

The expression and prognostic value of disulfidptosis progress in lung adenocarcinoma

Lina Ni^{1,*}, Huizhen Yang^{1,*}, Xiaoyu Wu¹, Kejin Zhou¹, Sheng Wang¹

¹Department of Respiratory, Jinhua Guangfu Cancer Hospital, Jinhua, Zhejiang 321200, China

*Equal contribution

Correspondence to: Sheng Wang, Kejin Zhou; email: wsh1992@alu.zcmu.edu.cn; 641124481@qq.com, <https://orcid.org/0009-0007-1888-9433>

Keywords: lung adenocarcinoma, disulfidptosis, prediction model, scRNA-seq, prognosis

Received: April 6, 2023

Accepted: July 18, 2023

Published: August 7, 2023

Copyright: © 2023 Ni 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

Disulfidptosis is a new cell death model caused by accumulating intracellular disulfides bonding to actin cytoskeleton proteins. This study aimed to investigate the expression and prognostic value of disulfidptosis-related genes (DRGs) in lung adenocarcinoma (LUAD). The data of expression profiles and scRNA-seq were collected from TCGA and GEO databases. The different expressions of DRGs between normal and LUAD tissues were compared. The LASSO analysis and multivariate Cox regression analysis were utilized to develop a DRGs model for the prognosis evaluation in LUAD. The model's predictive accuracy was evaluated with the area under the receiver operating characteristic curve (AUC) and C-index. Survival analysis, univariate and multivariate Cox regression analysis were used to assessing the predictive value of the DRGs model. ScRNA-seq data were analyzed with "Seurat" and "Monocle 2" packages. There were significant differences in 22 DRGs between normal and tumor tissues. A model with five DRGs (ACTB, FLNB, NCKAP1, SLC3A2, SLC7A11) was constructed. The AUC and C-index of the model were significantly higher than that based on clinical parameters. Survival analysis, univariate and multivariate Cox regression analysis demonstrated risk score was an independent prognostic predictor. In the scRNA-seq study, we identified 14 clusters and 11 cell types. Clusters 2, 8, and 13 were annotated into Epithelial cells. SLC7A11 and SLC3A2, NCKAP1 and FLNB, ACTB expressed most abundantly in Epithelial cells, Endothelial cells, Naive CD4 T, respectively. We explored the expression of DRGs in LUAD and constructed a predictive DRGs model, which was stable and reliable for predicting LUAD prognosis.

INTRODUCTION

It was reported that lung cancer (LC) was one of the most common malignant tumors worldwide and the leading cause of cancer-related mortality [1]. In 2020, over 2.2 million new cases were diagnosed as LC globally, and more than 1.8 million people died of LC. As the primary subtype of LC, lung adenocarcinoma (LUAD) accounts for more than 40% of LC cases [2, 3]. In recent years, new diagnostic techniques and treatment strategies, such as immunotherapy, have been emerging and prominently prolonged the survival time

of LUAD patients [3]. However, the 5-year survival rate of LUAD patients remains <20% [4, 5]. Therefore, there is an urgent need to identify more biomarkers for diagnosis and prognostic assessment in LUAD, which may be helpful to risk stratification and increase the long-term survival rate of LUAD patients.

Programmed cell death, such as apoptosis and autophagy, could enable cells to coordinate their end, thereby benefiting living organisms [6, 7]. Recently, Liu et al. proposed a novel form of cell death named disulfidptosis. The main mechanism of disulfidptosis is that accumulating

intracellular disulfides caused by cells starved of glucose in SLC7A11^{high} cells could bond in actin cytoskeleton proteins [8, 9]. In addition, excessive cystine uptake could couple with the insufficient supply of NADPH, causing NADPH depletion, aberrant disulfide bonding in actin cytoskeleton proteins, F-actin collapse, and subsequent cell death [8]. Distinct from apoptosis and ferroptosis, deleting BAX and BAK, fatty acid oxidation inhibitors, did not prevent death, and iron was not required [8]. The results reported a new cell death model, disulfide stress-mediated disulfidptosis, and a viable therapeutic strategy to target disulfidptosis in cancer treatment. However, the disulfidptosis process in LUAD and relevant disulfidptosis-related genes related to the prognosis remains to be further elucidated.

This study collected mRNA expression profiles from The Cancer Genome Atlas (TCGA; <https://www.cancer.gov>) and Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov>) to investigate the expression of disulfidptosis-related genes (DRGs). Then, we utilized disulfidptosis-related genes to construct a DRGs model for predicting the clinical outcome in LUAD. Lastly, we explored the expression profiles of disulfidptosis-related genes with Single-cell RNA sequencing (scRNA-seq).

RESULTS

The expression profile of 23 DRGs

As shown in Figure 1A, among 23 DRGs, 3 DRGs (GYS1, NDUFA11, ACTN4) were located at Chromosome 20. In addition, 3 DRGs (LRPPRC, NCKAP1, NDUFS1) were situated at Chromosome 2. 2 DRGs (FLNB, RPN1) were found at Chromosome 3. 2 DRGs (NUBPL, INF2) were located at Chromosome 14.

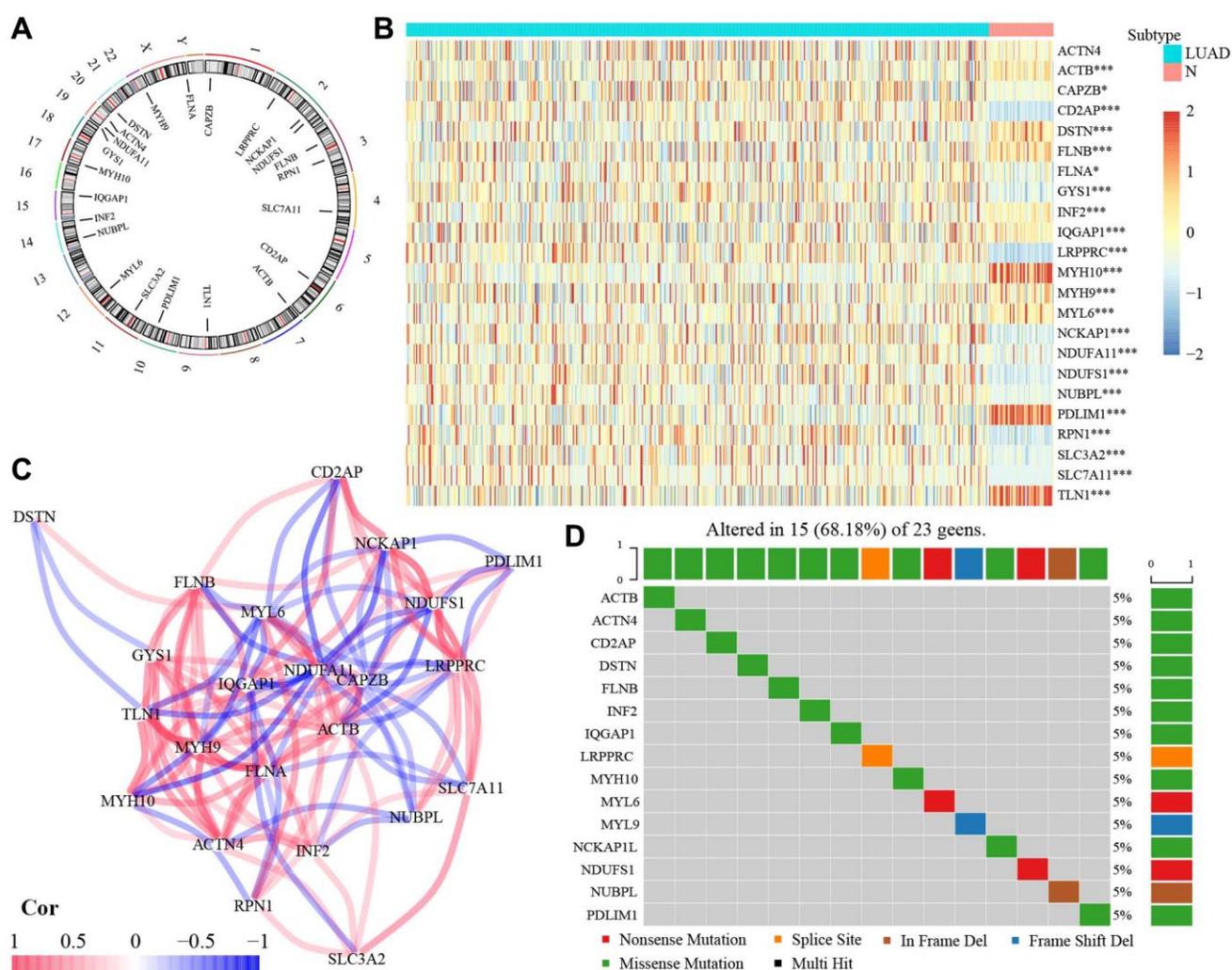


Figure 1. The expression of 23 DRGs in LUAD. (A) Location of 23 DRGs in chromosomes. (B) The different expression of DRGs between normal and LUAD tissues. (C) The correlation network of 23 DRGs. (D) The mutation frequency of 23 DRGs. Abbreviations: DRGs: Disulfidptosis-related genes; LUAD: Lung adenocarcinoma; N: normal tissues. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Others (CAPZB, SLC7A11, CD2AP, ACTB, TLN1, PDLIM1, SLC3A2, MYL6, IQGAP1, MYH10, DSTN, MYH9, FLNA) were located at Chromosome 4, 6, 7, 9, 10, 11, 12, 15, 17, 21, 22, and X, respectively. The Wilcoxon rank-sum test showed significantly different expressions in 22 DRGs, except for ACTN4, between normal and tumor tissues (Figure 1B). The correlation among 23 DRGs was presented in Figure 1C. Among 23 DRGs, 15 DRGs had gene mutation, although the mutation frequency for each DRGs was relatively low (Figure 1D).

