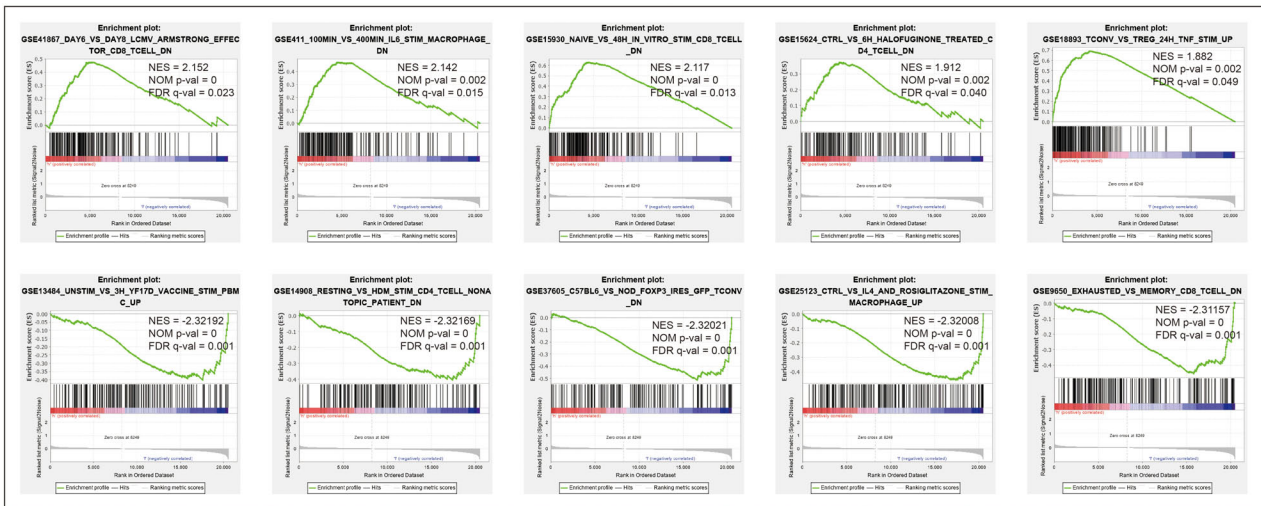
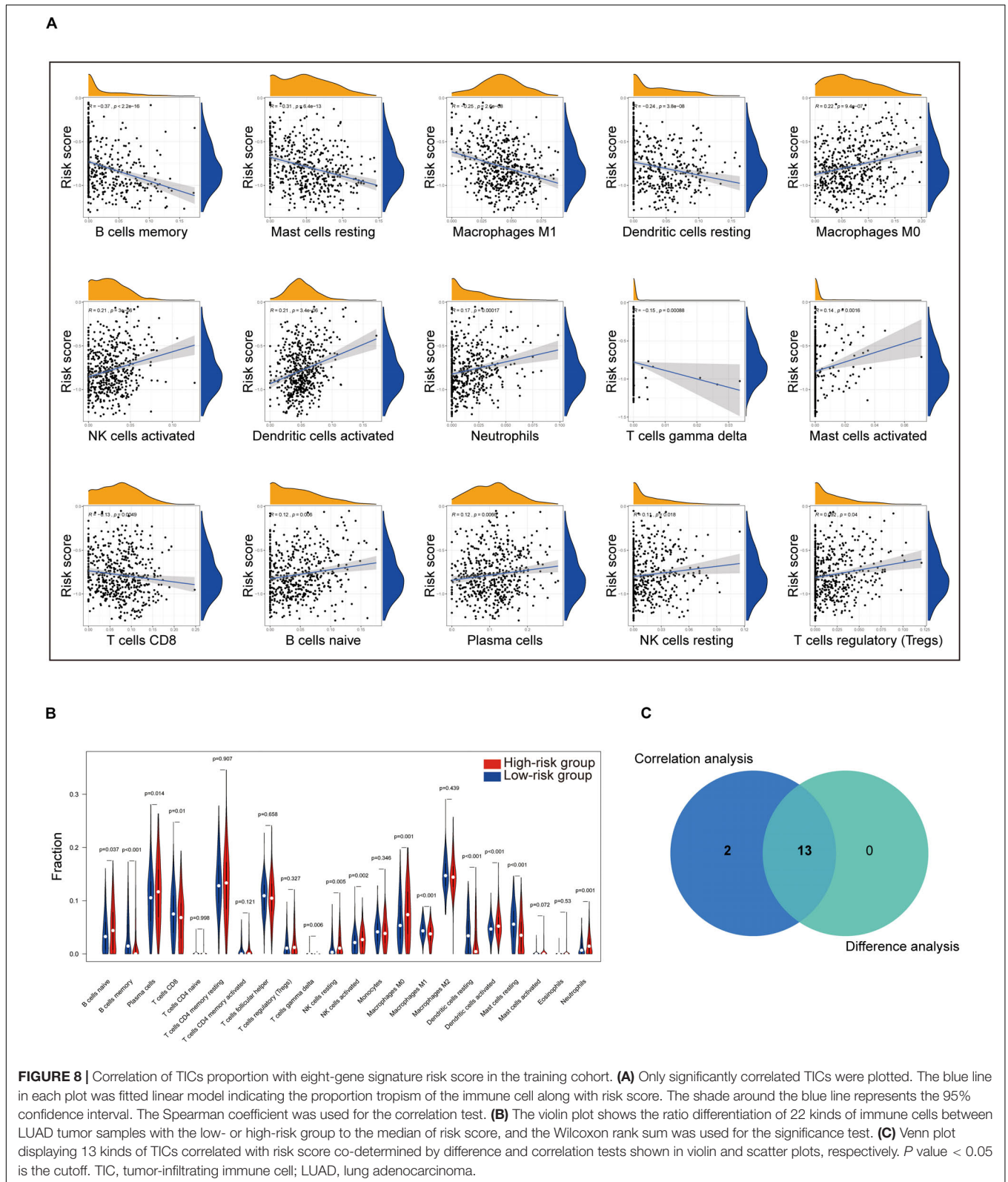


