

Prognostic analysis of lung adenocarcinoma based on cancer-associated fibroblasts genes using scRNA-sequencing

Han Zhang^{1,*}, Yuhang Wang^{1,*}, Kai Wang², Yun Ding¹, Xin Li², Shuai Zhao², Xiaoteng Jia¹, Daqiang Sun²

¹Clinical School of Thoracic, Tianjin Medical University, Tianjin, China

²Department of Thoracic Surgery, Tianjin Chest Hospital of Tianjin University, Tianjin, China

*Equal contribution and share first authorship

Correspondence to: Daqiang Sun; **email:** sdqmdsr@163.com, <https://orcid.org/0009-0000-6602-357X>

Keywords: scRNA-seq, CAFs markers, prognosis models, immunotherapy, drug sensitivity

Received: March 21, 2023

Accepted: June 9, 2023

Published: July 11, 2023

Copyright: © 2023 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Cancer-associated fibroblasts (CAFs) are an important component of the tumor microenvironment (TME). CAFs can promote tumor occurrence and metastasis by promoting cancer cell proliferation, angiogenesis, extracellular matrix (ECM) remodeling, and drug resistance. Nevertheless, how CAFs are related to Lung adenocarcinoma (LUAD) has not yet been revealed, especially since the CAFs-related prediction model has yet to be established. We combined Single-cell RNA-sequencing (scRNA-seq) and Bulk-RNA data to develop a predictive model of 8 CAFs-associated genes. Our model predicted LUAD prognosis and immunotherapy efficacy. TME, mutation landscape and drug sensitivity differences were also systematically analyzed between the LUAD patients of high- and low-risk. Moreover, the model prognostic performance was validated in four independent validation cohorts in the Gene expression omnibus (GEO) and the IMvigor210 immunotherapy cohort.

INTRODUCTION

Lung cancer has the highest prevalence and mortality among all malignancies [1], with 2.1 million diagnosed cases yearly, resulting in 1.8 million deaths [2]. Lung cancer comprises non-small (NSCLC; > 85%) and small cell lung cancer (SCLC; < 15%) [3]. LUAD is the primary NSCLC type [4]. Surgical resection offers the most efficient treatment option for LUAD patients at the early stage [5]. Furthermore, several therapies, including chemotherapy, targeted therapy, radiotherapy, and immunotherapy, have significantly enhanced LUAD patient survival rates [6]. LUAD treatment has significantly improved, while the prognosis remains dismal, especially for advanced cases [7]. Immune checkpoint inhibitors (ICIs) for PD-1, PD-L1, and CTLA4 have changed the treatment mode in advanced NSCLC. However, immunotherapy efficiency is limited [8], and in some cases, it leads to more rapid tumor growth than the patients who did not receive immuno-

therapy [9]. In summary, there are still many challenges in treating LUAD, and we must further investigate prognostic and efficacy predictors.

TME is the dynamic and highly heterogeneous environment in which cancerous cells interact with their surroundings, including immune cells, stromal cells, and blood vessels [10]. It has been demonstrated that CAFs are essential for TME [11–13]. CAFs can promote tumor occurrence and metastasis by promoting cancer cell proliferation, angiogenesis, ECM remodeling, and drug resistance [11, 14, 15]. CAFs in lung cancer upregulate *MiR-21* to induce calumenin protein secretion, thereby increasing tumor aggressiveness [16]. Detailed studies of the CAFs-immune microenvironment interactions, especially the CAFs-immune cell sophisticated mechanisms, may yield innovative approaches to target CAF-directed immunotherapy in the future [17]. The role of CAFs in LUAD has yet to be investigated.

TME cannot be characterized at high resolution using current bulk omics analyses. scRNA-seq facilitates massively parallel characterization of multiple cells at the transcriptome level, compensating for the limitations of traditional bulk-omic to explore TME characteristics more deeply and comprehensively [18–20]. Previous studies combining scRNA-seq and bulk RNA-seq successfully established prediction models to predict LUAD prognosis and immunotherapy efficacy based on B cells and natural killer (NK) cells [21, 22]. This study aims to analyze scRNA-seq data from 10 untreated primary LUAD cases systematically and identify 120 genes associated with CAFs. Finally, an 8-gene prognostic model was designated utilizing Lasso and stepwise multivariate COX regression on the above gene sets in The Cancer Genome Atlas (TCGA) database. TME, mutation landscape and drug sensitivity differences were systematically analyzed between the LUAD patients of high- and low-risk. Moreover, the model prognostic performance was validated in four independent validation cohorts in the GEO and the IMvigor210 immunotherapy cohort.

MATERIALS AND METHODS

The scRNA-seq files of 10 LUAD patients were downloaded from <https://codeocean.com/capsule/8321305/tree/v1>. The RNA sequencing profiles and clinical and mutation information of 517 LUAD patients were obtained from The Cancer Genome Atlas (TCGA; <https://portal.gdc.cancer.gov/>). The University of California, Santa Cruz, Xena browser (UCSC Xena; <https://xenabrowser.net/>) was also accessed to download LUAD patient survival data as a supplement. Additionally, GEO (<http://www.ncbi.nlm.nih.gov/geo/>) was assessed, and four microarray data, GSE31210 (n = 246), GSE37745 (n = 196), GSE50081 (n = 181), and GSE72094 (n = 442), were downloaded. The LUAD patient Tumor Immune Dysfunction and Exclusion (TIDE) scores were downloaded by accessing <http://tide.dfci.harvard.edu/>. Immunophenoscore (IPS) data of LUAD patients were downloaded from the Cancer Immunome Atlas (TCIA) database (<http://tcia.at/>), which is positively related to cancer immunogenicity with the predictive ability of the cancer patient immunotherapy response [23]. The immunotherapeutic cohort (IMvigor210) was acquired according to the guideline at <http://research-pub.gene.com/IMvigor210CoreBiologies/> [24].

Cancer-associated fibroblasts (CAFs) marker gene identification

‘Seurat (version 4.2.0)’ R package was utilized to read and quality control scRNA-seq files. The primary

steps were as follows: The function `Read-10x` was used to read the data, and the function ‘`CreatSeuratObject`’ was used to create the object. Then, the data were merged, and low-quality cells were removed with `nFeature > 10000` or `< 500`, `nCount > 100000` or `< 1000`, mitochondrial gene `> 30%`, and erythrocyte gene `> 5%`. Afterward, variance-stabilized UMI counts were normalized using the ‘`SCTransform`’ function [25]. SNN graphs and UMAP embeddings were constructed using the top 30 principal components. A canonical cell type marker score was used to identify the cell types in cell clusters. Markers were computed for each cell cluster using the ‘`FindAllMarkers`’ function with the following parameters: `only.pos = T`, `min.pct = 0.25`, `logfc.threshold = 1`. Thus, 120 marker genes related to CAFs were identified (Supplementary Table 1). The association was analyzed between individual cell subsets and 50 Hallmarkers by performing the R-package ‘`singleseqset` (version 1.1.0)’. The related gene expressions were also visualized using the ‘`FeaturePlot`’ function in different cell subsets.

Prognostic signature based on CAFs marker genes construction and validation

The ‘`survival` (version 3.3-1)’ and ‘`survivalminer` (version 0.4.9)’ R packages were used to perform a univariate COX regression for CAFs-related genes, with 32 genes exhibiting significant effects on overall survival (OS). Thirteen genes were selected for the Cox regression model using the ‘`glmnet` (version 4.1-4)’ R package and the minimum absolute contraction and LASSO method. Further gene screening was conducted using stepwise multiple COX regression for OS genes. Finally, a risk model was created based on eight gene mRNA expressions and their associated risk coefficients. Based on the median model score (4.842808), the TCGA patients were separated into high- and low-risk groups for each GEO-independent validation set. Kaplan-Meier was applied for the high-risk group survival analysis, and using the ‘`timeROC` (version 0.4)’ R package, the area under the curve (AUC) was calculated. The R package ‘`regplot` (version 1.1)’ was utilized for building the nomogram. Simultaneously, the model diagnostic capability was verified using four independent GEO datasets.

Enrichment analysis of CAFs

Besides converting the Gene ID into EntrezID, the 120 CAFs-related gene expression levels of LUAD patients were extracted from the TCGA database. Gene Ontology (GO) [26] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [27] were employed for the risk

groups using the ‘clusterProfiler (version 4.4.4)’ R package.

Somatic mutations analysis

The ‘maftools (version 2.12.05)’ R package was utilized to calculate and display the top 20 gene mutation landscape using TCGA database-tumor mutation data. The difference in tumor mutation burden (TMB) between the risk groups was illustrated as a violin plot. The Kaplan-Meier method revealed that high- and low-TMB and risk scores impact survival outcomes.

Estimate of tumor immune microenvironment

The TCGA database immune-infiltrating files for all tumors were acquired using the TIMER2.0 database (<http://timer.cistrome.org>). Then, the Wilcoxon rank-sum was applied to calculate the difference in immune cell infiltration between six algorithms, including TIMER [28], CIBERSORT [29], CIBERSORT-ABS [30], QUANTISEQ [31], MCPOUNTER [32], XCELL [33], and EPIC [34]. The above process was implemented using the ‘limma v 3.52.4’ and ‘pheatmap (v 1.0.12)’ R packages. Next, immune checkpoint expression and major histocompatibility complex (MHC) gene were compared between the two risk patients using the same method. Supplementary Table 2 depicts these genes. Other cancer-related gene sets, such as chemokines, growth factors and regulators, proteases and regulators, soluble or shed receptors or ligands, and interleukins, were also explored for differential expression between the two groups. Finally, the R package ‘estimate’ was conducted to further evaluate the LUAD patient immune microenvironment.

Immunotherapy efficacy prediction

The patient IPS was calculated to predict the immunotherapy response under various conditions. Next, the risk score ability was validated to predict the immunotherapy response in the IMvigor210 immunotherapy cohort. Finally, the risk group TIDE score was compared, revealing a positive correlation with the possibility of immune escape.

Prediction of chemotherapy response

The LUAD chemotherapy response data were downloaded from the Genomics of Drug Sensitivity in Cancer database (GDSC, <https://www.cancerrxgene.org>) [35]. The R package ‘oncoPredict (version 0.2)’ was utilized to analyze the two-group sensitivity to various chemotherapy drugs. Additionally, R packages ‘ggplot2 (version 3.4.0)’ and ‘ggpubr (version 0.5.0)’ were utilized to visualize the results.

Data availability statement

All data generated or analyzed during this study are included within this article.

RESULTS

Identification of CAFs marker genes

We obtained single-cell sequencing data from 10 surgically resected primary LAUDs (73,566 cells) without specific treatment [36]. After strict quality control (removing nFeature > 10,000 or 500, nCount > 100,000 or 1,000, mitochondrial gene > 30%, and erythrocyte gene > 5%), 62115 high-quality cells were obtained. Figure 1E depicts the quality control results for each patient as a violin plot. The SCTransform algorithm identified 3,000 hypervariable genes and data homogenization and normalization. For the 3,000 genes mentioned above, 30 PCs were used to reduce dimensionality, yielding 29 cell clusters (Figure 1A). Subsequently, annotation with canonical cell markers (Figure 1D) yielded seven cell subsets, with the 19th cluster annotated as CAFs. However, the cell distribution existed between different patients, but there was no significant batch effect (Figure 1C). Analysis of seven cell subsets and 50 Hallmarkers revealed a significant positive correlation between CAFs and epithelial-mesenchymal transition (EMT), myogenesis, and Wnt/β-catenin signaling (Figure 1F). Ultimately, the ‘FindAllMarkers’ function yielded 120 CAFs-related genes.

