Observational Study

Medicine

OPEN

Bioinformatics algorithm for lung adenocarcinoma based on macropinocytosis-related long noncoding RNAs as a reliable indicator for predicting survival outcomes and selecting suitable anti-tumor drugs

Hang Chen[a], Shuguang Xu[a], Zeyang Hu[a], Yiqing Wei[a], Youjie Zhu[b], Shenzhe Fang[a], Qiaoling Pan[a], Kaitai Liu[a], Ni Li[a], Linwen Zhu[a], Guodong Xu[a,*]

Abstract

As a highly conserved endocytic mechanism during evolution, macropinocytosis is enhanced in several malignant tumors, which promotes tumor growth by ingesting extracellular nutrients. Recent research has emphasized the crucial role of macropinocytosis in tumor immunity. In the present study, we established a new macropinocytosis-related algorithm comprising molecular subtypes and a prognostic signature, in which patients with lung adenocarcinoma (LUAD) were classified into different clusters and risk groups based on the expression of 16 macropinocytosis-related long noncoding RNAs. According to the molecular subtypes, we discovered that patients with LUAD in cluster1 had a higher content of stromal cells and immune cells, stronger intensity of immune activities, higher expression of *PD1*, *PDL1*, and *HAVCR2*, and a higher tumor mutational burden, while patients in cluster2 exhibited better survival advantages. Furthermore, the constructed prognostic signature revealed that low-risk patients showed better survival outcomes, earlier tumor stage, higher abundance of stromal cells and immune cells, higher immune activities, higher expression of *PD1*, *PDL1*, *CTLA4*, and *HAVCR2*, and more sensitivity to Paclitaxel and Erlotinib. By contrast, patients with high scores were more suitable for Gefitinib treatment. In conclusion, the novel algorithm that divided patients with LUAD into different groups according to their clusters and risk groups, which could provide theoretical support for predicting their survival outcomes and selecting drugs for chemotherapy, targeted therapy, and immunotherapy.

Abbreviations: CTLA-4 = cytotoxic T-lymphocyte associated protein 4, EGFR = epidermal growth factor receptor, EGFR-TKI = EGFR tyrosine kinase inhibitor, FDR = false discovery rate, GO = gene ontology, HAVCR-2 = hepatitis A virus cellular receptor 2, HRs = hazard ratios, IC_{50} = median inhibitory concentration, ICIs = immune checkpoint inhibitors, LASSO = least absolute shrinkage and selection operator, LUAD = lung adenocarcinoma, PD-1 = programmed cell death 1, PD-L1 = programmed cell death 1 ligand 1, qRT-PCR = quantitative real-time reverse transcription polymerase chain reaction, ssGSEA = single-sample gene-set enrichment analysis, TCGA = The Cancer Genome Atlas, TIME = tumor immune microenvironment, TMB = tumor mutation burden, VEGFA = vascular endothelial growth factor A.

Keywords: consensus cluster, lung adenocarcinoma, macropinocytosis, prognostic signature, tumor immunity

1. Introduction

As the most common malignant tumor, lung cancer is the leading cause of cancer-related deaths worldwide.[1] According to "Cancer Statistics, 2021" estimated by the American Cancer Society, among men and women, the incidence of lung cancer ranks second, and the mortality rate ranks first.[2] Furthermore, the National Central Cancer Registry of China estimated that lung cancer ranked first in morbidity and mortality of all malignant tumors in 2015.[3] Approximately 85% of lung cancer is

HC, SX, ZH, YW, YZ, SF, and QP contributed equally to this work.

The study was funded by Basic public welfare project of Ningbo (Grant No. 2019C50041), The Natural Science Foundation of Zhejiang (Grant No. LY22H160004), and The Natural Science Foundation of Ningbo. (Grant No. 2021J278)

Written informed consent for publication was obtained from all participants.

The authors declare that they have no conflicts of interest.

The datasets generated during and/or analyzed during the current study are publicly available. All data analysed during the current study are accessible from the TCGA database (https://portal.gdc.cancer.gov/).

The study was not applicable for ethical approval and consent to participate, in that the patients included in this study were all obtained from TCGA database.

[a]Department of Cardiothoracic Surgery, The Affiliated Lihuili Hospital, Ningbo University, Ningbo, Zhejiang, China, [b] Department of Respiratory Medicine, Ningbo Huamei Hospital, Chinese Academy of Sciences, Ningbo, Zhejiang, China.

**Correspondence: Guodong Xu, Department of Cardiothoracic Surgery, The Affiliated Lihuili Hospital, Ningbo University, Ningbo, Zhejiang, China (e-mail: xuguodong@nbu.edu.cn).*

How to cite this article: Chen H, Xu S, Hu Z, Wei Y, Zhu Y, Fang S, Pan Q, Liu K, Li N, Zhu L, Xu G. Bioinformatics algorithm for lung adenocarcinoma based on macropinocytosis-related long noncoding RNAs as a reliable indicator for predicting survival outcomes and selecting suitable anti-tumor drugs. Medicine 2022;101:38(e30543).

Received: 7 April 2022 / Received in final form: 9 August 2022 / Accepted: 10 August 2022

http://dx.doi.org/10.1097/MD.0000000000030543

classified clinically as nonsmall cell lung cancer (NSCLC), and approximately 60% is classified pathologically as lung adenocarcinoma (LUAD).[4] Recently, although breakthroughs in targeted therapy and immunotherapy have improved the quality of life and prolonged the overall survival of patients with advanced NSCLC, the 5-year survival rate for patients with NSCLC with distant metastasis is merely 7%.[5] Therefore, screening for novel and effective biomarkers is very important for precision medicine.

Macropinocytosis, a nonspecific actin-dependent endocytosis, allows cells to take up extracellular cargo, including proteins, pathogens, and cell debris, into large vesicles.[6] Recently, the role of macropinocytosis in clearing apoptotic cells, immune surveillance, virus invasion, and renewal of cell membranes, has been further investigated.[7] Increasing evidence has demonstrated that macropinocytosis plays a crucial role in physiological or pathological process of immune cells. For instance, a review by Canton[8] introduced the sentinel effect of macropinocytosis on innate immune cells in detail. Besides, Holt et al[9] proposed macropinocytosis endows dendritic cells with augmented immune surveillance. Furthermore, Doodnauth et al[10] discovered that macrophages can absorb low-density lipoprotein through macropinocytosis to form foam cells, leading to atherosclerosis.

The potential relationship between macropinocytosis and the development of malignant tumors has been explored in recent years. Li et al[11] demonstrated that breast cancer lung metastatic cells swallowed lipids in neutrophils via macropinocytosis, which promoted the proliferative ability of cancer cells. In addition, evidence has proved the close relationship between macropinocytosis and resistance of lung cancer cells to common antitumor drugs. For example, Wang et al[12] proposed that the ATP-abundant tumor microenvironment of lung cancer was closely associated with drug resistance. In addition, Yamazaki et al[13] reported that lung cancer cells with epidermal growth factor receptor (EGFR) mutations swallowed collagen type I through macropinocytosis, leading to resistance to EGFR tyrosine kinase inhibitor (EGFR-TKI). However, Takenaka proposed that patients with NSCLC with mutations that were sensitive to gefitinib could effectively inhibit macropinocytosis, which increased the antitumor effect of doxorubicin significantly.[14]

Thus, macropinocytosis in the tumor immune microenvironment (TIME) is closely related to the development and drug resistance of lung cancer, which seriously hinders the treatment of patients with lung cancer and affects their prognosis significantly. In the present study, we aimed to establish a macropinocytosis-related algorithm to classify patients with LUAD into different groups. We then assessed the TIME, prognosis, and median inhibitory concentration (IC_{50}) of common antitumor drugs using bioinformatic methods. Our findings and the constructed macropinocytosis-related algorithm represent a huge step forward for individualized precision medicine in lung cancer.

2. Materials and methods

2.1. Data acquisition

We followed the procedures shown in the flowchart (Fig. 1) to classify patients with LUAD into different clusters and risk groups, to provide a solid theoretical basis for more precise treatment. First, we acquired the transcriptome data, single nucleotide variation, and corresponding clinical information of patients with LUAD patients from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/)[15] database. The gene transfer format files were utilized to annotating long noncoding RNAs (lncRNAs) and mRNAs. Subsequently, 55 macropinocytosis-related genes (mrgenes) were obtained by searching for "macropinocytosis cancer" via PubMed (https://pubmed.ncbi.nlm.nih.gov/).[16] Then, the macropinocytosis-related lncRNAs (mrlncRNAs) were identified by Spearman correlation analysis between the mrgenes and lncRNAs ($|cor| > 0.4$ and $P < .001$). Subsequently, univariate Cox analysis was performed to filter the mrlncRNAs that were closely related to the prognosis of patients with LUAD ($P < .01$), which were prepared for subsequent modeling. The forest map, heatmap, and network diagram exhibited the P values, hazard ratios (HRs), expression, and correlation of mrlncRNAs related to prognosis, respectively, in which the former 2 maps were generated by the R software version x64-4.0.4 (https://www.R-project.org/), and the latter diagram were plotted using cytoscape version 3.8.2.[17]

2.2. Cluster analysis of mrlncrnas

Consensus clustering was performed using R-x64-4.0.4 ConsensusClusterPlus[18] package to classify patients with LUAD into different molecular subtypes, and clusters were set according to the optimal k value. A clinical heatmap was used to show the expression of mrlncRNAs related to prognosis, and illustrated the correlation between the clusters and clinicopathological characteristics. To explore the differences in survival outcomes of patients with LUAD between different clusters, we plotted a survival curve for visualization using Kaplan–Meier survival analysis. Furthermore, a correlation plot was generated to clearly exhibit the correlation coefficient between the 55 mrlncRNAs.