FIGURE 7 | Gene set enrichment analysis. **(A)** Top enriched gene sets annotated by the HALLMARK collection between the high- and low-risk groups in the training cohort. **(B)** Top enriched gene sets annotated by the C7 collection between the high- and low-risk groups in the training cohort. Only gene sets with $|NES| > 1$, $NOM\ p\text{-val} < 0.05$, and $FDR\ q\text{-val} < 0.25$ were considered significant.

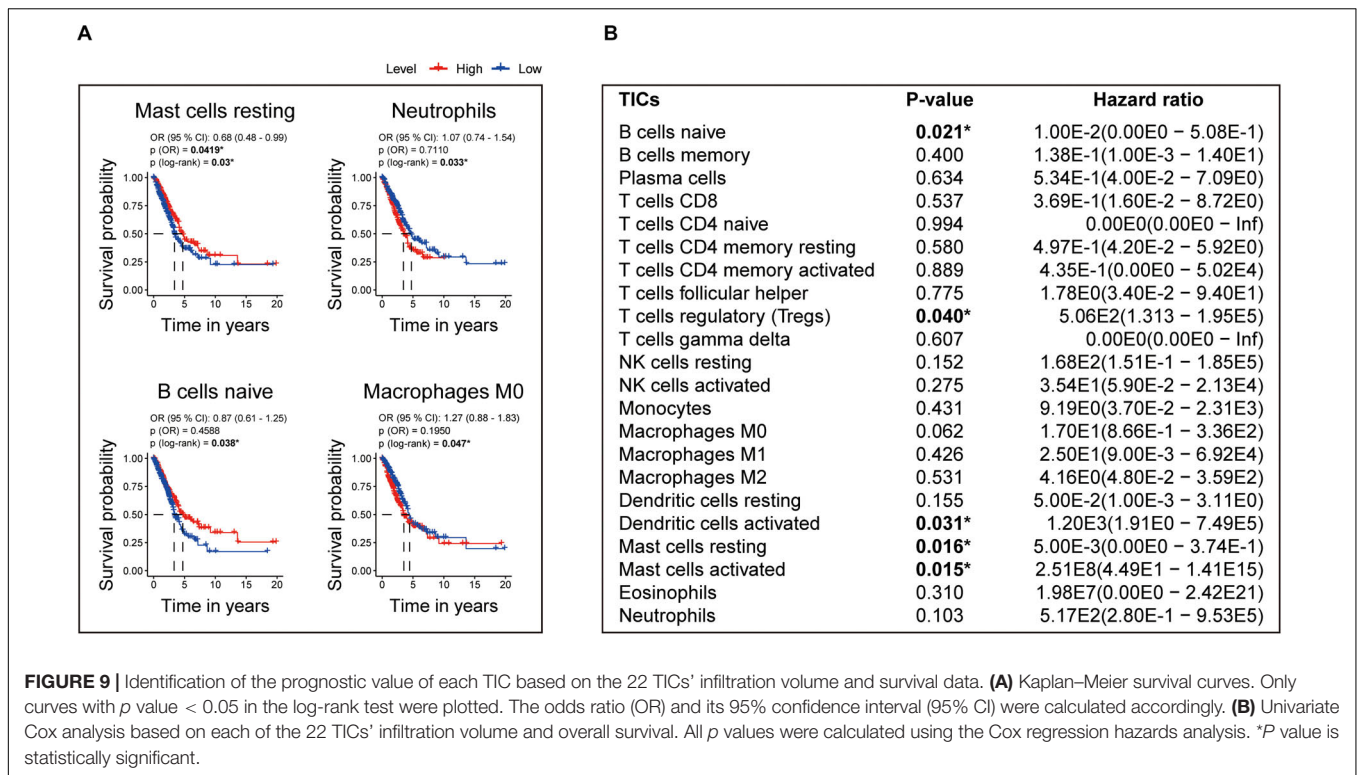
in the differentiation of B lymphocytes into plasma cells (Tam et al., 2008). Also, it has been reported that MS4A1 plays a vital role in the apoptosis of B-cell lymphoma Ramos cells (Kawabata et al., 2013). In one recent comprehensive bioinformatics research, MS4A1 was confirmed to be involved in the tumor microenvironment and TMB, and low expression of MS4A1 was related to the poor prognosis of ovarian cancer (Fan et al., 2020). A study confirmed that the expression of CD20 stromal lymphocytes led to the occurrence of MS4A1 dysregulation in

