



OPEN ACCESS

EDITED BY

Wenjie Shi,
Otto von Guericke University
Magdeburg, Germany

REVIEWED BY

Wei Guo,
Shanxi Provincial Cancer Hospital, China
Changfa Qu,
Harbin Medical Sciences University, China

*CORRESPONDENCE

Guofang Zhao
✉ guofangzhao2022@yeah.net

RECEIVED 02 March 2023

ACCEPTED 13 July 2023

PUBLISHED 31 July 2023

CITATION

Shu J, Jiang J and Zhao G (2023)
Identification of novel gene signature for
lung adenocarcinoma by machine learning
to predict immunotherapy and prognosis.
Front. Immunol. 14:1177847.
doi: 10.3389/fimmu.2023.1177847

COPYRIGHT

© 2023 Shu, Jiang and Zhao. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Identification of novel gene signature for lung adenocarcinoma by machine learning to predict immunotherapy and prognosis

Jianfeng Shu^{1,2}, Jinni Jiang^{1,2} and Guofang Zhao^{1*}

¹Department of Thoracic Surgery, Ningbo No.2 Hospital, Ningbo, China, ²Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, Ningbo, China

Background: Lung adenocarcinoma (LUAD) as a frequent type of lung cancer has a 5-year overall survival rate of lower than 20% among patients with advanced lung cancer. This study aims to construct a risk model to guide immunotherapy in LUAD patients effectively.

Materials and methods: LUAD Bulk RNA-seq data for the construction of a model, single-cell RNA sequencing (scRNA-seq) data (GSE203360) for cell cluster analysis, and microarray data (GSE31210) for validation were collected from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database. We used the Seurat R package to filter and process scRNA-seq data. Sample clustering was performed in the ConsensusClusterPlus R package. Differentially expressed genes (DEGs) between two groups were mined by the Limma R package. MCP-counter, CIBERSORT, ssGSEA, and ESTIMATE were employed to evaluate immune characteristics. Stepwise multivariate analysis, Univariate Cox analysis, and Lasso regression analysis were conducted to identify key prognostic genes and were used to construct the risk model. Key prognostic gene expressions were explored by RT-qPCR and Western blot assay.

Results: A total of 27 immune cell marker genes associated with prognosis were identified for subtyping LUAD samples into clusters C3, C2, and C1. C1 had the longest overall survival and highest immune infiltration among them, followed by C2 and C3. Oncogenic pathways such as VEGF, EFGR, and MAPK were more activated in C3 compared to the other two clusters. Based on the DEGs among clusters, we confirmed seven key prognostic genes including CPA3, S100P, PTTG1, LOXL2, MELTF, PKP2, and TMPRSS11E. Two risk groups defined by the seven-gene risk model presented distinct responses to immunotherapy and chemotherapy, immune infiltration, and prognosis. The mRNA and protein level of CPA3 was decreased, while the remaining six gene levels were increased in clinical tumor tissues.

Conclusion: Immune cell markers are effective in clustering LUAD samples into different subtypes, and they play important roles in regulating the immune microenvironment and cancer development. In addition, the seven-gene risk model may serve as a guide for assisting in personalized treatment in LUAD patients.

KEYWORDS

lung adenocarcinoma, single-cell analysis, immune cells, molecular subtyping, risk model, immunotherapy

Introduction

Lung cancer has the largest proportion among all cancer types (1). Lung adenocarcinoma (LUAD) accounts for over half of all lung cancer patients. The development of diagnosis and treatment improves the overall survival time and survival quality of LUAD patients. However, the 5-year survival of metastatic LUAD patients is lower than 20% (2). Worse still, the incidence and mortality of lung cancer are rising (3). Therefore, in order to identify new therapeutic targets and improve patient survival, there is an urgent need to further develop specific prognostic prediction methods for LUAD patients.

For these metastatic cancer patients, targeted therapy such as EGFR and ALK inhibitors have been developed, but they only show favorable efficiency to specific patients with EGFR and ALK mutations (4, 5). A monumental breakthrough in cancer treatment and immunotherapy has revolutionized the field of oncology by harnessing the body's natural defenses to eliminate malignant cells (6). For example, immunotherapy with immune checkpoint inhibitors such as programmed cell death ligand 1 (PD-L1) and programmed cell death protein 1 (PD-1) inhibitors has shown satisfying outcomes in multiple cancer types including lung cancer (7). Nevertheless, the response to anti-PD-1/PD-L1 therapy in metastatic cancer patients is unfavorable. Compared to chemotherapy and radiotherapy, immunotherapy has more benefits in treating cancer patients at the late stages, such as low side effects, low toxicity, and low resistance. Therefore, it is critical to improve the efficiency of immunotherapy.

Previous studies have shown that the response to immunotherapy is greatly affected by immune microenvironment

Abbreviations: AT, alveolar type; IC50, biochemical half maximal inhibitory concentration; CR, complete response; CDF, cumulative distribution function; DEGs, differentially expressed genes; FDR, false discovery rate; GO, Gene Ontology; IFN- γ , interferon- γ ; KEGG, Kyoto Encyclopedia of Genes and Genomes; Lasso, least absolute shrinkage and selection operator; LUAD, lung adenocarcinoma; MDSCs, myeloid-derived suppressor cells; NSCLC, non-small cell lung cancer; PR, partial response; PD-L1, programmed cell death ligand 1; PD-1, programmed cell death protein 1; PD, progressive disease; ROC, receiver operation characteristic; scRNA-seq, single-cell RNA sequencing; SD, stable disease; TCGA, The Cancer Genome Atlas; TIDE, Tumor Immune Dysfunction and Exclusion; TMB, tumor mutation burden.

and tumor heterogeneity (8, 9). In this study, we deciphered scRNA-seq data and the immune landscape of LUAD patients. In combination with bulk RNA-seq data, three molecular subtypes showing disparate prognosis, immune characteristics, and genomic features were categorized. In addition, a seven-gene risk model capable of dividing high-risk and low-risk LUAD patients was established. Two risk groups (high risk and low risk) manifested differential responses to immunotherapy and chemotherapy, prognosis, and immune microenvironment.

Materials and methods

Data collection of LUAD samples

From the Gene Expression Omnibus (GEO) database, scRNA-seq data (GSE203360) and microarray expression profiles (GSE31210) of LUAD samples were accessed (10–12). From the Cancer Genome Atlas (TCGA) database, Bulk RNA-seq data (transcripts per million, TPM) of LUAD samples were obtained on the Sangerbox platform (13, 14).

Preprocessing microarray and bulk RNA-seq data

For microarray data, we annotated probes into gene symbols according to the GPL570 platform. We excluded the probes that matched multiple genes. When one gene had multiple probes, the average expression was used. Totally, 246 LUAD samples in the GSE31210 dataset were included. For bulk RNA-seq data (named as TCGA dataset), the LUAD samples with a survival time > 0 and survival information were retained. Ensembl ID was transferred to the gene symbol. Protein-coding genes were included. Totally, 500 LUAD and 59 normal samples in the TCGA dataset were included.

Processing of scRNA-seq data

GSE203360 dataset includes scRNA-seq data of six LUAD samples. Quality control procedure was conducted to filter single cells: 1) Each gene expressed more than 200 genes and was

expressed in more than three cells; 2) The quantity of expressed genes in cells was between 100 and 5000; 3) The percentage of mitochondrial genes in one cell was less than 30%; 4) The number of unique molecular identifiers (UMI) was more than 100. Finally, a total of 18694 cells were included under the above filtering.

SEURAT R package (15) was utilized to process the scRNA-seq data. First, scRNA-seq data was log-normalized, and the “FindVariableFeatures” function was performed to discriminate highly variable genes. To eliminate the batch effects of six samples, “FindIntegrationAnchors” and “IntegrateData” functions were used. Then, we scaled the scRNA-seq data using the “ScaleData” function and implemented PCA to decrease dimensionality. Subsequently, “FindClusters” functions and “FindNeighbors” were conducted to cluster single cells into different subgroups under conditions of $\text{dim} = 40$ and $\text{resolution} = 0.2$. Finally, single cells were visualized by t-SNE plot through the “RunTSNE” function.

Next, we obtained cell markers of different cell types from CellMarker2.0 (16) and annotated single cells based on these markers. The “FindAllMarkers” function was used to filter DEGs of different cell types under the criteria of $\log_2\text{FC} = 0.5$ and $\text{Minpct} = 0.5$. DEGs with $P < 0.05$ were determined as marker genes of cell types. Only immune cells were remaining for further analysis. The WebGestaltR package (17) was employed to assess KEGG pathways and GO terms of marker genes.

Identification of molecular subtypes

First, DEGs were screened with the limma R package by comparing LUAD samples to normal samples in the TCGA dataset (18). DEGs with false discovery rates (FDR) < 0.05 and $|\log_2\text{FC}| > 1$ were retained. Univariate Cox regression analysis was performed on marker genes of immune cells screened from single-cell analysis. The marker genes with $P < 0.01$ were included for further analysis. Then, the intersection of DEGs and marker genes was selected as a basis for molecular subtyping.

Based on the expression profiles of intersected genes, the ConsensusClusterPlus R package (19) carried out unsupervised consensus clustering. A total of 500 bootstraps were conducted using the “hc” algorithm and “Pearson” distance. Each bootstrap contained 80% of TCGA samples. Cluster number k from two to 10 was set, and the optimal k could be determined based on the cumulative distribution function (CDF) and ConsensusClusterPlus method. The same procedure was performed in the GSE31210 dataset.

Immune characteristics and pathway analysis

We used multiple methodologies to estimate immune characteristics. CIBERSORT algorithm calculated the enrichment score for 22 kinds of immune cells (20). Stromal infiltration and immune infiltration were evaluated by ESTIMATE (21). We obtained gene signatures of 28 immune cells from previous studies and analyzed their enrichment scores using ssGSEA (22, 23). Moreover,

Microenvironment Cell Populations-counter (MCP-counter) was employed to assess stromal and immune cell infiltration (24). The enrichment of tumor-related pathways was assessed by the PROGENy (Pathway RespOnsive GENes) algorithm (25).

Establishment of a risk model

First of all, DEGs were distinguished from different clusters [$\text{FDR} < 0.05$ and $|\log_2\text{FC}| > \log_2(1.5)$]. We selected the intersected DEGs of different clusters for univariate Cox regression analysis, and the DEGs with $P < 0.05$ were screened. Then, Lasso regression compressed gene numbers in the glmnet R package (26). With the increase in the lambda value, the coefficients of genes were close to zero. A model was built by 10-fold cross-validation, and the optimal one was chosen according to the lambda value. Stepwise multivariate analysis (stepAIC) was further performed in the Mass R package to reduce the gene number (27). To reach the optimal model, the variables were deleted accordingly to reduce the AIC value. A sufficient fitting degree with fewer variables was ensured by StepAIC. The risk model was finally determined as $\text{risk score} = \sum(\text{exp } i^* \text{beta } i)$, where beta represents multivariate coefficients, exp represents expression levels, and i represents genes.