2.3. Analyses of tumor immunity in different molecular subtypes

Differential expression analyses were conducted between patients from different clusters to identify differentially expressed genes (false discovery rate < 0.05), which were evaluated by gene ontology (GO) functional enrichment analysis ($P < .05$ and false discovery rate < 0.05) to identify the top 10 enriched functional phenotypes in biological process, cellular component, and molecular function, respectively. According to the results of GO enrichment analysis, we speculated that the constructed molecular subtypes were closely related to tumor immunity. As calculated using the R-x64-4.0.4 estimate package, the StromalScore, ImmuneScore, and ESTIMATEScore (StromalScore + ImmuneScore) were recognized as reliable indicators to evaluate the tumor immune microenvironment, the differences of which in different molecular subtypes were explored using the Wilcoxon rank-sum test. The relative content of common immune cells and the relative intensity of immune activities were quantified using single-sample gene-set enrichment analysis (ssGSEA) algorithm, which were visualized using 2 multi-box plots. Wilcoxon rank-sum tests were employed to investigate whether the expression of common immune checkpoint inhibitors (ICIs) in patients with LUAD was statistically significantly different between the different clusters. The tumor mutation burden (TMB) was defined as the total number of substitutions, insertion mutations, and deletion mutations per million bases, which was considered as a reliable indicator to assess the efficacy of anti-programmed cell death 1 (PD-1) therapy.[19] The difference in TMB between different molecular subtypes was explored using Wilcoxon rank-sum tests.

2.4. Construction of the risk assessment model

Patients with LUAD were divided evenly into 2 groups randomly, and named the training group and the testing group. Least absolute shrinkage and selection operator (LASSO) regression analysis was performed on the training group to construct the ideal risk assessment model, and the mrlncRNAs included by the LASSO regression analysis and their corresponding

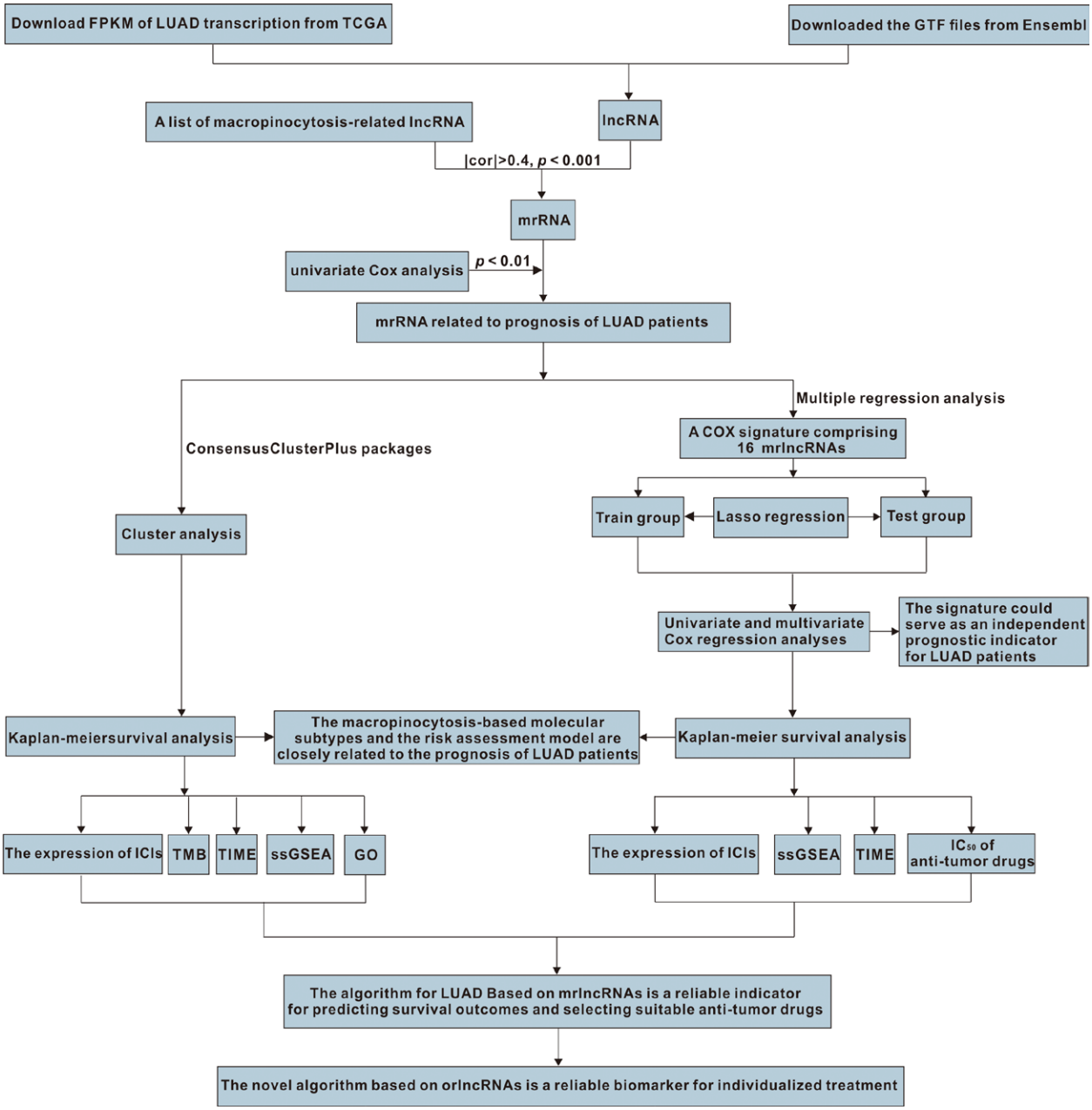

Figure 1. The procedure for establishing the macropinocytosis-based algorithm and exploring its clinical significance.

regression coefficients were exhibited in a table. The risk score of each patient in the training group was calculated based on the following formula:

$$\text{Risk score} = \sum_{i=1}^{16} \text{Coef(i)} \times \text{E(i)},$$

Coef (i) (Table 1) and E(i) represent the regression coefficients and the expression levels of the included mrlncRNAs, respectively. The median risk score in the training group was used as the cut-off point to classify the patients into high- and low-risk groups. Subsequently, 1-year receiver operating characteristic curves of the 2 risk groups were plotted and the corresponding area under the curve values were calculated. Then, survival curves for the 2 groups were generated using Kaplan–Meier survival analyses to explore whether the survival outcomes of patients with LUAD patients in the different risk groups were

Table 1

The regression coefficients of mrlncRNAs included by the lasso regression.

ID	Coef
CARD8-AS1	−0.000851001072322576
LINC00941	0.060640793514365
LINC01137	0.022200771858516
LINC01116	0.0349184545656507
AC010980.2	0.0691655030661195
LINC00324	−0.158673264800938
AL365203.2	0.0358159581572812
AL606489.1	0.0412286187481897
AC004687.1	−0.00643358868016261
HLA-DQB1-AS1	−0.0239731776316693
AL590226.1	−0.0736257017716491

statistically different. Furthermore, scatter plots were generated to visualize the potential relationship between the risk score and survival status. The heatmaps revealed the expression levels of the included mrlncRNAs in the patients in the 2 risk groups. To investigate whether the constructed prognostic signature could act as an independent prognostic indicator for patients with LUAD, forest maps were plotted using univariate and multivariate Cox regression analyses for validation. The clinical heatmap clearly exhibited the clinicopathological characteristics in different clusters and different risk groups, which were marked as: ***<0.001, **<0.01, and *<0.05. A series of box plots generated by Wilcoxon rank-sum tests were used to visualize the statistically different clinicopathological characteristics.

2.5. Exploration of the tumor immunity related to the prognostic signature

To better illustrate the potential relationship between the prognostic signature and the expression levels of common ICIs (e.g., *PD1*, *PDL1* (encoding programmed cell death 1 ligand 1), *CTLA4* (encoding cytotoxic T-lymphocyte associated protein 4), and *HAVCR2* (encoding Hepatitis A virus cellular receptor 2)), Wilcoxon rank-sum tests were used to analyze whether there were statistical differences in different risk groups. Subsequently, the types and abundance of immune cells were further explored using indicators of TIME and ssGSEA analyses, in which the statistical differences between different risk groups were detected by Wilcoxon rank-sum tests. In addition, the IC_{50} values of common chemotherapy and targeted drugs (e.g., Paclitaxel, Erlotinib, and Gefitinib) for lung cancer treatment were calculated using the pRRophetic[20] package, and compared using Wilcoxon rank-sum tests between different risk groups. A Sankey diagram clearly visualized the potential relationship between molecular subtypes and risk assessment models. Finally, Kaplan–Meier survival analysis was employed to evaluate the overall survival of patients with LUAD in different clusters and different risk groups.

3. Results

3.1. Identification of mrlncRNAs related to prognosis

From the TCGA, we obtained the data for 54 normal tissues and 497 LUAD tissues, differentiated 14,087 lncRNAs by gene transfer format files, and searched 55 mrgenes using PubMed. We filtered 1928 mrlncRNAs by coexpression analysis and identified 32 mrlncRNAs that were related to the prognosis of patients with LUAD using univariate Cox analysis, which were utilized for subsequent modeling. We found that a majority of the 32 mrlncRNAs were protective genes (Fig. 2A), and the 32 mrlncRNAs correlated with 14 mrgenes (Fig. 2B). Furthermore, 32 mrlncRNAs were all differential expressed between LUAD tissues and adjacent tissues (Fig. 2C).