asbestos-related lung squamous cell carcinoma (Wright et al., 2012). The KLRB1 gene encodes the CD161 receptor in natural killer cells. The gene is also expressed in the NKT cells (Pleshkan et al., 2007). The expression of KLRB1 was most frequently associated with favorable outcomes (Gentles et al., 2015). KLRB1 has been shown to play an essential prognostic role in pancreatic research (Gentles et al., 2015). Gentles et al. (2015) found that expression of favorably prognostic gene KLRB1 largely reflects tumor-associated leukocytes. Also, Pleshkan reported



that KLRB1 gene expression was suppressed in tumor tissues in 68% of patients with non-small-cell lung cancer in his study (Pleshkan et al., 2007). ACSM5 is a protein-coding gene. Among

its related pathways are Cytochrome P450 – arranged by substrate type and metabolism. Gene Ontology annotations related to this gene include GTP binding and butyrate-CoA ligase activity. An



important paralog of ACSM5 is ACSM4 (Stelzer et al., 2016). The role of ACSM5 in tumor progression is still in the initial stage of research and is worthy of further exploration. In the field of lung cancer or oncology, the genes in our eight-gene signature have not been extensively studied. However, the eight-gene signature has a significant role in predicting and diagnosing LUAD in our research, indicating it or each gene in it may be the potential specific directions for future research on LUAD.

Tumor mutational burden is defined as the total number of non-synonymous mutations in tumors, and it is an emerging independent indicator of the outcome of immunotherapy treatment of multiple tumor types (including lung cancer) (Carbone et al., 2017; Chalmers et al., 2017; Le et al., 2017; Yarchoan et al., 2017; Rizvi et al., 2018). A high mutation load correlates with an immunogenic tumor microenvironment with increased expression of tumor-specific neo-antigens that can be targeted by activated immune cells (Schumacher and Schreiber, 2015; Kim and Chen, 2016). According to reports, the relationship between TMB and prognosis of many cancers is not clear, and there are still many unknowns (Owada-Ozaki et al., 2018; Wu et al., 2019). For most cancer histologies, a higher TMB is pertinent to improved survival in patients receiving immune checkpoint inhibitors, although the definition of high TMB, or the cutpoints, varies markedly between diverse tumors (Goodman et al., 2017; Lin et al., 2020). Returning to our research, we found that the risk score had a weak relationship with TMB ($p = 4.8e-09$, $R = 0.26$, **Supplementary Figure S2**), which further measured the connection between our eight-gene signature and LUAD, and also provided more potential breakthroughs and inspiration for tumors TMB targeted therapies.

The GSEA found that gene sets enriched in HALLMARK collection in the high-risk group mainly related to MYC, E2F, and G2M checkpoint. For the C7 collection, we found that CD4, CD8, TNF, and IL 6 were mostly seen in the enriched gene sets. MYC is a family of regulatory genes and proto-oncogenes that encode transcription factors. It is often amplified in cells grown from lung tumors, and its transfection can enhance the *in vitro* proliferation and agar cloning of human small cell lung cancer cells (Barr et al., 2000). E2F is a group of genes that encodes a family of transcription factors (TF) in higher eukaryotes (Sun et al., 2018). Evidence from cell lines, mouse models, and human tissues indicates that TFs are implicated in lung cancer tumorigenesis (Sun et al., 2018). The G2M DNA damage checkpoint is an important cell cycle checkpoint in eukaryotes, ensuring that cells will not trigger mitosis until damaged or incompletely copied DNA is fully repaired (Lobrich and Jeggo, 2007). Deficient G2M checkpoints are associated with increased lung cancer risk (Xing et al., 2007). These results elaborated on the ways and means of eight-gene signature to participate in LUAD, which can help future targeted therapy research.