Evaluation of the risk model

According to the formula of the risk model, we calculated the risk score for each sample and divided them into groups of high and low risks using the median risk score. Kaplan-Meier survival analysis showed the prognosis difference between the two risk groups. Receiver operation characteristic (ROC) curve analysis evaluated the survival prediction efficiency (1-year, 3-year, and 5-year) of the risk model. A nomogram was developed combining the model with clinical features.

The risk model to predict chemotherapy and immunotherapy response

Tumor Immune Dysfunction and Exclusion (TIDE) (<http://tide.dfci.harvard.edu/>) predicted the response to immune checkpoint blockade (ICB) therapy. The estimated IC50 for assessing the response to eight chemotherapeutic drugs (Erlotinib, Rapamycin, PHA-665752, Roscovitine, Cisplatin, Paclitaxel, BI-2536, and Pyrimethamine) was calculated by pRRophetic R package. IMvigor210 dataset (28) containing the expression data of urothelial carcinoma patients treated by anti-programmed death-ligand 1 (PD-L1) therapy was used to further validate the performance of the risk model. IMvigor210 dataset contains 348 patients divided into complete response (CR), progressive disease (PD), partial response (PR), and stable disease (SD) groups.

Clinical samples

In this research, we collected five pairs of LUAD and adjacent normal lung tissues from Ningbo No.2 Hospital. Informed consent

was obtained from all participants. The study was approved by the Ethics Committee of Ningbo No.2 Hospital.

RNA preparation and real-time quantitative PCR (RT-qPCR) analysis

TRIzol reagent (Invitrogen, CA, USA) was applied to extract RNA from tissues. PrimeScript™ RT Master Mix (Takara, Dalian, China) was used to synthesize cDNAs to explore genes' mRNA expression detection according to the manufacturer's instructions. The RT-qPCR was conducted using a TB Green Premix Ex Taq™ II kit (Takara, Dalian, China) with a 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA) according to construction sequence. The relative expression of genes was performed using a $2^{-\Delta\Delta Ct}$ method using an internal reference gene, GAPDH. The primers of genes are shown in [Table 1](#).

Western blot assay

Proteins were extracted in RIPA lysis buffer (P0013B, Beyotime) and the concentration was determined using a BCA Protein assay kit (P0010S, Beyotime). Proteins were separated on sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGE) and then transferred to polyvinylidene fluoride (PVDF) membranes. The membranes were blocked with 5% non-fat dry milk and then incubated overnight with primary antibodies, including anti-CAP3 (1:1000, ARG58412, and Arigo Biolaboratories Corp.), anti-S100P (1:1000, 7677S, and CST), anti-PTTG1 (1:1000, 13445, and CST), anti-LOXL2 (1:1000, ab96233, and Abcam), anti-MELTF (1:1000, 10428-1-AP, and Proteintech), anti-PKP2 (1:1000, ab189323, and Abcam), anti-TMPRSS11E (1:1000, PA5-48775, and Invitrogen Antibodies), and anti- β -actin (1:10000, A5441, and Sigma). Horseradish peroxidase (HRP)-conjugated goat anti-mouse/rabbit IgG secondary antibody was used to incubate membranes at room temperature. The blots were detected by chemiluminescence and imaged on an AlphaView analysis system (ProteinSimple, USA). The quantification of individual protein bands was assessed by densitometry using ImageJ software.

TABLE 1 The primers of genes.

| Gene | Forward primer sequence (5'-3') | Reverse primer sequence (5'-3') |
|-----------|---------------------------------|---------------------------------|
| CPA3 | ATTCACGCACGAGAATGGGT | CCACATGCGGTCTTTGTCC |
| S100P | GAGACAGCCATGGGCATGAT | CGTCCAGGTCCTTGAGCAAT |
| PTTG1 | CCAAGGGACCCCTCAAACAA | GGCATCATCTGAGGCAGGAA |
| LOXL2 | CCAGTGTGGTCTGCAGAGAG | CCTGTGCACTGGATCTCGTT |
| MELTF | GGCACACAACCGTCTTTGAC | GGGGCACAGCAGTTTCATAGT |
| PKP2 | GAGATGACTCTGGAGCGAGC | AAGCTTGAGGATGCCACGAA |
| TMPRSS11E | CCTGATTGTCTGGCAGTGT | CCTCTCTGCCAAACTCAGCA |
| GAPDH | AGGGCATCCTGGGCTACAC | GCCAAATTCGTTGTCATACCAG |

Statistical analysis

All statistical analysis was done with the R program (v4.2.0). The Wilcoxon test evaluated the two-group difference. Differences among the three groups were analyzed by the Kruskal-Wallis test. Cox analysis and survival analysis both used the log-rank test.

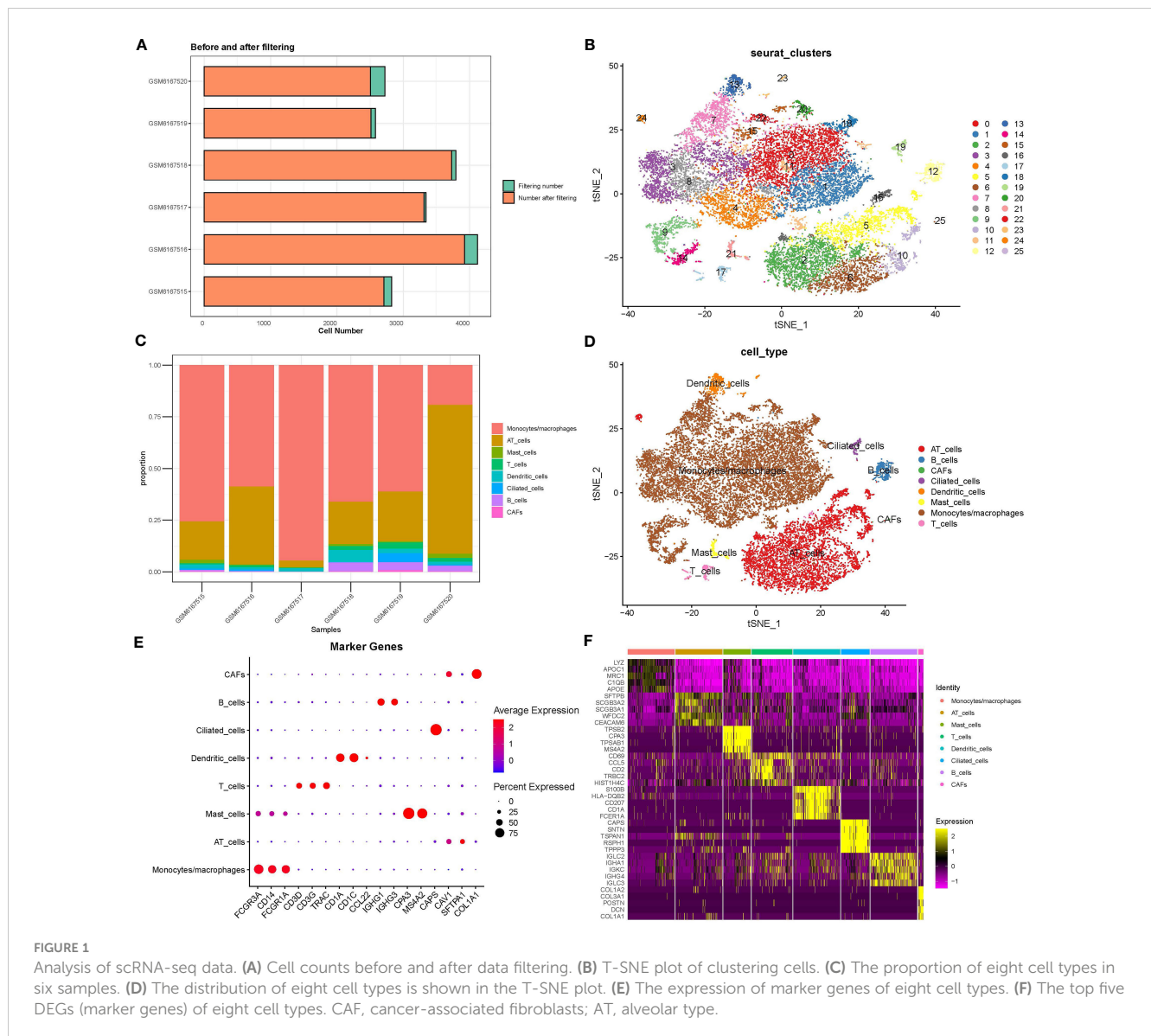
Results

Annotating cell types in single-cell data of LUAD patients

Single cells were filtered using the Seurat R package and 18694 cells were remaining ([Figures S1A, B](#) and [Figure 1A](#), see the filtering details in Materials and Methods). Then, the expression profiles of single cells were log-normalized and the batch effects of six samples were eliminated ([Figures S1C, D](#)). Gene expression levels were scaled and single cells were divided into 26 clusters ([Figure 1B](#)). Further, we annotated these clusters into eight cell types according to the markers from CellMarker2.0 (16) where monocytes/macrophages and AT cells (AT1 (alveolar type I cells) and AT2 (alveolar type II cells)) account for the majority of cells ([Figures 1C–E](#); [Table 2](#)). By assessing the difference in expression profiles among different cell types, we identified DEGs in each cell type and the top five DEGs were visualized ([Figure 1F](#)). Only the DEGs of five immune cells (dendritic cells, B cells, monocytes/macrophages, mast cells, and T cells) were retained for the subsequent analysis ([Table S1](#)). These DEGs were defined as the markers of five immune cells. Functional analysis showed that these markers were apparently enriched in immune-related pathways including phagosome and antigen processing and presentation ([Figure 2A](#)). Moreover, GO analysis revealed that immune-related biological pathways were significantly accumulated such as myeloid leukocyte, myeloid leukocyte activation, and involved myeloid leukocyte-mediated immunity in immune response ([Figure 2B](#)).