3.2. Construction of macropinocytosis-based molecular subtypes

According to the results of Consensus clustering, we divided the patients with LUAD patients into 2 clusters (Fig. 3A–C). A heatmap was used to visualize the expression levels of the 32 mrlncRNAs between the 2 clusters, which demonstrated clearly that the majority of the 32 mrlncRNAs were significantly highly expressed in cluster2 compared with that in cluster1 (Fig. 3D). The survival curve revealed the survival outcomes of patients with LUAD in the different clusters, which suggested that compared with the patients in cluster1, the patients in cluster2 exhibited a relative better survival advantage (Fig. 4A). The correlation plot showed that the majority of the 32 mrlncRNAs were positively correlated (Fig. 4B).

3.3. Significance of the macropinocytosis-based molecular subtypes

The functional phenotypes of the molecular subtypes were further investigated using GO function enrichment analysis, which indicated that the differentially expressed genes in the different clusters were enriched in several immune responses, including antimicrobial humoral immune response and antimicrobial humoral response (Fig. 4C). According to the analysis of the TIME-related indicators, we concluded that the abundance of stromal cells and immune cells around the tumor in cluster1 was relatively higher than that in cluster2 (Fig. 4D–F), which was validated by the results of ssGSEA, in which the relative content of common immune cells was relatively higher, and the relative intensity of common immune activities was relatively stronger in cluster1 than in cluster2 (Fig. 5A and B). The expression analyses indicated that the TMB (Fig. 5F) and the expression levels of common ICIs, including *PD1* (Fig. 5C), *PDL1* (Fig. 5D), and *HAVCR2* (Fig. 5E), in cluster1 were significantly higher than those in cluster2. In summary, the results above suggested that the patients with LUAD in cluster1 had relatively active tumor immunity, which rendered them more sensitive to immunotherapy targeting PD-1, PD-L1, and HAVCR-2.

3.4. Establishment of the risk assessment model

We assigned the patients with LUAD into a training group and a testing group randomly, and performed LASSO regression analysis on the patients in training group to construct the ideal prognostic signature, in which 16 mrlncRNAs were included in the modeling process (Fig. 5G and H). The 1-year area under the curve of the training group and testing group were 0.803, and 0.749, respectively, which suggested that the constructed signature had a good predictive ability after verification (Fig. 6A and B). Subsequently, the survival curves and scatter plots of the training group and testing group revealed and validated the relatively better survival advantage of low-risk patients (Fig. 6C–F). The 2 heatmaps showed the expression levels of the 16 mrlncRNAs in the training group and testing group (Fig. 7A,B). Forest maps constructed using the results of univariate and multivariate Cox regression analyses proved that tumor stage and the risk score could act as independent prognostic indicators for patients with LUAD after verification (Fig. 7C–F). The clinical heatmap revealed that the risk groups were closely correlated with N, T, stage, ImmuneScore, and cluster (Fig. 8A). Furthermore, the patients with LUAD in the advanced N stage (N1–3, Fig. 8B, $P < .01$), T (T3–4, Fig. 8C, $P < .01$), and stage (stage III–IV, Fig. 8D, $P < .001$) had a relatively higher risk scores, while the risk scores of patients in the low ImmuneScore group (Fig. 8E, $P < .001$) and cluster1 (Fig. 8F, $P < .001$) were relatively higher.

3.5. Exploration of significance in the treatment for patients with LUAD

The National Comprehensive Cancer Network Guidelines (NCCN Guidelines)[21] advocate that a comprehensive treatment integrating chemotherapy (including Paclitaxel), targeted therapy (including Erlotinib and Gefitinib), and immunotherapy (including anti-PD-1, and anti-PD-1 therapies) would be of great significance to treat patients with NSCLC. The expression analyses of common ICIs indicated that the expression of *PD1* (Fig. 8G), *PDL1* (Fig. 9A), *CTLA4* (Fig. 9B), and *HAVCR2* (Fig 9C) in the low-risk

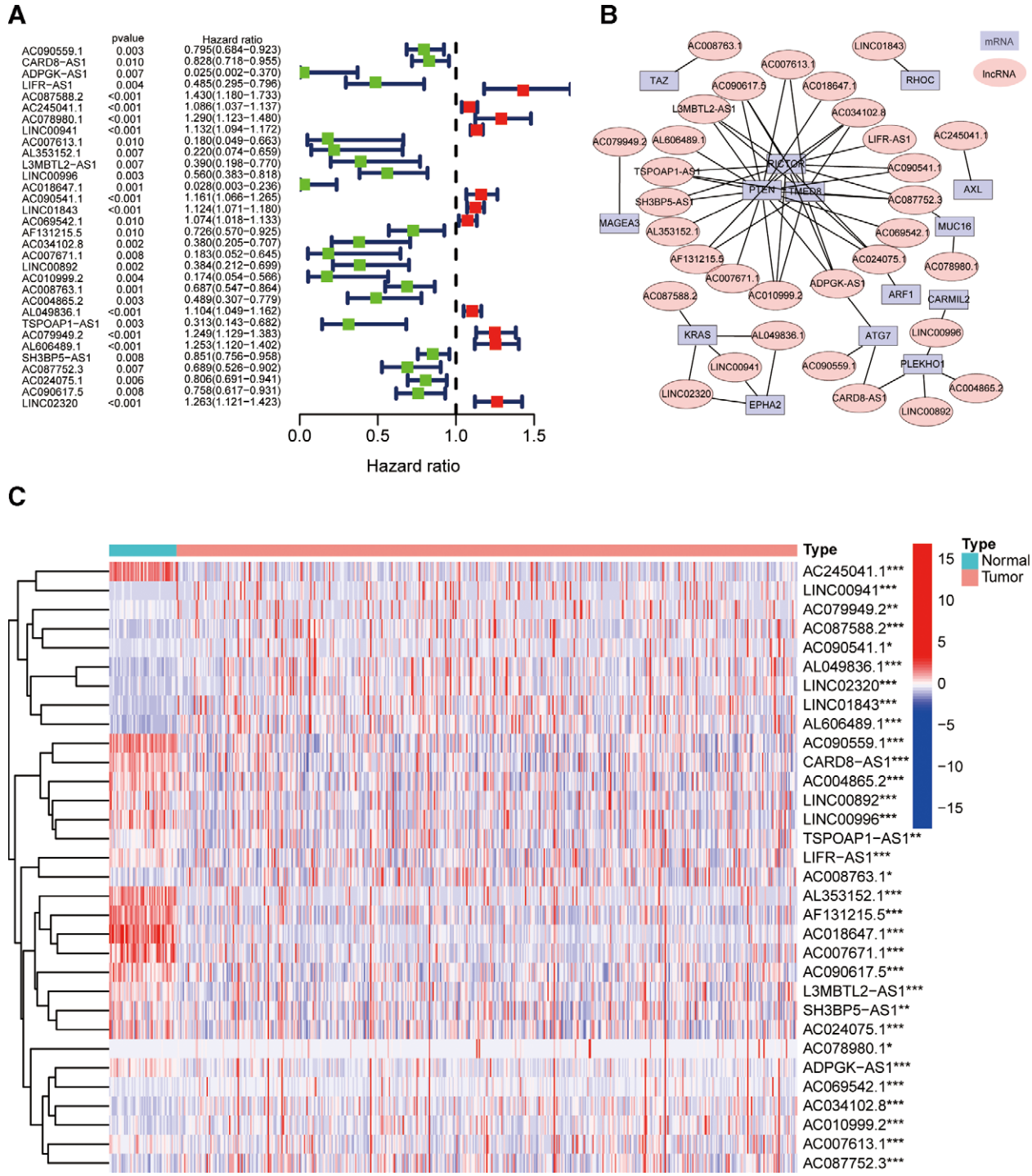

Figure 2. Identification of mrlncRNAs related to survival outcomes of patients with LUAD. (A) Univariate Cox analysis screening of 32 mrlncRNAs closely associated with the prognosis of patients with LUAD, in which red cubes represent risk mrlncRNAs, while green cubes represent protective mrlncRNAs. (B) A network diagram visualizing the connection between the mrgenes and mrlncRNAs, in which the blue cubes represent mrgenes, while the red circles represent mrlncRNAs. (C) The 32 mrlncRNAs are all differentially expressed in LUAD tissues and normal tissues. LUAD = lung adenocarcinoma.

group were significantly higher than those in the high-risk group, which could provide the theoretical basis for clinical immunotherapy for patients with LUAD. Moreover, we determined that the contents of stromal cells and immune cells in the low-risk patients was higher than those in the high risk patients, based on the results of TIME analysis (Fig. 9D–F), which was validated by ssGSEA, i.e., common immune cells were relatively abundant, and common immune activities were relatively active (Fig. 9G and H). The results of IC_{50} analysis using the pRRophetic package showed that the low-risk patients were more sensitive to Paclitaxel (Fig. 10A) and Erlotinib (Fig 10B), while the patients in the high risk group were relatively sensitive to Gefitinib (Fig. 10C). The Sankey diagram indicated that a majority of high-risk patients were in cluster1, and a majority of low risk patients were in cluster2 (Fig. 10D). Subsequently, the survival curve suggested that high-risk patients in cluster1 had statistically significant worse survival outcomes than those in cluster2 (Fig. 10E).