In addition, the TICs' analysis based on the CIBERSORT algorithm found that Mast cells resting and B cells naive have prognostic value in LUAD, and a significant correlation with the risk score, indicating that the infiltration of Mast cells resting and B cells naive play potential roles influencing the prognostic capacity of the eight-gene signature. Mast cells are immune cells that accumulate in the tumors and their microenvironment during disease progression (Maciel et al., 2015). Mast cells play an essential role in type I hypersensitivity and also during the early

stages of the innate immune response to pathogens (Campillo-Navarro et al., 2014). Mast cells can trigger different mechanisms that contribute to the homeostasis and adequate function of the lungs (Campillo-Navarro et al., 2014). In 2000, Imada reported that stage I LUAD patients had an increased number of mast cells in tumor areas around blood vessels, and those patients with higher mast cell counts had a worse prognosis (Imada et al., 2000). B cells infiltrating lung cancer have a unique role in anti-tumor immunity. Recent studies have shown that proliferating B cells can be observed in approximately 35% of lung cancers (Gottlin et al., 2011). Besides, B lymphocytes that infiltrate the tumor can be observed at all stages of human lung cancer development (Dieu-Nosjean et al., 2014), and their performance differs between stage and histological subtype (Banat et al., 2015; Kurebayashi et al., 2016), indicating that B cells play a crucial role in the development of lung cancer (Wang et al., 2019). From our results, Mast cells resting and B cells naive have the potential to target the eight-gene signature for the treatment of LUAD, and more efforts need to be implemented to validate our results further.

Our research also has some limitations. The eight-gene signature came from retrospective data, and more prospective data are needed for proving the clinical utility of it. Also, because of the limited clinical characteristics of patients included in the TCGA cohort, we could not perform specific clinical subgroup analyses. Besides, there is currently no wet experimental data explaining the relationship between these eight genes and their mechanism in LUAD samples. Therefore, between the eight-gene signature and the prognosis of LUAD, more effort is needed to clarify the potential relationship.

CONCLUSION

In conclusion, our research defined a robust eight-gene signature in LUAD. This signature was related to the prognosis of LUAD and can accurately identify the prognostic risk of patients. Notably, we evaluated the reliability and accuracy of the signature by examining in a validation cohort and determined the crucial roles of Mast cells resting and B cells naive in the prognosis of the gene signature, which could potentially promote the development of new therapies for LUAD treatment.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: TCGA: <https://portal.gdc.cancer.gov/>; GEO: <https://www.ncbi.nlm.nih.gov/geo/>.

AUTHOR CONTRIBUTIONS

CM organized and wrote the article. HL produced the figures and visualized the data. JC revised the article. XZ, JZ, YZ, and ZF contributed to the literature search for the article. All authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité – Universitätsmedizin Berlin.

ACKNOWLEDGMENTS

HL and CM thank Zhengzhou University Overseas Virtual Research Institute. CM thanks China Scholarship Council (No. 201708410121).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.571641/full#supplementary-material>

Supplementary Figure 1 | Characteristics of the eight-gene signature. **(Upper and middle)** The distribution of risk score and patient's survival time, as well as status for training cohort based on overall survival **(A)** and validation cohort based on overall survival **(B)**. The black dotted line is the median cutoff dividing patients into low-risk and high-risk groups. **(Bottom)** Heatmap of the eight-gene expression profiles in prognostic signature for training cohort based on overall survival **(A)** and validation cohort based on overall survival **(B)**.

Supplementary Figure 2 | The correlation between the eight-gene signature and tumor mutational burden in LUAD. The blue line in each plot was fitted linear model indicating the proportion tropism of the copy number along with risk score. The shade around the blue line represents the 95% confidence interval. The Spearman coefficient was used for the correlation test. LUAD, lung adenocarcinoma.

Supplementary Figure 3 | TIC profile in tumor samples in the training cohort and correlation analysis. **(A)** Bar plot showing the proportion of 22 kinds of TICs in LUAD tumor samples in the training cohort. The column names of the plot were sample ID. **(B)** Heatmap showing the correlation between 22 kinds of TICs. The numeric and shade of each small color box indicate the coefficient between two kinds of cells. X shape covered coefficient is not statistically significant. The Pearson coefficient was used for the significance tests. P value < 0.05 is the cutoff. TIC, tumor-infiltrating immune cell; LUAD, lung adenocarcinoma.

Supplementary Table 1 | Immune and stromal scores of each sample of the TCGA-LUAD cohort.

Supplementary Table 2 | Two hundred and forty-four genes were significantly predicting prognosis of LUAD patients by Kaplan–Meier analysis ($p < 0.05$).