Prognostic signature based on CAFs marker genes construction and validation

Univariate COX regression analysis of 120 CAFs-related genes in the TCGA database revealed that 32 genes were significantly related to OS, with seven genes representing protective factors and 25 genes representing risk factors (Supplementary Figure 1A). These 32 genes were further analyzed using LassoCOX regression, and 13 were obtained according to the optimal λ value (logλ = -4). Then, the abovementioned 13 genes were used for stepwise multivariable COX regression analysis. Finally, eight genes, *TIMP1*, *TPM2*, *NR2F2*, *MFAP4*, *SOD3*, *CAVI*, *SERPINH1*, and *FMO2*, were used to build a prognosis model. The ‘Featureplot’ function was applied to display these eight gene expressions in different cell subsets. Online TIMER (<http://timer.comp-genomics.org>) was accessed to investigate these eight abnormal gene expressions in pan-cancer (Supplementary Figure 1D, 1E). The final model combined with the coefficient of each gene is: Risk = (0.2403052**TIMP1* expression) + (0.1475574**TPM2* expression) + (0.2686883**NR2F2*

expression) + (-0.1635410**MFAP4* expression) + (-0.1219534**SOD3* expression) + (0.2009316**CAVI* expression) + (0.1449876**SERPINH1* expression) + (-0.1816038**FMO2* expression).

An increased risk score indicates an increase in the number of patients who have died, indicating a high risk and poor OS positive association. The high-risk group had significant *TIMP1*, *TPM2*, *NR2F2*, *CAVI*, and *SERPINH1* expressions, whereas the low-risk group had significant *MFAP4*, *SOD3*, and *FMO2* expressions (Figure 2A). The Kaplan-Meier (K-M) survival curve demonstrated that high-risk patients had significantly lower OS ($p < 0.001$, Figure 2B), PFS ($p = 0.012$), and DSS ($p < 0.001$, Figure 2C, 2E). Despite this, the two groups did not differ significantly in DFS ($p = 0.056$). The ‘timeROC’ R package was employed to evaluate the model diagnostic efficacy and related clinical characteristics for survival outcomes, revealing that pathological stage (AUC = 0.705) and risk (AUC = 0.688, Figure 2F) were the two indicators with good diagnostic efficiency. After 10-fold cross-validation, the model’s performance was evaluated, and its mean AUC values were 0.688, 0.680, and 0.646 over one, three, and five years, respectively.

Next, four independent GEO datasets verified the model’s validity and stability. In GSE37745, the high-

risk patient had a worse prognosis ($p < 0.001$, Figure 3A), with AUC values of 0.554, 0.616, and 0.631 at one-, three-, and five-year, respectively (Figure 3B). In GSE72094, the low-risk patient OS was significantly prolonged ($p < 0.001$, Figure 3C), with AUC values of 0.664, 0.709, and 0.638 at one-, three-, and five-year, respectively (Figure 3D). In GSE50081, high-risk patients had poor OS ($p = 0.01$, Figure 3E), with AUC values of 0.649, 0.572, and 0.584 at one-, three-, and five-year, respectively (Figure 3F). In GSE31210, the high-risk patient had poor OS ($p = 0.016$, Figure 3G) and AUC values of 0.876, 0.649, and 0.600 with the one-, three-, and five-year, respectively (Figure 3H). In conclusion, high-risk patients had poor long-term survival outcomes in validation and training sets. Furthermore, our model has a good prediction efficiency for the patient survival outcome.

The model diagnostic and prognostic ability combined with clinical features

Univariate COX regression was applied in the TCGA database by five clinical features, including age, gender, stage, tumor site, and smoking status. Besides the model risk score, OS risk factors revealed stage ($p < 0.001$, HR = 1.763, 95% CI = 1.441–1.942) with risk score ($p < 0.001$, HR = 2.787, 95% CI = 2.091–3.713, Figure 4A). Meanwhile, Multivariate COX regression findings

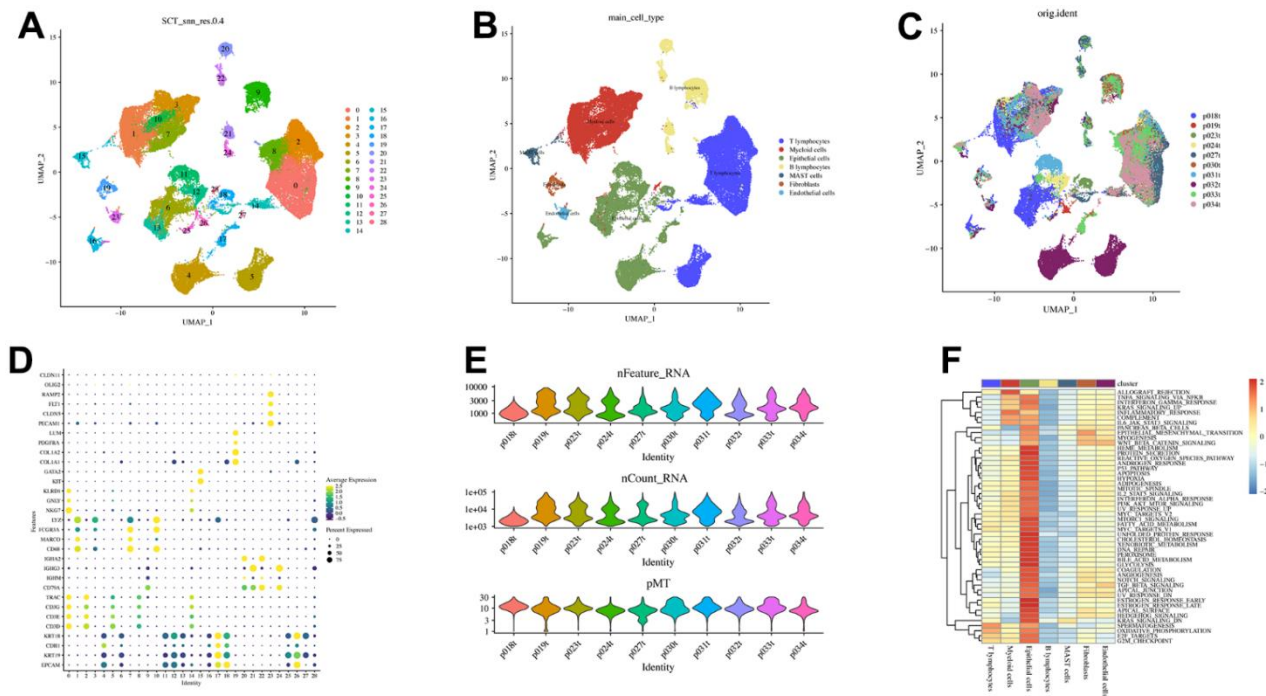


Figure 1. Visualization of scRNA-seq data from 10 LUAD patients. (A) The U-MAP algorithm identified 29 cell subsets. (B) Seven cell types were identified based on marker genes. (C) UMAP plot of 62,115 cells, colored by patients. (D) Marker genes for different cell subsets. (E) Quality control results. (F) Association between cell subsets and Hallmarks.

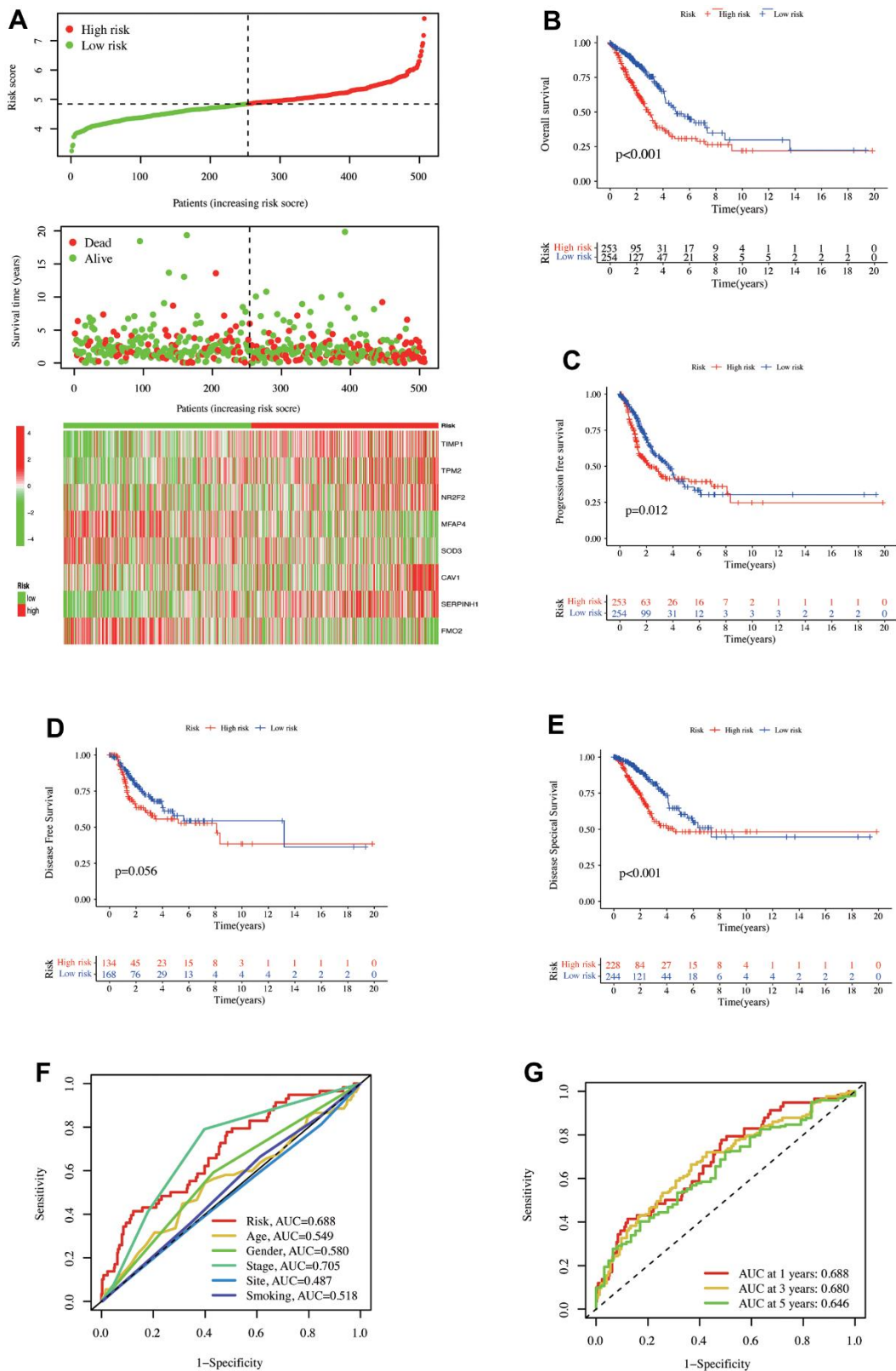


Figure 2. The prediction model of CAFs-related genes was established in the TCGA database. (A) Distribution of risk scores and OS status. 8 gene expressions were involved in the model construction in the high- and low-risk groups. Kaplan-Meier curve illustrated the differences between high- and low-risk groups for **(B)** OS, **(C)** PFS, **(D)** DFS, and **(E)** DSS. **(F)** ROC curves for risk scores and clinical characteristics. **(G)** ROC curves for one-, three-, and five-year OS of risk scores in the TCGA database.