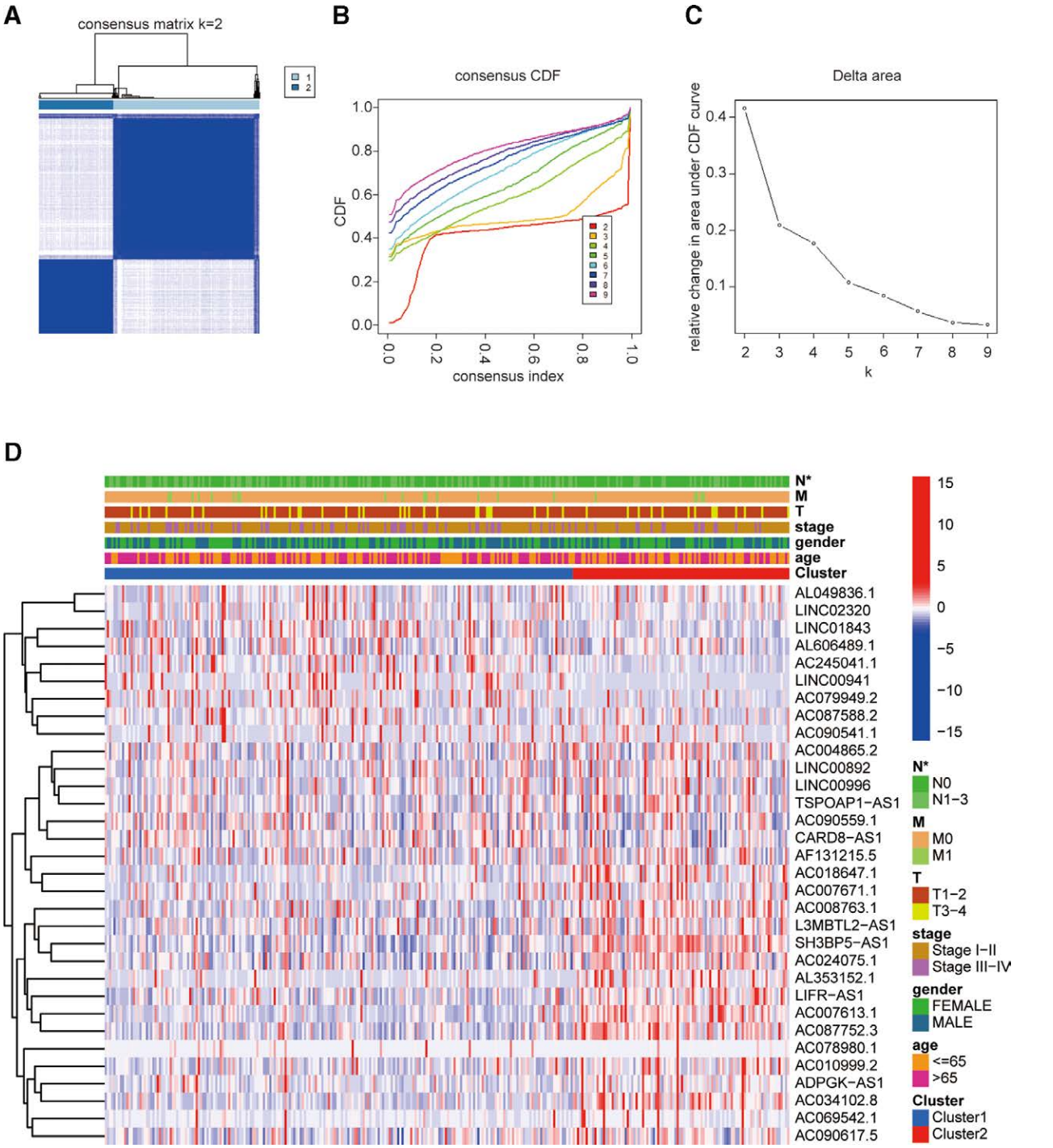

Figure 3. Cluster analysis based on the 32 mrlncRNAs. (A–C) In the process of generating the consensus matrix, when the *k* value was equal to 2, the optimal molecular subtypes were obtained, in which the slope of the cation diffusion facilitator (CDF) was the lowest, and the delta area under the CDF curve was the largest. (D) A heatmap showing the expression levels of the 32 mrlncRNAs in different molecular subtypes, which indicates that the molecular subtypes are closely related to the N stage.

4. Discussion

Recently, advances in bioinformatic analyses have greatly contributed to the emergence of prognostic signatures for patients with cancer,[22] using which doctors can easily predict the prognosis, tumor immunity, and even the IC_{50} of antitumor drugs for patients with cancer merely by detecting the expression of a few mRNAs or lncRNAs.

In the present study, we constructed an algorithm, i.e., a combination of molecular subtypes and a risk assessment model, to classify patients LUAD into 4 groups: cluster1 + low risk, cluster1 + high risk, cluster2 + low risk, and cluster2 + high risk, in which the prognosis, tumor immunity, and IC_{50} of antitumor drugs were assessed in patients with LUAD in the different groups. The results indicated that in terms of molecular subtypes, patients with LUAD in cluster1 had a higher content of stromal cells and immune cells, stronger intensity of immune activities, higher expression of *PD1*, *PDL1*, and *HAVCR2*, and a higher TMB than those in cluster2. The patients in cluster2 exhibited a better survival

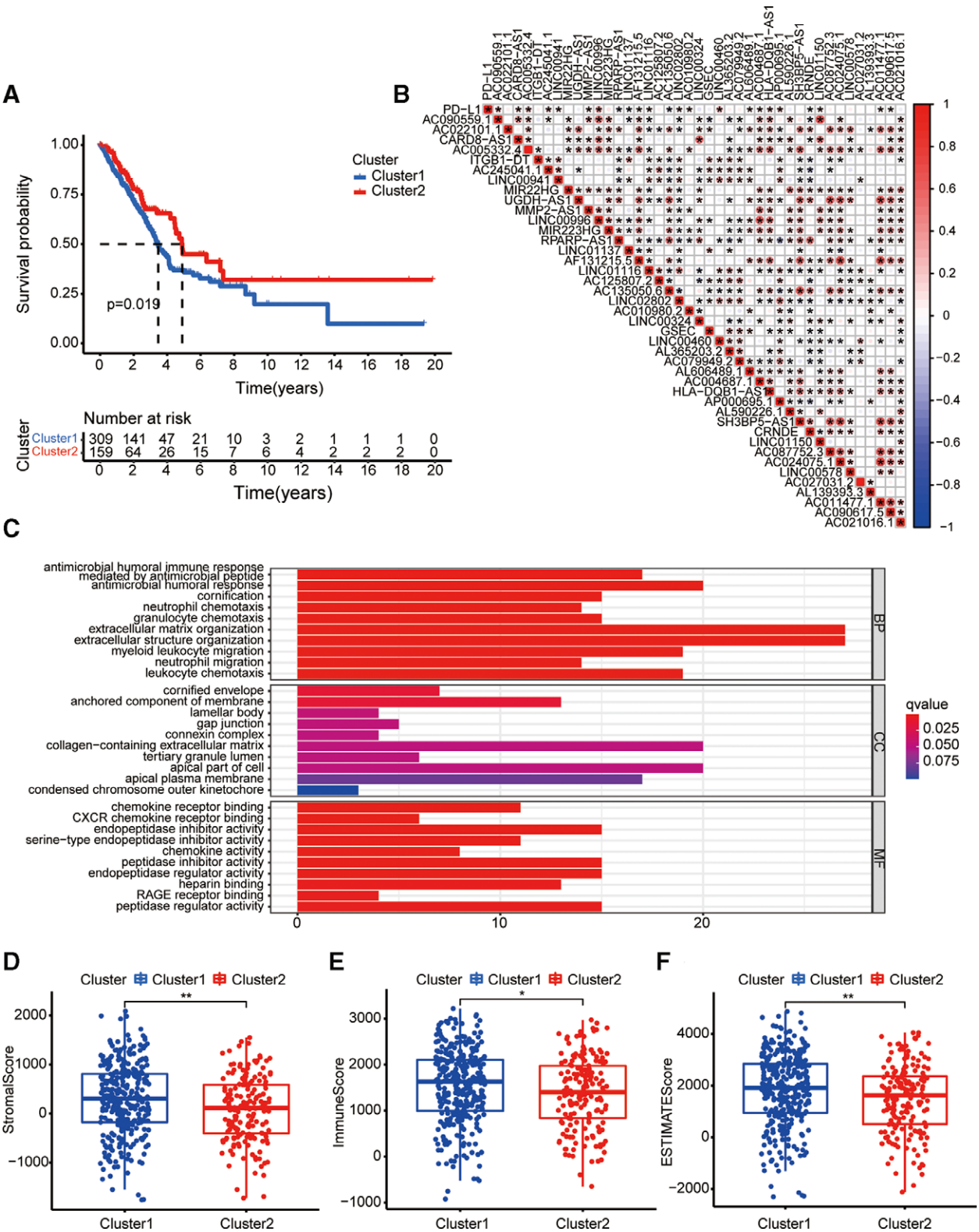

Figure 4. The clinical significance of the molecular subtypes. (A) The patients in cluster1 always have a poor prognosis. (B) The majority of the 32 mrlncRNAs are positively correlated. (D) The differential expressed genes in different clusters are enriched in several immune responses, including antimicrobial humoral immune response and antimicrobial humoral response. (D–F) The abundance of stromal cells (D) and immune cells (E) around the tumor in cluster1 is relatively higher than that in cluster2 (F).

advantage than those in cluster1. According to the LASSO regression-constructed prognostic signature, low-risk patients showed better survival outcomes, earlier tumor stage, higher abundance of stromal cells and immune cells, more immune activities, higher expression of *PD1*, *PDL1*, *CTLA4*, and *HAVCR2*, and higher sensitivity to Paclitaxel and Erlotinib, while the patients with high scores were more suitable for Gefitinib treatment.

