Identification of lung adenocarcinoma subtypes and a prognostic signature based on activity changes of the hallmark and immunologic gene sets

Shunkai Zhou  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Dehua Zeng  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Meiqing Zhang  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Mengmeng Chen  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Yaming Liu  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Qiqiang Chen  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Zhenya Lin  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Shengsheng Yang  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Zhichao Fu  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Wenmin Ying  
ywm15080015210@163.com

Fuding hospital

Duohuang Lian  
the 900th Hospital of the joint logistics support force of the Chinese people's Liberation Army

Research Article

Keywords: Lung adenocarcinoma, subtype, hallmark and immunologic gene sets, prognostic signature, immune microenvironment, prognosis

Posted Date: November 18th, 2022
**DOI:** https://doi.org/10.21203/rs.3.rs-2248895/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.

**Version of Record:** A version of this preprint was published at Heliyon on March 1st, 2024. See the published version at https://doi.org/10.1016/j.heliyon.2024.e28090.
Abstract

Background: Lung adenocarcinoma (LUAD) has a complex tumor heterogeneity. This study aimed to identify LUAD subtypes and build a reliable prognostic signature based on the activity changes of the hallmark and immunologic gene sets.

Methods: Changes in the activities of the hallmark and immunologic gene sets were analyzed based on The Cancer Genome Atlas (TCGA)-LUAD dataset, followed by identification of prognosis-related differential gene sets (DGSs) and their related LUAD subtypes. Survival analysis, correlation with clinical characteristics, and immune microenvironment assessment for subtypes were performed. Moreover, the DGSs among different subtypes were identified, followed by the construction and evaluation of a prognostic risk score model and nomogram. The tumor mutation burden (TMB) of different risk groups was compared.

Results: Two LUAD subtypes were identified based on the activity changes of the hallmark and immunologic gene sets. Cluster 2 had worse prognosis, more advanced tumor and clinical stages, and higher immune infiltration than cluster 1. Moreover, a prognostic risk score signature was established using two LUAD subtype-related DGSs, which could stratify patients at different risk levels. A shorter survival time and higher TMB levels were observed in the high-risk patients. The established nomogram accurately predicted the survival outcomes.

Conclusions: Our findings revealed that our constructed prognostic signature could accurately predict the survival outcomes and immune microenvironment of patients with LUAD, which was helpful in predicting the prognosis and guiding personalized therapeutic strategies for LUAD.

Highlights

1. Two molecular subtypes of LUAD were identified.
2. Cluster 2 showed worse prognosis and higher immune infiltration than cluster 1.
3. A prognostic risk score model was established using two subtype-related DGSs.
4. The high-risk group had worse survival and higher TMB than the low-risk group.
5. Nomogram constructed based on the risk score and clinical parameters can predict survival.

Introduction

Lung cancer is responsible for a large number of cancer-related deaths worldwide. According to a published report in 2020, the global lung cancer burden has increased, with 2.2 million (11.44%) new cases and 1.8 million (18%) deaths. Moreover, approximately 75% of lung cancer cases are diagnosed at the metastatic or locally advanced stages, resulting in poor overall survival (OS) (6% and 37%, respectively) (1). Non-small cell lung cancer (NSCLC) is the main lung cancer type, of which lung adenocarcinoma (LUAD) is the most common subtype having complex tumor heterogeneity (2). Currently,
the available therapeutic regimens include surgery, chemotherapy, and radiotherapy, either alone or in combination. However, using chemotherapy or targeted drugs is almost always accompanied by multiple adverse effects, such as drug resistance and multi-organ toxicity (3). For patients with cancer that exhibit different clinical phenotypes and expected prognoses, individualized treatment strategies will help to achieve good clinical outcomes (4). Therefore, identification of LUAD subtypes and a reliable prognostic signature will improve patient's clinical outcomes through personalized treatment and accurate prognosis assessment.