Establishment and evaluation of a DRGs model

We utilized the least absolute shrinkage and selection operator (LASSO) regression (LASSO) analysis to analyze 23 DRGs and determined 8 DRGs highly related to the overall survival (OS) of LUAD patients in the TCGA set (Figure 2A). Then, those 8 DRGs were subjected to the multivariate Cox regression analysis. 5 candidate DRGs (ACTB, FLNB, NCKAP1, SLC3A2, SLC7A11) were screened out and used to build a

prognostic model (Figure 2B) and a risk score formula to calculate the risk score of all patients. The risk score formula was presented as follows. The Kaplan–Meier plots demonstrated that all 5 DRGs were unfavorable factors for the clinical outcomes of LUAD patients (Figure 2C–G). Figure 2H showed the correlation among 5 DRGs. According to the optimal cut-off risk score: 0.987, patients were divided into low- and high-risk groups. The principal component analysis (PCA) revealed that the samples in two risk groups were distributed in different areas (Figure 2I).

$$\begin{aligned} \text{Risk score} = & (0.430 \times \text{exp of ACTB}) \\ & + (0.248 \times \text{exp of FLNB}) \\ & + (0.408 \times \text{exp of NCKAP1}) \\ & + (0.336 \times \text{exp of SLC3A11}) \\ & + (0.153 \times \text{exp of SLC7A11}) \end{aligned}$$

Next, we calculated the area under the curve (AUC) value and the c-index in each set to estimate the

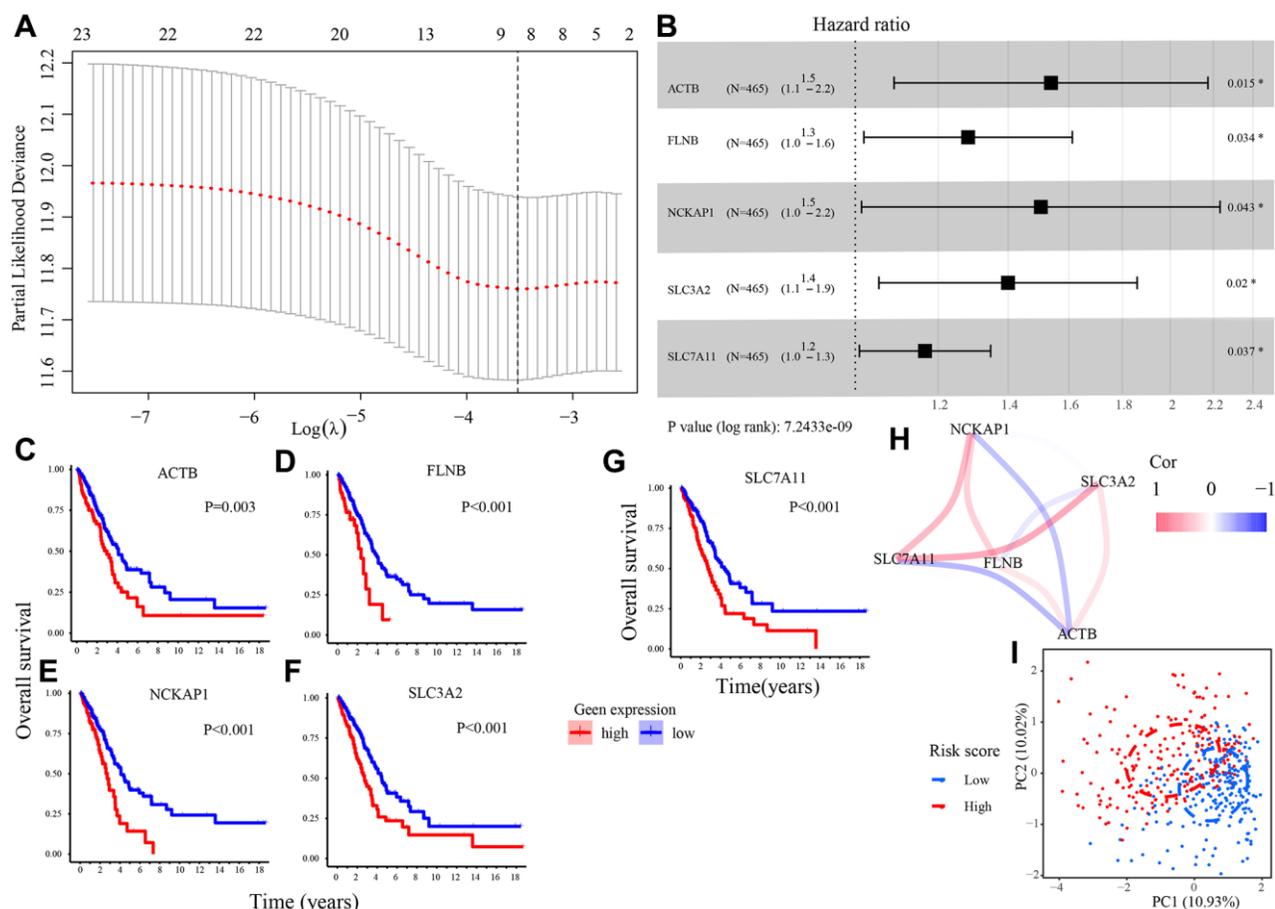


Figure 2. Construction of a DRGs model in TCGA set. (A) The LASSO regression analysis to filter out prognosis-related DRGs. (B) The multivariate Cox regression analysis to develop a DRGs model. (C–G) Survival analysis of 5 candidate DRGs (ACTB, FLNB, NCKAP1, SLC3A2, SLC7A11). (H) The correlation network of 5 candidate DRGs. (I) Patients in different risk groups gathering in two areas in PCA analysis. Abbreviations: LASSO: Least absolute shrinkage and selection operator; PCA: Principal component analysis.

predictive accuracy of the DRGs model. The results showed the AUC values and c-index based on risk score in all sets were higher than that based on clinical indexes (age, gender, smoking, TNM stage), hinting that the DRGs model performed well in predicting clinical outcomes of LUAD patients (Figure 3A, 3B). In addition, the calibration plot in each set revealed that when the line angle is 45°, it represents the best prediction result, indicating that the DRGs model had a good predictive ability (Figure 3C).

Correlation between DRGs model and clinical characteristics

The distributions of samples with different TNM stage ($P = 0.003$), tumor size ($P < 0.003$), and lymph node metastasis ($P = 0.025$) were significantly diverse between different risk groups (Figure 4A). Moreover, there was a significant difference in the risk scores among patients with varying tumor sizes ($P < 0.001$, Figure 4E). Similar results were observed among patients with different TNM stages (Figure 4H). The risk score significantly increased in female patients ($P = 0.042$, Figure 4C) or with lymph node metastasis ($P = 0.009$, Figure 4F). However, we did not find the effect of age

(Figure 4B), smoking (Figure 4D) and distant metastasis (Figure 4G) on the risk score.

The DRGs model is an independent prognostic factor for the prognosis of LUAD

The Kaplan–Meier plot in the TCGA set illustrated that the high-risk score predicted poorer clinical outcomes ($P < 0.001$, Figure 5A). It was confirmed in GEO set 1 ($P < 0.001$, Figure 5B), GEO set 2 ($P < 0.001$, Figure 5C), and GEO set 3 ($P < 0.001$, Figure 5D). Furthermore, stratification analyses demonstrated the low-risk patients had a better prognosis in each subgroup (Supplementary Figure 1).

To compare the DRGs model with clinical parameters (age, gender, stage, tumor size, lymph node metastasis, and distance metastasis), we performed the univariate and multivariate Cox regression model. The univariate Cox regression model showed that the DRGs model was an essential prognosis-related influence factor (HR = 3.423, 95% CI (2.431, 4.820), $P < 0.001$, Figure 5E). The multivariate Cox regression model demonstrated that the DRGs model was an independent prognostic factor for the OS of LUAD patients (HR = 3.316,

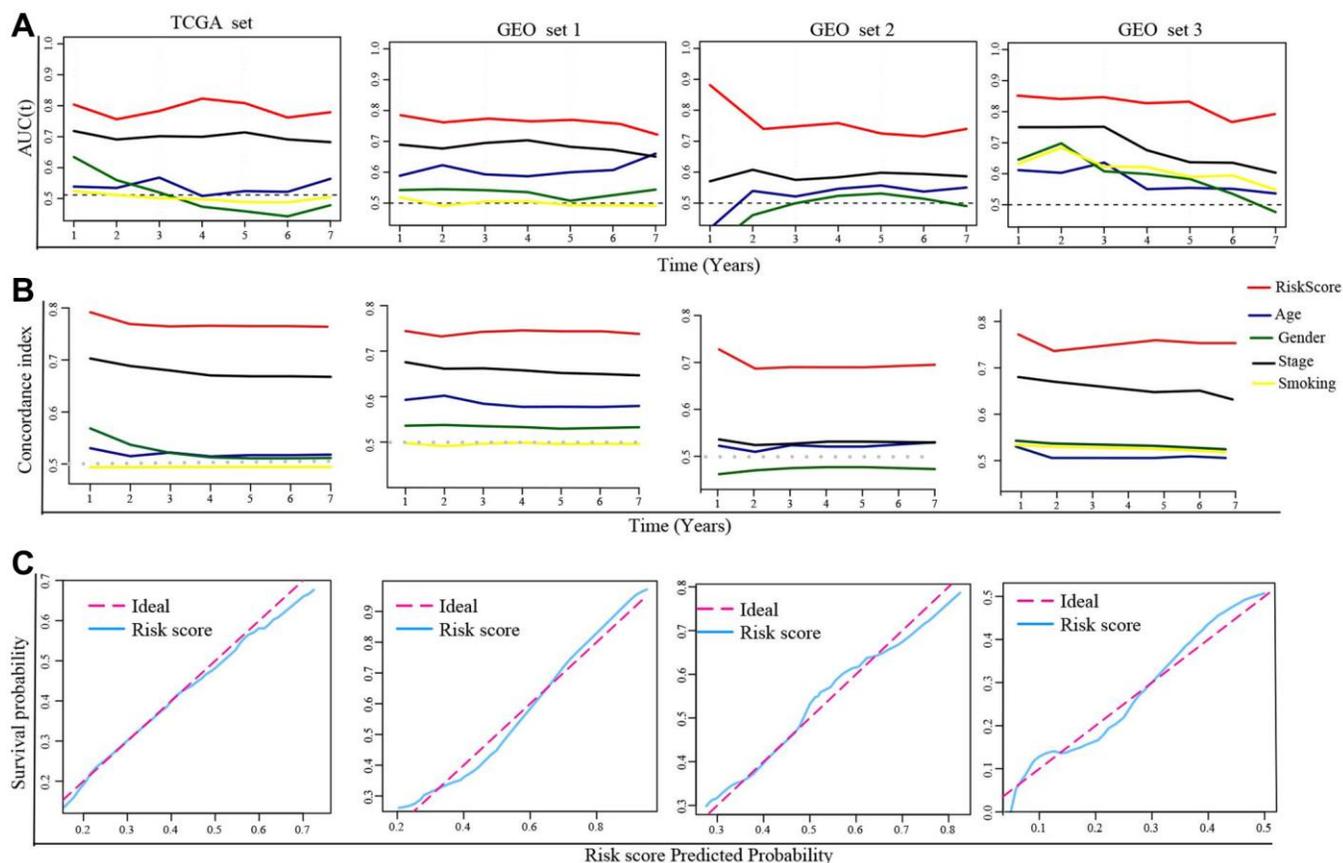


Figure 3. Evaluation of the DRGs model. (A) The time-dependent AUC value in each set. (B) The time-dependent C-index in each set. (C) The calibration plots of each set. Abbreviations: AUC: The area under the curve; C-index: concordance index.