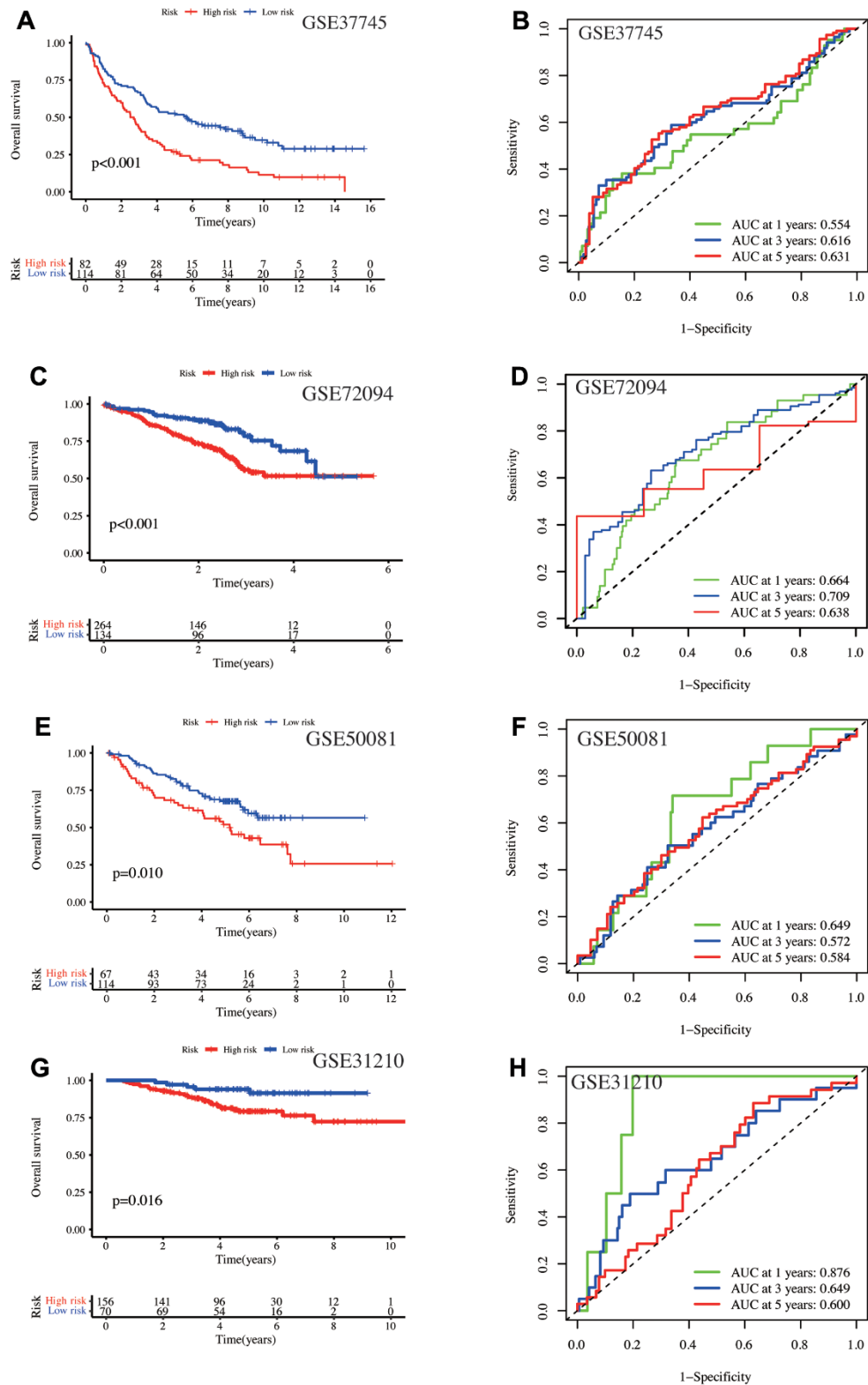


Figure 3. Validation of the model using four independent GEO database cohorts. In the GSE37745 cohort, (A) Kaplan-Meier curves depict the prognosis for the high- and low-risk groups, and (B) the AUC values at one-, three-, and five-year. In the GSE72094 cohort, (C) Kaplan-Meier curves show the prognosis of the high- and low-risk groups, and (D) the AUC values at one-, three-, and five-year. In the GSE50081 cohort, (E) K-M curves display the prognosis for the high- and low-risk groups, and (F) the AUC values at one-, three-, and five-year. In the GSE31210 cohort, (G) Kaplan-Meier curves depict the prognosis of the high- and low-risk groups, and (H) the AUC values at one-, three-, and five-year.

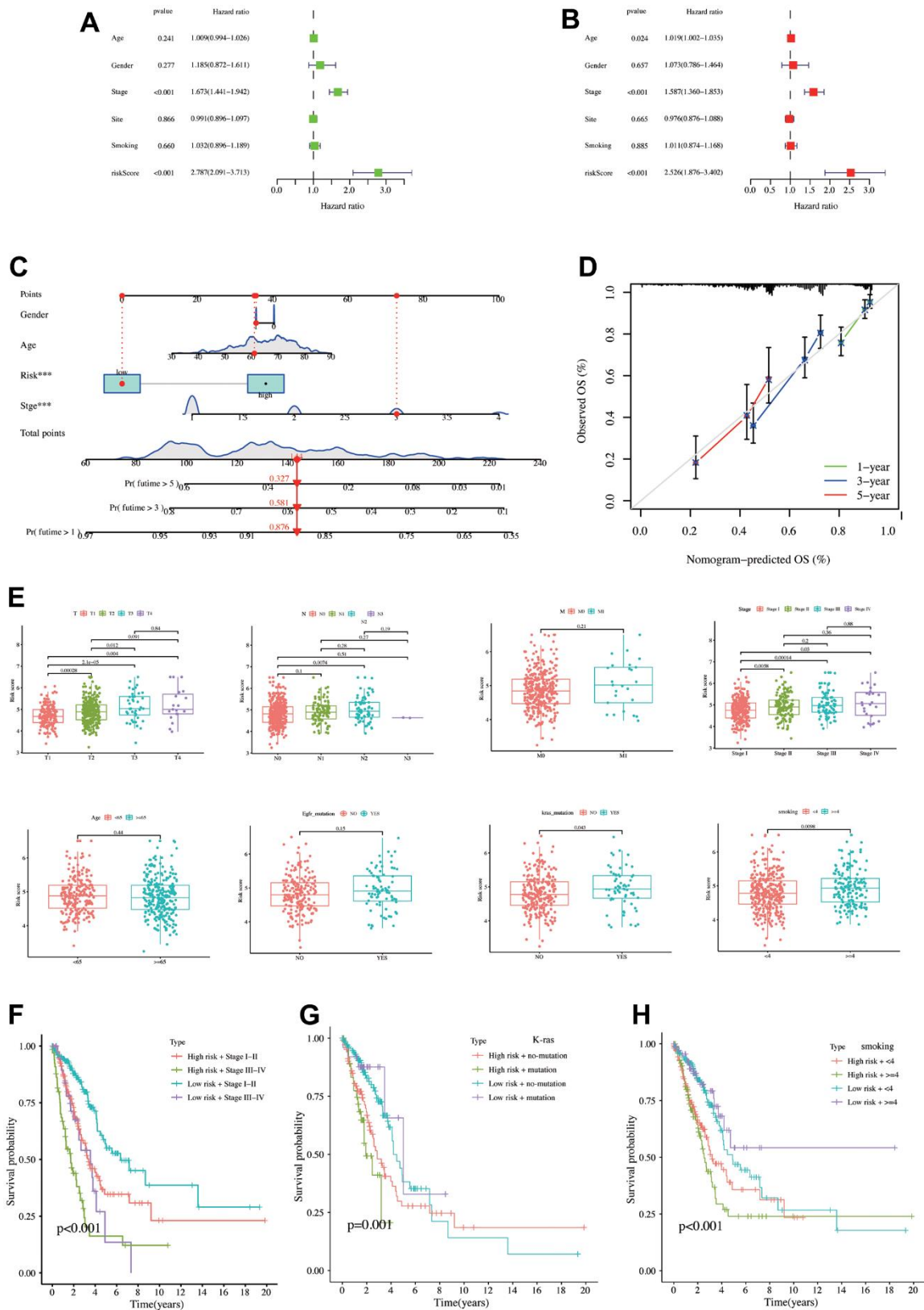


Figure 4. Establishment and validation of nomograms, risk score differences between different clinical characteristics. (A) Univariate COX regression analysis. (B) Multivariate COX regression analysis. (C) Nomogram constructed using risk score and clinical characteristics. (D) Calibration curve of the nomogram. (E) Risk score differences between different clinical characteristics. The K-M analysis curves for the patients stratified by (F) risk score and stage, (G) risk score and K-Ras mutation status, (H) risk score, and smoking level.

revealed that the OS risk factors were age ($p = 0.024$, HR = 1.019, 95% CI = 1.002–1.035), stage ($p < 0.001$, HR = 1.587, 95% CI = 1.360–1.853), and risk score ($p < 0.001$, HR = 2.526, 95% CI = 1.876–3.402, Figure 4B). The model risk score was a survival outcome risk factor in both COX analyses. Next, the constructed nomogram revealed that a patient's gender, age, risk score, and pathology stage are as follows: male, 60 years old, low risk, and stage III. Then, the patient's total score was 144, and the one-, three-, and five-year survival rates were 0.876, 0.581, and 0.327, respectively (Figure 4C). The calibration curve exhibited a high correlation between predicted and actual values, revealing good predictive nomogram performance (Figure 4D). Figure 4E demonstrates that patients with a later pathological stage had a higher risk score, partly explaining the poor prognosis of the high-risk patient. Risk scores did not differ significantly between patients aged 65 years and those aged more than or less than 65 years ($p = 0.44$) and between EGFR mutation and non-mutation groups ($p = 0.15$). Patients with a K-RAS mutation exhibited a higher risk score ($p = 0.043$), as did patients with a higher smoking score ($p = 0.0098$). Next, patients were stratified using the statistically significant clinical characteristics and risk scores described above. High-risk + stage III–IV exhibited a poor prognosis, while low-risk + stage I–II exhibited the best prognosis (Figure 4F). Meanwhile, the K-RAS mutation + high-risk group had poor survival prognosis than the low-risk and K-RAS non-mutation groups (Figure 4G). Finally, the high-risk + high-smoking score patients also have a poor survival prognosis (Figure 4H).

Enrichment analysis

GO analysis indicated the significance of the 120 CAFs-related genes enrichment in ECM organization and structural constituent besides the collagen-containing ECM and other pathways (Figure 5A). Moreover, KEGG analysis indicated that these genes significantly correlate with protein digestion and absorption, ECM-receptor interaction, complement and coagulation cascades, and other pathways (Figure 5B), indicating the CAF importance in ECM remodeling, consistent with former studies [13–15]. KEGG and GO analyses were also conducted to explore functional differences between the risk groups. According to Gene Set Enrichment Analysis (GSEA) analysis, the significant high-risk group enrichment in DNA replication, mitotic nuclear division, and other pathways (Figure 5C) indicates its significant correlation with cell replication pathways. Meanwhile, the significant low-risk group enrichment in the B cell receptor (BCR) and immunoglobulin complex pathways (Figure 5E) indicates its significant enrichment in immune-related pathways.

GSEA analysis also indicated the significant enrichment of cell cycle, ECM receptor interaction, and focal adhesion pathway in the high-risk group (Figure 5D). Meanwhile, the low-risk group enrichment appeared in allograft rejection, asthma, and autoimmune thyroid disorder pathway (Figure 5F).

Somatic mutation analysis

TMB is a quantitative biomarker that reflects the total tumor cell mutation count, the total count of somatic gene coding errors, base substitution, gene insertion, or deletion errors detected per million bases [37]. Higher TMB levels correlated with longer OS after immunotherapy across multiple cancers. Figure 6A, 6B depict a waterfall map of the top 20 genes mutation landscape with the highest mutation frequency. The high-risk patients exhibited increased TMB levels ($p = 6.1e-05$), suggesting their further benefit from the immunotherapy (Figure 6C). Meanwhile, Figure 6D depicts that the risk score positively correlates with the TMB level ($R = 0.2$, $p = 5.9e-06$). The median TMB level was utilized as the boundary to divide TMB into high and low groups. According to the K-M curve, the high TMB group has a better survival prognosis than the low TMB group (Figure 6E). When combined with the risk model, the low TMB+ high-risk group has the worst survival prognosis (Figure 6F).