Cancer is a complex disease involving complex regulations between the tumor and the immune system (5). The immune system plays an essential role in cancer initiation and progression, and evading immune destruction is considered a cancer hallmark. The breakthrough in immunotherapy has greatly benefited some patients with lung cancer (6). Recently, many research studies have focused on discovering immune-related prognostic gene sets in various solid tumors that are related to immunotherapy response and cancer prognosis (7). For instance, a prognostic signature constructed from 10 immune-related genes shows great performance in predicting disease prognosis and can improve the immunotherapy management in LUAD (8). Song et al. also constructed a prognostic signature comprising 30 immune-related genes, which had a high prognostic value in patients with LUAD and promoted individualized treatment (9). In addition, cancer complexity is represented by numerous cancer hallmarks that allow cancer cells to proliferate and metastasize. Cancer hallmark genes are responsible for crucial phenotypic features associated with the malignant development of cancers, and signatures established by cancer hallmark genes are important in various cancer types (10). A hallmark-based six-gene expression signature can reflect the genetic hallmarks of colorectal cancer cells and can be used to assess the recurrence risk of colorectal cancer (11). Moreover, a robust six-gene prognostic signature that is remarkably associated with cancer hallmarks is reliable for predicting OS in NSCLC (12). These data suggest the potential of immunologic and hallmark gene sets as promising biomarkers for cancer prognosis. However, the crucial hallmark and immunologic gene sets associated with LUAD prognosis have not yet been fully investigated.

Changes in the activities of the hallmark and immunologic gene sets may have a crucial impact on lung cancer prognosis by affecting the colonization of circulating tumor cells in blood and their proliferation after therapeutic interventions, such as pulmonary celllectomy. Nevertheless, few studies have been conducted to identify LUAD subtypes or prognostic gene sets based on activity changes of the hallmark and immunologic gene sets. Herein, we analyzed the activity changes of the hallmark and immunologic gene sets on basis of the The Cancer Genome Atlas (TCGA)-LUAD dataset, followed by identification of prognosis-related differential gene sets (DGSs) and their related LUAD subtypes. Survival analysis, correlation with clinical characteristics, and immune microenvironment assessment for the subtypes were performed. Moreover, the DGSs among different subtypes were identified, followed by the construction and evaluation of a prognostic risk score model and nomogram. Additionally, two Gene Expression Omnibus (GEO) datasets (GSE37745 and GSE68465) were used to validate the predictive performance of the prognostic model. Furthermore, the tumor mutation burden (TMB) levels of the different risk groups
were compared. Our findings provide more in-depth insight to accurately assess the prognosis and improve the therapeutic interventions for LUAD.

**Materials And Methods**

**Data acquisition**

The gene expression RNA-sequencing (RNA-seq) and clinical data of LUAD were downloaded from TCGA database. This dataset included 504 cancerous and 59 normal samples.

Gene expression profiles of lung cancer with prognostic information, namely GSE37745 and GSE68465, were downloaded from the GEO database. By de-batch integration, 638 lung cancer samples from the two datasets were used for subsequent analyses.

Furthermore, 4922 hallmark and immunologic gene sets were downloaded from the Molecular Signatures Database (MSigDB) (13) using Gene set enrichment analysis (GSEA) (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp).

**Gene set variation analysis (GSVA)**

GSVA was applied to observe the variation in the pathway activities of a set of genes under a particular biological condition by estimating the relative enrichment of a gene set over a sample population (14). Based on the sample information in TCGA-LUAD dataset, the enrichment score of 4922 hallmark and immunologic gene sets were analyzed using GSVA (14) to uncover changes in the pathway activities of the hallmark and immunologic gene sets between the tumor and normal samples.

**Identification of prognosis-related DGSs**

Based on the enrichment scores of gene sets in samples in TCGA-LUAD dataset, the DGSs between tumor and normal samples were analyzed using the t-test in R4.1.2, with a cut-off value of p < 0.01. Moreover, prognosis-related gene sets were screened using univariate Cox regression analysis in the survival package (version 3.2.13) (15) in R4.1.2, with a threshold value of p < 0.01. By intersection analysis of DGSs and prognosis-related gene sets, prognosis-related DGSs were obtained.

**Subtype analysis using consensus clustering**

Based on prognosis-related DGSs, lung cancer samples were clustered using an unsupervised clustering method with the ConsensusClusterPlus package (version 1.58.0) (16) in R4.1.2. The number of clusters (k) was set as 2–9. Moreover, the stability of the clustering results was verified using the proportion of ambiguous clustering (PAC); the minimum PAC was considered the optimal k value.

**Survival analyses for subtypes**

Based on the survival information of samples, Kaplan–Meier (KM) survival analyses for each subtype were conducted using the survival package (version 3.2.13) (15) in R4.1.2, with the aim of comparing the
prognostic differences among the subtypes.