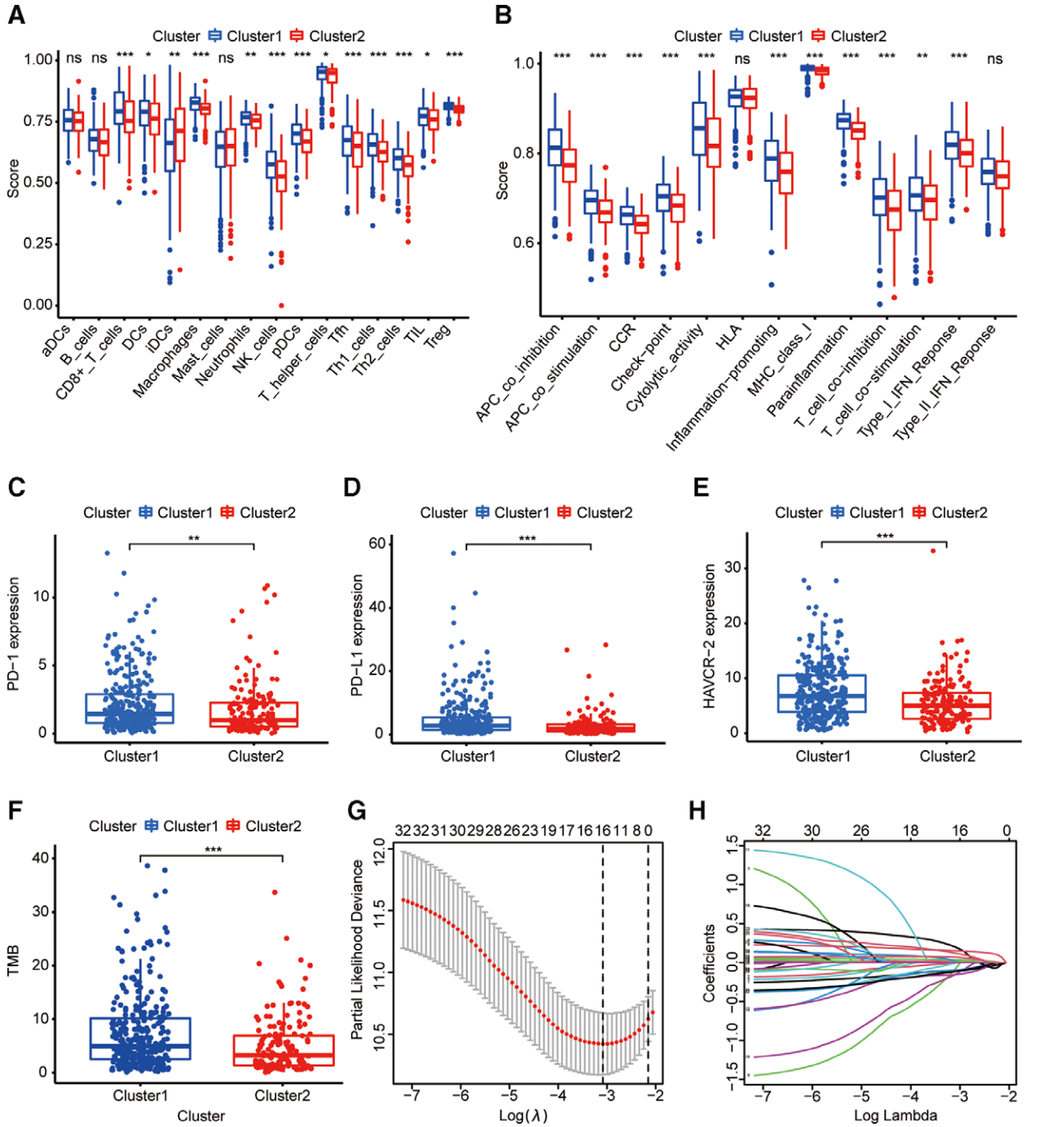

Figure 5. The clinical significance of the molecular subtypes and the construction of the 16 mrlncRNAs signature. (A,B) In cluster1, the relative content of common immune cells is relatively higher (A), and the relative intensity of common immune activities is relatively stronger (B) than in cluster2. (C–E) The expression levels of PD-1 (C), PD-L1 (D), and HAVCR-2 (E) in cluster1 are significantly higher than those in cluster2. (F) Compared with the patients with LUAD in cluster2, those in cluster1 tend to have a higher TMB. (G,H) The 2 plots exhibiting the process of LASSO regression. LUAD = lung adenocarcinoma, TMB = tumor mutation burden.

It is universally acknowledged that when conditions permit, surgery is still the first choice for patients with early stage NSCLC; however, patients with advanced lung cancer often have distant metastases and are therefore not suitable for surgery.[23,24] Currently, comprehensive treatments such as chemotherapy, targeted therapy, and immunotherapy are playing an irreplaceable role in treating patients with advanced LUAD.[25,26] Lung cancer is a very heterogeneous malignant tumor[27]; therefore, it is inadvisable for doctors to provide all patients with the same treatment plan. Thus, choosing the right treatment plan is of great significance to improve the overall quality of life of patients. In terms of clinical significance, the developed algorithm is a simpler, cheaper, and more reliable indicator than current mainstream genetic testing. For example, doctors are advised to provide low-risk patients with LUAD in cluster1 with anti-PD-1 and anti-PD-L1 immunotherapy, while high-risk patients with LUAD in cluster2 are not recommended for this treatment. Moreover, paclitaxel and erlotinib are recommended for low-risk patients with LUAD, while gefitinib is recommended for high-risk patients

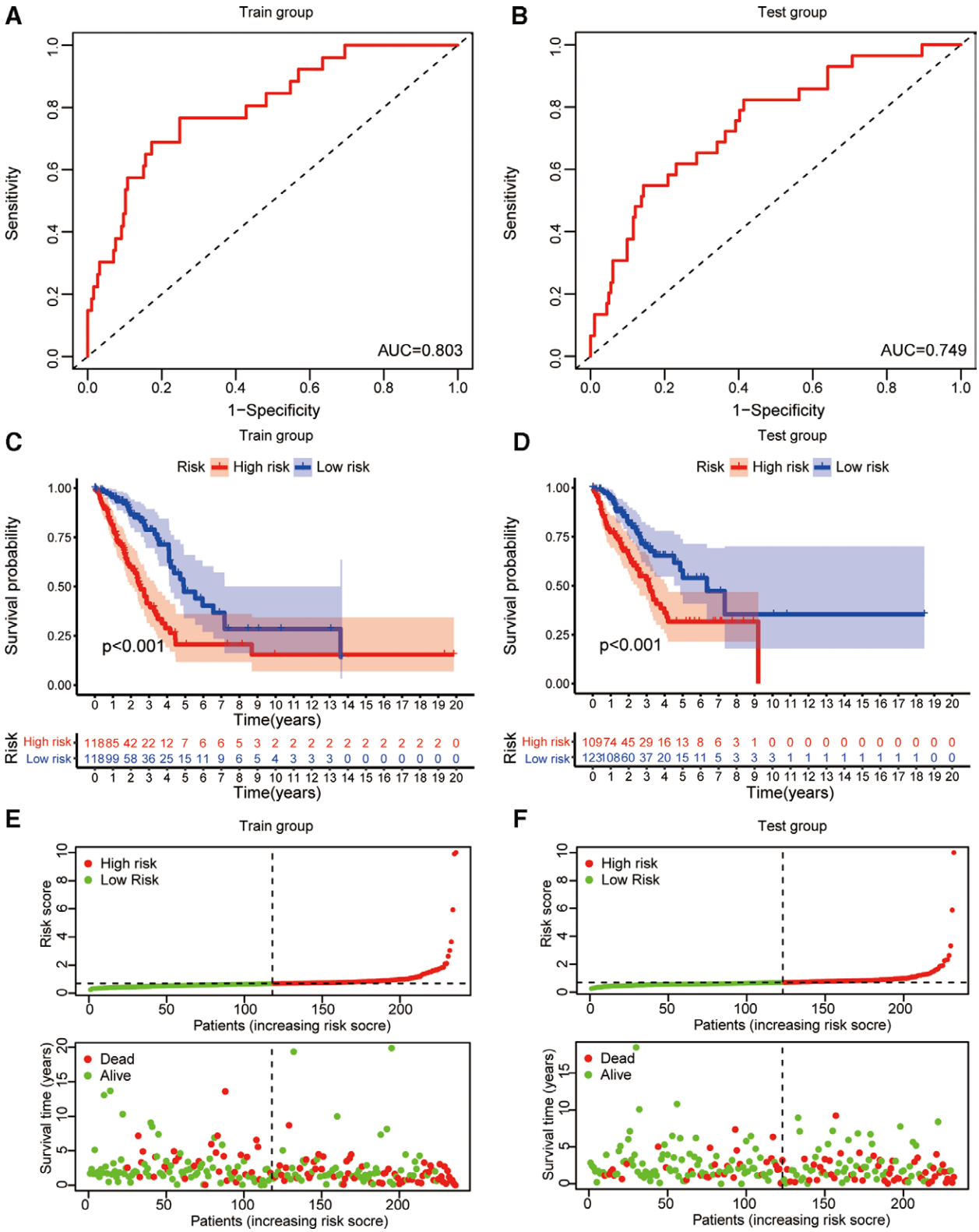

Figure 6. Evaluation of predictive ability of the macropinocytosis-related prognostic signature. (A,B) 1-year receiver operating characteristic curves indicating that the 1-year area under the curve of the training group and testing group were 0.803 (A), and 0.749 (B), respectively. (C–F) The survival curves and scatter plots of the training group (C and E) and the testing group (D and F) revealed and validated the relatively better survival advantages of low-risk patients.

with LUAD. Using the algorithm, we can select an effective treatment plan and provide personalized treatment for each patient with LUAD. In terms of survival outcomes, the LUAD patients in the low-risk group in cluster1 and cluster2 always showed a relatively better survival advantage, while the high-risk patients in cluster1 tended to have poor prognosis, with a 5-year survival rate of <25%.