Tumor immune microenvironment estimate

The immune cell composition and abundance in the TME strongly influenced tumor progression and immunotherapy effectiveness. A heat map demonstrates the immune cells with significant differences ($p < 0.05$, Figure 7A). Overall, the low-risk group had a significant abundance of immune cell infiltration, including T and B cells, in the TIMER and XCELL algorithms. Next, the common risk group immune checkpoint gene expressions were assessed (Supplementary Table 2). Twenty genes displayed significant differences, with 15 upregulated in the low-risk group and 5 upregulated in the high-risk group (Figure 7B). The differential expression of common MHC molecules was also analyzed (Supplementary Table 2). Interestingly, all 15 results with significant differences were high in the low-risk group (Figure 7B).

Moreover, the other tumor-related gene differential expression was analyzed (Figure 7C). The low-risk group had higher immune and estimate scores through the 'estimate' algorithm, but the stromal score did not differ significantly (Figure 7D). Supplementary Table 3 provides 'estimate' score details for each LUAD patient in the TCGA data. Moreover, tumor purity was reduced in the low-risk group (Figure 7F). The risk score and

tumor purity exhibited a significant positive association (Figure 7G). In summary, the low-risk group had greater immune infiltration and upregulation of immune checkpoint-related and MHC genes.

Immunotherapy efficacy prediction

LUAD patient IPS was downloaded from the TCIA database to assess the immunotherapy response,

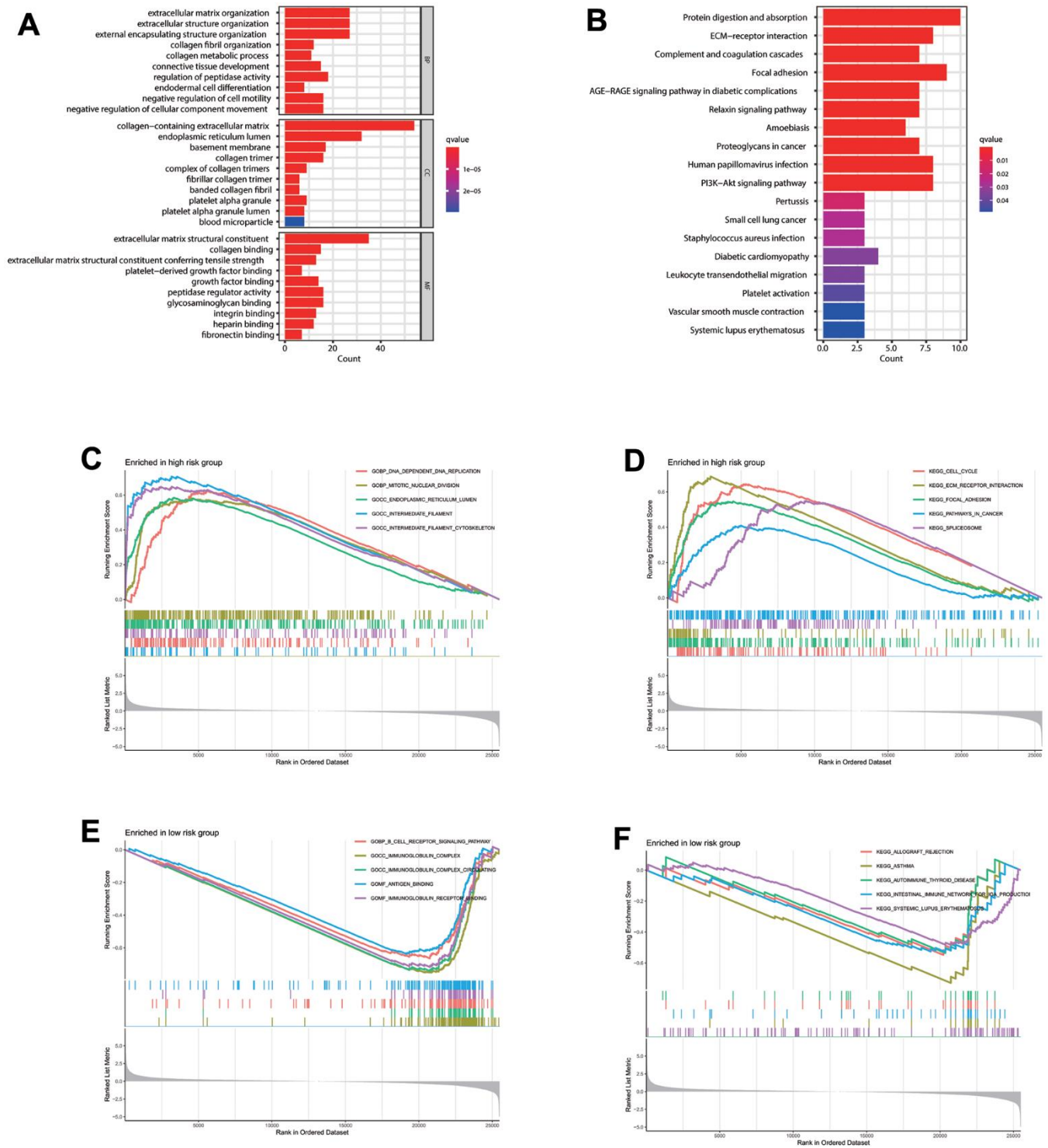


Figure 5. Enrichment analysis. (A) GO and (B) KEGG enrichment analysis of 120 CAFs-related genes. Gene Set Enrichment Analysis (GSEA) analysis for patients in the (C, D) high- and (E, F) low-risk groups.

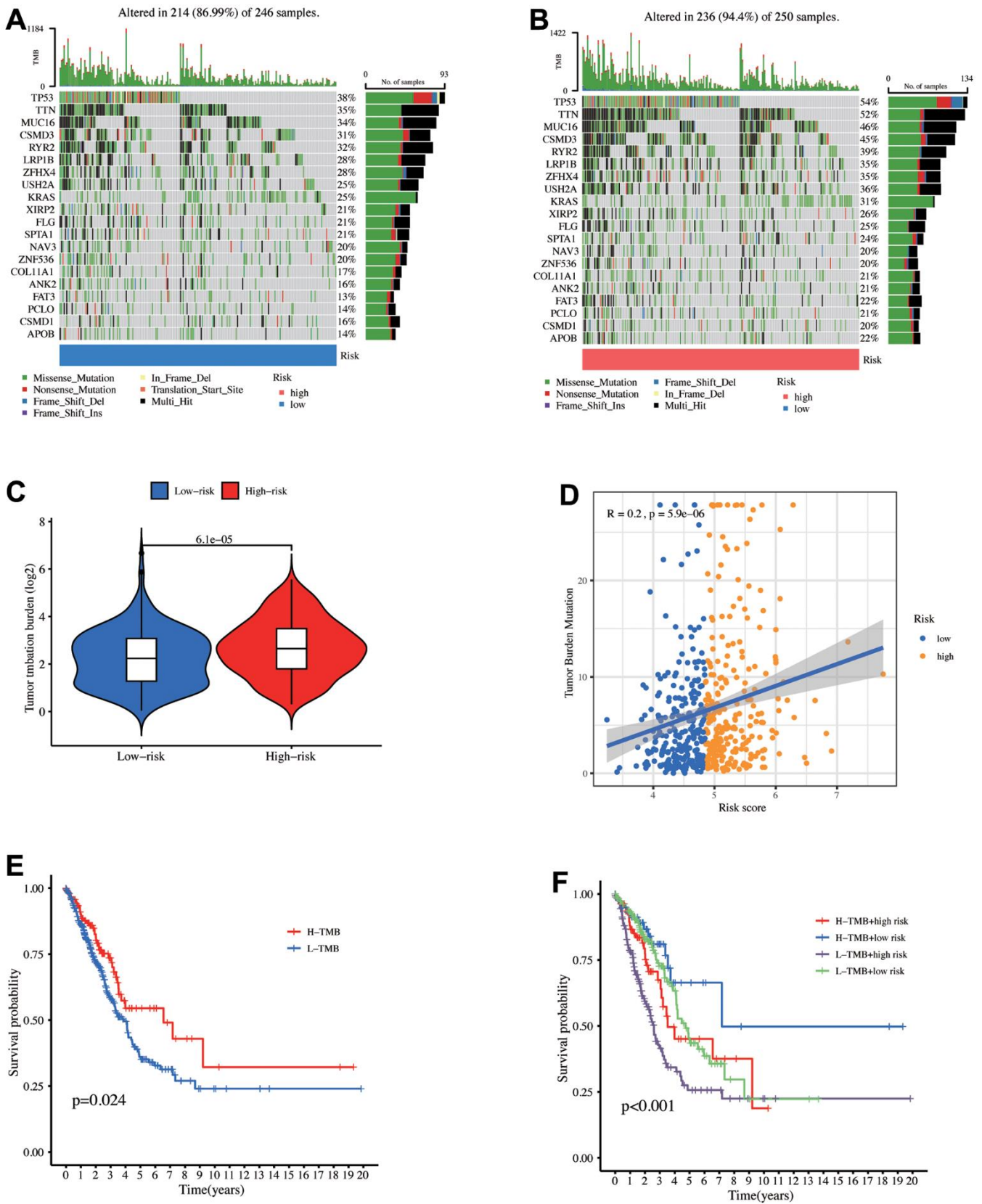


Figure 6. Mutation analysis between the high- and low-risk groups. The mutational landscapes of the (A) high- and (B) low-risk groups. (C) TMB level comparison between high- and low-risk groups. (D) Correlation between risk score and TMB. (E) K-M curves for high and low TMB levels. (F) K-M curves for the patients stratified by risk score and TMB.

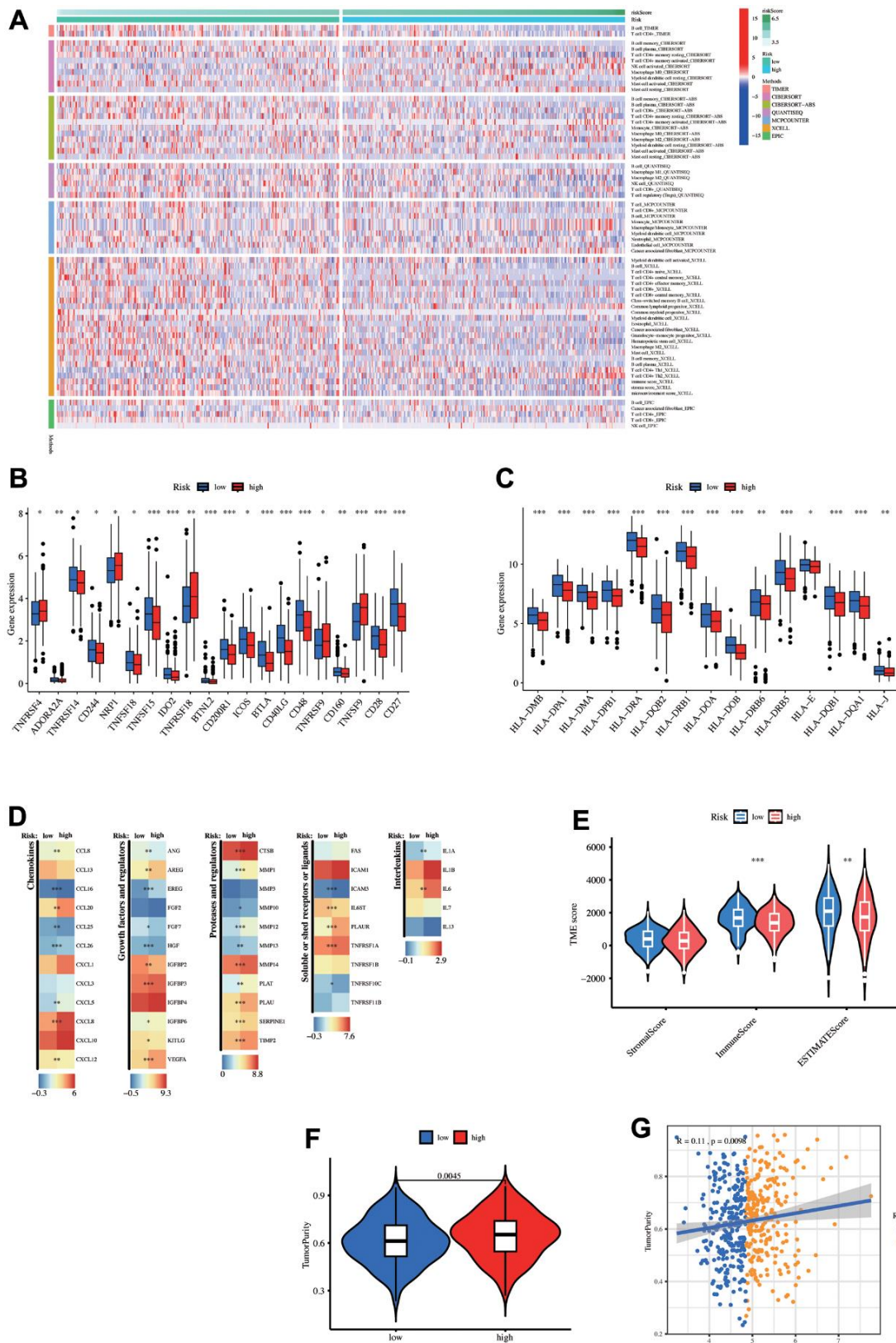


Figure 7. Tumor microenvironment (TME) analysis. (A) The heatmap displays the differences in the number of immune cells among the eight algorithms. Differential expression of (B) immune checkpoint-related genes, (C) MHC-related genes, and other (D) Tumor-related gene sets between high- and low-risk groups. (E) Estimate score differences between the high- and low-risk groups. (F) Tumor purity differences between the high- and low-risk groups. (G) Correlation between risk-score and tumor purity.