Supplementary Table 3 | Two hundred and ninety-four genes were significantly predicting the prognosis of LUAD patients by univariate Cox regression analysis ($p < 0.05$).

Supplementary Table 4 | Two hundred and fourteen genes were significantly predicting the prognosis of LUAD patients by both Kaplan–Meier and univariate Cox regression analyses ($p < 0.05$).

Supplementary Table 5 | Enriched gene sets in HALLMARK collection in high-risk group ($|NES| > 1$, $NOM\ p\text{-val} < 0.05$, and $FDR\ q\text{-val} < 0.25$).

Supplementary Table 6 | Enriched gene sets in HALLMARK collection in low-risk group ($|NES| > 1$, $NOM\ p\text{-val} < 0.05$, and $FDR\ q\text{-val} < 0.25$).

Supplementary Table 7 | Enriched gene sets in C7 collection in high-risk group ($|NES| > 1$, $NOM\ p\text{-val} < 0.05$, and $FDR\ q\text{-val} < 0.25$).

Supplementary Table 8 | Enriched gene sets in C7 collection in low-risk group ($|NES| > 1$, $NOM\ p\text{-val} < 0.05$, and $FDR\ q\text{-val} < 0.25$).

Supplementary Table 9 | Correlations of risk score with 22 kinds of TICs.