Molecular subtyping based on the markers of immune cells

We performed univariate Cox regression analysis to ascertain the markers that were significantly associated with LUAD prognosis

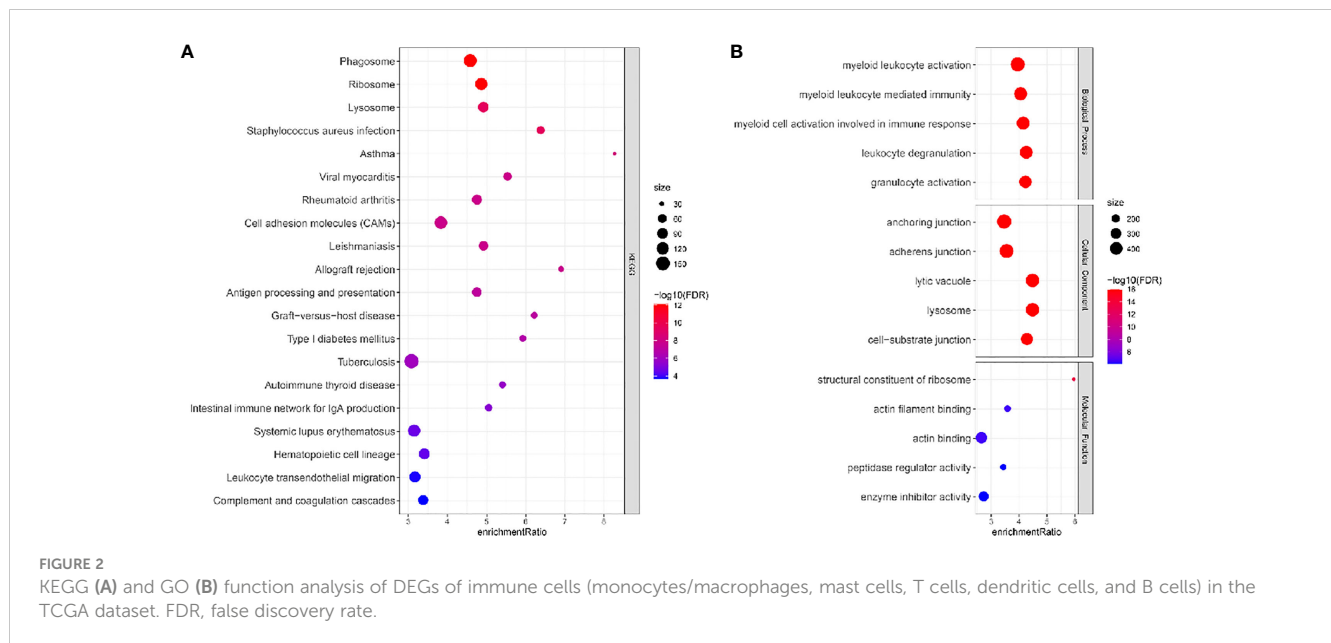


in the TCGA dataset. A total of 77 marker genes were identified with $P < 0.01$ (Table S2). We further screened dysregulated genes of LUAD by comparing LUAD samples with normal samples in the TCGA dataset. To ensure which markers had a remarkable

contribution to LUAD, we selected the intersection between prognostic marker genes and dysregulated genes. Venn plot showed that a total of 27 marker genes were intersected and these genes were used as a basis for the following analysis (Figures 3A, B).

TABLE 2 Twenty-six clusters annotated by marker genes of eight cell types.

| Cell type | marker gene | seurat_clusters |
|-----------------------|----------------------|---------------------------------|
| Monocytes/macrophages | FCGR3A, CD14, FCGR1A | 0,1,3,4,7,8,9,11,14,15,18,20,22 |
| T_cells | CD3D, CD3G, TRAC | 17 |
| Dendritic_cells | CD1A, CD1C, CCL22 | 13,23 |
| B_cells | IGHG1, IGHG3 | 12 |
| Mast_cells | CPA3, MS4A2 | 21 |
| Ciliated_cells | CAPS | 19 |
| AT cell | CAV1, SFTPA1 | 2,5,6,10,16,24 |
| CAFs | COL1A1 | 25 |



Next, we conducted consensus clustering based on gene expression profiles of 27 marker genes. According to CDF curves and the area under CDF curves, we confirmed cluster number $k = 3$ and classified LUAD samples into three clusters (Figure S2). Kaplan-Meier survival analysis uncovered that three clusters (C1, C2, and C3) had distinct overall survival in both TCGA ($P < 0.0001$) and GSE31210 ($P = 0.00033$) datasets (Figures 3C, D). C1 had the longest survival while C3 had the worst survival. The distribution of clinical characteristics in three clusters showed corresponding results with their prognosis. The proportion of patients with late stages was obviously higher than those with early stages in C3 (Figure 3E). For example, C3 consisted of higher percentages of stages II to IV than C1 and C2. Not surprisingly, the proportion of dead patients was also significantly higher in C3 compared with that in C1. The results suggested that the subtyping based on the 27 marker genes was reliable.

Analysis of immune characteristics and biological pathways

The tumor microenvironment can indicate an immune response to cancer cells, and a high infiltration of cytotoxic lymphocytes commonly suggests a high activation of anti-cancer response and a relatively favorable outcome. We used several strategies to evaluate the immune infiltration of three clusters in the TCGA dataset. CIBERSORT analysis showed that most immune cells, such as activated memory CD4 T cells, resting memory CD4 T cells, macrophages, and CD8 T cells, were differently enriched in three clusters (Figures 4A, B). Compared with C2 and C3, C1 had higher immune infiltration and stromal infiltration (Figure 4C). Interferon- γ (IFN- γ) is an essential factor in anti-cancer response and immune regulation. We found that C1 had the highest expression of IFN- γ followed by C3 (Figure 4D), indicating that the immune response was relatively active in C1. Moreover, both MCP-counter and ssGSEA revealed the consistent result that most

immune cells were more enriched in C1 compared with the other two clusters (Figures 4E, F). To figure out the possible molecular mechanisms of different prognoses in three clusters, we evaluated 11 tumor-related pathways using the PROGENy algorithm (25). As a result, most pathways such as PI3K, VEGF, hypoxia, and EGFR were more activated in C3 compared with C1 and C2 (Figure 4G). Notably, the pro-apoptotic signaling, Trail, was the most enriched in C1, which may lead to its favorable prognosis.

Variation characteristics of three clusters

Genomic characteristics have been reported to be linked with cancer development, prognosis, and response to immunotherapy. We obtained several genomic features from previous studies (29, 30). C3 showed the highest tumor mutation burden (TMB) while C1 had the lowest tumor purity (Figures 5A, B). In addition, C1 also had the lowest scores of aneuploidy, homologous recombination defect, fraction altered, and number of segments (Figure 5C). Furthermore, we assessed the frequencies of gene mutation and screened the significantly mutated genes in three clusters. The top 15 mutated genes of three clusters were listed (Figures 5D–F). The mutation frequencies of the top 15 genes varied largely among three clusters. Especially, TP53 had the highest mutation rate in C3 with a mutation frequency of 66%, while the frequencies were 36% and 35% in C1 and C2, respectively. Notably, KRAS presented an extremely higher mutation rate in C2 (41%) compared with C1 (18%) and C3 (25%). To figure out the influence of mutations in oncogenic pathways, we employed maftools (31) to calculate the mutation frequencies of genes of 10 oncogenic pathways. Six of 10 pathways were evidently affected in LUAD samples, including RTK-RAS, WNT, NOTCH, Hippo, PI3K, and TP53 (Figures 5G–I). C3 had the largest proportion of affected pathways compared to the other two clusters, indicating that the biological functions of these oncogenic pathways may be largely interfered with.

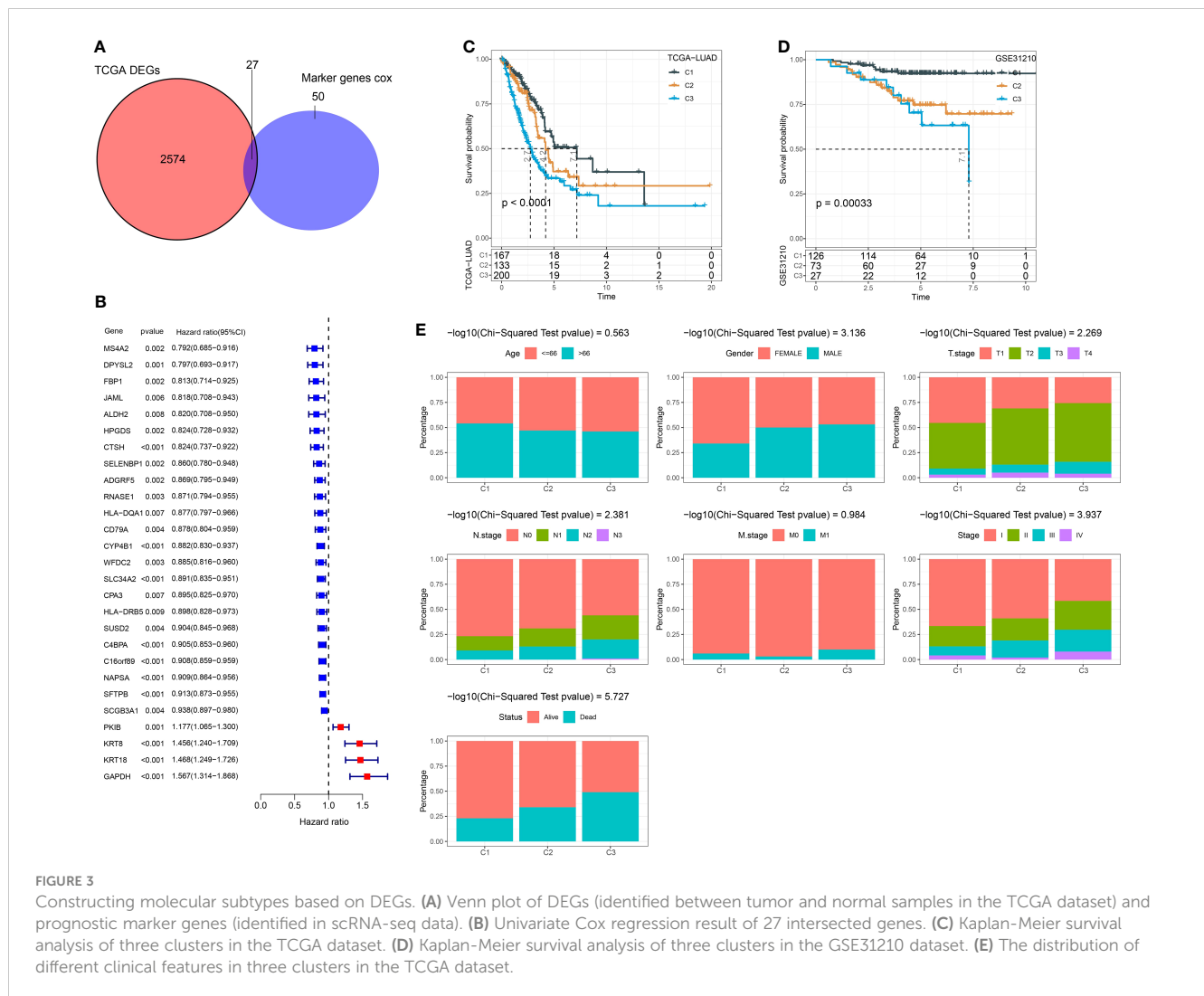


FIGURE 3

Constructing molecular subtypes based on DEGs. (A) Venn plot of DEGs (identified between tumor and normal samples in the TCGA dataset) and prognostic marker genes (identified in scRNA-seq data). (B) Univariate Cox regression result of 27 intersected genes. (C) Kaplan-Meier survival analysis of three clusters in the TCGA dataset. (D) Kaplan-Meier survival analysis of three clusters in the GSE31210 dataset. (E) The distribution of different clinical features in three clusters in the TCGA dataset.