95% CI (2.300, 4.782), $P < 0.001$, Figure 5F). Similar results were observed in GEO set 1, 2, and 3 (Supplementary Figure 2).

Gene set enrichment analysis (GSEA)

Then, we performed the GSEA to explore the altered biological roles and signaling pathways between the low- and high-risk group. A total of 27 pathways were enriched, including pathways in cancer, P53 signaling pathway, and so on, that were highly associated with tumorigenesis and development (Figure 6A). Figure 6B, 6C presented the top 5 pathways enriched in high- and low-risk groups.

Tumor mutational burden (TMB)

The mutational landscape showed that the frequent mutation events in the low-risk group (96.53%) were

more than that in the high-risk group (90.99%) (Figure 7A, 7B). In the high-risk group, the most frequently mutated gene was TTN (50%), followed by TP53 (50%) and MUC56 (44%). In the low-risk group, the top 3 frequently mutated genes were TP53 (45%), TTN (43%), and CSMD3 (38%). However, there was no difference in TMB between the two risk groups ($P = 0.180$, Figure 7C). Survival analysis revealed that high TMB had a better prognosis in LUAD patients ($P = 0.008$, Figure 7D). Furthermore, it could better predict prognosis when combining the TMB and the risk score (Figure 7E).

Investigation of the relation of the DRGs model to the immune statue

We applied the single-sample gene set enrichment analysis (ssGSEA) to analyze the level of 15 tumor-infiltrating immune cells (TIICs) and the activity of 13

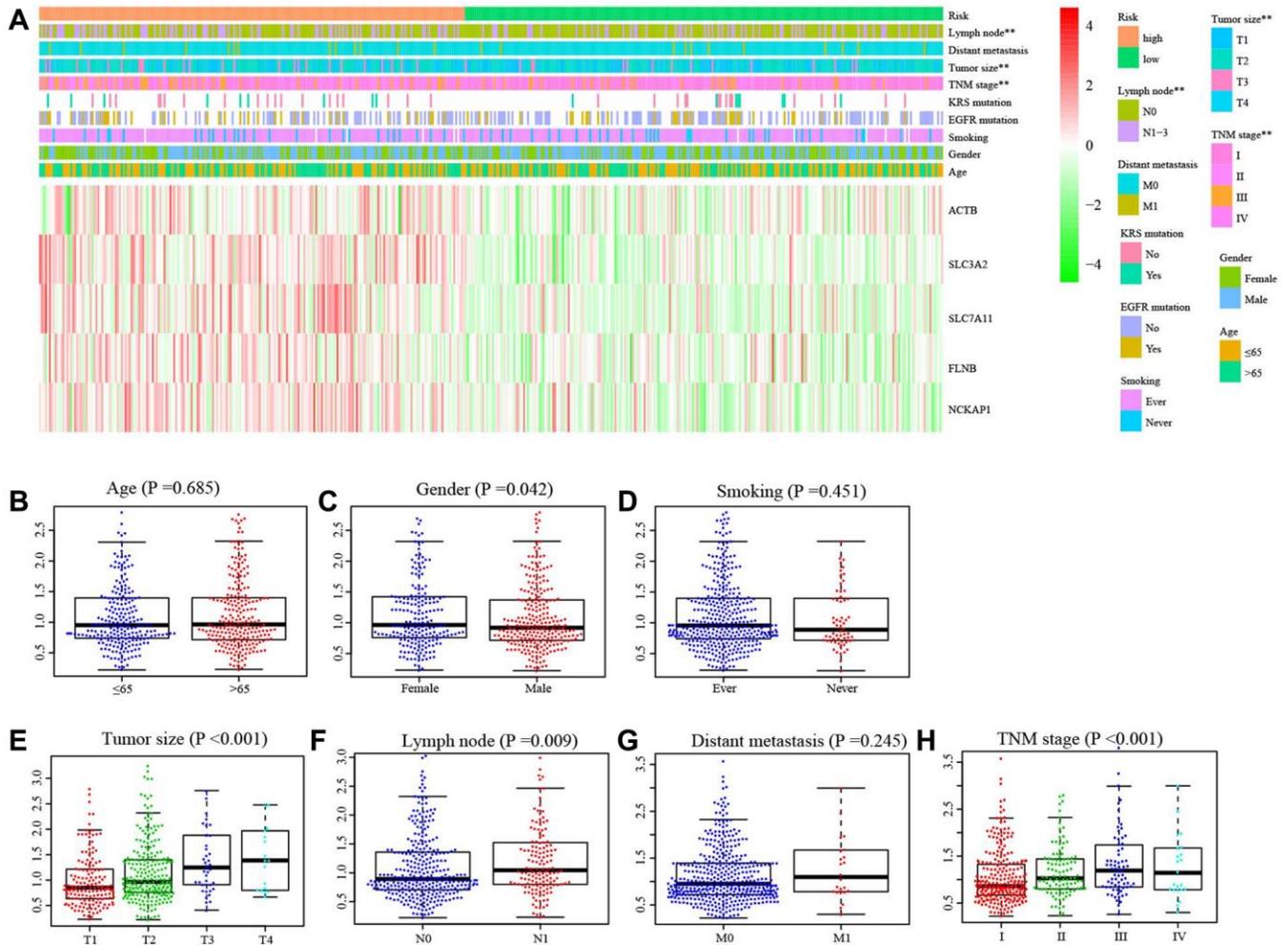


Figure 4. Correlation between risk score and clinical characteristics. (A) The distribution of clinical characteristics between high- and low-risk groups. (B) The difference of risk score between Age ≤ 65 and Age > 65 , (C) different gender, (D) smoking and non-smoking, (E) tumor size, (F) with and without lymph node metastasis, (G) with and without distance metastasis, (H) different TNM stage. ** $P < 0.01$.

immune-related functions. The results demonstrated that the enrichment scores of 9 TIICs (aDCs, B cells, CD8 T cells, iDCs, Mast cells, Neutrophils, Tfh, Th1 cells, TIL) in the low-risk group were significantly enhanced (Figure 8A). Similar results were observed in 4 immune-related functions (Check-point, HLA, T cell co-stimulation, Type II IFN response, Figure 8B). The immune score in the high-risk group was significantly higher than in the low-risk group (Figure 8C). The TIDE score and Exclusion score in the high-risk group were increased (Figure 8D), whereas the IPS values of PD-1/PD-L1 and CTLA4 were decreased (Figure 8E). Drug sensitivity analysis manifested a significant

difference in the sensitivity of 47 anti-tumor drugs between high- and low-risk groups (Supplementary Figure 3).

scRNA-seq analysis

Before filtering, there were 45,632 features for 23,747 cells in the 8 LUAD samples. Then, we performed data standardization and quality control and finally selected 18,977 cells and the top 2000 highly expressed and variable genes for further analysis. The PCA for reducing data dimensionality reduction revealed no noticeable separation trend of cells. Nonlinear

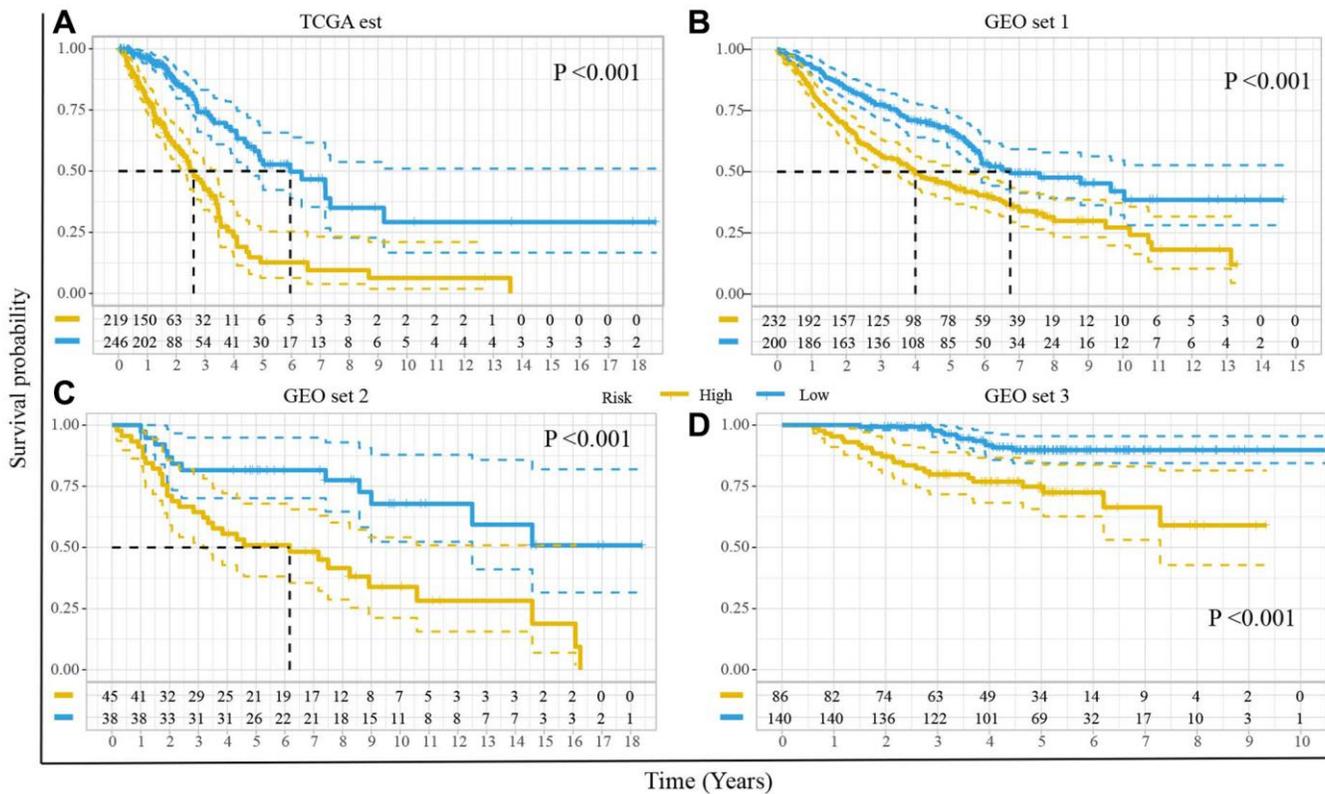


Figure 5. The DRGs model is an independent prognostic factor for the prognosis of LUAD. Survival difference between high- and low-risk groups (A) in TCGA set, (B) GEO set 1, (C) GEO set 2, (D) GEO set 3. (E) The univariate Cox regression analyses of risk score and clinical characteristics in TCGA set. (F) The multivariate Cox regression analyses of risk score and clinical characteristics in TCGA set.

dimension reduction was performed with the t-Stochastic Neighbor Embedding (t-SNE) algorithm, which successfully clustered the cells into 14 clusters (Figure 9A). Then, we annotated all clusters, and 11 cell types were identified (Figure 9B). In addition, the expression level of SLC7A11 and SLC3A2 were most abundant in Epithelial cells (Figure 9D). NCKAP1 and FLNB had the highest expression in Endothelial cells. ACTB was expressed in most cell types. However, the expression level in Naive CD4 T was the richest. In this study, cluster 2, 8, and 13 were annotated into Epithelial cells, which mainly contained cancer cells and cancer stem cells. Thence, we included the Epithelial cells in the pseudo-time cell differentiation trajectory analysis. The results were presented in Figure 9F and demonstrated the evolutionary pattern of Epithelial cells. Based on the ordering of pseudotime, Epithelial cells (cluster 2) were divided into Epithelial cells (cluster 8) and Epithelial cells (cluster 13). Figure 9G showed the 5 candidate DRGs expression in different developmental states. Finally, we further analyzed the Epithelial cells with the t-SNE algorithm and clustered Epithelial cells into 6 sub-clusters (Figure 9C). The expression of SLC7A11, SLC3A2, NCKAP1, and ACTB in sub-cluster 4 was

highest, whereas FLNB in sub-cluster 2 was highest (Figure 9E).