indicating the higher IPS of the low-risk patients regardless of whether CTLA-4 and PD1 were expressed or not and further benefit from immunotherapy (Figure 8A–8D). Moreover, the complete response (CR) + partial response (PR) group risk score was decreased than the stable disease (SD) + progression disease (PD) group (Figure 8E), revealing the strongest diagnostic power of neoantigen (AUC = 0.778), followed by TMB (AUC = 0.718). The risk score also showed a certain predictive power (AUC = 0.578, Figure 8F). K-M analysis demonstrated that the high-risk patients had poor survival outcomes in the IMvigor210 immunotherapy cohort (Figure 8G). The low-risk patients had higher TIDE scores, indicating a high susceptibility to immune escape (Figure 8H). Our study discovered immunotherapy benefits for low-risk patients besides better long-term survival after immunotherapy.

Prediction of chemotherapy response

Further analysis determined whether IC₅₀ levels differed between the LUAD patient risk groups. We selected 20 LUAD-related chemotherapy drugs with significant differences in drug sensitivity, revealing that the low-risk group had high sensitivity to the following 17 drugs: Gemcitabine, 5-Fluorouracil, Epirubicin, Savolitinib, AZD6738, Alisertib, AZD1332, I-BET-762, Ulixertinib, Trametinib, Cisplatin, Cediranib, Talazoparib, BI-2536, Crizotinib, Cytarabine, and Dasatinib (Figure 9A). In contrast, three drugs (Axitinib, ABT737, and AZD8055) demonstrated higher sensitivity in high-risk patients (Figure 9B). The above results indicate that our model shows promising potential in distinguishing the sensitivity of LUAD patients to chemotherapy drugs, thereby providing new avenues for future treatment strategies in LUAD. However, further validation through clinical trials and animal experiments is necessary to confirm the accuracy of our drug sensitivity predictions.

DISCUSSION

Lung cancer has the highest prevalence and mortality among all malignancies [1, 2]. LUAD, a primary subtype of lung cancer, has a five-year OS of less than 20% [38]. Immunotherapy has improved LUAD patient prognosis to a certain extent, but with some challenges, such as poor response and resistance [8, 9, 39]. Thus, a model that can predict LUAD prognosis and immunotherapy efficacy must be established. CAFs are the most important component of TME [11–15]. They can promote immunosuppressed TME by secreting cytokines, growth factors, chemokines, exosomes, and other methods, allowing tumors to escape the immune system [17]. Nevertheless, how

CAFs are related to LUAD has been unrevealed, especially since the CAFs-related prediction model has yet to be established.

Our study constructed an eight-gene prognostic model based on CAFs-associated genes in combination with scRNA-Seq and bulkRNA-seq, demonstrating good prognostic and diagnostic capabilities in the TCGA database and four independent GEO validation cohorts. The high-risk group patient LUAD had a poor prognosis associated with a later pathological stage. Concurrently, eight genes involved in the model construction were also associated with LUAD pathogenesis or prognosis. The glycoprotein tissue inhibitor of metalloproteinase -1 (*TIMP-1*) primarily affects fibroblasts and keratinocytes proliferation and participates in ECM turnover [40, 41]. The interaction between *TIMP1* and *CD63* can promote LUAD progression [42], and *TIMP1* overexpression was correlated to poor LUAD prognosis [42, 43]. *TPM2* is an actin filament-binding protein whose primary function is stabilizing and integrating actin filaments [44, 45]. *TPM2* overexpression in LUAD may inhibit tumor proliferation and invasiveness, resulting in a better prognosis for patients [46]. Nuclear receptor subfamily 2 group F member 2 (*NR2F2*) is a nuclear orphan receptor primarily involved in tumor progression, stem cell differentiation, energy metabolism, and other processes [47, 48]. *NR2F2*-related signaling pathway can promote platinum resistance in lung cancer with brain metastasis via high glutathione (GLH) consumption [49]. The ECM protein Microfibrillar-associated protein 4 (*MFAP4*) locates primarily in the vascular wall ECM. It is involved in tissue repair and angiogenesis [50, 51]. *miR147b* can regulate *MFAP4* to promote LUAD progression [52]. Superoxide dismutase 3 (*SOD3*), a secreted antioxidant enzyme [53], primarily functions to maintain REDOX homeostasis in tissues [54]. LUAD has lower *SOD3* expression than normal tissues, and patients with elevated *SOD3* expression have a poor prognosis [55]. Caveolin-1 (*CAVI*) is a membrane protein predominantly involved in ECM remodeling, cell migration, cancer signaling, and endocytosis [56]. *CAVI*-derived peptides can inhibit idiopathic pulmonary fibrosis (IPF) progression [57], while *CAVI* overexpression can promote lung cancer progression and increase radiotherapy resistance [56]. The Serpin peptidase inhibitor clade H member 1 (*SERPINH1*) gene is a member of the serine proteinase inhibitor serpin superfamily [58]. *SERPINH1* is highly expressed in IPF and lung cancer patients, suggesting that it may be involved in both disease pathogenesis [59]. Flavin Containing Dimethylaniline Mono-oxygenase 2 (*FMO2*) gene is primarily present in mammalian lung tissues; its abundance in LUAD tissue is lower than in normal tissue [60].

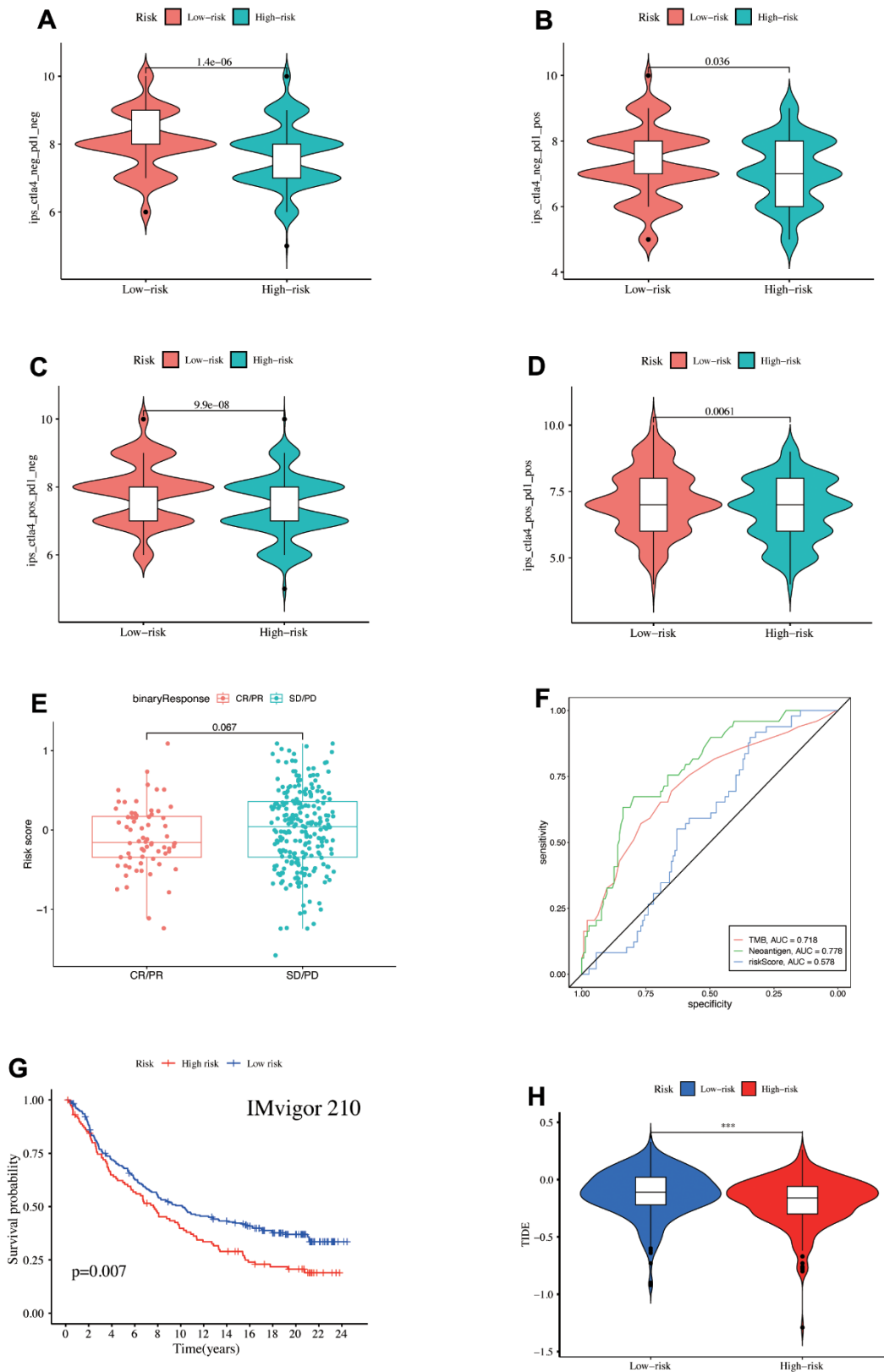


Figure 8. (A–D) IPS score differences. Validation of the model's ability to predict immunotherapy response in the IMvigor210 cohort, (E) Comparison of the risk score differences between CR + PR and SD + PD groups, (F) ROC curve illustrated the ability of TMB, Neoantigen, and risk-score to predict the efficacy of immunotherapy, (G) survival prognosis of high- and low-risk groups. (H) TIDE score differences in the TCGA database.