Construction of a risk model based on the DEGs in three clusters

Given that the three clusters had different prognoses, immune microenvironments, mutation characteristics, and enrichment of biological pathways, we next attempted to dig out potential key genes contributable to their outcomes. Therefore, we used the limma R package to excavate the DEGs in C1 vs. other, C2 vs. other, and C3 vs. other [FDR < 0.05 and |log2FC| > log2(1.5)]. As a result, we identified 1587 DEGs in C1, 900 DEGs in C2, and 1437 DEGs in C3 (Figures S3A–C). Venn plot showed that the DEGs of three groups shared 133 overlapped ones (Figure S3D). KEGG analysis showed that the DEGs of C3 were significantly enriched in the cell cycle, p53 signaling pathway, and microRNAs in cancer (Figure S3E). The 133 common DEGs were also enriched in the cell cycle (Figure S3F), suggesting that these DEGs may have an influence on different prognoses by regulating the cell cycle pathway.

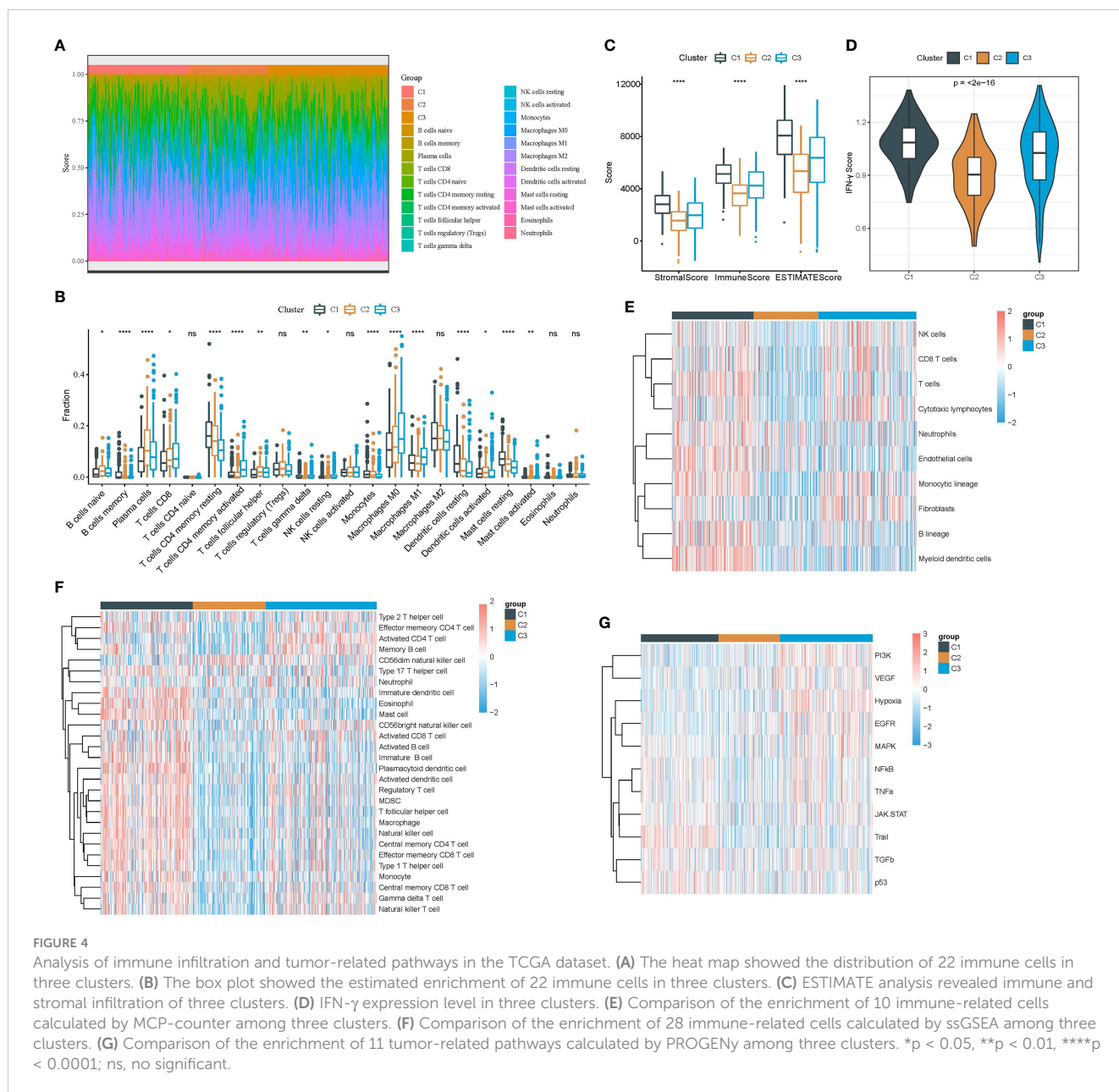
To further explore the potential key DEGs among 133 DEGs, we performed univariate Cox regression analysis and identified 109 DEGs associated with overall survival (Table S3). Then, using Lasso regression analysis, we compressed the coefficients of DEGs to

retain the important DEGs. The Lasso model reached the optimal when the lambda value = 0.0315, and a total of 15 DEGs were remaining (Figures 6A, B). To further reduce the number of genes, stepAIC was employed to ensure a sufficient fitting degree with the least number of genes. As a result, seven key prognostic genes were determined to construct a risk model (Figure 6C).

$$\text{Risk Score} = -0.073 \times \text{CPA3} + 0.077 \times \text{S100P} + 0.128 \times \text{PTTG1} + 0.158 \times \text{LOXL2} + 0.152 \times \text{MELTF} + 0.119 \times \text{PKP2} + 0.058 \times \text{Tmprss11e}$$

Assessment of the seven-gene risk model

We determined the risk score for each tumor sample in the TCGA dataset using the risk model. The median value of the risk score was used to identify the high-risk and low-risk groups (Figure 6D). More samples in the high-risk group had a dead status than in the low-risk group (Figures 6D, G). With AUCs of 0.72, 0.72, and 0.65, respectively, the ROC curve showed that the risk model was effective for predicting 1-year, 3-year, and 5-year survival (Figure 6E). The two risk groups had varied overall survival rates, as indicated by the Kaplan-Meier survival curve



($P < 0.0001$; Figure 6F). We observed consistent results in the GSE31210 dataset (Figures 6H–K), demonstrating the robustness and reliability of the seven-gene risk model. In addition, we evaluated the performance of the risk model in the samples with different clinical characteristics. Survival analysis showed that the risk model was effective in distinguishing high-risk and low-risk samples with different clinical features including TNM stage, age, and gender (Figure S4).

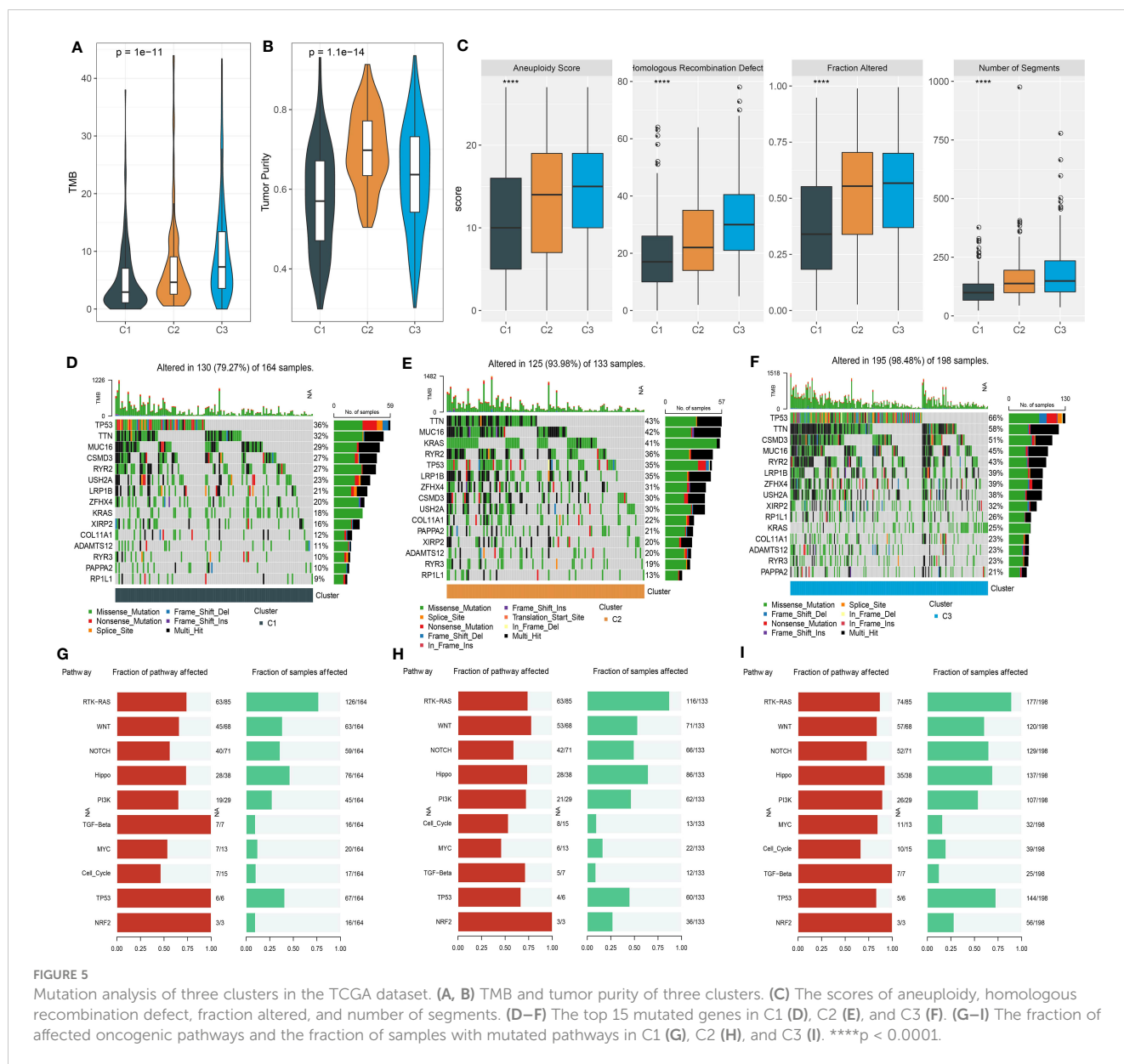
Establishing a nomogram to optimize the application of the risk model

Risk score and stage were identified as independent risk factors by univariate and multivariate analyses (Figures 7A, B). A nomogram based on risk score and stage was therefore created

(Figure 7C). The risk score showed a more significant effect on the total points compared to the stage. The calibration curve presented that the predicted 1-year, 3-year, and 5-year survival was close to the observed ones (Figure 7D). Decision curve analysis validated that patients could obtain more benefits from the nomogram compared to the risk score (Figure 7E). Also, compared with clinical characteristics including age, gender, and stage, both risk score and nomogram showed higher AUC of 1-year, 3-year, and 5-year survival (Figures 7F–H).