Among the 16 mrlncRNAs identified by the LASSO regression analysis, several of them, including *AC090559.1*,[28,29] *CARD-8AS1*,[30] *LINC01843*,[31,32] *AF131215.5*,[33,34] *AC079949.2*,[35,36]

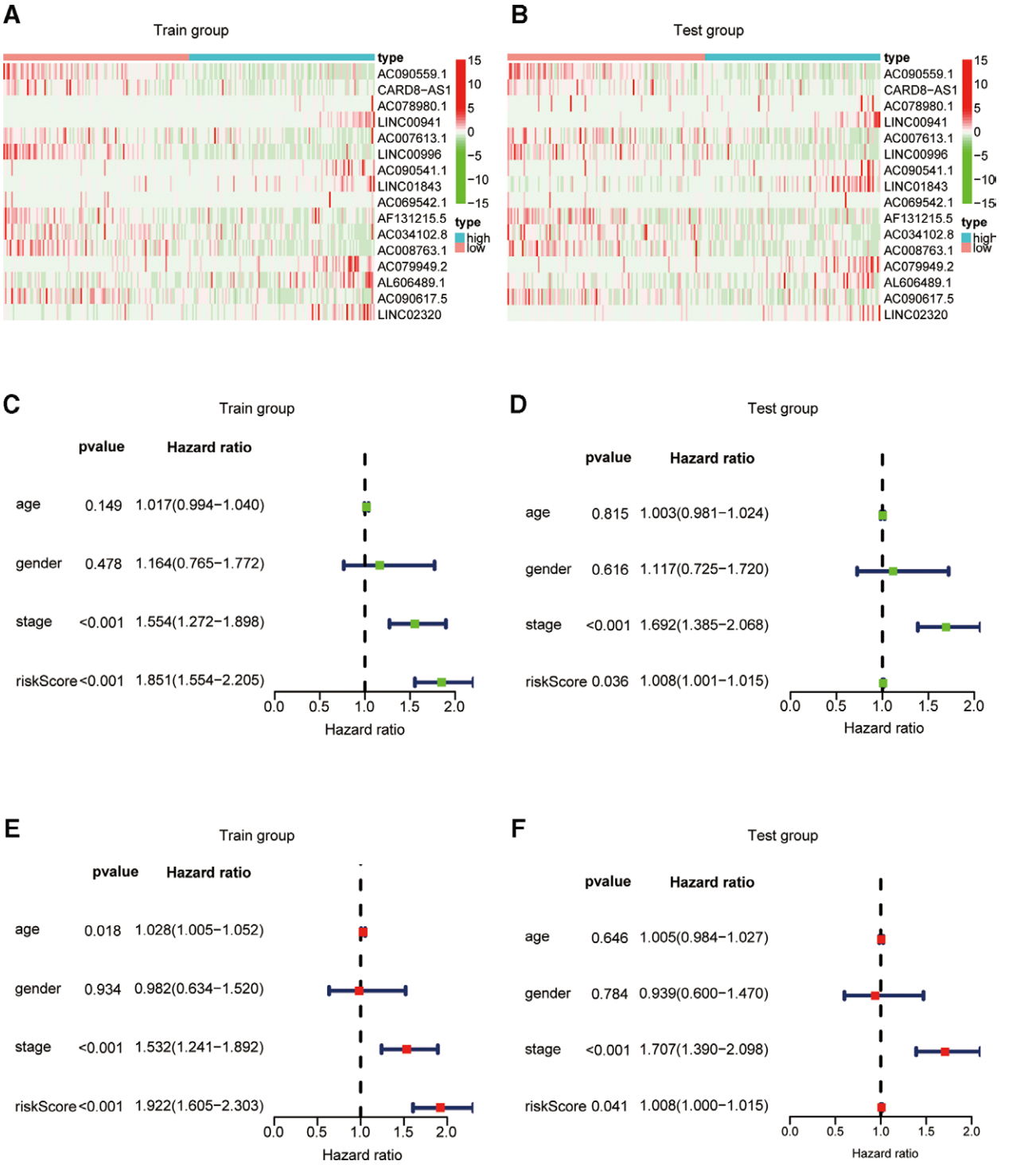

Figure 7. Evaluation of the predictive ability of the macropinocytosis-related prognostic signature. (A,B) The expression levels of the 16 mrlncRNAs in the training group (A) and testing group B). (C–F) Risk score (training group: hazard ratio [HR] = 1.851, confidence interval [CI] = 1.554–2.205, $P < .001$, testing group: HR = 1.008, [CI] = 1.001–1.015, $P = .036$), and stage (training group: HR = 1.554, [CI] = 1.272–1.898, $P < .001$, testing group: HR = 1.692, [CI] = 1.385–2.068, $P < .001$), could act as independent prognostic indicators for patients with LUAD after verification.

AL606489.1,[28,29] have been included in prognostic signatures for a variety of malignant tumors. Some of the other mrlncRNAs have been confirmed to be closely related to various malignant tumors. For instance, Ren et al[37] demonstrated that *LINC00941* promotes the progression of NSCLC by mediating the miR-877-3p/vascular endothelial growth factor A (VEGFA) axis. Yan et al[38] identified *LINC00996* as a potential therapeutic target by constructing a prognostic model and predicting downstream targets. However, other mrlncRNAs have not been reported, such as *AC078980.1*, *AC007613.1*, *AC090541.1*, *AC069542.1*, *AC034102.8*, *AC008763.1*, *AC090617.5*, and *LINC02320*, which could represent promising biomarkers for further research.

The present study was the first to construct a novel macropinocytosis-based algorithm using bioinformatics, in which we evaluated in detail the survival of patients with LUAD, tumor immunity, and drug sensitivity to common antitumor drugs. The constructed prognostic signature was subjected to