REFERENCES

- Adriaensen, D., and Timmermans, J. P. (2004). Purinergic signalling in the lung: important in asthma and COPD? *Curr. Opin. Pharmacol.* 4, 207–214. doi: 10.1016/j.coph.2004.01.010
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Alsaab, H. O., Sau, S., Alzhrani, R., Tatiparti, K., Bhise, K., Kashaw, S. K., et al. (2017). PD-1 and PD-L1 checkpoint signaling inhibition for cancer immunotherapy: mechanism, combinations, and clinical outcome. *Front. Pharmacol.* 8:561. doi: 10.3389/fphar.2017.00561
- Banat, G. A., Tretny, A., Pullamsetti, S. S., Wilhelm, J., Weigert, A., Olesch, C., et al. (2015). Immune and inflammatory cell composition of human lung cancer stroma. *PLoS One* 10:e0139073. doi: 10.1371/journal.pone.0139073
- Barr, L. F., Campbell, S. E., Diette, G. B., Gabrielson, E. W., Kim, S., Shim, H., et al. (2000). c-Myc suppresses the tumorigenicity of lung cancer cells and down-regulates vascular endothelial growth factor expression. *Cancer Res.* 60, 143–149.
- Bi, K. W., Wei, X. G., Qin, X. X., and Li, B. (2020). BTK has potential to be a prognostic factor for lung adenocarcinoma and an indicator for tumor microenvironment remodeling: a study based on TCGA data mining. *Front. Oncol.* 10:424. doi: 10.3389/fonc.2020.00424
- Brandt, B., Kemming, D., Packeisen, J., Simon, R., Helms, M., Feldmann, U., et al. (2005). Expression of early placenta insulin-like growth factor in breast cancer cells provides an autocrine loop that predominantly enhances invasiveness and motility. *Endocr. Relat. Cancer* 12, 823–837. doi: 10.1677/erc.1.00975
- Brandt, B., Roetger, A., Bidart, J. M., Packeisen, J., Schier, K., Mikesch, J. H., et al. (2002). Early placenta insulin-like growth factor (pro-EPIL) is overexpressed and secreted by c-erbB-2-positive cells with high invasion potential. *Cancer Res.* 62, 1020–1024.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492
- Campillo-Navarro, M., Chavez-Blanco, A. D., Wong-Baeza, I., Serafín-López, J., Flores-Mejía, R., Estrada-Parra, S., et al. (2014). Mast cells in lung homeostasis: beyond type I hypersensitivity. *Curr. Respir. Med. Rev.* 10, 115–123. doi: 10.2174/1573398x10666141024220151
- Cancer Genome Atlas Research and Network. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550. doi: 10.1038/nature13385
- Carbone, D. P., Reck, M., Paz-Ares, L., Creelan, B., Horn, L., Steins, M., et al. (2017). First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N. Engl. J. Med.* 376, 2415–2426.
- Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 9:34.
- Chen, S., Feng, W., Yang, X., Yang, W., Ru, Y., Liao, J., et al. (2014). Functional expression of P2X family receptors in macrophages is affected by microenvironment in mouse T cell acute lymphoblastic leukemia. *Biochem. Biophys. Res. Commun.* 446, 1002–1009. doi: 10.1016/j.bbrc.2014.03.048
- Dieu-Nosjean, M. C., Goc, J., Giraldo, N. A., Sautes-Fridman, C., and Fridman, W. H. (2014). Tertiary lymphoid structures in cancer and beyond. *Trends Immunol.* 35, 571–580. doi: 10.1016/j.it.2014.09.006
- Fan, L., Lei, H., Lin, Y., Zhou, Z., Shu, G., Yan, Z., et al. (2020). CXCL13, FCRLA, PLA2G2D and MS4A1 Were Identified as Prognostic Biomarkers and Correlated with Immune Infiltration in Ovarian Cancer by Integrating TMB and Microenvironment Analysis. Available online at: <https://ssrn.com/abstract=3514638> (accessed January 6, 2020).
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22.
- Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W., Kim, D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* 21, 938–945. doi: 10.1038/nm.3909
- Gibney, G. T., Weiner, L. M., and Atkins, M. B. (2016). Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol.* 17, e542–e551.
- Goeman, J. J. (2010). L1 penalized estimation in the Cox proportional hazards model. *Biomed. J.* 52, 70–84.
- Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., et al. (2017). Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* 16, 2598–2608. doi: 10.1158/1535-7163.mct-17-0386