Assessment of immune characteristics

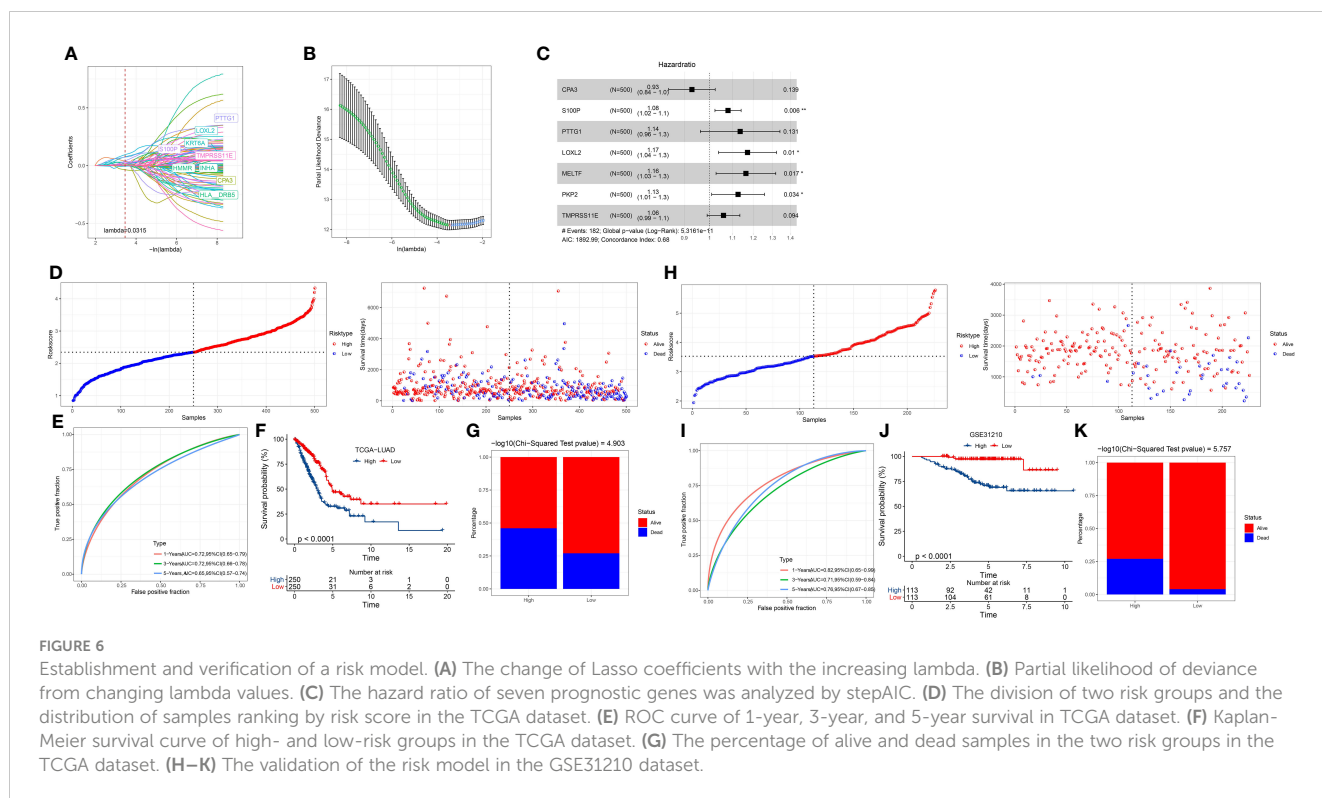
The immune microenvironment in high-risk and low-risk groups was assessed with different tools and methodologies. ESTIMATE demonstrated that the low-risk group had obviously



higher immune scores and stromal scores than the high-risk group, indicating higher stromal infiltration and immune infiltration in the low-risk group (Figure 8A). CIBERSORT algorithm revealed that several immune cells were differentially enriched between the two risk groups, such as dendritic cells, M0 macrophages, mast cells, memory CD4 T cells, and M1 macrophages (Figure 8B). Using the marker genes of five immune cells obtained by scRNA analysis, the ssGSEA finding indicated that the low-risk group had a higher immune cell score (Figure S5). This demonstrated that the two risk groups had differential immune infiltration. Correlation analysis further supported an association of the risk score with immune cell infiltration (Figures 8C–F). For instance, B cell lineage (including memory B cells, monocytes, activated B cells, and immature B cells) was negatively correlated with risk score. M0 macrophages, M1 macrophages, and fibroblasts were positively related to risk scores.

The potential of the risk score in predicting immunotherapy and chemotherapy

TIDE tool can predict the response to immune checkpoint inhibitors (ICIs). We evaluated the enrichment score of myeloid-derived suppressor cells (MDSCs), T cell dysfunction, and T cell exclusion and calculated the TIDE score. We observed that risk score was positively related to TIDE score ($P < 0.001$, $R = 0.34$, Figure 8G), suggesting that patients with high-risk scores were more vulnerable in the response to ICIs than those with low-risk scores. Notably, the risk score was also positively associated with T cell exclusion ($P < 0.001$, $R = 0.43$) and MDSC ($P < 0.001$, $R = 0.62$), which may contribute to a negative response to ICIs. The correlation analysis between RiskScore and immune checkpoint genes showed that RiskScore was positively correlated with immune checkpoint genes, indicating that patients in



the high-risk group were more suitable for ICB therapy (Figure S6). In addition, we used a pRRophetic package to predict the response to chemotherapeutic drugs. The two risk groups had distinct sensitivity to different chemotherapeutic drugs (Figure 8H). The high-risk group was more sensitive to Cisplatin, Paclitaxel, BI-2536, and Pyrimethamine, the while low-risk group was more sensitive to Erlotinib, Rapamycin, PHA-665752, and Roscovitine.

The performance of the seven-gene risk model in immunotherapy-treated patients

We further examined the performance of the risk model in the patients administrated by PD-1 inhibitors (IMvigor210 dataset). The same procedure was performed to divide patients into high- and low-risk groups. Survival analysis manifested that the low-risk group had markedly longer overall survival than the high-risk group (Figure 9A). Notably, the grouping efficiency was higher in the patients with early stages (I+II) than those with late stages (III+IV) (Figures 9B, C). In the IMvigor210 dataset, patients were divided into four groups of different responses to PD-1 inhibitors including CR, PR, SD, and PD where CR and PR indicated positive responses to immunotherapy. We compared the risk score between CR/PR and SD/PD groups. Unsurprisingly, the CR/PR group had a significantly lower risk score than the SD/PD group (P = 0.00017, Figure 9D). TIDE score was unable to significantly distinguish between high-risk and low-risk groups (P = 0.095, Figure 9E). The AUC of the risk score reached 0.63 while it was 0.58 in the TIDE score (Figure 9F).

Seven key genes were validated by RT-qPCR and Western blot assay

To know the expression profile of CPA3, S100P, PTTG1, LOXL2, MELTF, PKP2, and TMPRSS11E, we used RT-qPCR and Western blot assay to explore the expression in clinical patient tissues. As shown in Figure 10, the mRNA level of CAP3 was downregulated, while S100P, PTTG1, LOXL2, MELTF, PKP2, and TMPRSS11E mRNA levels were upregulated in LUDA tissues in comparison to paracarcinoma. Surprisingly, the protein levels of seven genes had a similar trend in tumor tissues (Figure 11).

Discussion

In this study, we utilized scRNA-seq data and bulk RNA-seq data to explore prognostic genes associated with LUAD and immunity. We identified three clusters (C1, C2, and C3) with distinct clinical and molecular features. Through differential analysis between different clusters, we detected a series of DEGs. We further excavated seven key prognostic genes from these DEGs to establish a risk model. The risk model was effective in discriminating between high-risk and low-risk patients, which could provide guidance for clinical treatment.

C1 had the longest overall survival whereas C3 had the worst prognosis. Consistent with clinical characteristics, samples with advanced stages had a higher proportion in C3 compared with the other two clusters. In addition, C3 had the largest number of dead samples. To clarify the molecular mechanisms causing the