**Correlation between subtypes and clinical characteristics**

According to the clinical data of samples in each type, the correlation between subtypes and different clinical characteristics, including gender, TNM staging, and clinical stage, was analyzed using the Chi-square test.

**Immune microenvironment assessment for subtypes**

According to the expression levels in each sample in the TCGA-LUAD dataset, the proportion of 22 immune cell types was estimated using the CIBERSORT algorithm (17). Differential immune cells (DICs) were compared among the different subtypes. In addition, the estimate package (18) in R4.1.2 was applied to assess the immune, stromal, and ESTIMATE scores, as well as tumor purity of each sample. The differences in stromal and immune scores among the different subtypes were analyzed using the Wilcoxon test.

**Identification of DGSs among subtypes**

Based on tumor subtype information, the DGSs among different subtypes were analyzed using the limma package (version 3.50.0) (19) in R4.1.2. The cut-off values were $p < 0.01$ and $|\log_2 FC| > 0.25$.

**Construction and evaluation of a prognostic risk score model**

For subtype-related DGSs, least absolute shrinkage and selector operation (LASSO) regression analysis was performed to select feature gene sets using the glmnet package (version 4.1.3) (20) in R4.1.2. Subsequently, according to the enrichment scores of the feature gene sets, they were divided into high- and low-enrichment score groups. Survival analysis was performed using the survival package (version 3.2.13) to analyze the association between the enrichment scores of gene sets and prognosis of the disease. The gene sets with $p < 0.01$ were subjected to stepwise Cox regression analysis, followed by establishment of a prognostic risk score model with the following formula:

$$\text{Risk score} = h_0(t) \times \exp (\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n)$$

where $\beta$ is the regression coefficient, $h_0(t)$ is the benchmark risk function, and $h(t, X)$ is the risk function related to the $X$ covariable at time $t$.

Using this formula, the risk score of each sample in TCGA-LUAD and GEO datasets was calculated. Survival analysis was then performed to analyze the prognostic differences between different risk groups that were divided based on their median risk scores.

**Nomogram construction**
To investigate the prognostic factors, risk scores and clinical factors (age, sex, pathologic M, N, and T, and clinical stage) in TCGA-LUAD dataset were subjected to univariate and multivariate Cox regression analyses. Variables with \( p < 0.01 \) were selected as independent prognostic factors. Then, a nomogram was established using the rms package (version 6.2.0) (21) to predict the 1-, 3-, and 5-year survival probabilities of patients with lung cancer.

**Analysis of the mutation differences between different risk groups**

Using the samples in TCGA-LUAD dataset, the mutation information of each gene in the sample was counted, and the top 20 significantly mutated genes with large mutation numbers were selected. Moreover, the mutation frequencies of these significantly mutated genes were analyzed in all LUAD samples using the maftools package (version 2.8.0) in R4.1.2. Moreover, the TMB levels of all tumor samples were calculated, and the differences between different risk groups were compared.

**Results**

**Activity changes of gene sets in lung cancer samples**

To reveal variation in the activities of the hallmark and immunologic gene sets in lung cancer and normal samples, GSVA was conducted to calculate the enrichment score of 4922 hallmark and immunologic gene sets. The enrichment score differences of hallmark and immunologic gene sets in tumor and non-tumor samples are shown in Fig. 1.

**Identification of prognosis-related DGSs**

Based on the enrichment score of 4922 gene sets, 4048 DGSs between tumor and normal samples were identified. Moreover, 1054 prognosis-related gene sets were obtained. By further intersection analysis, 955 prognosis-related DGSs were obtained.

**Two subtypes were clustered based on prognosis-related DGSs, which had different survival and clinical phenotypes**

Based on prognosis-related DGSs, lung cancer samples were clustered into two subtypes using unsupervised consensus clustering analysis (Fig. 2A-B). Clusters 1 and 2 contained 244 and 260 lung cancer samples, respectively. Further PAC analysis showed that the optimal k value was 2 (Fig. 2C).

According to the clinical information of samples clustered in different subtypes, KM survival analysis was performed to determine whether there were any differences in the prognoses of the two subtypes. The results showed that the samples from cluster 2 showed worse prognosis than those from cluster 1 (\( p = 0.004 \), Fig. 2D). Then, association analysis of the subtypes with clinical phenotypes was performed. There was no significant difference in pathologic M between the two subtypes (Fig. 2E). The clustering of
subtypes was significantly related to gender, pathologic N, pathologic T, and clinical stages. Compared with cluster 1, cluster 2 contained more male samples, more samples with N2–N3 stage, more samples with T2–T4 stage, and more samples with II–IV clinical stage (Fig. 2F-I). These data suggest that the samples of cluster 2 had relatively high levels of malignancy compared to those of cluster 1, which possibly resulted in their poor prognosis.