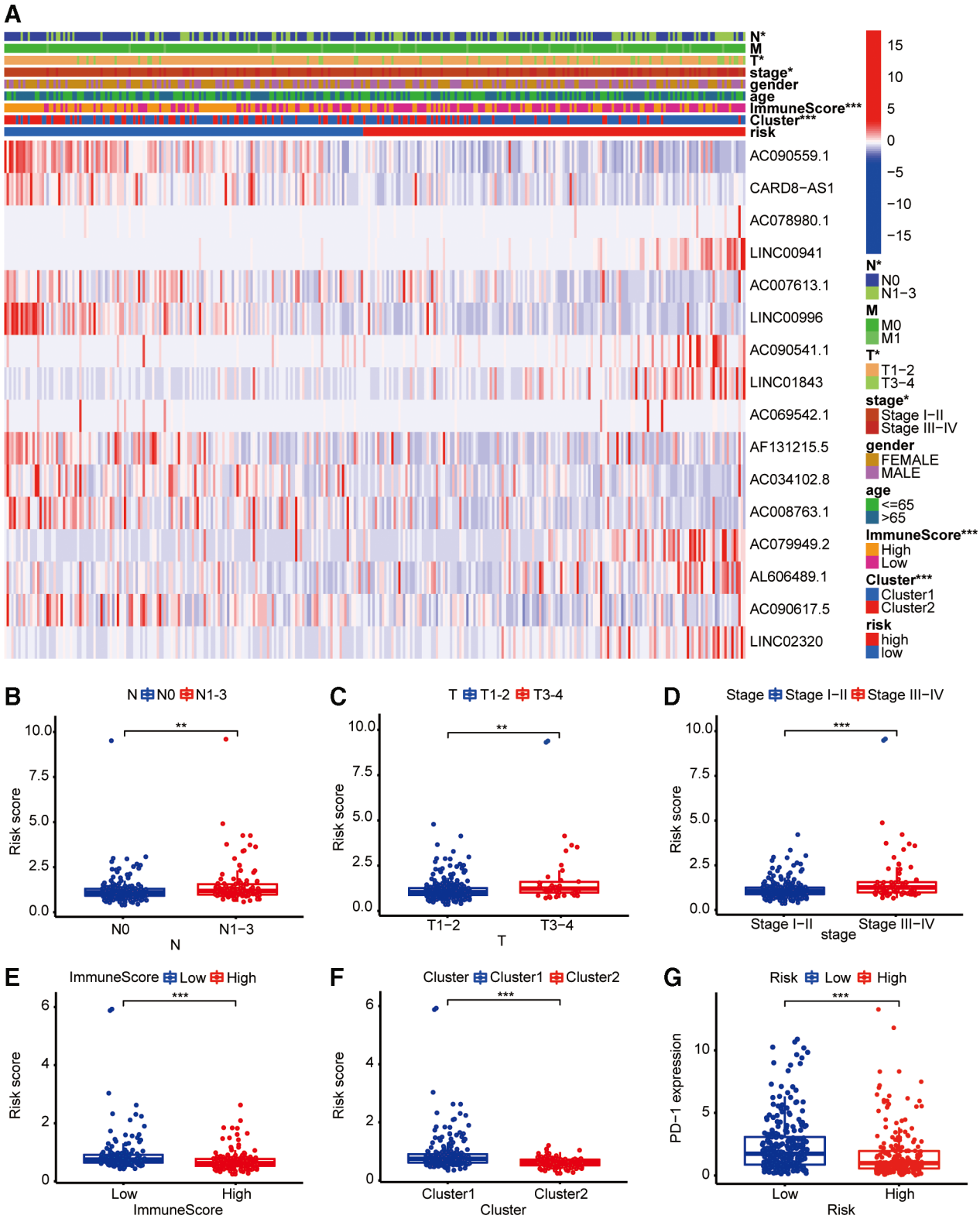

Figure 8. The potential relationship between the risk score and clinicopathological characteristics. (A) The risk groups are closely correlated with N, T, Stage, ImmuneScore, and Cluster. (B–F) The patients with LUAD with advanced stage of N (B), T (C), Stage (D) had a relatively higher risk score, while the risk scores of patients in the low-ImmuneScore group (E) and cluster1 (F) are relatively higher. (G) Patients in the low-risk group always have relatively higher expression of PD-1.

internal verification by dividing the patients with LUAD randomly into a training group and a testing group, which proved the reliability of the signature. Moreover, compared with the traditional prognostic signature that divided patients into 2 risk groups,[39] we divided the patients into 4 groups based on molecular subtypes and the prognostic signature, which could offer patients with LUAD more precise treatment. We included chemotherapy (Paclitaxel), targeted therapy (Erlotinib, Gefitinib),

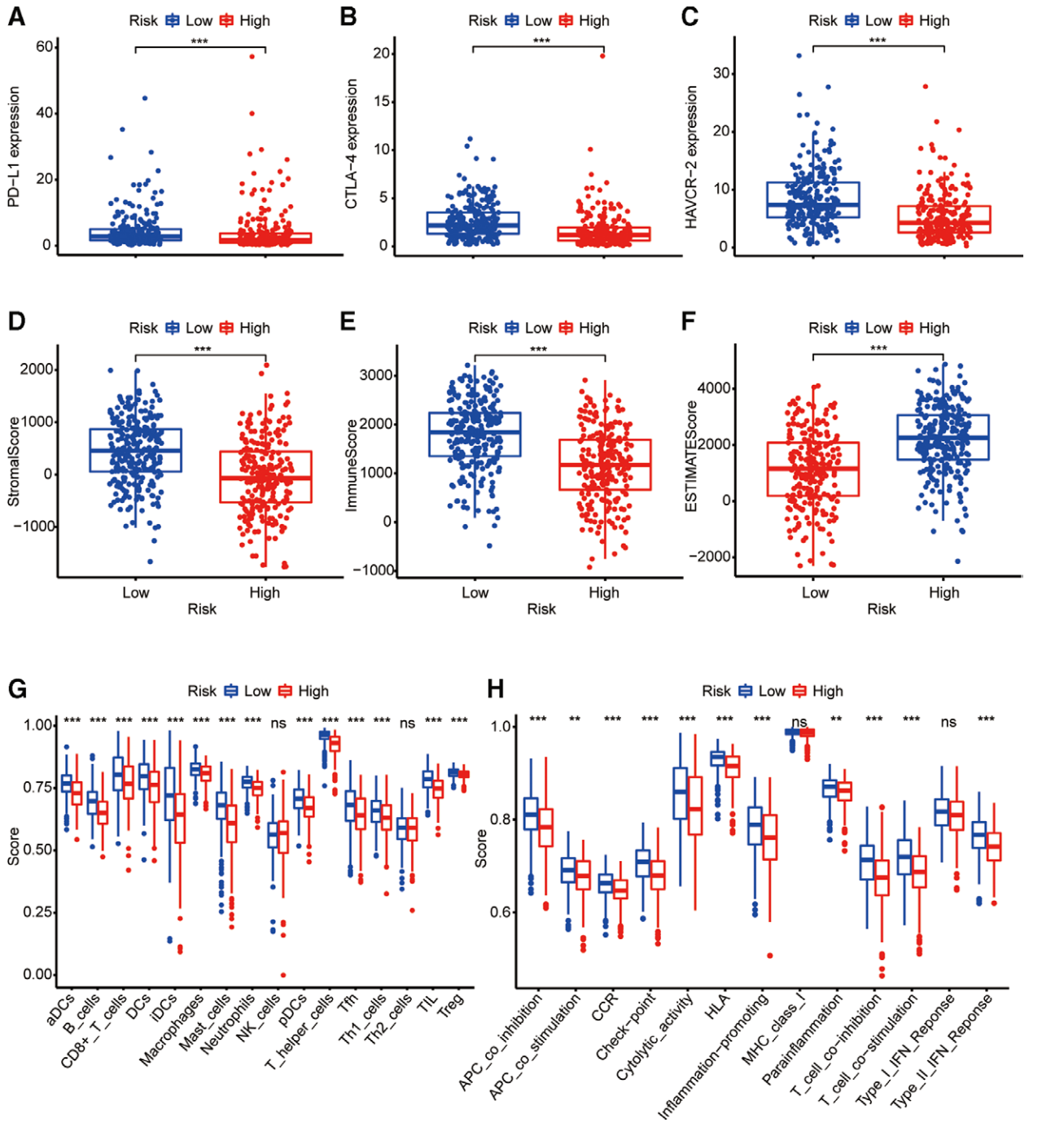

Figure 9. Exploration of tumor immunity of the 16 mrlncRNAs signature. (A–C) The expression levels of PD-L1 (A), CTLA-4 (B), and HAVCR-2 (C) in the low-risk group are significantly higher than those in the high-risk group. (D–F) The StromalScore (D), ImmuneScore (E), and ESTIMATEScore (F) in the low-risk patients are relatively higher than those in the high risk patients. (G, H) The scores of common immune cells (G) and immune activities (H) of low-risk patients are significantly higher than those of high risk patients.

and immunotherapy (PD-1, PD-L1, CTLA-4, and HAVCR-2 therapy), which are currently recommended for patients with advanced lung cancer.

However, our research had limitations. For instance, considering that the data was all downloaded from the same public database, we failed to eliminate bias from the profile analyzed. In addition, the analyses were limited to bioinformatic analyses of the TCGA database, and further external verification is needed, such as quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR), on clinical specimens to verify the reliability of the algorithm.

In conclusion, the novel algorithm based on mrlncRNAs demonstrated promising clinical predictive ability. The algorithm can easily predict the prognosis of patents with LUAD and select the appropriate treatment only by detecting the expression of 16 mrlncRNAs, which is of great significance for the clinical treatment of patients with advanced LUAD for whom surgery is not appropriate.