Construction of a prognostic nomogram

Using clinical characteristics (age, gender, smoking, TNM stage) and risk score, we develop a prognostic nomogram in TCGA set for predicting 1-, 3-, and 5-year OS of LUAD patients (Figure 10A). The AUCs and c-index in each set were more than 0.7 (Figure 10B, 10C). In addition, the calibration plot showed an optimally fit with the ideal model (Figure 10D).

DISCUSSION

It has been reported that the incidence rate of LC is continuously rising because of smoking, air pollution, and other factors. Also, the morbidity of LUAD, the most common pathological subtype of LC, is increasing [1, 10]. Despite emerging advanced diagnostic and therapeutic means which have extensively upgraded the long-term quality of life and survival rate of LUAD patients, the 5-year survival rate is still <20% [4, 5]. The most commonly used method for clinical treatment

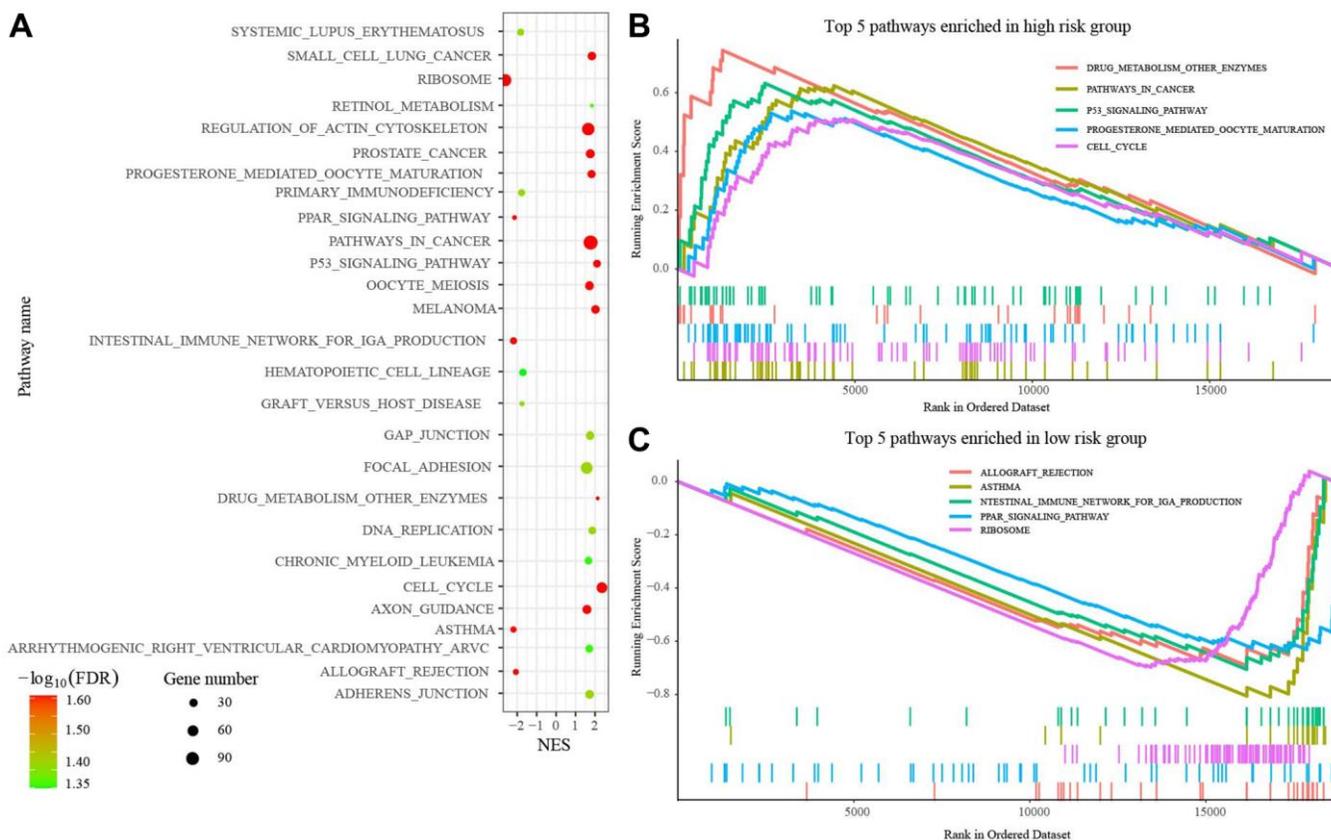


Figure 6. Gene set enrichment analysis (GSEA) between high- and low-risk groups. (A) 27 pathways enriched in GSEA. **(B)** Top 5 pathways enriched in high-risk group. **(C)** Top 5 pathways enriched in low-risk group. Abbreviations: FDR: False discovery rate; NES: Normalized enrichment score.

and prognosis prediction in LUAD is the TNM stage. Nevertheless, numerous studies have shown that, although patients had the same TNM stage and therapy strategy, the prognosis was different, indicating that prognosis prediction only relying on the TNM stage maybe not be enough to evaluate the clinical outcomes in LC [11]. Therefore, identifying more accurate tumor-specific biomarkers of LUAD for further guiding clinical decisions and prognostic evaluation remains a top priority.

Recently, a novel cell death model: disulfidptosis, has been proposed, which was induced by disulfide stress and markedly distinguished from any of the commonly studied forms of regulated cell death [8, 9]. In the cell death model, the excessive accumulation of intracellular disulfide molecules causes disulfide stress and bond to actin cytoskeleton proteins, subsequently leading to actin network collapse and cell death. In the process, NADPH could suppress disulfidptosis via resolving disulfide bonding to actin cytoskeleton proteins. There is no doubt that the concept would benefit a further and more comprehensive understanding of cell death mechanisms,

which may provide a practicable therapeutic means by targeting disulfidptosis in cancer treatment. However, up to now, no research has been reported on the role of disulfidptosis in LUAD. Herein, we would investigate the expression of DRGs in LUAD and the predictive value of DRGs in LUAD.

Firstly, we analyzed the differential expression of DRGs between normal and LUAD tissues and found the expression of 22 DGRs except for ACTN4 in LUAD tissues was significantly increased/decreased compared with normal tissues, implying the process of disulfidptosis in LUAD underwent drastically changed. Then, we took advantage of LASSO analysis and multivariate Cox regression analysis to develop a DRGs model for predicting the prognosis of LUAD patients with 5 candidate DRGs (ACTB, FLNB, NCKAP1, SLC3A2, SLC7A11). The AUC value and C-index based on the DRGs model in 4 sets were all more than 0.7 and significantly superior to that based on clinical parameters (age, gender, TNM stage, and smoking). Furthermore, the calibration plot demonstrated the consistency between the survival rate predicted by the

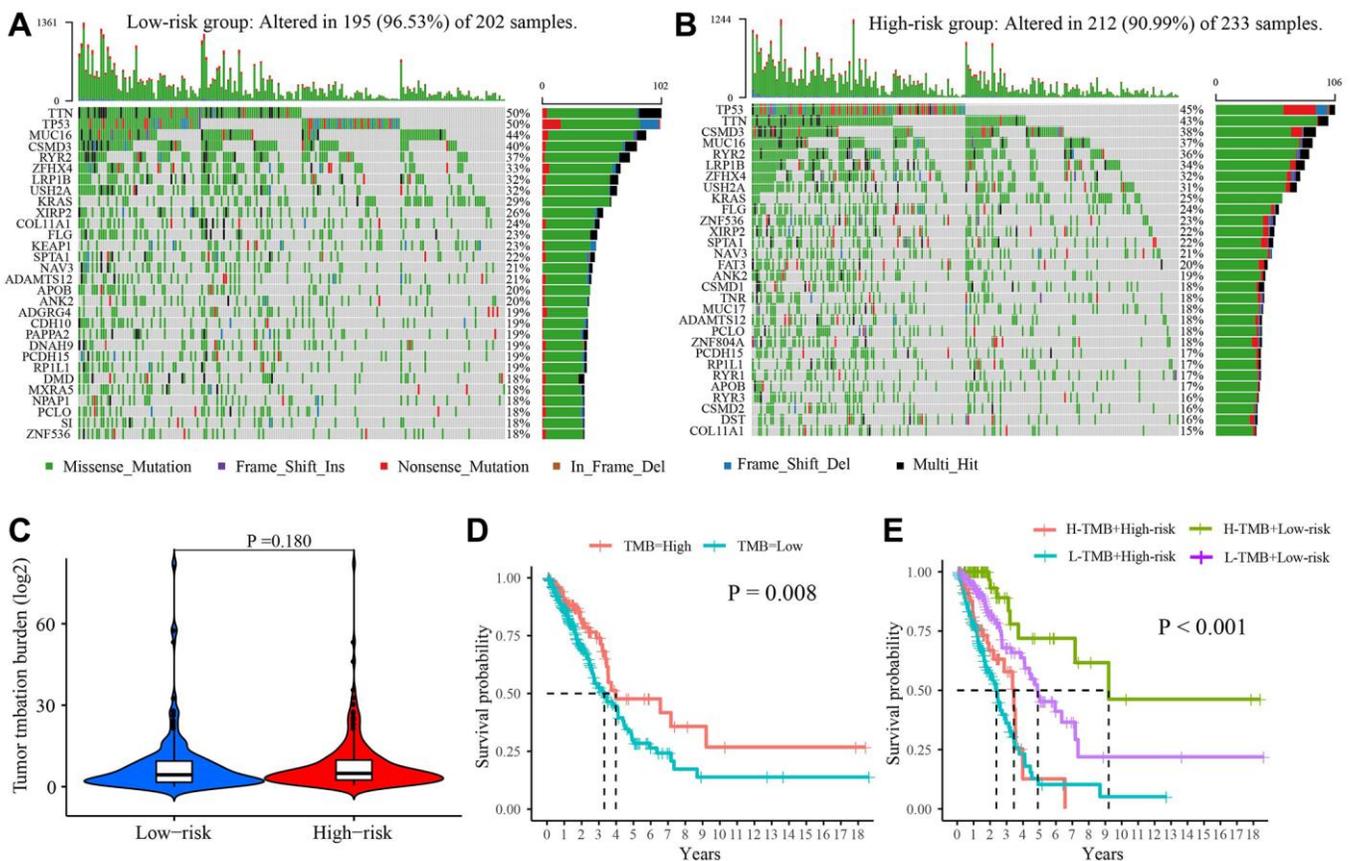


Figure 7. The tumor mutation burden characteristics in low- and high-risk group. (A) Mutational landscape in the low-risk group. (B) Mutational landscape in the high-risk group. (C) The difference in TMB between two groups. (D) Survival difference between high- and low-TMB. (E) Survival analysis of TMB along with risk score. Abbreviation: TMB: Tumor mutational burden.