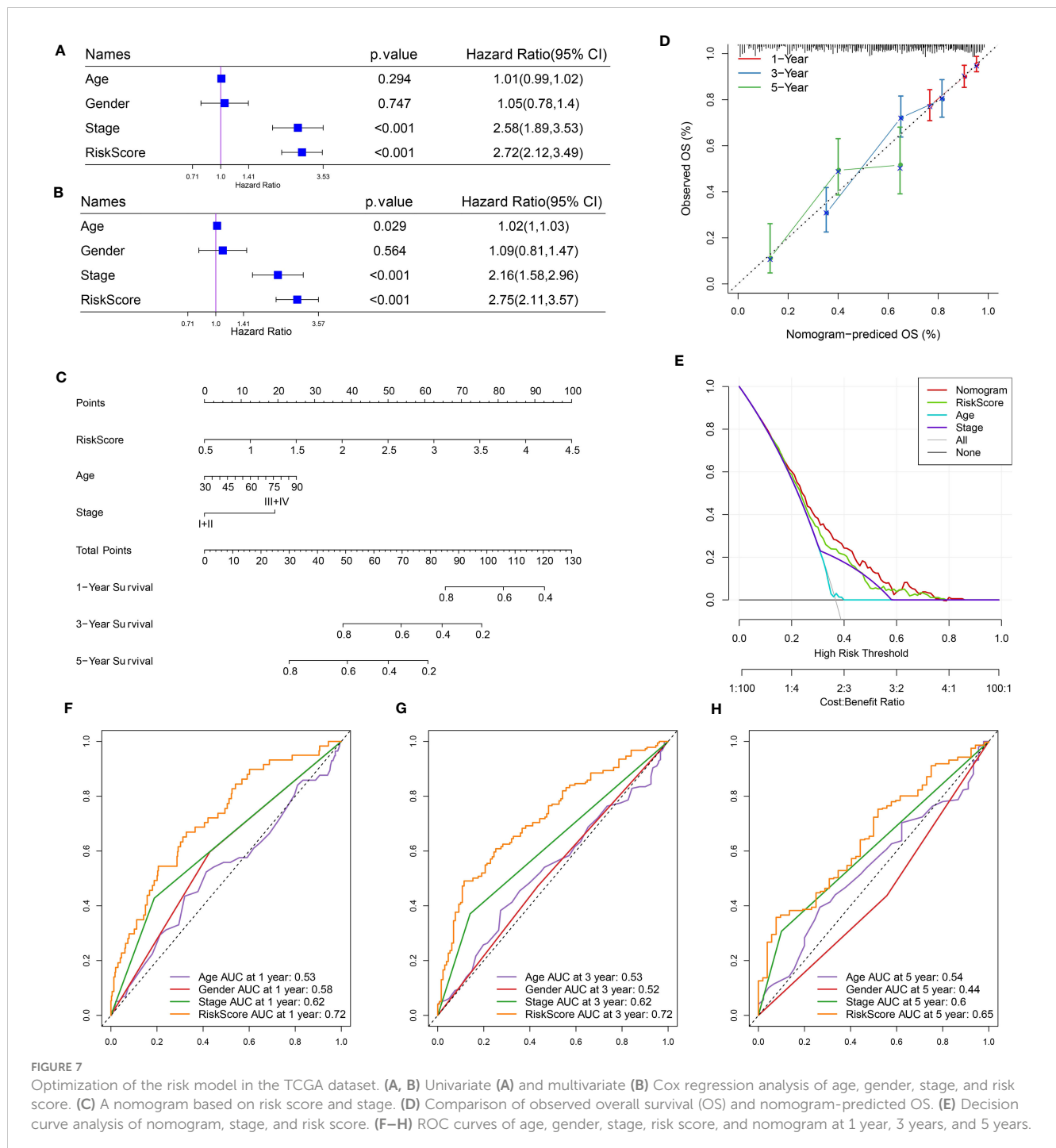


FIGURE 7

Optimization of the risk model in the TCGA dataset. (A, B) Univariate (A) and multivariate (B) Cox regression analysis of age, gender, stage, and risk score. (C) A nomogram based on risk score and stage. (D) Comparison of observed overall survival (OS) and nomogram-predicted OS. (E) Decision curve analysis of nomogram, stage, and risk score. (F–H) ROC curves of age, gender, stage, risk score, and nomogram at 1 year, 3 years, and 5 years.

different prognoses of the three clusters, we analyzed their immune characteristics, biological pathways, and genomic features. C1 had the highest immune infiltration while C3 had the lowest immune infiltration. The immune microenvironment has been demonstrated to be extraordinarily associated with anti-cancer immune response and prognosis (32). Previous research has shown that high immune infiltration is associated with a good prognosis in lung cancer (33, 34), which is consistent with our result. However, C3 showed higher immune infiltration than C2. We further analyzed the enrichment of tumor-related pathways. Not surprisingly, C3 manifested higher enrichment of oncogenic

pathways such as VEGF, hypoxia, EGFR, and MAPK than the other two clusters, which were responsible for the worse prognosis of C3. In addition, C3 had high levels of TMB, aneuploidy, homologous recombination defect, fraction altered, and number of segments, suggesting the genomic instability of C3. Genomic instability is one of cancer hallmarks, which is closely associated with cancer development (35, 36). Molecular analysis of the three clusters explained their different prognosis and the reliability of molecular subtyping based on immune cell markers.

Furthermore, we assessed DEGs between different clusters, and a total of 133 DEGs were identified. KEGG analysis revealed that

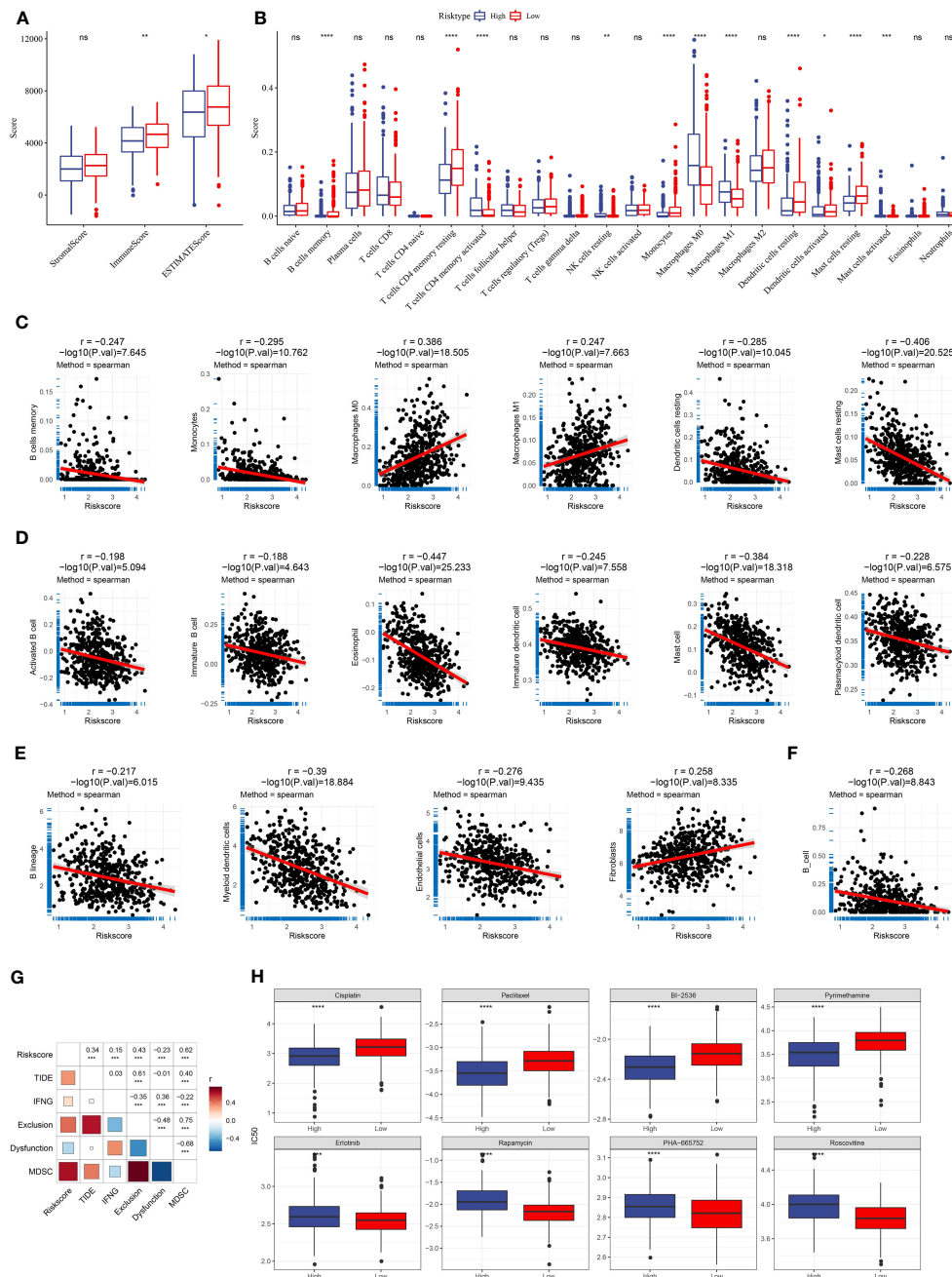


FIGURE 8
 The relation of risk score to immune infiltration and chemotherapeutic response. **(A)** Immune and stromal infiltration of the two risk groups were calculated by the ESTIMATE algorithm. **(B)** The estimated enrichment of 22 immune cells in the two risk groups was calculated by CIBERSORT. **(C–F)** Spearman correlation analysis of risk score with the enrichment of different immune cells and stromal cells was calculated by different tools [CIBERSORT (C), ssGSEA (D), MCP-counter (E), and TIMER (F)]. Significant correlation pairs were presented. **(G)** Correlation heatmap of risk score with TIDE, IFN- γ , T cell exclusion, T cell dysfunction, and MDSC. **(H)** The estimated IC50 of eight chemotherapeutic drugs in the high- and low-risk groups. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$; ns, no significant.

these DEGs were significantly enriched in the cell cycle, implying that the dysregulation of the cell cycle pathway greatly contributed to the different outcomes of the three clusters. Using Lasso and stepAIC algorithms, we confirmed seven key prognostic genes including CPA3, S100P, PTTG1, LOXL2, MELTF, PKP2, and TMPRSS11E to establish a risk model. CPA3 belongs to zinc metalloproteases, which may contribute to the degradation of endogenous proteins and the inactivation of venom-associated

peptides. CPA3 was reported as a prognostic biomarker in colon cancer (37), but its role has not been fully investigated in cancers. S100P is an EF-hand calcium-binding protein and its overexpression has been detected in many cancer types (38). The overexpression of S100P promotes cancer progression and metastasis through extracellular signaling via the RAGE receptor or through intracellular interaction with ezrin (38). S100P serves as a potential therapeutic target in cancer treatment. Chien et al.