**Cluster 2 had higher immune infiltration than cluster 1**

To investigate whether there was a significant difference in the tumor immune microenvironment between the two subtypes, the proportion of 22 types of immune cells was quantified based on the expression data in TCGA-LUAD. Eleven DICs were identified between the two subtypes. The abundance of plasma cells, activated memory cluster of differentiation (CD)-4 T cells, activated natural killer (NK) cells, M₀ macrophages, M₁ macrophages, resting dendritic cells, and neutrophils was significantly elevated in cluster 2, whereas that of resting memory CD4 T cells, memory B cells, monocytes, and resting mast cells was remarkably increased in cluster 1 (Fig. 3A). Moreover, the stromal, immune, and ESTIMATE scores were significantly decreased in cluster 2 than in cluster 1 (Fig. 3B-D), while the tumor purity was dramatically increased in cluster 2 (Fig. 3E).

**The prognostic risk score model was built by two subtype-related DGSs**

Based on the tumor subtype information, 331 DGSs between clusters 1 and 2 were identified (Fig. 4A). For subtype-related DGSs, LASSO regression analysis was performed, and five feature gene sets were obtained (Fig. 4B-C). Survival analysis further revealed that the high enrichment scores of GSE1460_NAIVE_CD4_TCELL_CORD_BLOOD_VS_THYMIC_STROMAL_CELL_DN, GSE20715_0H_VS_48H_OZONE_LUNG_DN, and GSE45365_HEALTHY_VS_MCMV_INFECTION_CD11B_DC_DN were associated with poor OS, whereas the high enrichment scores of GSE16450_CTRL_VS_IFNA_6H_STIM_IMMATURE_NEURON_CELL_LINE_DN and GSE22886_NAIVE_BCELL_VS_BLOOD_PLASMA_CELL_UP were associated with better OS (Fig. 4D-E). Further stepwise Cox regression analysis identified two gene sets, GSE1460_NAIVE_CD4_TCELL_CORD_BLOOD_VS_THYMIC_STROMAL_CELL_DN and GSE16450_CTRL_VS_IFNA_6H_STIM_IMMATURE_NEURON_CELL_LINE_DN (Fig. 4F). Based on the regression coefficient and enrichment scores in TCGA-LUAD dataset, a prognostic risk score model was developed.

**The prognostic signature had good accuracy in predicting OS**

The risk score of each patient in TCGA-LUAD training dataset was calculated, and patients were stratified into high- and low-risk groups. The patients with lung cancer in the high-risk group had a shorter OS than those in the low-risk group (Fig. 5A-B). The area under the receiver operating characteristic (ROC) curve
(AUC) value reached 0.749 (95% CI: 0.706–0.792), suggesting that the risk score had a good predictive performance for OS (Fig. 5C). In addition, the GEO validation dataset was used to test the robustness of the risk score model. Likewise, the survival time of patients was significantly reduced in the high-risk group (Fig. 5D-E). The AUC value of the risk score was 0.713 (95% CI: 0.672–0.754) (Fig. 5F).

A nomogram was constructed by the risk score and clinical factors

The independent prognostic factors were further investigated. Univariate Cox regression analysis showed that the risk score, pathologic N, pathologic T, and clinical stages were significantly associated with OS (p < 0.001, Fig. 6A). Further multivariate Cox regression analysis revealed the risk score (p = 0.004) and clinical stage (p < 0.001) as independent prognostic factors for patients with lung cancer (Fig. 6B). Then, a nomogram was established, which could accurately predict the 1-, 3-, and 5-year survival probabilities (Fig. 6C). Calibration curve analysis confirmed the good predictive value of the nomogram (Fig. 6D).

Risk score was positively correlated with TMB

Based on the mutation information of samples in TCGA-LUAD dataset, top 20 significantly mutated genes were obtained. TP53 and titin (TTN) had significantly higher mutation frequencies in high-risk samples than in low-risk samples (Fig. 7A). Moreover, the TMB of all tumor samples was calculated, and the TMB in the high-risk samples was significantly higher than that in the low-risk samples (p < 0.05, Fig. 7B). Correlation analysis revealed that the TMB levels in samples were significantly positively correlated with the risk score (p = 1.6e-13, Fig. 7C).