DRGs model and the actual survival rate. All data suggested that the established prognostic model is stable enough to estimate the prognosis of LUAD patients.

We divided all samples into low- and high-risk groups according to the risk-score formula. The survival analysis illustrated that patients in the high-risk group exerted poorer clinical outcomes. And stratification analyses confirmed the results. Moreover, the univariate and multivariate Cox regression model indicated that the risk score was an independent prognostic factor in LUAD patients. The above data manifested that disulfidptosis was highly associated with the prognosis of LUAD patients, and the DRGs model could provide a good prediction of the clinical outcomes.

To determine the potential biological process changes between the low- and high-risk groups, we executed a GSEA analysis. The results showed 27 pathways were enriched. Of note, several immune-related pathways were identified in the low-risk group, including Primary immunodeficiency, Intestinal immune network for IgA production, etc. For further investigating the relation of the DRGs model to the immune status, we appraised the difference in the level of TIICs, the activity of immune-related functions, and the response to immunotherapy between the two groups. Interestingly, in the low-risk group, patients had a higher level of TIICs and better immune function, which may be why those patients had better treatment responses to immune therapy and better prognosis.

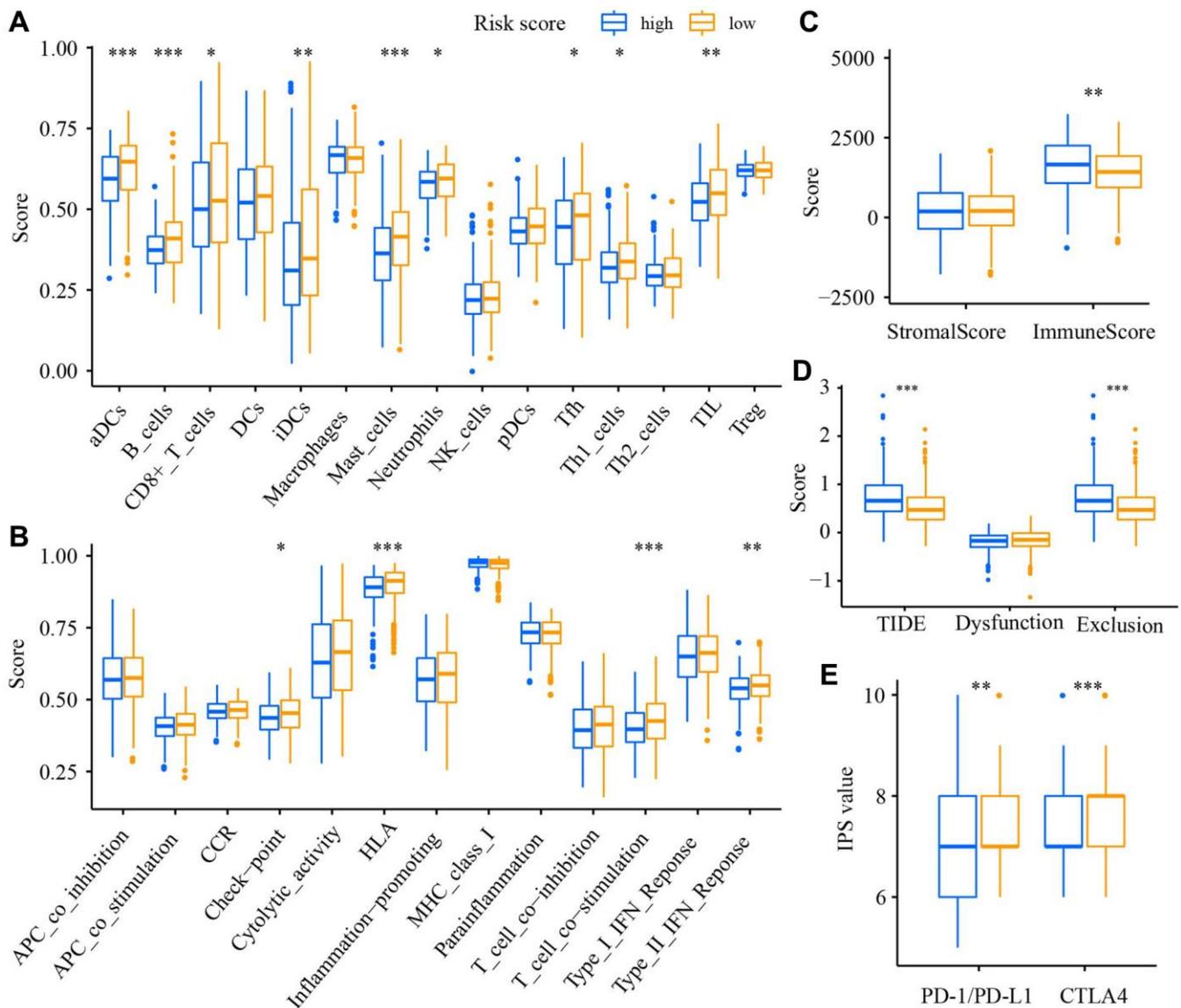


Figure 8. The relation of the DRGs model to immune status. Comparison of the enrichment score of (A) TIICs, (B) immune-related functions, (C) TME, (D) TIDE score, (E) The immunophenoscore value between low- and high-risk groups. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

The application of Next Generation Sequencing (NGS) provides a practicable means for exploring human cell transcription patterns [12]. Traditional method is cost-effective and measures the total RNA of all cells. Still,

it fails to consider the heterogeneity of cancer cells, which poses substantial challenges in diagnosing and treating cancer [13]. To uncover the expression profiles of DRGs in different cell types in LUAD tissue, we

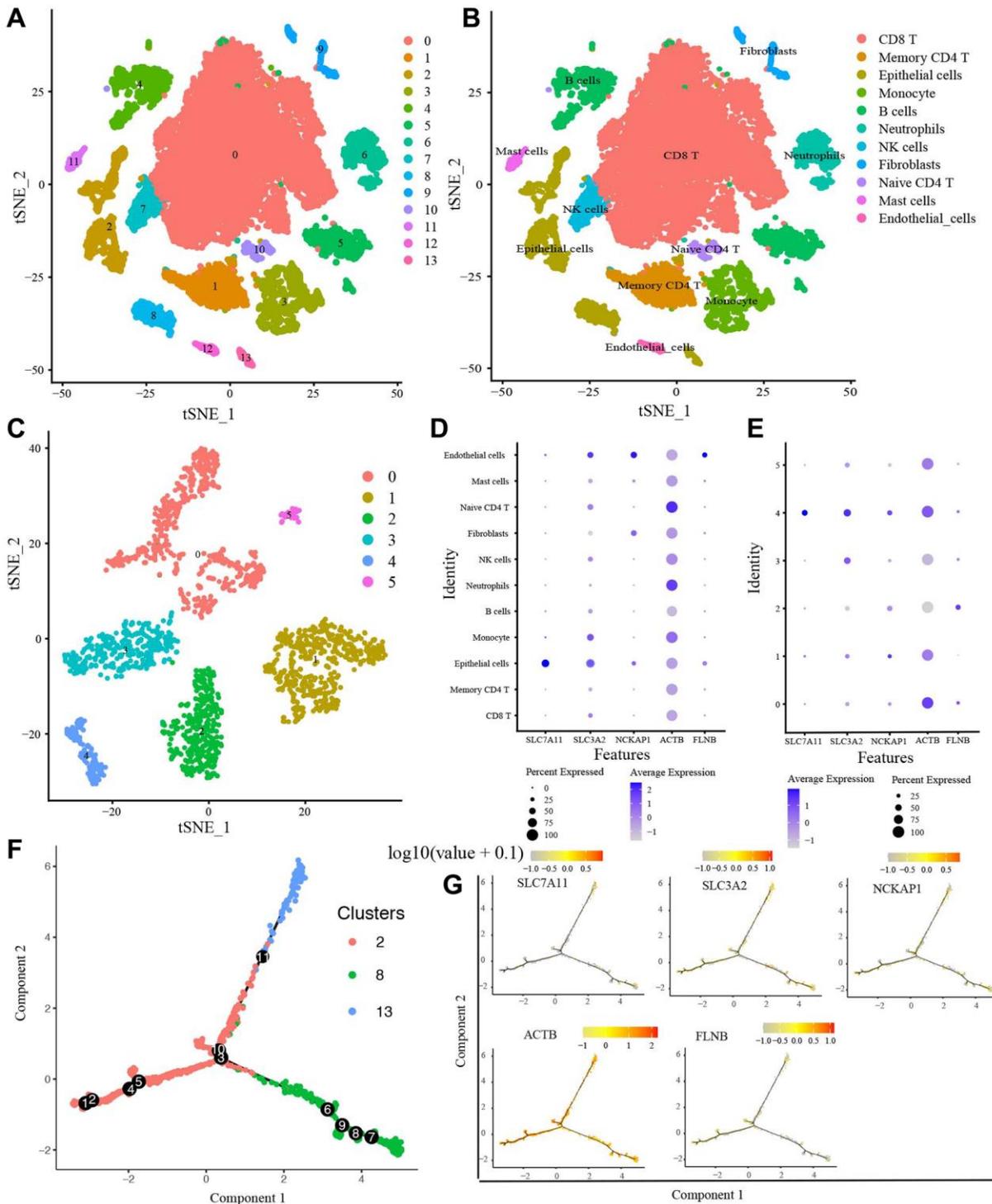


Figure 9. scRNA-seq data analysis. (A) The t-SNE algorithm divided the cells into 14 clusters by 20 principal components. (B) The tSNE plot revealing 14 clusters was annotated into 11 different cell types. (C) The t-SNE algorithm divided Epithelial cells into 6 sub-clusters. (D) The expression of 5 candidate DRGs in 11 cell types, (E) in 6 Epithelial cells sub-clusters. (F) The trajectory analysis of Epithelial cells with top 10 marker genes. (G) 5 candidate DRGs expression in different developmental states. Abbreviations: ScRNA-seq: Single-cell RNA sequencing; t-SNE: t-Stochastic Neighbor Embedding.