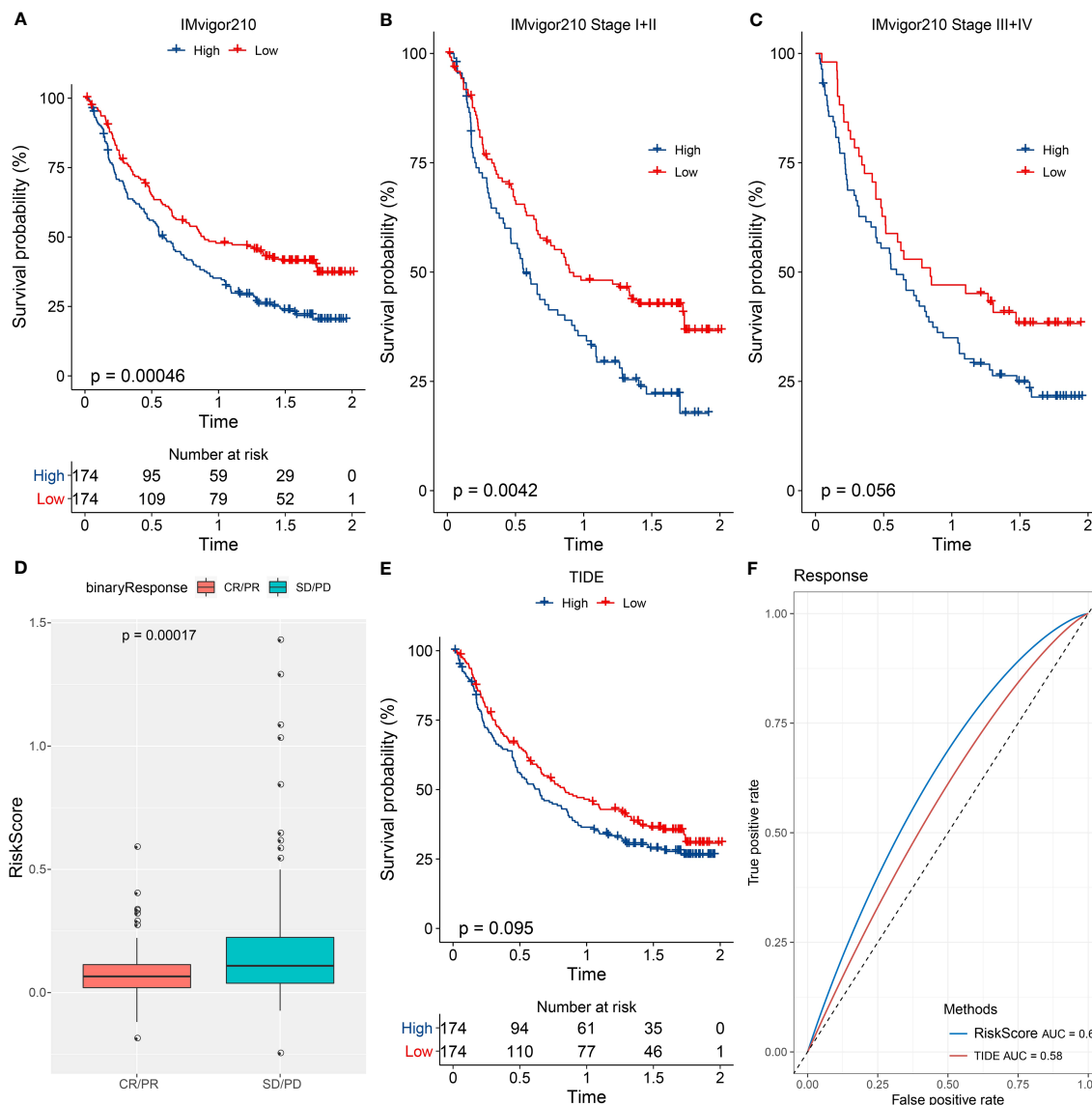
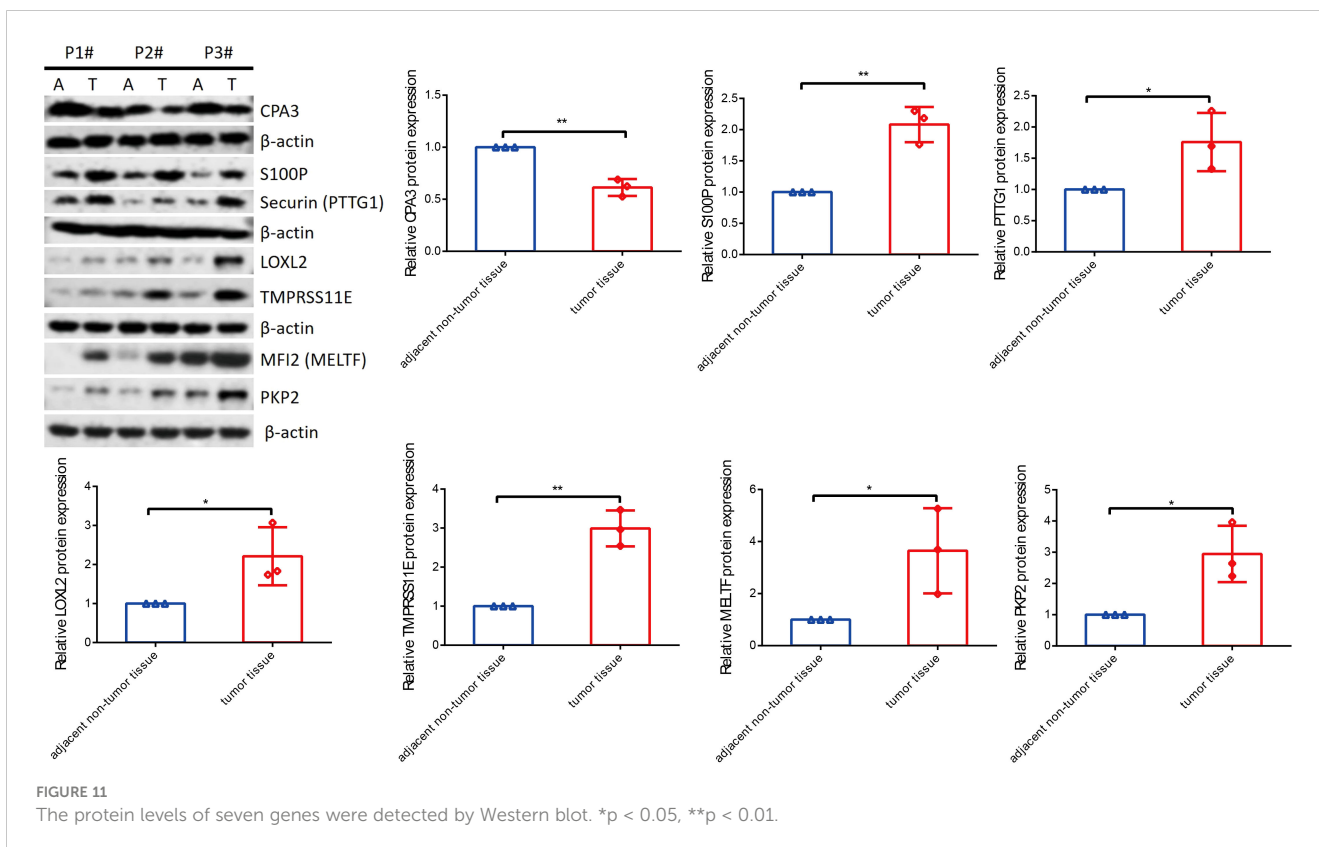
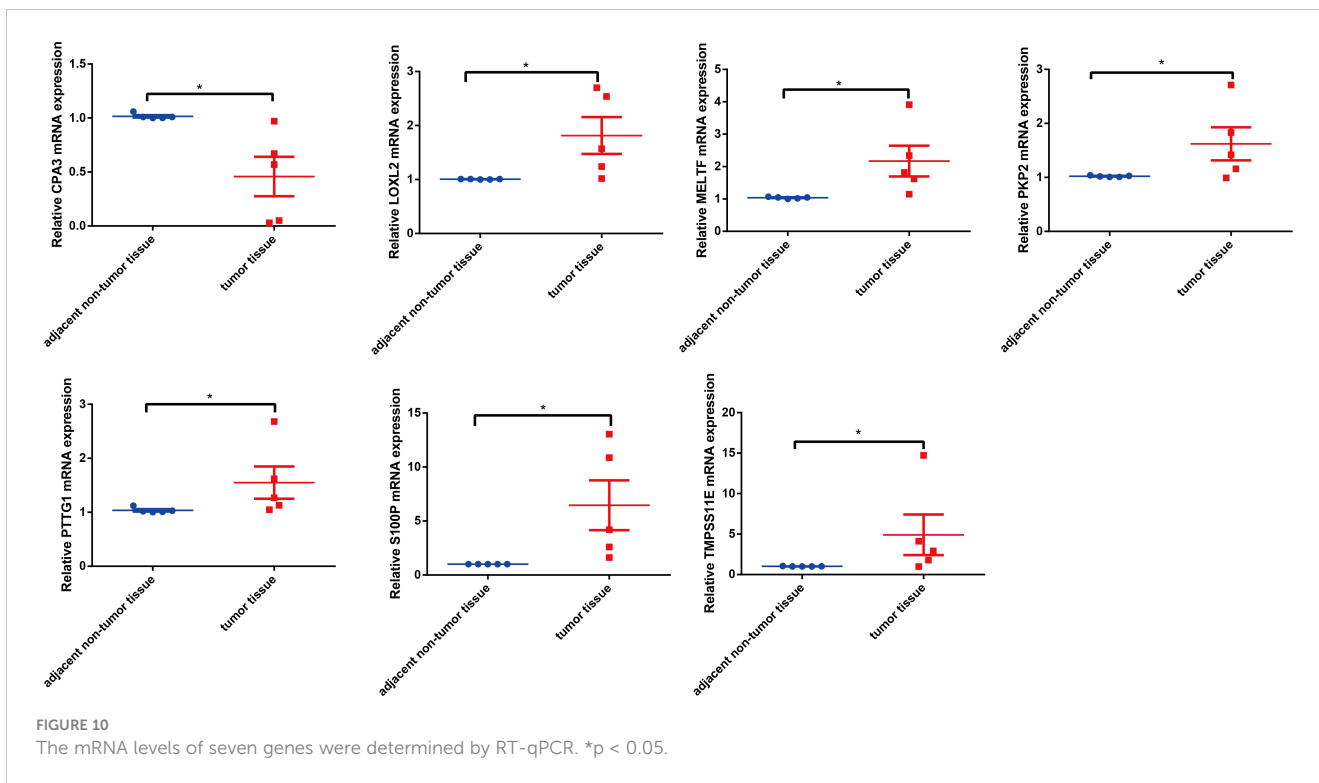


FIGURE 9 The performance of the risk model in immunotherapy-treated patients. (A–C) Kaplan-Meier survival analysis of the two risk groups in all patients (A), patients of early stages (I+II) (B), and patients of late stages (III+IV) (C) in the IMvigor210 dataset. (D) Comparison of the risk score in CR/PR and SD/PD groups. (E) Kaplan-Meier survival analysis of the high- and low-TIDE score groups. (F) ROC curve of TIDE score and risk score in the response to immunotherapy.

revealed that targeting S100P inhibited cell motility and migration in invasive non-small cell lung cancer (NSCLC) cells (39). PTTG1 is also detected to be overexpressed in multiple cancers such as breast cancer, lung cancer, and gastric cancer (40). PTTG1 is involved in regulating the transcription of basic fibroblast growth factor, activating c-Myc, and promoting apoptosis of specific cell types (41, 42). Knockdown of PTTG1 strengthens the anti-cancer immune response in LUAD (43), implying PTTG1 is a potential target. LOXL2 belongs to the lysyl oxidase (LOX) family and increasing evidence has shown its role in cancer cell invasion and metastasis (44). High expression of LOXL2 is associated with poor prognosis in NSCLC patients (45). Melanotransferrin (MELTF) is identified as a serological biomarker in gastric cancer (46), colorectal cancer (47), and lung cancer (48). Plakophilins 2

(PKP2) was found to promote LUAD cell proliferation and migration through epithelial-mesenchymal transition and EGFR signaling (49, 50). Few papers have discussed TMPRSS11E’s function in malignancies.