Discussion

LUAD is a heterogeneous disease, and hence, patients with the same pathological stage may exhibit different prognoses; some can overcome the disease, while others may show a relapse. Treatment regimens targeting anaplastic lymphoma kinase and epidermal growth factor receptor have benefited only a small percentage of patients with LUAD over the past decade (22, 23). Identification of the molecular subtypes of LUAD is urgently needed for its precise treatment. In this study, we uncovered two molecular subtypes of LUAD based on the activity changes of the hallmark and immunologic gene sets in LUAD samples from TCGA-LUAD cohort. Cluster 2 had worse prognosis, more advanced TN and clinical stages, and higher immune infiltration than cluster 1. Moreover, we established a prognostic signature using two LUAD subtype-related DGSs, which could stratify patients into different risk groups, and patients in the high-risk group had a reduced survival time and high TMB levels.

Several studies have focused on the exploration of LUAD subtypes due to the significant heterogeneity of this cancer type. For example, Xu et al. identified three LUAD subtypes by analyzing the immunogenomic profiling of 29 immune signatures, which provide potential strategies for LUAD treatment (24). Wang et al. identified two distinct subtypes of LUAD based on 433 immune-related prognostic genes, which
exhibited significantly different survival outcomes (25). Based on the activity changes of the hallmark and immunologic gene sets, we identified two LUAD subtypes, and cluster 2 had worse prognosis, more advanced TN and clinical stages, and higher immune infiltration than cluster 1. Immune infiltration of the tumor microenvironment is associated with survival in LUAD (26). In patients with LUAD, the proportion of tumor-infiltrating immune cells, such as the plasma cells, activated memory CD4 T cells, activated NK cells, resting dendritic cells, and M1 macrophages, was increased in the tumor tissues compared to non-tumor tissues, suggesting a potential association between immune infiltration and LUAD development and clinical outcome (27). Taken together, different prognosis, clinical characteristics, and immune infiltration of LUAD subtypes illustrate the necessity of studying different molecular subtypes and the significance of developing novel therapeutic strategies.

Based on two subtype-related DSGs, we built a prognostic signature to calculate the risk scores in LUAD. Worse OS was found in patients with high-risk scores, suggesting that more regular follow-up examinations are required to monitor the disease occurrence in these patients. In addition, the risk score could independently predict OS in patients with LUAD, hinting at its significant value in clinical decision making. Moreover, a nomogram can assess the survival of patients with cancer [28]. For a more intuitive clinical application, we constructed a nomogram that incorporated the risk scores and clinical features, including clinical stage, which exhibited good accuracy in predicting prognosis. This nomogram provides a convenient method for predicting survival and developing individualized treatment strategies for patients.

Furthermore, our constructed prognostic signature could divide patients into different risk groups, and high TMB levels were observed in the high-risk patients. TMB is an indicator for predicting responses to immune checkpoint blockade (ICB) treatment (28, 29). An integrated analysis of 27 cancer types has revealed a positive correlation between TMB levels and ICB treatment (30). Three previous clinical trials, including CHECKMATE-026, KEYNOTE-001, and CHECKMATE-227, revealed that patients with NSCLC with high TMB levels could benefit more from ICB treatment (31–33). In current clinical practice, few patients are found to benefit from ICB therapy; therefore, it is necessary to develop a new approach to screen patients sensitive to ICB treatment. Our results revealed higher TMB levels in high-risk patients, implying that these patients might be more sensitive to ICB therapy. In addition, TP53 and TTN were found to be frequently mutated in LUAD samples, and their mutation frequencies were higher in the high-risk samples than the low-risk samples. TP53 mutation has been revealed to be a poor prognostic factor in patients with LUAD (34). Moreover, TP53-mutated tumors exhibit remarkably increased TMB in LUAD, suggesting that patients with LUAD with TP53 mutations might derive clinical benefits from ICB therapy (35). TTN mutations have been reported to frequently occur in many cancer types and are associated with TMB status (36, 37). Xue et al. suggested that TTN/TP53 mutations could function as predictors of chemotherapy response in LUAD (38). Thus, we speculate that TTN/TP53 mutations may be responsible for the response to ICB therapy in patients with LUAD with high-risk scores.