performed scRNA-seq analysis. All cells were clustered into 14 clusters and annotated to 11 cell types. CD8 T cell was the most abundant cell. CD8 T cells are classified into naive like, cytotoxic, and dysfunctional CD8 T cells based on their differentiation status. CD8 T cells have the ability to selectively detect and eradicate cancer cells. However, the sustained antigen stimulation of the tumor itself is considered one of the main driving factors for T cell dysfunction. In order to eliminate the tumor cells, T cells will enter the tumor microenvironment from the peripheral circulation. Because of this, the proportion of CD8 T cells in the tumor microenvironment will significantly increase. In addition, three clusters (2, 8, 13) were annotated to Epithelial cells. Tumor Epithelial cells mainly contain cancer cells and cancer stem cells. Then, we pictured

the transcriptional differentiation trajectory of the epithelial cells. The results demonstrated that Epithelial cells (clusters 2) differentiate into two clusters (cluster 8 and 13), implying the heterogeneity of cancer cells.

Herein, we developed a DRGs model to predict the OS in LUAD patients. However, some limitations still need to be addressed. (1) All data were gathered from two public databases: TCGA and GEO, and lacking sufficient experimental evidence and animal models to verify the results, which may lead to potential selection bias. (2) This study was retrospective, and prospective validations were still needed. (3) All data used in the study were from microarray expression and RNA-seq (RNA sequencing). And the specific role of disulfidptosis progress remains relatively enigmatic and

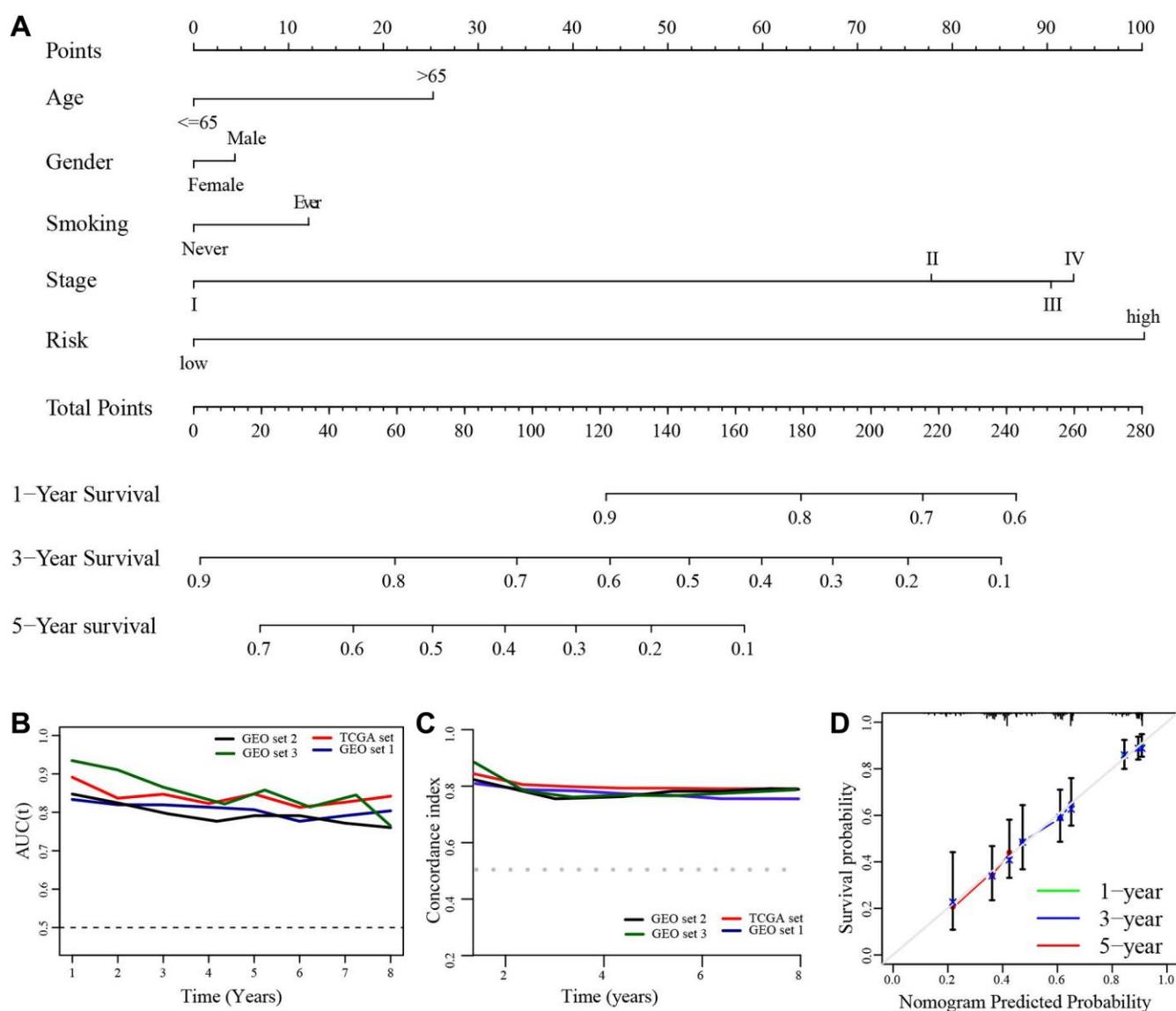


Figure 10. Construction of a prognostic nomogram. (A) A nomogram for predicting the 1-, 3-, and 5-year overall survival rate in LUAD patients. (B) The time-dependent AUC value. (C) The time-dependent C-index. (D) The calibration plot.

warrants further investigation. (4) Herein, we performed stratification analyses and determined the significant survival difference between the low- and high-risk groups. However, due to lacking information of therapies like surgery, targeted therapy, and immunotherapy in most patients, we could not homogenize the treatment and evaluate the predictive effectiveness and accuracy of the IRGPs signature in patients with surgery, targeted therapy, and immunotherapy. It may bring biased prognosis predictions.

Together, in the study, we investigated the expression profile of DRGs and constructed a predictive DRGs model for predicting the prognosis of LUAD patients. Evidence showed that the model was stable and reliable for predicting LUAD prognosis. Our research provided a novel understanding of the predictive value of disulfidptosis progress in LUAD and will be helpful for the prognosis evaluation of LUAD patients.

MATERIALS AND METHODS

Patients set

Four independent sets, including 1306 samples, were collected from TCGA and GEO databases in the study. TCGA set ($n = 465$) was from TCGA-LUAD and level 3 RNA sequencing. GEO set 1 ($n = 432$), GEO set 2 ($n = 83$), and GEO set 3 ($n = 226$) were from GSE68465 [14], GSE30219 [15], and GSE31210 [16], respectively. The gene expression profile in GSE68465, GSE30219, GSE31210 was raw data and normalized with the robust multi-array average algorithm via the R package Affy (3.17). The corresponding clinical data of all patients were also collected. Patients lacking pathologic diagnosis, survival time, or survival status would be removed. All patients' detailed demographic and baseline information was presented in Table 1. ScRNA-seq data from 8 human LUAD samples were gathered from the GSE171145 [17] dataset in the GEO database. Disulfidptosis-related genes were obtained from previous literature [8, 9, 18, 19].

Establishment and evaluation of a disulfidptosis-related genes model (DRGs) model

In the TCGA set, we investigated the difference in the expression of 23 DRGs between normal and LUAD tissues with the Wilcoxon rank-sum test. Then, we preliminary identified prognosis-related genes among 23 DRGs using the least absolute shrinkage and selection operator (LASSO) regression analysis with R package “glmnet” (4.0.2). Under the minimum lambda, genes with coefficient $\neq 0$ were chosen for screening genes. The iteration was set as 1,000 and 10-fold cross-validations to avoid overfitting. Then, the screened genes were subjected to the multivariate Cox regression

analysis to filter candidate genes and develop a DRGs model for predicting the prognosis of LUAD. The risk score was calculated as Risk score = $h_0(t) \times \exp(\beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n)$, where β refers to the regression coefficient; X represented the gene expression level; $h_0(t)$ is the benchmark risk function. Also, a risk score formula was formed, and patients' risk scores were calculated. Finally, we used the “surv cutpoint” functions of the package “Survminer” (0.4.9) in R to determine the optimal cut-off value of risk score, which divided patients into low- and high-risk groups. Based on the cut-off value, patients in GEO set 1, 2, and 3 were also divided into two groups.

The AUC value and the c-index in four sets were calculated to assess the predictive capacity of the DRGs model. The calibration plot with a boot-strapping set of 1,000 resamples was also pictured to investigate the predictive accuracy of the DRGs model.

GSEA

The GSEA was performed to explore the altered biological roles and signaling pathways between the low- and high-risk group with the R package “clusterProfiler” (4.2.2). The “c2.cp.v7.2.symbols.gmt (Curated)” was selected as annotated gene set. The false discovery rate (FDR) < 0.05 was set as the threshold.

Assessment of the relation of the DRGs model to the immune statute and drug sensitivity

The R package: GSVA (3.1.7) was applied to carry out the ssGSEA reflecting the level of TIICs and the activity of immune-related functions. The R package: “ESTIMATE” (1.0.13) was used to calculate the immune and stromal score in the tumor microenvironment. The immunophenoscore (IPS) value of samples was gathered from The Cancer Immunome Atlas (TCIA, <https://tcia.at>) to present the response to immune checkpoint inhibitors treatment [20]. The TIDE score was collected from Tumor Immune Dysfunction and Exclusion (TIDE, <http://tide.dfci.harvard.edu>). The higher TIDE score represents the worse response to immune therapy.

We took advantage of the R package “pRRophetic” (version 6) [21] to calculate the semi-inhibitory concentration values of commonly used anti-tumor drugs for LUAD to look for variations in the efficacy of medicines between the two groups. $P < 0.01$ was set as the cut-off value.

TMB

The DNA somatic mutation data of corresponding LUAD patients were also downloaded from TCGA and

Table 1. The baseline characteristics of lung adenocarcinoma patients in this study.