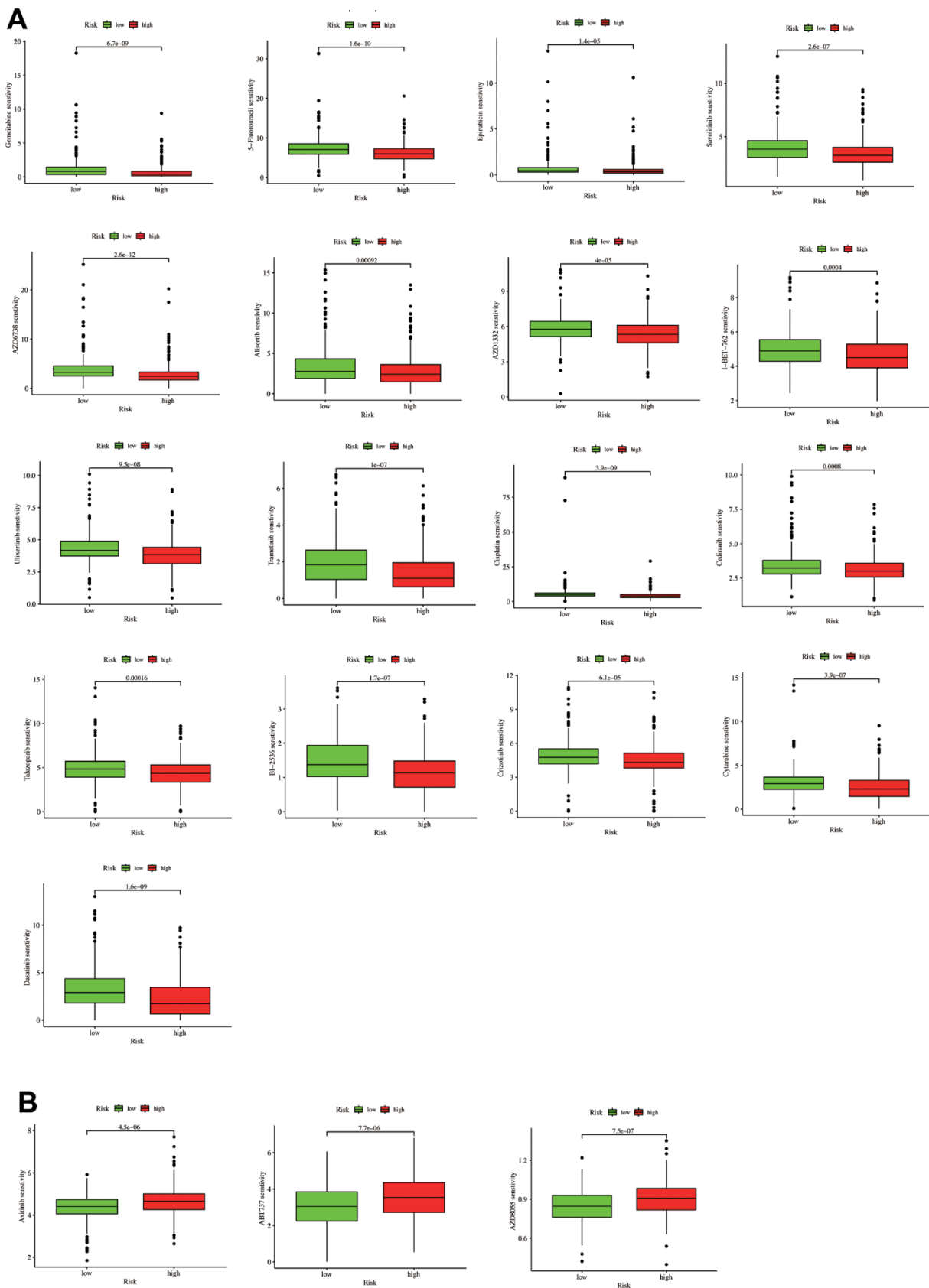


Figure 9. A drug sensitivity analysis. (A) Drugs with higher sensitivity in the low-risk group. (B) Drugs with higher sensitivity in the high-risk group.

Next, enrichment analyses were conducted to explore the biological roles of CAFs involved in LUAD. scRNA-seq analysis between CAFs cell subsets and 50 Hallmarkers revealed a positive correlation between CAFs and EMT and Wnt/ β -catenin signaling. EMT is a cellular program that can promote tumorigenesis, metastatic tumor ability, and tumor resistance to anti-tumor therapy by reshaping cell-cell and cell-ECM interactions [61]. Wnt/ β -catenin signaling is essential for embryonic development and tissue homeostasis. Abnormal WNT/ β -catenin pathway activation is closely related to tumor progression and poor prognosis [62, 63]. Wnt/ β -catenin signaling activation can regulate disease progression and metastasis in LUAD [64]. The enrichment analysis of 120 CAFs-related genes revealed a significant enrichment of these genes in ECM organization, protein digestion and absorption, and other ECM-related pathways. The enrichment analysis revealed significant enrichment in DNA replication and cell cycle pathways in the high-risk group, which may result in abnormal cell cycle regulation and DNA replication, leading to a poor prognosis. Meanwhile, the low-risk group had significant enrichment in immune-related pathways, such as the B cell receptor (BCR) signaling pathway, allograft rejection, and abundant immune cell infiltration, indicating a better prognosis.

The EGFR mutation group has higher risk scores. EGFR mutations are a common type of NSCLC mutation that affect approximately 40% to 55% of Asian LUAD patients [1, 65]. Although LUAD patients with EGFR mutations exhibited sensitivity to tyrosine kinase inhibitor (TKI) treatment, their overall prognosis was poor due to the tendency to develop drug resistance later in treatment [66, 67]. Patients with PD-L1 overexpression and EGFR mutations have a decreased immunotherapy response rate, possibly due to lower tumor-infiltrating T-cell activity in patients with EGFR mutations [68]. This is consistent with our IMvigor210 cohort study, where the low-risk group had better PD-1 therapy response and prognosis.

Afterward, the mutational landscape of the risk groups revealed a higher TP53 gene mutation frequency in the high-risk group (54% vs 38%). Patients with low TP53 mutation burden exhibited immune cell infiltration and were more probably to profit from immune checkpoint inhibitors (ICIs) therapy [69, 70]. In addition to TBM's significant positive relation to the risk score and mutation-associated neoantigens (MANA), it was overexpressed in the high-risk group and involved in the CD8+T process of recognizing tumor cells during immunotherapy [37, 71]. It is assumed that high TMB-level patients exhibit more sensitivity to immunotherapy. However, TMB also has certain shortcomings in predicting the efficacy of immunotherapy. A study finding did not predict PFS, complete pathologic remission (CRP), and main pathological response

(MPR) in stage IIIA NSCLC patients treated with immunotherapy [72]. Multiple factors, such as tumor antigens, lymphocyte infiltration, and antigen-presenting cells (APCs), influence tumor immunotherapy response [71]. It is a complex and dynamic process that needs comprehensive evaluation.

The TME difference was systematically analyzed between the two groups to explore whether the risk score can anticipate the immunotherapy response. Six different algorithms were employed to assess the differences in immune cell infiltration between the two groups, revealing a greater immune cell abundance in the low-risk patients. According to the TME evaluated by the 'ESTIMATE' algorithm, the high immune and estimate scores indicated the high immune cell infiltration that makes the tumor less likely to undergo immune escape, explaining this group's better prognosis consisting of our enrichment analysis, thereby revealing its significant enrichment in immune-related pathways. In contrast, the high-risk group had higher tumor purity. Moreover, patients with high immune cell infiltration are more probably to profit from immunotherapy [73].

Furthermore, the relationship between risk score and immune-related genes revealed significant immune checkpoint-related gene abundance in the low-risk group. Meanwhile, all MHC molecules were significantly overexpressed in the low-risk group. T cells must recognize tumor antigens in MHC presence to generate anti-tumor immune responses. The immune checkpoint-related genes and the MHC gene overexpression were correlated to better immunotherapy response [74]. The TME analysis suggests the increased immunotherapy benefit for the low-risk group.

Subsequently, we utilized IPS [23] and TIDE [75] scores to predict immunotherapy response in the risk groups. We observed the performance of risk scores in immunotherapy cohorts (IMvigor210) treated with anti-PD-L1. Moreover, the IPS score correlates positively with immunotherapy efficacy [23]. We discovered that immunotherapy benefits patients in the low-risk group with higher IPS scores regardless of CTLA4 and PD1 expression. The CR + PR group patients had lower risk scores after IMvigor210 immunotherapy, indicating high immunotherapy sensitivity. Our risk score also had some predictive power for immunotherapy response (AUC = 0.578). In the IMvigor210 cohort, the high-risk group prognosis was poor, demonstrating the model accuracy and consistency for the LUAD patient prognosis prediction, consistent with the TCGA and GEO database results.

In contrast, the low-risk patients had a lower TIDE score and were more likely to have an immune escape. Despite

having a lower TMB level and a higher TIDE score, the immune infiltration degree was higher, more immune-related pathways were enriched, and more immune-related genes were highly expressed. Finally, they had a better prognosis and immunotherapy effects. Therefore, higher immune cell infiltration level patients in LUAD are more likely to benefit from immunotherapy. A previous study revealed that the immune cell infiltration level was more predictive than the TMB level [21].

We also examined the drug susceptibility differences between the risk groups to better guide the risk score clinical practice. We revealed that the low-risk group had higher sensitivity to 17 drugs: Gemcitabine, 5-Fluorouracil, Epirubicin, Savolitinib, AZD6738, Alisertib, AZD1332, I-BET-762, Ulixertinib, Trametinib, Cisplatin, Cediranib, Talazoparib, BI-2536, Crizotinib, Cytarabine, and Dasatinib (Figure 9A). In contrast, three drugs (Axitinib, ABT737, and AZD8055) had higher sensitivity in high-risk patients.

We combined scRNA-seq and Bulk-RNA data to develop a predictive model of 8 CAFs-associated genes. Our model predicted LUAD prognosis and immunotherapy efficacy. The low-risk group patients had better survival prognoses and immunotherapy sensitivity. We also conducted enrichment, drug susceptibility, and mutational landscape analyses. We still have some drawbacks. Our analysis is a public database-dependent, and we need to conduct more *in vivo* and *in vitro* experiments. Some issues must be discussed in greater depth. In the future, we must continue to conduct in-depth study series to investigate the role of CAFs in LUAD more systematically and comprehensively and to provide innovative insights to treat LUAD precisely.

AUTHOR CONTRIBUTIONS

Han Zhang: conceived the study. Han Zhang, Yuhang Wang and Kai Wang: analyzed the data and made charts. Shuai Zhao, Xiaoteng Jia and Yun Ding: wrote the manuscript. XL and DS: revised and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

CONFLICTS 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.

FUNDING

This study was supported by the Natural Funding Project of Tianjin Science and Technology Bureau (No. 20JCYBJC01350), Tianjin Health Science and

Technology Project Key Projects (ZD20023) and Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-018A).