- Gottlin, E. B., Bentley, R. C., Campa, M. J., Pisetsky, D. S., Herndon, J. E. II, and Patz, E. F. Jr. (2011). The association of intratumoral germinal centers with early-stage non-small cell lung cancer. *J. Thorac. Oncol.* 6, 1687–1690. doi: 10.1097/jto.0b013e3182217bec
- Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., et al. (2015). Non-small-cell lung cancer. *Nat. Rev. Dis. Primers* 1:15009.
- Guo, H., Cai, J., Wang, X., Wang, B., Li, X., Qu, X., et al. (2020). Prognostic values of a novel multi-mRNA signature for predicting relapse of cholangiocarcinoma. *Int. J. Biol. Sci.* 16, 869–881. doi: 10.7150/ijbs.38846
- Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747
- Hellmann, M. D., Rizvi, N. A., Goldman, J. W., Gettinger, S. N., Borghaei, H., Brahmer, J. R., et al. (2017). Nivolumab plus ipilimumab as first-line treatment for advanced non-small-cell lung cancer (CheckMate 012): results of an open-label, phase 1, multicohort study. *Lancet Oncol.* 18, 31–41. doi: 10.1016/s1473-0163(16)30624-6
- Herbst, R. S., Morgensztern, D., and Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature* 553, 446–454.
- Hui, L., and Chen, Y. (2015). Tumor microenvironment: sanctuary of the devil. *Cancer Lett.* 368, 7–13. doi: 10.1016/j.canlet.2015.07.039
- Imada, A., Shijubo, N., Kojima, H., and Abe, S. (2000). Mast cells correlate with angiogenesis and poor outcome in stage I lung adenocarcinoma. *Eur. Respir. J.* 15, 1087–1093. doi: 10.1034/j.1399-3003.2000.01517.x
- Kawabata, K. C., Ehata, S., Komuro, A., Takeuchi, K., and Miyazono, K. (2013). TGF-beta-induced apoptosis of B-cell lymphoma Ramos cells through reduction of MS4A1/CD20. *Oncogene* 32, 2096–2106. doi: 10.1038/nc.2012.219
- Kim, J. M., and Chen, D. S. (2016). Immune escape to PD-L1/PD-1 blockade: seven steps to success (or failure). *Ann. Oncol.* 27, 1492–1504. doi: 10.1093/annonc/mdw217
- Kurebayashi, Y., Emoto, K., Hayashi, Y., Kamiyama, I., Ohtsuka, T., Asamura, H., et al. (2016). Comprehensive immune profiling of lung adenocarcinomas reveals four immunosubtypes with plasma cell subtype a negative indicator. *Cancer Immunol. Res.* 4, 234–247. doi: 10.1158/2326-6066.cir-15-0214
- Laurent, A., Rouillac, C., Delezoide, A. L., Giovangrandi, Y., Vekemans, M., Bellet, D., et al. (1998). Insulin-like 4 (INSL4) gene expression in human embryonic and trophoblastic tissues. *Mol. Reprod. Dev.* 51, 123–129. doi: 10.1002/(sici)1098-2795(199810)51:2<123::aid-mrd1>3.0.co;2-s
- Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., et al. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357, 409–413.
- Li, X., Xu, Z., Du, W., Zhang, Z., Wei, Y., Wang, H., et al. (2014). Aiols promotes anchorage independence by silencing p66Shc transcription in cancer cells. *Cancer Cell.* 25, 575–589. doi: 10.1016/j.ccr.2014.03.020
- Lin, C., Shi, X., Zhao, J., He, Q., Fan, Y., Xu, W., et al. (2020). Tumor mutation burden correlates with efficacy of chemotherapy/targeted therapy in advanced non-small cell lung cancer. *Front. Oncol.* 10:480. doi: 10.3389/fonc.2020.00480
- Liu, J., Nie, S., Wu, Z., Jiang, Y., Wan, Y., Li, S., et al. (2020). Exploration of a novel prognostic risk signatures and immune checkpoint molecules in endometrial carcinoma microenvironment. *Genomics* 112, 3117–3134. doi: 10.1016/j.ygeno.2020.05.022
- Liu, Y., Li, X., Chang, R., Chen, Y., and Gao, Y. (2020). PLEK2 and SCN7A: novel biomarkers of non-small cell lung cancer. *Res. Square*
- Liu, Y., Liu, B., Jin, G., Zhang, J., Wang, X., Feng, Y., et al. (2019a). An integrated three-long non-coding RNA signature predicts prognosis in colorectal cancer patients. *Front. Oncol.* 9:1269. doi: 10.3389/fonc.2019.01269
- Liu, Y., Wu, L., Ao, H., Zhao, M., Leng, X., Liu, M., et al. (2019b). Prognostic implications of autophagy-associated gene signatures in non-small cell lung cancer. *Aging* 11, 11440–11462. doi: 10.18632/aging.102544