Parameter	TCGA cohort	GEO cohort 1	GEO cohort 2	GEO cohort 3
Database	TCGA-LUAD	GSE68465	GSE30219	GSE31210
Age				
≤65	235 (50.53%)	224 (51.85%)	60 (72.29%)	176 (77.88%)
>65	230 (49.46%)	208 (48.15%)	23 (27.71%)	50 (22.12%)
Gender				
Female	248 (53.33%)	214 (49.54%)	36 (43.37%)	121 (53.54%)
Male	217 (46.67%)	218 (50.46%)	47 (56.63%)	105 (46.46%)
Smoking				
Never	62 (13.33%)	50 (11.57%)	0	115 (55.88%)
Ever	384 (82.58%)	290 (67.13%)	0	111 (49.12%)
NA	19 (4.08%)	92 (21.30%)	83 (100%)	0
TNM stage				
I	264 (56.77%)	149 (34.50%)	64 (77.11%)	114 (50.44%)
II	103 (22.15%)	216 (50.00%)	13 (15.66%)	78 (34.51%)
III	73 (15.70%)	67 (15.50%)	6 (7.23%)	34 (15.05%)
IV	25 (5.38%)	0	0	0
Tumor size				
T1	162 (34.84%)	148 (34.26%)	56 (67.47%)	NA
T2	241 (51.83%)	245 (56.71%)	17 (20.48%)	NA
T3	44 (9.46%)	28 (6.48%)	10 (12.05%)	NA
T4	18 (3.87%)	11 (2.55%)	0	NA
NA	0	0	0	226 (100%)
Lymph node				
N0	316 (67.96%)	298 (68.98%)	73 (87.95%)	NA
N1-3	149 (32.04%)	134 (31.02%)	10 (12.05%)	NA
NA	0	0	0	226 (100%)
Metastasis				
M0	442 (95.05%)	432 (100%)	83 (100%)	NA
M1	23 (4.95%)	0	0	NA
NA	0	0	0	226 (100%)
Survival status				
Alive	305 (65.60%)	203 (46.99%)	40 (48.19%)	191 (84.51%)
Dead	160 (34.40%)	229 (53.01%)	43 (51.81%)	35 (15.49%)
Risk score				
Low	219 (47.10%)	232 (53.70%)	45 (54.22%)	86 (38.05%)
High	246 (52.90%)	200 (46.30%)	38 (45.78%)	140 (61.95%)
Total	465 (100%)	432 (100%)	83 (100%)	226 (100%)

Abbreviations: TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; NA: represents information not available.

further analyzed with the “maftools” R package (3.17). The ChromPlot function in the “maftools” R package was used to visualize the output results. The results were presented with waterfall diagram to show the variation distribution of genes with high somatic mutation frequency in LUAD samples and two different groups. Further, it was used to evaluate the difference in survival of the

patients classified in the high-TMB and low-TMB groups.

scRNA-seq data processing

The R package: “Seurat” (3.4) [22] was used to analyze the transcript count matrix for quality control and

preliminary data exploration. The filtering threshold was set as follows:

Excluding genes detected in less than 3 cells
Excluding cells with <50 genes detected
Excluding cells with >10% mitochondrial gene expression

Then, the expression profiles were normalized with the Log Normalization algorithm and were subsequently normalized using a linear regression model. The top 2000 highly expressed and variable genes were selected for PCA to determine significant and influential dimensions. The t-SNE algorithm was used to reduce the dimension of the top 20 principal components and gather major cell clusters. The marker genes between difference clusters were identified with $|\log_2(\text{fold change})| > 1$ and adjusted P value < 0.05 as the threshold. Cell annotation was carried out with the “SingleR” package (3.17) [23] and reports from the literature [24–26]. Finally, single-cell trajectory analysis was performed with the “Monocle 2 algorithm” [27].

Statistical analysis

The categorical data were presented as Numbers and compared with the chi-square test. The measurement data were presented as Mean \pm standard deviation (SD) and compared with the Wilcoxon rank-sum test. Correlation analysis was performed with the Spearman correlation test. Survival analysis was performed with the Kaplan–Meier plot and compared by the log-rank method. Finally, the univariate and multivariate Cox regression analysis determined the independent prognostic predictors. $P < 0.05$ was set as the cut-off value. All statistical analyses were conducted with R 4.1.1 (<https://www.r-project.org>).

Availability of data and materials

Bulk RNA-seq data were analyzed in this study, this data can be found at: The Cancer Genome Atlas (TCGA-LUAD, <https://portal.gdc.cancer.gov>) and Gene Expression Omnibus (GSE68465, GSE31210, and GSE30219; <https://www.ncbi.nlm.nih.gov>). ScRNA-seq data were collected from GEO-GSE171145.

Abbreviations

AUC: The area under the curve; C-index: concordance index; DRGs: Disulfidptosis-related genes; FDR: False discovery rate; GEO: Gene Expression Omnibus; GSEA: Gene set enrichment analysis; HR: Hazard ratio; IPS value: immunophenoscore value; LASSO: Least absolute shrinkage and selection operator; LC: Lung cancer; LUAD: Lung adenocarcinoma; OS: overall survival;

PCA: Principal component analysis; RNA-seq: RNA sequencing; ROC: Receiver operating characteristic curve; ScRNA-seq: Single-cell RNA sequencing; ssGSEA: Single-sample gene set enrichment analysis; TCGA: The Cancer Genome Atlas; TIICs: Tumor-infiltrating immune cells; TMB: Tumor mutational burden; t-SNE: t-Stochastic Neighbor Embedding.

AUTHOR CONTRIBUTIONS

XN and XD designed the study. HY, SW and XW analyzed the data. HY and SW contributed to interpretation of the results. HY and SW drafted the manuscript. XN and XD revised the final version of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors gratefully acknowledge contributions from TCGA database and GEO database.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

Project was supported by Jinhua City Science and Technology Plan Project (Grant No. 2022-3-063) and the Foundation of Key Laboratory of Cancer Prevention and Therapy Combining Traditional Chinese and Western Medicine of Zhejiang Province, Hangzhou, China (Grant No. 202206WS, 202206SJ).

REFERENCES

1. Barta JA, Powell CA, Wisnivesky JP. Global Epidemiology of Lung Cancer. *Ann Glob Health*. 2019; 85:8. <https://doi.org/10.5334/aogh.2419> PMID:30741509
2. Miller KD, Fidler-Benaoudia M, Keegan TH, Hipp HS, Jemal A, Siegel RL. Cancer statistics for adolescents and young adults, 2020. *CA Cancer J Clin*. 2020; 70:443–59. <https://doi.org/10.3322/caac.21637> PMID:32940362
3. 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>

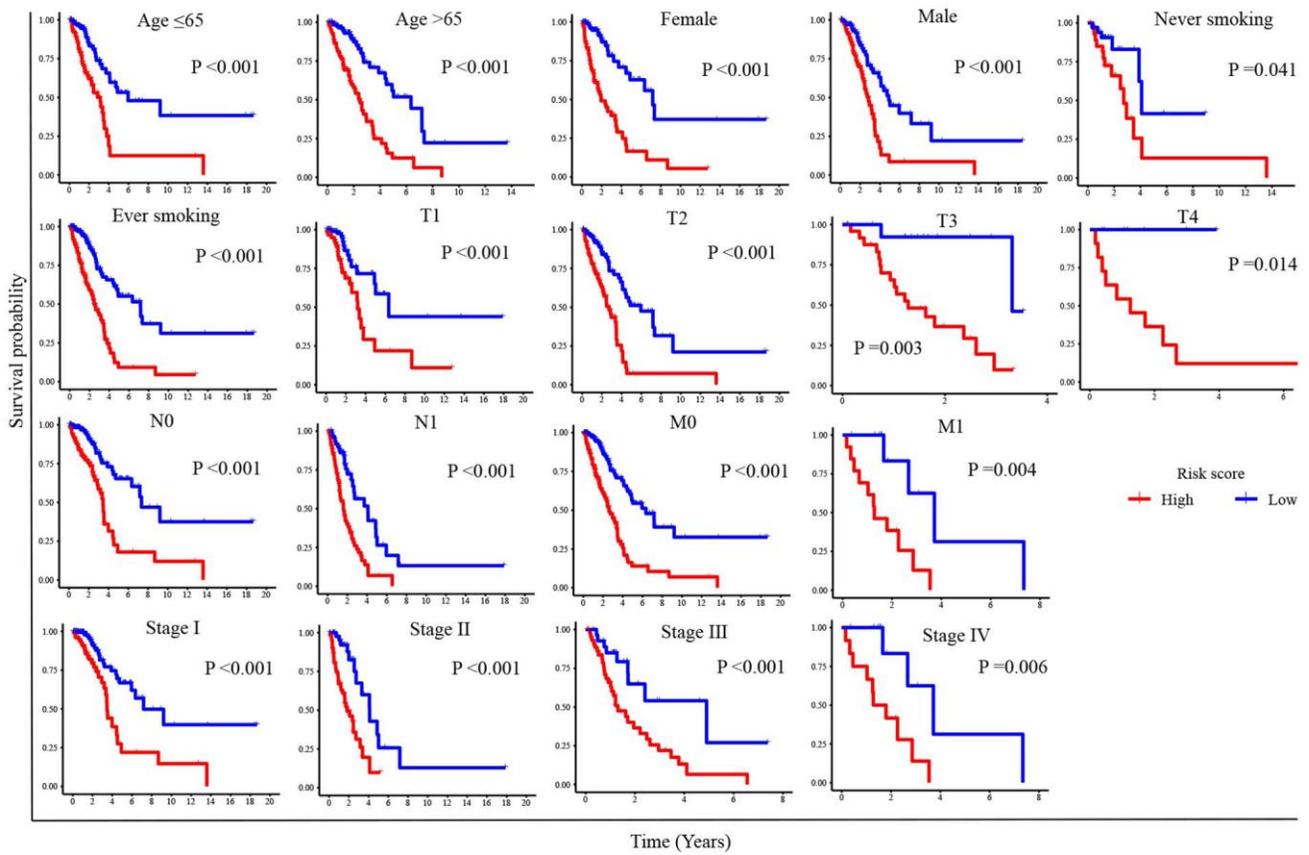
PMID:[33538338](#)