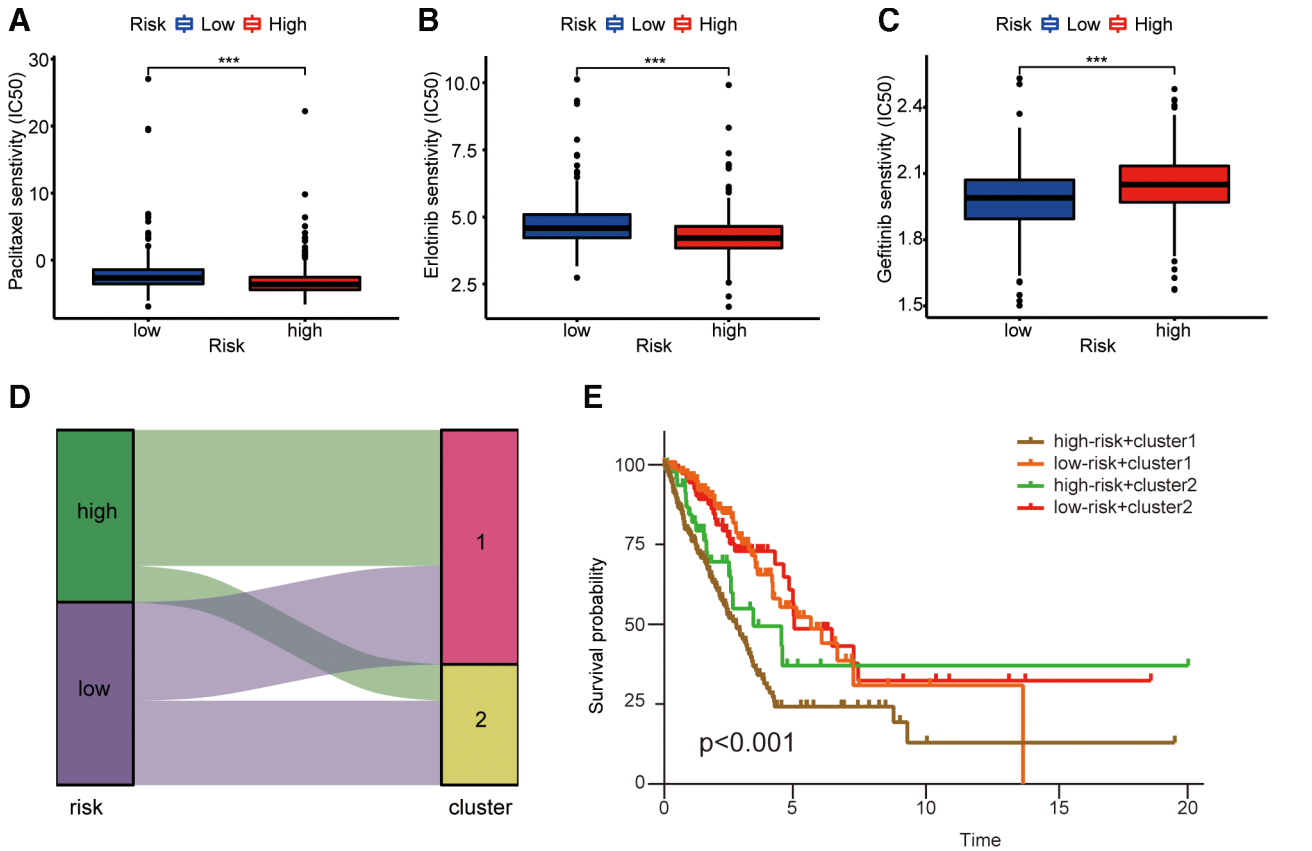

Figure 10. The clinical significance of the 16 mrlncRNAs signature. (A–C) The IC_{50} of Paclitaxel and Erlotinib of low-risk patients are significantly higher, while Gefitinib is the opposite. (D) A majority of high-risk patients are in cluster1, and a majority of patients in cluster2 have low risk scores. (E) High-risk patients in cluster1 exhibit worst survival outcomes, and are statistically significant.

Authors contribution

HC and GX contributed to the conception of the study; SX, ZH, YW, and YZ performed the R language; SF and QP contributed significantly to analysis and manuscript preparation; HC performed the data analyses and wrote the manuscript; KL, NL, and LZ helped perform the analysis with constructive discussions.

Conceptualization: Guodong Xu, Hang Chen.
Formal analysis: Qiaoling Pan, Shenzhe Fang, Software: Shuguang Xu, Yiqing Wei, Youjie Zhu, Zeyang Hu.
Writing – original draft: Hang Chen.
Writing – review & editing: Kaitai Liu, Linwen Zhu, Ni Li.

Acknowledgement

We thank the TCGA database for generously sharing a large amount of data.

References

[1] Nasim F, Sabath B, Eapen G. Lung cancer. Med Clin North Am. 2019;103:463–73.
[2] Siegel R, Miller K, Fuchs H, Jemal A. Cancer statistics, 2021. CA: A Cancer J Clin. 2021;71:7–33.
[3] Chen W, Zheng R, Baade P, et al. Cancer statistics in China. Cancer Stat China. 2016;66:115–32.
[4] Shi J, Wang L, Wu N, et al. Clinical characteristics and medical service utilization of lung cancer in China, 2005-2014: overall design and results from a multicenter retrospective epidemiologic survey. Lung Cancer. 2019;128:91–100.
[5] American Cancer Society. Non-small cell lung cancer survival rates by stage [OL]. Available at: www.cancer.org/cancer/non-small-cell-lung-cancer/detection-diagnosis-staging/survival-rates.html [Access date March 20, 2021].
[6] Hanada K, Kawada K, Nishikawa G, et al. Dual blockade of macropinocytosis and asparagine bioavailability shows synergistic anti-tumor effects on KRAS-mutant colorectal cancer. Cancer Lett. 2021;522:129–41.
[7] Saito N, Sawai S. Three-dimensional morphodynamic simulations of macropinocytic cups. iScience. 2021;24:103087.
[8] Canton J. Macropinocytosis: new insights into its underappreciated role in innate immune cell surveillance. Front Immunol. 2018;9:2286.
[9] Holt J, Bottomly K, Mooseker M. Assessment of myosin II, Va, VI and VIIa loss of function on endocytosis and endocytic vesicle motility in bone marrow-derived dendritic cells. Cell Motil Cytoskeleton. 2007;64:756–66.
[10] Doodnauth S, Grinstein S, Maxson M. Constitutive and stimulated macropinocytosis in macrophages: roles in immunity and in the pathogenesis of atherosclerosis. Philos Trans R Soc B. 2019;374:20180147.
[11] Li P, Lu M, Shi J, et al. Lung mesenchymal cells elicit lipid storage in neutrophils that fuel breast cancer lung metastasis. Nat Immunol. 2020;21:1444–55.
[12] Wang X, Li Y, Qian Y, et al. Extracellular ATP, as an energy and phosphorylating molecule, induces different types of drug resistances in cancer cells through ATP internalization and intracellular ATP level increase. Oncotarget. 2017;8:87860–77.
[13] Yamazaki S, Su Y, Maruyama A, et al. Uptake of collagen type I via macropinocytosis cause mTOR activation and anti-cancer drug resistance. Biochem Biophys Res Commun. 2020;526:191–8.
[14] Takenaka T, Nakai S, Katayama M, et al. Effects of gefitinib treatment on cellular uptake of extracellular vesicles in EGFR-mutant non-small cell lung cancer cells. Int J Pharm. 2019;572:118762.
[15] Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemporary Oncol. 2015;19:A68–77.
[16] Fiorini N, Lipman D, Lu Z. Towards PubMed 2.0. eLife 2017;6:e28801.
[17] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

[18] Wilkerson M, Hayes D. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3. (2017).
[19] Yarchoan M, Hopkins A, Jaffee E. Tumor mutational burden and response rate to PD-1 Inhibition. N Engl J Med. 2017;377:2500–1.
[20] Geeleher P, Cox N, Huang R. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS One. 2014;9:e107468.
[21] Ettinger D, Wood D, Aisner D, et al. Non-small cell lung cancer, version 5.2017, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2017;15:504–35.
[22] Chen H, Shen W, Ni S, et al. Construction of an immune-related lncRNA signature as a novel prognosis biomarker for LUAD. Aging. 2021;13:20684–97.
[23] Liu X, Xing H, Liu H, Chen J. Current status and future perspectives on immunotherapy in neoadjuvant therapy of resectable non-small cell lung cancer. Asia-Pacific J Clin Oncol. 2021;18:335–43.
[24] Li J, Wang Y, Li J, Cao S, Che G. Meta-analysis of lobectomy and sublobar resection for stage I non-small cell lung cancer with spread through air spaces. Clin Lung Cancer. 2022;23:268–213.
[25] Reuss J, Gosa L, Liu S. Antibody drug conjugates in lung cancer: state of the current therapeutic landscape and future developments. Clin Lung Cancer. 2021;22:483–99.
[26] Stein J, Rivera M, Weiner A, et al. Sociodemographic disparities in the management of advanced lung cancer: a narrative review. J Thoracic Disease 2021;13:3772–800.
[27] Wu J, Lou Y, Ma Y, Xu J, Shi T. A novel risk-score model with eight MiRNA signatures for overall survival of patients with lung adenocarcinoma. Front Genet. 2021;12:741112.
[28] Wu L, Wen Z, Song Y, Wang L. A novel autophagy-related lncRNA survival model for lung adenocarcinoma. J Cell Mol Med. 2021;25:5681–90.
[29] Guo Y, Qu Z, Li D, et al. Identification of a prognostic ferroptosis-related lncRNA signature in the tumor microenvironment of lung adenocarcinoma. Cell Death Discov. 2021;7:190.
[30] Lin X, Jiang T, Bai J, et al. Characterization of transcriptome transition associates long noncoding RNAs with glioma progression. Mol Ther Nucleic Acids. 2018;13:620–32.
[31] Zheng Z, Zhang Q, Wu W, et al. Identification and validation of a ferroptosis-related long non-coding RNA signature for predicting the outcome of lung adenocarcinoma. Front Genet. 2021;12:690509.
[32] Li J, Li R, Liu X, et al. A seven immune-related lncRNAs model to increase the predicted value of lung adenocarcinoma. Front Oncol. 2020;10:560779.
[33] Hou J, Yao C. Potential prognostic biomarkers of lung adenocarcinoma based on bioinformatic analysis. Front Oncol. 2021;2021:8859996.
[34] Wang E, Li Y, Ming R, et al. The prognostic value and immune landscapes of a mA/mC/mA-Related LncRNAs signature in head and neck squamous cell carcinoma. Front Cell Dev Biol. 2021;9:718974.
[35] Shao J, Zhang B, Kuai L, Li Q. Integrated analysis of hypoxia-associated lncRNA signature to predict prognosis and immune microenvironment of lung adenocarcinoma patients. Bioengineered. 2021;12:6186–200.
[36] You J, Fang W, Zhao Q, Chen L, Chen L, Chen F. Identification of a RNA-Seq based prognostic signature with seven immune-related lncRNAs for lung adenocarcinoma. Clin Lab. 2021;67.
[37] Ren M, Chen S, Wang L, Rui W, Li P. LINC00941 promotes progression of non-small cell lung cancer by sponging miR-877-3p to regulate VEGFA expression. Front Oncol. 2021;11:650037.
[38] Yan T, Ma G, Wang K, Liu W, Zhong W, Du J. The immune heterogeneity between pulmonary adenocarcinoma and squamous cell carcinoma: a comprehensive analysis based on lncRNA model. Front Immunol. 2021;12:547333.
[39] Chen H, Hu Z, Sang M, et al. Identification of an autophagy-related lncRNA Prognostic signature and related tumor immunity research in lung adenocarcinoma. Front Genet. 2021;12:767694.