- Lobrich, M., and Jeggo, P. A. (2007). The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nat. Rev. Cancer* 7, 861–869. doi: 10.1038/nrc2248
- Lu, C., Lam, H. N., and Menon, R. K. (2005). New members of the insulin family: regulators of metabolism, growth and now reproduction. *Pediatr. Res.* 57(5 Pt 2), 70R–73R.
- Maciel, T. T., Moura, I. C., and Hermine, O. (2015). The role of mast cells in cancers. *F1000Prime Rep.* 7:09.
- Menezes, J., Salgado, R. N., Acquadro, F., Gómez-López, G., Carralero, M. C., Barroso, A., et al. (2013). ASXL1, TP53 and IKZF3 mutations are present in the chronic phase and blast crisis of chronic myeloid leukemia. *Blood Cancer J.* 3:e157. doi: 10.1038/bcj.2013.54
- Ostasiewicz, B., Ostasiewicz, P., Dus-Szchniewicz, K., Ostasiewicz, K., and Ziolkowski, P. (2016). Quantitative analysis of gene expression in fixed colorectal carcinoma samples as a method for biomarker validation. *Mol. Med. Rep.* 13, 5084–5092. doi: 10.3892/mmr.2016.5200
- Owada-Ozaki, Y., Muto, S., Takagi, H., Inoue, T., Watanabe, Y., Fukuhara, M., et al. (2018). Prognostic impact of tumor mutation burden in patients with completely resected non-small cell lung cancer: brief report. *J. Thorac. Oncol.* 13, 1217–1221. doi: 10.1016/j.jtho.2018.04.003
- Pleshkan, V. V., Zinov'Eva, M. V., Vinogradova, T. V., and Sverdlov, E. D. (2007). Transcription of the KLRB1 gene is suppressed in human cancer tissues. *Molekuliarnaia Genetika, Mikrobiologiya i Virusologiya* 4, 3–7.
- Pollak, M. (2012). The insulin and insulin-like growth factor receptor family in neoplasia: an update. *Nat. Rev. Cancer* 12, 159–169. doi: 10.1038/nrc3215
- Pourabdollah, M., Bahmanyar, M., Atenafu, E. G., Reece, D., Hou, J., and Chang, H. (2016). High IKZF1/3 protein expression is a favorable prognostic factor for survival of relapsed/refractory multiple myeloma patients treated with lenalidomide. *J. Hematol. Oncol.* 9:123.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rizvi, H., Sanchez-Vega, F., La, K., Chatila, W., Jonsson, P., Halpenny, D., et al. (2018). Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-small-cell lung cancer profiled with targeted next-generation sequencing. *J. Clin. Oncol.* 36, 633–641.
- Robinson, D. R., Wu, Y. M., Lonigro, R. J., Vats, P., Cobain, E., Everett, J., et al. (2017). Integrative clinical genomics of metastatic cancer. *Nature* 548, 297–303.
- Sauerbrei, W., Royston, P., and Binder, H. (2007). Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat. Med.* 26, 5512–5528. doi: 10.1002/sim.3148
- Schumacher, T. N., and Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. *Science* 348, 69–74.
- Shi, J., Hou, Z., Yan, J., Qiu, W., Liang, L., Meng, M., et al. (2020). The prognostic significance of fibroblast activation protein- α in human lung adenocarcinoma. *Ann. Transl. Med.* 8:224. doi: 10.21037/atm.2020.01.82
- Steghs, E. M. P., Bakker, M., Hoogkamer, A. Q., Boer, J. M., Hartman, Q. J., Stalpers, F., et al. (2018). High STAP1 expression in DUX4-rearranged cases is not suitable as therapeutic target in pediatric B-cell precursor acute lymphoblastic leukemia. *Sci. Rep.* 8:693.
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.* 54, 13031–13033.
- Sun, C. C., Zhou, Q., Hu, W., Li, S. J., Zhang, F., Chen, Z. L., et al. (2018). Transcriptional E2F1/2/5/8 as potential targets and transcriptional E2F3/6/7 as new biomarkers for the prognosis of human lung carcinoma. *Aging* 10, 973–987. doi: 10.18632/aging.101441
- Tam, C. S., Otero-Palacios, J., Abruzzo, L. V., Jorgensen, J. L., Ferrajoli, A., Wierda, W. G., et al. (2008). Chronic lymphocytic leukaemia CD20 expression is dependent on the genetic subtype: a study of quantitative flow cytometry and fluorescent in-situ hybridization in 510 patients. *Br. J. Haematol.* 141, 36–40. doi: 10.1111/j.1365-2141.2008.07012.x
- Terada, L. S., and Liu, Z. (2014). Aiolos and lymphocyte mimicry in lung cancer. *Mol. Cell Oncol.* 1:e29912. doi: 10.4161/mco.29912
- Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* 16, 385–395. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3
- Verrecchia, F., and Redini, F. (2018). Transforming growth factor-beta signaling plays a pivotal role in the interplay between osteosarcoma cells and their microenvironment. *Front. Oncol.* 8:133. doi: 10.3389/fonc.2018.00133
- Wang, S. S., Liu, W., Ly, D., Xu, H., Qu, L., and Zhang, L. (2019). Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer. *Cell Mol. Immunol.* 16, 6–18. doi: 10.1038/s41423-018-0027-x
- Wareham, K., Vial, C., Wykes, R. C., Bradding, P., and Seward, E. P. (2009). Functional evidence for the expression of P2X1, P2X4 and P2X7 receptors in human lung mast cells. *Br. J. Pharmacol.* 157, 1215–1224. doi: 10.1111/j.1476-5381.2009.00287.x
- Wright, C. M., Savarimuthu Francis, S. M., Tan, M. E., Martins, M. U., Winterford, C., Davidson, M. R., et al. (2012). MS4A1 dysregulation in asbestos-related lung squamous cell carcinoma is due to CD20 stromal lymphocyte expression. *PLoS One* 7:e34943. doi: 10.1371/journal.pone.0034943
- Wu, H. X., Wang, Z. X., Zhao, Q., Chen, D. L., He, M. M., Yang, L. P., et al. (2019). Tumor mutational and indel burden: a systematic pan-cancer evaluation as prognostic biomarkers. *Ann. Transl. Med.* 7, 640. doi: 10.21037/atm.2019.10.116
- Xing, J., Spitz, M. R., Lu, C., Zhao, H., Yang, H., Wang, W., et al. (2007). Deficient G2-M and S checkpoints are associated with increased lung cancer risk: a case-control analysis. *Cancer Epidemiol. Biomarkers Prev.* 16, 1517–1522. doi: 10.1158/1055-9965.epi-07-0111
- Xu, X., Huang, Z., Zheng, L., and Fan, Y. (2018). The efficacy and safety of anti-PD-1/PD-L1 antibodies combined with chemotherapy or CTLA4 antibody as a first-line treatment for advanced lung cancer. *Int. J. Cancer* 142, 2344–2354. doi: 10.1002/ijc.31252
- Yang, R., Li, S. W., Chen, Z., Zhou, X., Ni, W., Fu, D. A., et al. (2018). Role of INSL4 signaling in sustaining the growth and viability of LKB1-inactivated lung cancer. *J. Natl. Cancer Inst.* 111, 664–674. doi: 10.1093/jnci/djy166
- Yarchoan, M., Hopkins, A., and Jaffee, E. M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* 377, 2500–2501. doi: 10.1056/NEJMcl1713444
- Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4:2612. doi: 10.1038/ncomms3612
- Yuan, J., Hegde, P. S., Clynes, R., Foukas, P. G., Harari, A., Kleen, T. O., et al. (2016). Novel technologies and emerging biomarkers for personalized cancer immunotherapy. *J. Immunother. Cancer* 4:3. doi: 10.1186/s40425-016-0107-3
- Zhou, C., Wang, S., Zhou, Q., Zhao, J., Xia, X., Chen, W., et al. (2019). A long non-coding RNA signature to improve prognostic prediction of pancreatic ductal adenocarcinoma. *Front. Oncol.* 9:1160. doi: 10.3389/fonc.2019.01160

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ma, Luo, Cao, Zheng, Zhang, Zhang and Fu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.