4. 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](#)
5. de Groot PM, Wu CC, Carter BW, Munden RF. The epidemiology of lung cancer. *Transl Lung Cancer Res*. 2018; 7:220–33.
<https://doi.org/10.21037/tlcr.2018.05.06>
PMID:[30050761](#)
6. Kroemer G, El-Deiry WS, Golstein P, Peter ME, Vaux D, Vandenabeele P, Zhivotovsky B, Blagosklonny MV, Malorni W, Knight RA, Piacentini M, Nagata S, Melino G, and Nomenclature Committee on Cell Death. Classification of cell death: recommendations of the Nomenclature Committee on Cell Death. *Cell Death Differ*. 2005 (Suppl 2); 12:1463–7.
<https://doi.org/10.1038/sj.cdd.4401724>
PMID:[16247491](#)
7. Zhang X, Wang L, Li H, Zhang L, Zheng X, Cheng W. Crosstalk between noncoding RNAs and ferroptosis: new dawn for overcoming cancer progression. *Cell Death Dis*. 2020; 11:580.
<https://doi.org/10.1038/s41419-020-02772-8>
PMID:[32709863](#)
8. Liu X, Nie L, Zhang Y, Yan Y, Wang C, Colic M, Olszewski K, Horbath A, Chen X, Lei G, Mao C, Wu S, Zhuang L, et al. Actin cytoskeleton vulnerability to disulfide stress mediates disulfidptosis. *Nat Cell Biol*. 2023; 25:404–14.
<https://doi.org/10.1038/s41556-023-01091-2>
PMID:[36747082](#)
9. Machesky LM. Deadly actin collapse by disulfidptosis. *Nat Cell Biol*. 2023; 25:375–6.
<https://doi.org/10.1038/s41556-023-01100-4>
PMID:[36918690](#)
10. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020; 70:7–30.
<https://doi.org/10.3322/caac.21590>
PMID:[31912902](#)
11. Liu Y, Ye G, Huang L, Zhang C, Sheng Y, Wu B, Han L, Wu C, Dong B, Qi Y. Single-cell transcriptome analysis demonstrates inter-patient and intra-tumor heterogeneity in primary and metastatic lung adenocarcinoma. *Aging (Albany NY)*. 2020; 12:21559–81.
<https://doi.org/10.18632/aging.103945>
PMID:[33170151](#)
12. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol*. 2018; 14:479–92.
<https://doi.org/10.1038/s41581-018-0021-7>
PMID:[29789704](#)
13. Zhang D, Lu W, Cui S, Mei H, Wu X, Zhuo Z. Establishment of an ovarian cancer omentum metastasis-related prognostic model by integrated analysis of scRNA-seq and bulk RNA-seq. *J Ovarian Res*. 2022; 15:123.
<https://doi.org/10.1186/s13048-022-01059-0>
PMID:[36424614](#)
14. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, Chang AC, Zhu CQ, Strumpf D, et al, and Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med*. 2008; 14:822–7.
<https://doi.org/10.1038/nm.1790>
PMID:[18641660](#)
15. Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, Moro-Sibilot D, Brichon PY, Lantuejoul S, Hainaut P, Laffaire J, de Reyniès A, Beer DG, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med*. 2013; 5:186ra66.
<https://doi.org/10.1126/scitranslmed.3005723>
PMID:[23698379](#)
16. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, Watanabe S, Sakamoto H, Kumamoto K, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res*. 2012; 72:100–11.
<https://doi.org/10.1158/0008-5472.CAN-11-1403>
PMID:[22080568](#)
17. Yang L, He YT, Dong S, Wei XW, Chen ZH, Zhang B, Chen WD, Yang XR, Wang F, Shang XM, Zhong WZ, Wu YL, Zhou Q. Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J Immunother Cancer*. 2022; 10:e003534.
<https://doi.org/10.1136/jitc-2021-003534>
PMID:[35140113](#)
18. Zhong Z, Zhang C, Ni S, Ma M, Zhang X, Sang W, Lv T, Qian Z, Yi C, Yu B. NFATc1-mediated expression of SLC7A11 drives sensitivity to TXNRD1 inhibitors in osteoclast precursors. *Redox Biol*. 2023; 63:102711.
<https://doi.org/10.1016/j.redox.2023.102711>
PMID:[37148740](#)
19. Zhao S, Wang L, Ding W, Ye B, Cheng C, Shao J, Liu J, Zhou H. Crosstalk of disulfidptosis-related subtypes, establishment of a prognostic signature and immune infiltration characteristics in bladder cancer based on

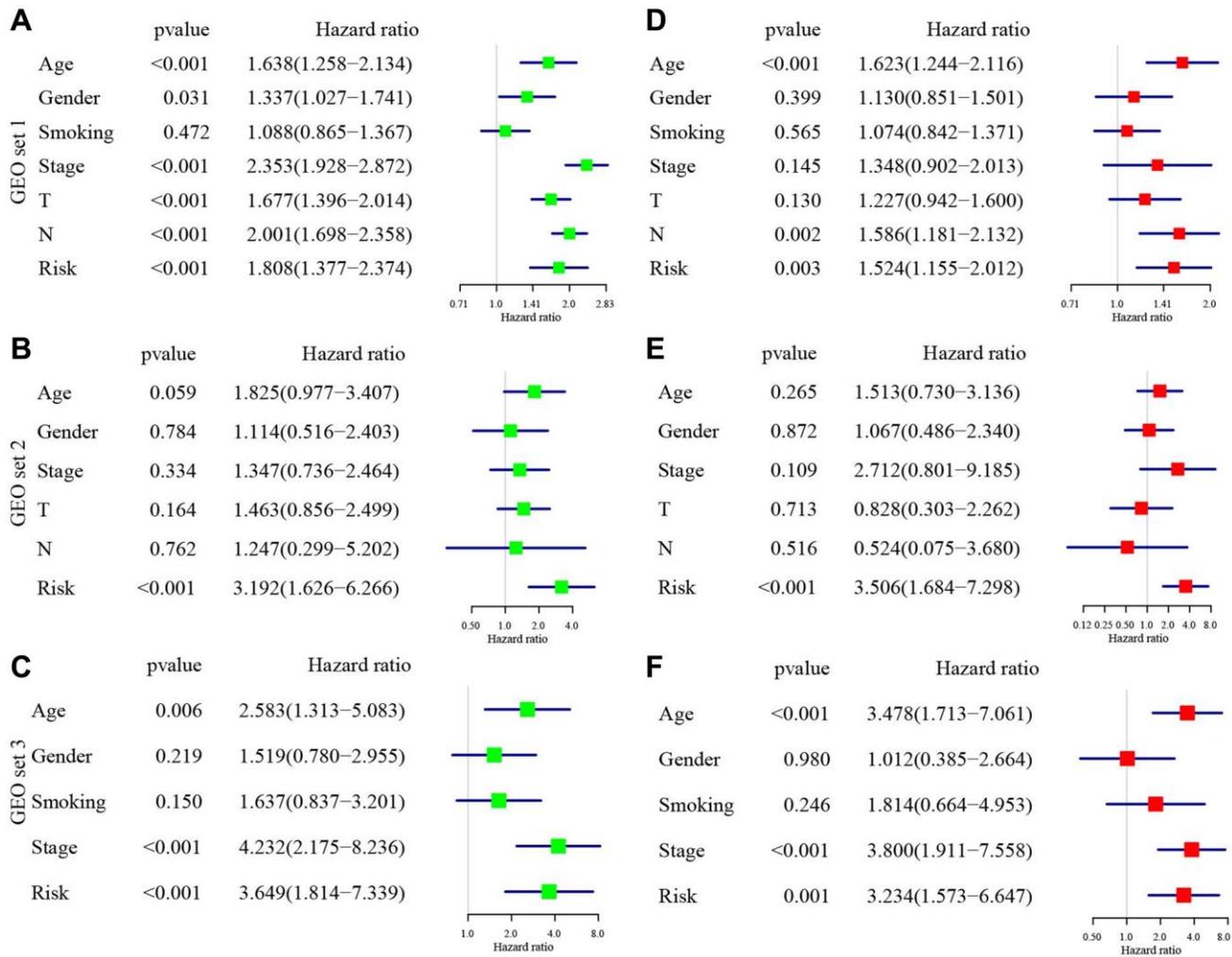
- a machine learning survival framework. *Front Endocrinol (Lausanne)*. 2023; 14:1180404.
<https://doi.org/10.3389/fendo.2023.1180404>
PMID:[37152941](https://pubmed.ncbi.nlm.nih.gov/37152941/)
20. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hui-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, et al. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell*. 2016; 165:35–44.
<https://doi.org/10.1016/j.cell.2016.02.065>
PMID:[26997480](https://pubmed.ncbi.nlm.nih.gov/26997480/)
 21. Geleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One*. 2014; 9:e107468.
<https://doi.org/10.1371/journal.pone.0107468>
PMID:[25229481](https://pubmed.ncbi.nlm.nih.gov/25229481/)
 22. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021; 184:3573–87.e29.
<https://doi.org/10.1016/j.cell.2021.04.048>
PMID:[34062119](https://pubmed.ncbi.nlm.nih.gov/34062119/)
 23. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. 2019; 20:163–72.
<https://doi.org/10.1038/s41590-018-0276-y>
PMID:[30643263](https://pubmed.ncbi.nlm.nih.gov/30643263/)
 24. Wang Z, Wang Y, Chang M, Wang Y, Liu P, Wu J, Wang G, Tang X, Hui X, Liu P, Guo X, Xing B, Wang Y, et al. Single-cell transcriptomic analyses provide insights into the cellular origins and drivers of brain metastasis from lung adenocarcinoma. *Neuro Oncol*. 2023; 25:1262–74.
<https://doi.org/10.1093/neuonc/noad017>
PMID:[36656750](https://pubmed.ncbi.nlm.nih.gov/36656750/)
 25. Wu F, Fan J, He Y, Xiong A, Yu J, Li Y, Zhang Y, Zhao W, Zhou F, Li W, Zhang J, Zhang X, Qiao M, et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat Commun*. 2021; 12:2540.
<https://doi.org/10.1038/s41467-021-22801-0>
PMID:[33953163](https://pubmed.ncbi.nlm.nih.gov/33953163/)
 26. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, Bassez A, Decaluwé H, Pircher A, Van den Eynde K, Weynand B, Verbeken E, De Leyn P, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med*. 2018; 24:1277–89.
<https://doi.org/10.1038/s41591-018-0096-5>
PMID:[29988129](https://pubmed.ncbi.nlm.nih.gov/29988129/)
 27. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014; 32:381–6.
<https://doi.org/10.1038/nbt.2859>
PMID:[24658644](https://pubmed.ncbi.nlm.nih.gov/24658644/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. Stratification analyses of overall survival between high- and low-risk patients in different subgroup.



Supplementary Figure 2. The univariate Cox regression analysis of risk score and clinical characteristics in (A) GEO set 1, (B) GEO set 2, (C) GEO set 3. The multivariate Cox regression analysis of risk score and clinical characteristics in (D) GEO set 1, (E) GEO set 2, (F) GEO set 3.