REFERENCES

1. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, Paz-Ares L. Lung cancer: current therapies and new targeted treatments. *Lancet*. 2017; 389:299–311. [https://doi.org/10.1016/S0140-6736\(16\)30958-8](https://doi.org/10.1016/S0140-6736(16)30958-8) PMID:27574741
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021; 71:209–49. <https://doi.org/10.3322/caac.21660> PMID:33538338
3. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature*. 2018; 553:446–54. <https://doi.org/10.1038/nature25183> PMID:29364287
4. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res*. 2016; 5:288–300. <https://doi.org/10.21037/tlcr.2016.06.07> PMID:27413711
5. Quan YH, Oh CH, Jung D, Lim JY, Choi BH, Rho J, Choi Y, Han KN, Kim BM, Kim C, Park JH, Kim HK. Evaluation of Intraoperative Near-Infrared Fluorescence Visualization of the Lung Tumor Margin With Indocyanine Green Inhalation. *JAMA Surg*. 2020; 155:732–40. <https://doi.org/10.1001/jamasurg.2020.1314> PMID:32579150
6. Kichenadasse G, Miners JO, Mangoni AA, Rowland A, Hopkins AM, Sorich MJ. Association Between Body Mass Index and Overall Survival With Immune Checkpoint Inhibitor Therapy for Advanced Non-Small Cell Lung Cancer. *JAMA Oncol*. 2020; 6:512–8. <https://doi.org/10.1001/jamaoncol.2019.5241> PMID:31876896
7. Denisenko TV, Budkevich IN, Zhivotovsky B. Cell death-based treatment of lung adenocarcinoma. *Cell Death Dis*. 2018; 9:117. <https://doi.org/10.1038/s41419-017-0063-y> PMID:29371589
8. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017; 541:321–30. <https://doi.org/10.1038/nature21349> PMID:28102259

9. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R. Hyperprogressors after Immunotherapy: Analysis of Genomic Alterations Associated with Accelerated Growth Rate. *Clin Cancer Res.* 2017; 23:4242–50.
<https://doi.org/10.1158/1078-0432.CCR-16-3133>
PMID:[28351930](https://pubmed.ncbi.nlm.nih.gov/28351930/)
10. Xu Z, Qian J, Meng C, Liu Y, Ding Q, Wu H, Li P, Ran F, Liu GQ, Wang Y, Ling Y. TME-targeting theranostic agent uses NIR tracking for tumor diagnosis and surgical resection and acts as chemotherapeutic showing enhanced efficiency and minimal toxicity. *Theranostics.* 2022; 12:2535–48.
<https://doi.org/10.7150/thno.68074> PMID:[35401826](https://pubmed.ncbi.nlm.nih.gov/35401826/)
11. LeBleu VS, Kalluri R. A peek into cancer-associated fibroblasts: origins, functions and translational impact. *Dis Model Mech.* 2018; 11:dmm029447.
<https://doi.org/10.1242/dmm.029447>
PMID:[29686035](https://pubmed.ncbi.nlm.nih.gov/29686035/)
12. Ziani L, Chouaib S, Thiery J. Alteration of the Antitumor Immune Response by Cancer-Associated Fibroblasts. *Front Immunol.* 2018; 9:414.
<https://doi.org/10.3389/fimmu.2018.00414>
PMID:[29545811](https://pubmed.ncbi.nlm.nih.gov/29545811/)
13. Öhlund D, Elyada E, Tuveson D. Fibroblast heterogeneity in the cancer wound. *J Exp Med.* 2014; 211:1503–23.
<https://doi.org/10.1084/jem.20140692>
PMID:[25071162](https://pubmed.ncbi.nlm.nih.gov/25071162/)
14. Barrett R, Puré E. Cancer-associated fibroblasts: key determinants of tumor immunity and immunotherapy. *Curr Opin Immunol.* 2020; 64:80–7.
<https://doi.org/10.1016/j.coi.2020.03.004>
PMID:[32402828](https://pubmed.ncbi.nlm.nih.gov/32402828/)
15. Barbazán J, Matic Vignjevic D. Cancer associated fibroblasts: is the force the path to the dark side? *Curr Opin Cell Biol.* 2019; 56:71–9.
<https://doi.org/10.1016/j.ceb.2018.09.002>
PMID:[30308331](https://pubmed.ncbi.nlm.nih.gov/30308331/)
16. Kunita A, Morita S, Irida TU, Goto A, Niki T, Takai D, Nakajima J, Fukayama M. MicroRNA-21 in cancer-associated fibroblasts supports lung adenocarcinoma progression. *Sci Rep.* 2018; 8:8838.
<https://doi.org/10.1038/s41598-018-27128-3>
PMID:[29892003](https://pubmed.ncbi.nlm.nih.gov/29892003/)
17. Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, Zhang B, Meng Q, Yu X, Shi S. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. *Mol Cancer.* 2021; 20:131.
<https://doi.org/10.1186/s12943-021-01428-1>
PMID:[34635121](https://pubmed.ncbi.nlm.nih.gov/34635121/)
18. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, Nainys J, Wu K, Kisieliovas V, Setty M, Choi K, Fromme RM, Dao P, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell.* 2018; 174:1293–308.e36.
<https://doi.org/10.1016/j.cell.2018.05.060>
PMID:[29961579](https://pubmed.ncbi.nlm.nih.gov/29961579/)
19. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, Deschler DG, Varvares MA, Mylvaganam R, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell.* 2017; 171:1611–24.e24.
<https://doi.org/10.1016/j.cell.2017.10.044>
PMID:[29198524](https://pubmed.ncbi.nlm.nih.gov/29198524/)
20. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, Lee JI, Suh YL, Ku BM, Eum HH, Choi S, Choi YL, Joung JG, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun.* 2020; 11:2285.
<https://doi.org/10.1038/s41467-020-16164-1>
PMID:[32385277](https://pubmed.ncbi.nlm.nih.gov/32385277/)
21. Song P, Li W, Guo L, Ying J, Gao S, He J. Identification and Validation of a Novel Signature Based on NK Cell Marker Genes to Predict Prognosis and Immunotherapy Response in Lung Adenocarcinoma by Integrated Analysis of Single-Cell and Bulk RNA-Sequencing. *Front Immunol.* 2022; 13:850745.
<https://doi.org/10.3389/fimmu.2022.850745>
PMID:[35757748](https://pubmed.ncbi.nlm.nih.gov/35757748/)
22. Song P, Li W, Wu X, Qian Z, Ying J, Gao S, He J. Integrated analysis of single-cell and bulk RNA-sequencing identifies a signature based on B cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma. *Cancer Immunol Immunother.* 2022; 71:2341–54.
<https://doi.org/10.1007/s00262-022-03143-2>
PMID:[35152302](https://pubmed.ncbi.nlm.nih.gov/35152302/)
23. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* 2017; 18:248–62.
<https://doi.org/10.1016/j.celrep.2016.12.019>
PMID:[28052254](https://pubmed.ncbi.nlm.nih.gov/28052254/)
24. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE II, Koeppen H, Astarita JL, Cubas R, Jhunjunwala S, Banchereau R, Yang Y, et al. TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature.* 2018; 554:544–8.
<https://doi.org/10.1038/nature25501> PMID:[29443960](https://pubmed.ncbi.nlm.nih.gov/29443960/)

25. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019; 20:296.
<https://doi.org/10.1186/s13059-019-1874-1>
PMID:[31870423](https://pubmed.ncbi.nlm.nih.gov/31870423/)
26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–504.
<https://doi.org/10.1101/gr.1239303>
PMID:[14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
27. Blake JA, Dolan M, Drabkin H, Hill DP, Li N, Sitnikov D, Bridges S, Burgess S, Buza T, McCarthy F, Peddinti D, Pillai L, Carbon S, et al., and Gene Ontology Consortium. Gene Ontology annotations and resources. *Nucleic Acids Res.* 2013; 41:D530–5.
<https://doi.org/10.1093/nar/gks1050> PMID:[23161678](https://pubmed.ncbi.nlm.nih.gov/23161678/)
28. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, Liu XS. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020; 48:W509–14.
<https://doi.org/10.1093/nar/gkaa407> PMID:[32442275](https://pubmed.ncbi.nlm.nih.gov/32442275/)
29. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol.* 2018; 1711:243–59.
https://doi.org/10.1007/978-1-4939-7493-1_12
PMID:[29344893](https://pubmed.ncbi.nlm.nih.gov/29344893/)
30. Tamminga M, Hiltermann TJN, Schuurin E, Timens W, Fehrmann RS, Groen HJ. Immune microenvironment composition in non-small cell lung cancer and its association with survival. *Clin Transl Immunology.* 2020; 9:e1142.
<https://doi.org/10.1002/cti2.1142>
PMID:[32547744](https://pubmed.ncbi.nlm.nih.gov/32547744/)
31. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D, Sopper S, Ijsselsteijn M, Brouwer TP, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 2019; 11:34.
<https://doi.org/10.1186/s13073-019-0638-6>
PMID:[31126321](https://pubmed.ncbi.nlm.nih.gov/31126321/)
32. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016; 17:218.
<https://doi.org/10.1186/s13059-016-1070-5>
PMID:[27765066](https://pubmed.ncbi.nlm.nih.gov/27765066/)
33. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017; 18:220.
<https://doi.org/10.1186/s13059-017-1349-1>
PMID:[29141660](https://pubmed.ncbi.nlm.nih.gov/29141660/)
34. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife.* 2017; 6:e26476.
<https://doi.org/10.7554/eLife.26476> PMID:[29130882](https://pubmed.ncbi.nlm.nih.gov/29130882/)
35. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform.* 2021; 22:bbab260.
<https://doi.org/10.1093/bib/bbab260>
PMID:[34260682](https://pubmed.ncbi.nlm.nih.gov/34260682/)
36. Bischoff P, Trinks A, Obermayer B, Pett JP, Wiederspahn J, Uhlitz F, Liang X, Lehmann A, Jurmeister P, Elsner A, Dziodzio T, Rückert JC, Neudecker J, et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene.* 2021; 40:6748–58.
<https://doi.org/10.1038/s41388-021-02054-3>
PMID:[34663877](https://pubmed.ncbi.nlm.nih.gov/34663877/)
37. Sha D, Jin Z, Budczies J, Kluck K, Stenzinger A, Sinicrope FA. Tumor Mutational Burden as a Predictive Biomarker in Solid Tumors. *Cancer Discov.* 2020; 10:1808–25.
<https://doi.org/10.1158/2159-8290.CD-20-0522>
PMID:[33139244](https://pubmed.ncbi.nlm.nih.gov/33139244/)
38. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell.* 2012; 150:1107–20.
<https://doi.org/10.1016/j.cell.2012.08.029>
PMID:[22980975](https://pubmed.ncbi.nlm.nih.gov/22980975/)
39. Yi M, Niu M, Wu Y, Ge H, Jiao D, Zhu S, Zhang J, Yan Y, Zhou P, Chu Q, Wu K. Combination of oral STING agonist MSA-2 and anti-TGF- β /PD-L1 bispecific antibody YM101: a novel immune cocktail therapy for non-inflamed tumors. *J Hematol Oncol.* 2022; 15:142.
<https://doi.org/10.1186/s13045-022-01363-8>
PMID:[36209176](https://pubmed.ncbi.nlm.nih.gov/36209176/)
40. Jung KK, Liu XW, Chirco R, Fridman R, Kim HR. Identification of CD63 as a tissue inhibitor of metalloproteinase-1 interacting cell surface protein. *EMBO J.* 2006; 25:3934–42.
<https://doi.org/10.1038/sj.emboj.7601281>
PMID:[16917503](https://pubmed.ncbi.nlm.nih.gov/16917503/)
41. Brew K, Nagase H. The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with