Our study has several limitations. First, our analysis was concluded using online public data, and the robustness of our constructed prognostic signature warrants further verification in more cohorts. Second,
our identified prognostic gene sets were not validated in clinical samples. Lastly, the prognostic value of the nomogram is required to be validated in more prospective clinical cohorts.

In conclusion, two LUAD subtypes were identified based on their prognostic hallmarks and immunologic gene sets. Our constructed prognostic signature could accurately predict the survival outcomes and immune microenvironment of patients with LUAD. Furthermore, the nomogram may facilitate prognosis prediction and guide personalized therapeutic strategies for patients with LUAD.

Declarations

**Ethics approval and consent to participate**

TCGA belong to public databases. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. Our study is based on open source data, so there are no ethical issues and other conflicts of interest.

**Consent for publication**

Not Applicable.

**Availability of data and materials**

Previously reported gene expression and clinical data were used to support this study and are available at the Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). These prior studies (and datasets) are cited at relevant places within the text as references.

**Competing interests**

This is an original article, and has not been simultaneously presented to any other periodical.

The authors have declared that no competing interests exist.

All listed authors participated meaningfully in the study and that we have been and approved the final manuscript.

**Funding**

Not Applicable.

**Authors’ contributions**

Among the authors, Wenmin Ying and Duohuan Lian are responsible for the conceptualization and design of the research; Shunkai Zhou plays a guiding role in methodology and data curation; Dehua Zeng mainly takes charge of writing and researching; Meiqing Zhang and Mengmeng Chen are responsible for investigation and acquisition of data; Yaming Liu and Qiqiang Chen take the responsibility of statistical
analysis; Zhenya Lin and Shengsheng Yang are responsible for software and validation; Zhichao Fu takes charge of visualization.

**Acknowledgements**

We acknowledge TCGA database for providing their platforms and contributors for uploading their meaningful datasets.

**References**


Figures
Figure 1

Heatmap of the enrichment scores of 4922 hallmark and immunologic gene sets in tumor and normal samples. Gene set variation analysis (GSVA) was conducted based on The Cancer Genome Atlas (TGCA)-lung adenocarcinoma (LUAD) dataset.
Figure 2


Figure 3

Analysis of the tumor immune microenvironment of two subtypes. A: Comparison of the infiltration levels of 22 types of immune cells between two subtypes. *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. B: Comparison of the immune scores between two subtypes. C: Comparison of the immune scores between two subtypes. D: Comparison of the stromal score between two subtypes. E: Comparison of the ESTIMATE score between two subtypes. F: Comparison of the tumor purity between two subtypes.
Figure 4

Identification of subtype-related differential gene sets (DGSs) and construction of a prognostic risk score model. A: Identification of subtype-related DGSs. B: Least absolute shrinkage and selection operator (LASSO) coefficient distribution. C: Likelihood deviation of LASSO coefficient distribution and the two dotted vertical lines represent lambda.min (left, red line) and lambda.1se (right, black line). D: Forest plots of univariate COX regression analysis of five gene sets. E: Kaplan–Meier (KM) survival analyses of the...
association of enrichment scores of five gene sets and OS. F: Forest plots of multivariate COX regression analysis of two gene sets in the prognostic risk score model.

Figure 5

The prognostic risk score model has good efficiency in predicting prognosis. A: Risk scores were calculated in samples in TCGA-LUAD dataset. B: KM survival analysis of high-risk and low-risk groups based on TCGA-LUAD dataset. C: Reactive oxygen species (ROC) analysis of the value of risk score in predicting the survival of patients in TCGA-LUAD dataset. D: Risk scores were calculated in samples in GEO dataset. E: KM survival analysis of high-risk and low-risk groups based on GEO dataset. F: ROC analysis of the value of risk score in predicting the survival of patients in GEO dataset.
Figure 6

The risk score was an independent prognostic factor and nomogram was constructed based on the risk scores and clinical factors. A: Forest plots of univariate Cox regression analysis. B: Forest plots of multivariate Cox regression analysis. C: The constructed nomogram can predict the survival of patients with lung cancer. D: Calibration curve of nomogram for predicting the 1-, 3-, and 5-year survival probabilities. X-axis represents the nomogram-predicted probability and y-axis represents the actual survival.
Figure 7

The risk score was positively correlated with the tumor mutation burden (TMB). A: The mutational landscape of top 20 significantly mutated genes in different risk groups. B: Comparison of TMB levels of high- and low-risk groups. C: Correlation analysis of TMB levels with risk score.