In the TCGA and GSE31210 datasets, the seven-gene risk model performed satisfactorily in predicting the prognosis of LUAD. Compared to patients with a potentially high risk, more immunological infiltration and a better prognosis were found in the low-risk group. The risk score and TIDE score were positively correlated, showing that high-risk groups had a higher likelihood of avoiding immunotherapy. In response to chemotherapy, the two risk groups also showed different IC50. The high-risk group was more sensitive to Cisplatin, Paclitaxel, BI-2536, and Pyrimethamine, while the low-risk group was more sensitive to Erlotinib, Rapamycin, PHA-



665752, and Roscovitine. These results supported that the risk model may provide a direction for treating LUAD patients with immunotherapy and chemotherapeutic drugs.

Conclusions

In conclusion, this study integrated the analysis of scRNA-seq data and bulk RNA-seq data to identify molecular subtypes for LUAD patients. We focused on immune cell marker genes and confirmed three clusters with different prognoses, clinical features, immune infiltration, and genomic features. Importantly, we constructed a seven-gene risk model which was reliable and effective to predict prognosis and guide personalized therapies for LUAD patients.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Ningbo No.2 Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

The study was designed by JS and JJ. GZ participated in the literature review. JJ performed the data analysis and interpretation. The original draft of the manuscript was done by GZ. JS reviewed and edited the manuscript. The manuscript was reviewed and approved by all authors.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* (2018) 391(10125):1023–75. doi: 10.1016/S0140-6736(17)33326-3
3. Bade BC, Dela Cruz CS. Lung cancer 2020: epidemiology, etiology, and prevention. *Clinics chest Med* (2020) 41(1):1–24. doi: 10.1016/j.ccm.2019.10.001
4. Lin JJ, Cardarella S, Lydon CA, Dahlberg SE, Jackman DM, Jänne PA, et al. Five-year survival in EGFR-mutant metastatic lung adenocarcinoma treated with EGFR-TKIs. *J Thorac Oncology: Off Publ Int Assoc Study Lung Cancer* (2016) 11(4):556–65. doi: 10.1016/j.jtho.2015.12.103

Funding

This work was supported by the National Natural Science Foundation of China (82003199 to JS); the Natural Science Foundation of Zhejiang Province (LY22H160013 to JS); the Natural Science Foundation of Ningbo (202003N4205 to JS); the Research Foundation of Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences (2020YJY0209 to JS); the Ningbo Health Branding Subject Fund (PPXK2018-05); the Natural Science Foundation of Zhejiang Province (LHDMD23H160001); the Science and technology innovation guidance fund project of Hangzhou Medical College (grant no. CX2022006); the Hwamei Fund (grant no. 2022HMKY37); the Medical Health Science and Technology Project of Zhejiang Provincial Health Commission (Grant No. 2021KY1009 and 2022KY1138); the Key discipline of Hwamei Hospital, University of Chinese Academy of Science (grant no. 2020ZDXK03).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1177847/full#supplementary-material>

5. Sullivan I, Planchard D. ALK inhibitors in non-small cell lung cancer: the latest evidence and developments. *Ther Adv Med Oncol* (2016) 8(1):32–47. doi: 10.1177/1758834015617355
6. Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol* (2020) 17(8):807–21. doi: 10.1038/s41423-020-0488-6
7. Shields MD, Marin-Acevedo JA, Pellini B. Immunotherapy for advanced non-small cell lung cancer: A decade of progress. *Am Soc Clin Oncol Educ book Am Soc Clin Oncol Annu Meeting* (2021) 41:1–23. doi: 10.1200/EDBK_321483
8. Sadeghi Rad H, Monkman J, Warkiani ME, Ladwa R, O'Byrne K, Rezaei N, et al. Understanding the tumor microenvironment for effective immunotherapy. *Medicinal Res Rev* (2021) 41(3):1474–98. doi: 10.1002/med.21765
9. Frankel T, Lanfranca MP, Zou W. The role of tumor microenvironment in cancer immunotherapy. *Adv Exp Med Biol* (2017) 1036:51–64. doi: 10.1007/978-3-319-67577-0_4

10. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol* (2016) 1418:93–110. doi: 10.1007/978-1-4939-3578-9_5
11. He Y, Yu F, Tian Y, Hu Q, Wang B, Wang L, et al. Single-cell RNA sequencing unravels distinct tumor microenvironment of different components of lung adenocarcinoma featured as mixed ground-glass opacity. *Front Immunol* (2022) 13:903513. doi: 10.3389/fimmu.2022.903513
12. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* (2012) 72(1):100–11. doi: 10.1158/0008-5472.CAN-11-1403
13. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol* (2015) 19(1a):A68–77. doi: 10.5114/wo.2014.47136
14. Shen W, Song Z, Xiao Z, Huang M, Shen D, Gao P, et al. Sangerbox: A comprehensive, interaction-friendly clinical bioinformatics analysis platform. *iMeta* (2022) 1(3):e36. doi: 10.1002/imt2.36
15. Gribov A, Sill M, Lück S, Rücker F, Döhner K, Bullinger L, et al. SEURAT: visual analytics for the integrated analysis of microarray data. *BMC Med Genomics* (2010) 3:21. doi: 10.1186/1755-8794-3-21
16. Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* (2023) 51(D1):D870–d6. doi: 10.1093/nar/gkac947
17. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* (2019) 47(W1):W199–w205. doi: 10.1093/nar/gkz401
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
19. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26(12):1572–3. doi: 10.1093/bioinformatics/btq170
20. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12
21. Yoshihara K, Shahmoradgol M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
22. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
23. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* (2009) 462(7269):108–12. doi: 10.1038/nature08460
24. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* (2016) 17(1):218. doi: 10.1186/s13059-016-1070-5
25. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun* (2018) 9(1):20. doi: 10.1038/s41467-017-02391-6
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Software* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01
27. Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Trans Med* (2016) 4(7):136. doi: 10.21037/atm.2016.03.35
28. Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet (London England)* (2017) 389(10064):67–76. doi: 10.1016/S0140-6736(16)32455-2
29. Bagaev A, Kotlov N, Nomic K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* (2021) 39(6):845–65.e7. doi: 10.1016/j.ccell.2021.04.014
30. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity* (2018) 48(4):812–30.e14. doi: 10.1016/j.immuni.2018.03.023
31. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118
32. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Cancer Res* (2019) 79(18):4557–66. doi: 10.1158/0008-5472.CAN-18-3962
33. Liu X, Wu S, Yang Y, Zhao M, Zhu G, Hou Z. The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. *Biomedicine pharmacotherapy = Biomedecine pharmacotherapie* (2017) 95:55–61. doi: 10.1016/j.biopha.2017.08.003
34. Guan M, Jiao Y, Zhou L. Immune infiltration analysis with the CIBERSORT method in lung cancer. *Dis markers* (2022) 2022:3186427. doi: 10.1155/2022/3186427
35. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability—an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* (2010) 11(3):220–8. doi: 10.1038/nrm2858
36. Duijff PHG, Nanayakkara D, Nones K, Srihari S, Kalimutho M, Khanna KK. Mechanisms of genomic instability in breast cancer. *Trends Mol Med* (2019) 25(7):595–611. doi: 10.1016/j.molmed.2019.04.004
37. Fang Z, Xu S, Xie Y, Yan W. Identification of a prognostic gene signature of colon cancer using integrated bioinformatics analysis. *World J Surg Oncol* (2021) 19(1):13. doi: 10.1186/s12957-020-02116-y
38. Gibadulinova A, Tothova V, Pastorek J, Pastorekova S. Transcriptional regulation and functional implication of S100P in cancer. *Amino Acids* (2011) 41(4):885–92. doi: 10.1007/s00726-010-0495-5
39. Chien MH, Lee WJ, Hsieh FK, Li CF, Cheng TY, Wang MY, et al. Keap1-Nrf2 interaction suppresses cell motility in lung adenocarcinomas by targeting the S100P protein. *Clin Cancer research: an Off J Am Assoc Cancer Res* (2015) 21(20):4719–32. doi: 10.1158/1078-0432.CCR-14-2880
40. Rehfeld N, Geddert H, Atamna A, Rohrbeck A, Garcia G, Kliszewski S, et al. The influence of the pituitary tumor transforming gene-1 (PTTG-1) on survival of patients with small cell lung cancer and non-small cell lung cancer. *J Carcinogenesis* (2006) 5:4. doi: 10.1186/1477-3163-5-4
41. Pei L. Identification of c-myc as a down-stream target for pituitary tumor-transforming gene. *J Biol Chem* (2001) 276(11):8484–91. doi: 10.1074/jbc.M009654200
42. Bernal JA, Luna R, Espina A, Lázaro I, Ramos-Morales F, Romero F, et al. Human securin interacts with p53 and modulates p53-mediated transcriptional activity and apoptosis. *Nat Genet* (2002) 32(2):306–11. doi: 10.1038/ng997
43. Chen Z, Cao K, Hou Y, Lu F, Li L, Wang L, et al. PTTG1 knockdown enhances radiation-induced antitumor immunity in lung adenocarcinoma. *Life Sci* (2021) 277:119594. doi: 10.1016/j.lfs.2021.119594
44. Wu L, Zhu Y. The function and mechanisms of action of LOXL2 in cancer (Review). *Int J Mol Med* (2015) 36(5):1200–4. doi: 10.3892/ijmm.2015.2337
45. Zhan P, Lv XJ, Ji YN, Xie H, Yu LK. Increased lysyl oxidase-like 2 associates with a poor prognosis in non-small cell lung cancer. *Clin Respir J* (2018) 12(2):712–20. doi: 10.1111/crj.12584
46. Sawaki K, Kanda M, Umeda S, Miwa T, Tanaka C, Kobayashi D, et al. Level of melanotransferrin in tissue and sera serves as a prognostic marker of gastric cancer. *Anticancer Res* (2019) 39(11):6125–33. doi: 10.21873/anticancer.13820
47. Shin J, Kim HJ, Kim G, Song M, Woo SJ, Lee ST, et al. Discovery of melanotransferrin as a serological marker of colorectal cancer by secretome analysis and quantitative proteomics. *J Proteome Res* (2014) 13(11):4919–31. doi: 10.1021/pr500790f
48. Lei Y, Lu Z, Huang J, Zang R, Che Y, Mao S, et al. The membrane-bound and soluble form of melanotransferrin function independently in the diagnosis and targeted therapy of lung cancer. *Cell Death Disease* (2020) 11(10):933. doi: 10.1038/s41419-020-03124-2
49. Wu Y, Liu L, Shen X, Liu W, Ma R. Plakophilin-2 promotes lung adenocarcinoma development via enhancing focal adhesion and epithelial-mesenchymal transition. *Cancer Manage Res* (2021) 13:559–70. doi: 10.2147/CMAR.S281663
50. Hao XL, Tian Z, Han F, Chen JP, Gao LY, Liu JY. Plakophilin-2 accelerates cell proliferation and migration through activating EGFR signaling in lung adenocarcinoma. *Pathology Res Practice* (2019) 215(7):152438. doi: 10.1016/j.prp.2019.152438