- structural and functional diversity. *Biochim Biophys Acta*. 2010; 1803:55–71.
<https://doi.org/10.1016/j.bbamcr.2010.01.003>
PMID:20080133
42. Duch P, Díaz-Valdivia N, Ikemori R, Gabasa M, Radisky ES, Arshakyan M, Gea-Sorlí S, Mateu-Bosch A, Bragado P, Carrasco JL, Mori H, Ramírez J, Teixidó C, et al. Aberrant TIMP-1 overexpression in tumor-associated fibroblasts drives tumor progression through CD63 in lung adenocarcinoma. *Matrix Biol*. 2022; 111:207–25.
<https://doi.org/10.1016/j.matbio.2022.06.009>
PMID:35787446
43. Fong KM, Kida Y, Zimmerman PV, Smith PJ. TIMP1 and adverse prognosis in non-small cell lung cancer. *Clin Cancer Res*. 1996; 2:1369–72.
PMID:9816309
44. Meng LB, Shan MJ, Qiu Y, Qi R, Yu ZM, Guo P, Di CY, Gong T. TPM2 as a potential predictive biomarker for atherosclerosis. *Aging (Albany NY)*. 2019; 11:6960–82.
<https://doi.org/10.18632/aging.102231>
PMID:31487691
45. Meng LB, Xu HX, Shan MJ, Hu GF, Liu LT, Chen YH, Liu YQ, Wang L, Chen Z, Li YJ, Gong T, Liu DP. A Potential Target for Clinical Atherosclerosis: A Novel Insight Derived from TPM2. *Aging Dis*. 2022; 13:373–8.
<https://doi.org/10.14336/AD.2021.0926>
PMID:35371599
46. Duan P, Cui J, Li H, Yuan L. Tropomyosin 2 exerts anti-tumor effects in lung adenocarcinoma and is a novel prognostic biomarker. *Histol Histopathol*. 2023; 38:669–80.
<https://doi.org/10.14670/HH-18-514> PMID:36102257
47. Miao W, Chen M, Chen M, Cui C, Zhu Y, Luo X, Wu B. Nr2f2 Overexpression Aggravates Ferroptosis and Mitochondrial Dysfunction by Regulating the PGC-1 α Signaling in Diabetes-Induced Heart Failure Mice. *Mediators Inflamm*. 2022; 2022:8373389.
<https://doi.org/10.1155/2022/8373389>
PMID:36081650
48. Lin FJ, Qin J, Tang K, Tsai SY, Tsai MJ. Coup d'Etat: an orphan takes control. *Endocr Rev*. 2011; 32:404–21.
<https://doi.org/10.1210/er.2010-0021> PMID:21257780
49. Liu W, Zhou Y, Duan W, Song J, Wei S, Xia S, Wang Y, Du X, Li E, Ren C, Wang W, Zhan Q, Wang Q. Glutathione peroxidase 4-dependent glutathione high-consumption drives acquired platinum chemoresistance in lung cancer-derived brain metastasis. *Clin Transl Med*. 2021; 11:e517.
<https://doi.org/10.1002/ctm2.517> PMID:34586745
50. Schlosser A, Pilecki B, Hemstra LE, Kejlving K, Kristmannsdottir GB, Wulf-Johansson H, Moeller JB, Füchtbauer EM, Nielsen O, Kirketerp-Møller K, Dubey LK, Hansen PB, Stubbe J, et al. MFAP4 Promotes Vascular Smooth Muscle Migration, Proliferation and Accelerates Neointima Formation. *Arterioscler Thromb Vasc Biol*. 2016; 36:122–33.
<https://doi.org/10.1161/ATVBAHA.115.306672>
PMID:26564819
51. Torp N, Israelsen M, Madsen B, Lutz P, Jansen C, Strassburg C, Mortensen C, Knudsen AW, Sorensen GL, Holmskov U, Schlosser A, Thiele M, Trebicka J, Krag A. Level of MFAP4 in ascites independently predicts 1-year transplant-free survival in patients with cirrhosis. *JHEP Rep*. 2021; 3:100287.
<https://doi.org/10.1016/j.jhepr.2021.100287>
PMID:34041469
52. Feng YY, Liu CH, Xue Y, Chen YY, Wang YL, Wu XZ. MicroRNA-147b promotes lung adenocarcinoma cell aggressiveness through negatively regulating microfibril-associated glycoprotein 4 (MFAP4) and affects prognosis of lung adenocarcinoma patients. *Gene*. 2020; 730:144316.
<https://doi.org/10.1016/j.gene.2019.144316>
PMID:31884109
53. Carmona-Rodríguez L, Martínez-Rey D, Mira E, Mañes S. SOD3 boosts T cell infiltration by normalizing the tumor endothelium and inducing laminin- α 4. *Oncoimmunology*. 2020; 9:1794163.
<https://doi.org/10.1080/2162402X.2020.1794163>
PMID:32934887
54. Kim JH, Jeong HD, Song MJ, Lee DH, Chung JH, Lee ST. SOD3 Suppresses the Expression of MMP-1 and Increases the Integrity of Extracellular Matrix in Fibroblasts. *Antioxidants (Basel)*. 2022; 11:928.
<https://doi.org/10.3390/antiox11050928>
PMID:35624792
55. Zhang Y, Lu X, Zhang Y, Zhao D, Gong H, Du Y, Sun H. The Effect of Extracellular Superoxide Dismutase (SOD3) Gene in Lung Cancer. *Front Oncol*. 2022; 12:722646.
<https://doi.org/10.3389/fonc.2022.722646>
PMID:35356201
56. Leiser D, Samanta S, Eley J, Strauss J, Creed M, Kingsbury T, Staats PN, Bhandary B, Chen M, Dukic T, Roy S, Mahmood J, Vujaskovic Z, Shukla HD. Role of caveolin-1 as a biomarker for radiation resistance and tumor aggression in lung cancer. *PLoS One*. 2021; 16:e0258951.
<https://doi.org/10.1371/journal.pone.0258951>
PMID:34762666
57. Marudamuthu AS, Bhandary YP, Fan L, Radhakrishnan V, MacKenzie B, Maier E, Shetty SK, Nagaraja MR, Gopu V, Tiwari N, Zhang Y, Watts AB, Williams RO 3rd, et al. Caveolin-1-derived peptide limits development of pulmonary fibrosis. *Sci Transl Med*. 2019; 11:eaat2848.

- <https://doi.org/10.1126/scitranslmed.aat2848>
PMID:31826982
58. Hu L, Liu Y, Wei C, Jin H, Mei L, Wu C. SERPINH1, Targeted by miR-29b, Modulated Proliferation and Migration of Human Retinal Endothelial Cells Under High Glucose Conditions. *Diabetes Metab Syndr Obes*. 2021; 14:3471–83.
<https://doi.org/10.2147/DMSO.S307771>
PMID:34377003
59. Kamikawaji K, Seki N, Watanabe M, Mataka H, Kumamoto T, Takagi K, Mizuno K, Inoue H. Regulation of LOXL2 and SERPINH1 by antitumor microRNA-29a in lung cancer with idiopathic pulmonary fibrosis. *J Hum Genet*. 2016; 61:985–93.
<https://doi.org/10.1038/jhg.2016.99>
PMID:27488440
60. Siddens LK, Krueger SK, Henderson MC, Williams DE. Mammalian flavin-containing monooxygenase (FMO) as a source of hydrogen peroxide. *Biochem Pharmacol*. 2014; 89:141–7.
<https://doi.org/10.1016/j.bcp.2014.02.006>
PMID:24561181
61. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol*. 2019; 20:69–84.
<https://doi.org/10.1038/s41580-018-0080-4>
PMID:30459476
62. Yu F, Yu C, Li F, Zuo Y, Wang Y, Yao L, Wu C, Wang C, Ye L. Wnt/ β -catenin signaling in cancers and targeted therapies. *Signal Transduct Target Ther*. 2021; 6:307.
<https://doi.org/10.1038/s41392-021-00701-5>
PMID:34456337
63. Liu J, Xiao Q, Xiao J, Niu C, Li Y, Zhang X, Zhou Z, Shu G, Yin G. Wnt/ β -catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduct Target Ther*. 2022; 7:3.
<https://doi.org/10.1038/s41392-021-00762-6>
PMID:34980884
64. Pan J, Fang S, Tian H, Zhou C, Zhao X, Tian H, He J, Shen W, Meng X, Jin X, Gong Z. lncRNA JPX/miR-33a-5p/Twist1 axis regulates tumorigenesis and metastasis of lung cancer by activating Wnt/ β -catenin signaling. *Mol Cancer*. 2020; 19:9.
<https://doi.org/10.1186/s12943-020-1133-9>
PMID:31941509
65. Seow WJ, Matsuo K, Hsiung CA, Shiraishi K, Song M, Kim HN, Wong MP, Hong YC, Hosgood HD 3rd, Wang Z, Chang IS, Wang JC, Chatterjee N, et al. Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from Western populations. *Hum Mol Genet*. 2017; 26:454–65.
<https://doi.org/10.1093/hmg/ddw414> PMID:28025329
66. Cortot AB, Jänne PA. Molecular mechanisms of resistance in epidermal growth factor receptor-mutant lung adenocarcinomas. *Eur Respir Rev*. 2014; 23:356–66.
<https://doi.org/10.1183/09059180.00004614>
PMID:25176972
67. He D, Wang D, Lu P, Yang N, Xue Z, Zhu X, Zhang P, Fan G. Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. *Oncogene*. 2021; 40:355–68.
<https://doi.org/10.1038/s41388-020-01528-0>
PMID:33144684
68. Toki MI, Mani N, Smithy JW, Liu Y, Altan M, Wasserman B, Tuktamyshev R, Schalper K, Syrigos KN, Rimm DL. Immune Marker Profiling and Programmed Death Ligand 1 Expression Across NSCLC Mutations. *J Thorac Oncol*. 2018; 13:1884–96.
<https://doi.org/10.1016/j.jtho.2018.09.012>
PMID:30267840
69. Wang S, Xie T, Xing P, Li J. P57.05 Low Variant Allele Frequency of Tp53 as a Biomarker for Pd-1/Pd-L1 Inhibitors in Lung Adenocarcinoma. *J Thorac Oncol*. 2021; 16:S1138–S9.
<https://doi.org/10.1016/j.jtho.2021.08.577>
70. Skoulidis F, Goldberg ME, Greenawalt DM, Hellmann MD, Awad MM, Gainor JF, Schrock AB, Hartmaier RJ, Trabucco SE, Gay L, Ali SM, Elvin JA, Singal G, et al. STK11/LKB1 Mutations and PD-1 Inhibitor Resistance in KRAS-Mutant Lung Adenocarcinoma. *Cancer Discov*. 2018; 8:822–35.
<https://doi.org/10.1158/2159-8290.CD-18-0099>
PMID:29773717
71. Bonaventura P, Shekarian T, Alcazer V, Valladeau-Guilemond J, Valsesia-Wittmann S, Amigorena S, Caux C, Depil S. Cold Tumors: A Therapeutic Challenge for Immunotherapy. *Front Immunol*. 2019; 10:168.
<https://doi.org/10.3389/fimmu.2019.00168>
PMID:30800125
72. Pan Y, Fu Y, Zeng Y, Liu X, Peng Y, Hu C, Deng C, Qiu Z, Zou J, Liu Y, Wu F. The key to immunotherapy: how to choose better therapeutic biomarkers for patients with non-small cell lung cancer. *Biomark Res*. 2022; 10:9.
<https://doi.org/10.1186/s40364-022-00355-7>
PMID:35255999
73. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol*. 2013; 14:1014–22.
<https://doi.org/10.1038/ni.2703> PMID:24048123

74. Rodig SJ, Gusenleitner D, Jackson DG, Gjini E, Giobbie-Hurder A, Jin C, Chang H, Lovitch SB, Horak C, Weber JS, Weirather JL, Wolchok JD, Postow MA, et al. MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. *Sci Transl Med.* 2018; 10:ear3342.
<https://doi.org/10.1126/scitranslmed.aar3342>
PMID:[30021886](https://pubmed.ncbi.nlm.nih.gov/30021886/)
75. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018; 24:1550–8.
<https://doi.org/10.1038/s41591-018-0136-1>
PMID:[30127393](https://pubmed.ncbi.nlm.nih.gov/30127393/)

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1, 3.

Supplementary Table 1. Details of the 120 CAFs-associated genes in scRNA-seq.

Supplementary Table 2. Information on 51 immune checkpoint-related genes and 23 MHC-related genes.

Checkpoints	MHC
PD-L1	HLA-E
PD1	HLA-DPB2
TIM3	HLA-C
CTLA4	HLA-J
B7-H4	HLA-DQB1
B7-H3	HLA-DQB2
BTLA	HLA-DQA1
VISTA	HLA-A
IDO1	HLA-DMA
PSGL-1	HLA-DOB
LAG3	HLA-DRB1
PD-L2	HLA-H
OX40	HLA-B
IDO2	HLA-DRB5
TNFRSF8	HLA-DOA
CD27	HLA-DPB1
ICOS	HLA-DRA
TNFRSF18	HLA-DRB6
TIGIT	HLA-L
TNFRSF9	HLA-F
TNFRSF14	HLA-G
TNFRSF4	HLA-DMB
CD28	HLA-DPA1
LGALS9	
CD70	
CD80	
LGALS9	
CD70	
CD80	
TNFSF15	
NRP1	
BTNL2	
HHLA2	
ICOSLG	
CD40	
TNFSF9	
TNFSF14	
CD86	

KIR3DL1
CD200
ADORA2A
TNFRSF25
CD244
CD48
LAIR1
CD40LG
TMIGD2
CD200R1
TNFSF18
CD44
CD160

Supplementary Table 3. 'Estimate' score for each LUAD patient in TCGA data.