

# Construction of a prognostic model for lung adenocarcinoma based on m6A/m5C/m1A genes

Hao Ding<sup>1</sup>, Yuanyuan Teng<sup>1</sup>, Ping Gao<sup>1</sup>, Qi Zhang<sup>1</sup>, Mengdi Wang<sup>1</sup>, Yi Yu<sup>2</sup>, Yueping Fan<sup>3,\*</sup>, Li Zhu<sup>4,\*</sup>

<sup>1</sup>Department of Respiratory Disease, Affiliated People's Hospital of Jiangsu University, NO. 8 Dianli Road, Runzhou District, Zhenjiang City, Jiangsu Province 212002, China

<sup>2</sup>Department of General Practice, Jiankang Road Community Health Service Center, NO. 239 Zhongshan East Road, Jingkou District, Zhenjiang City, Jiangsu Province 212008, China

<sup>3</sup>Department of Respiratory, Jurong Branch Hospital, Affiliated Hospital of Jiangsu University, NO. 8 Huayang South Road, Jurong City, Zhenjiang City, Jiangsu Province 212400, China

<sup>4</sup>Department of Nephrology, Affiliated People's Hospital of Jiangsu University, NO. 8 Dianli Road, Runzhou District, Zhenjiang City, Jiangsu Province 212002, China

\*Corresponding authors. Department of Nephrology, Affiliated People's Hospital of Jiangsu University, NO.8 Dianli Road, Runzhou District, Zhenjiang City, Jiangsu Province, 212002, China. E-mail: lizzzzhu@163.com; Department of Respiratory, Jurong Branch Hospital, Affiliated Hospital of Jiangsu University, NO.8 Huayang South Road, Jurong City, Zhenjiang City, Jiangsu Province, 212400, China. E-mail: ypfanny@163.com

## Abstract

**Background:** Developing a prognostic model for lung adenocarcinoma (LUAD) that utilizes m6A/m5C/m1A genes holds immense importance in providing precise prognosis predictions for individuals.

**Methods:** This study mined m6A/m5C/m1A-related differential genes in LUAD based on public databases, identified LUAD tumor subtypes based on these genes, and further built a risk prognostic model grounded in differential genes between subtypes. The immune status between high- and low-risk groups was investigated, and the distribution of feature genes in tumor immune cells was analyzed using single-cell analysis. Based on the expression levels of feature genes, a projection of chemotherapeutic and targeted drugs was made for individuals identified as high-risk. Ultimately, cell experiments were further verified.

**Results:** The 6-gene risk prognosis model based on differential genes between tumor subtypes had good predictive performance. Individuals classified as low-risk exhibited a higher ( $P < 0.05$ ) abundance of infiltrating immune cells. Feature genes were mainly distributed in tumor immune cells like CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and regulatory T cells. Four drugs with relatively low IC<sub>50</sub> values were found in the high-risk group: Elesclomol, Pyrimethamine, Saracatinib, and Temsirolimus. In addition, four drugs with significant positive correlation ( $P < 0.001$ ) between IC<sub>50</sub> values and feature gene expression were found, including Alectinib, Estramustine, Brigatinib, and Elesclomol. The low expression of key gene NTSR1 reduced the IC<sub>50</sub> value of irinotecan.

**Conclusion:** Based on the m6A/m5C/m1A-related genes in LUAD, LUAD patients were divided into 2 subtypes, and a m6A/m5C/m1A-related LUAD prognostic model was constructed to provide a reference for the prognosis prediction of LUAD.

**Keywords:** m6A/m5C/m1A; lung adenocarcinoma; prognosis; immune; drug prediction

## Introduction

As the primary cause of cancer-related mortalities, lung cancer covers 18.0% of all cancer deaths. Approximately one in every eight cancer individuals is diagnosed with lung cancer, and about two-thirds of all lung cancer deaths worldwide are attributed to smoking [1]. Lung adenocarcinoma (LUAD) is a prevalent histological subtype of non-small cell lung cancer (NSCLC). Despite the availability of various therapeutic modalities, such as surgical resection, radiation therapy, chemotherapy, and targeted therapy, the prognosis of individuals with LUAD remains unfavorable [2]. Recent studies have demonstrated the crucial role of methylation modification regulatory factors in the development and progression of LUAD, and their association with patient prognosis. Therefore, the development of a prognostic model targeting methylation modification regulatory factors is becoming a promising approach to guide the clinical prognosis and treatment of LUAD individuals [3].

Mounting evidence suggests that RNA modifications are key players in diverse RNA-related processes, such as translation,

transcription, processing, stability, and splicing [4, 5]. Thus far, over 170 distinct RNA modifications have been authenticated [6]. RNA methylation is a significant form of epigenetic modification, with N6-methyladenosine (m6A), 5-methylcytosine (m5C), and N1-methyladenosine (m1A) being some of the most prominent types of RNA methylation [7]. The regulation of these methylation modifications is governed by three distinct classes of enzymes in a dynamic manner, including writers (methylases), erasers (demethylases), and readers (binding proteins) [8–10]. While RNA modifications are not traditionally classified as cancer-driving factors, mounting evidence supports their involvement in the progression of cancer. The deviant expression of these modifications has been associated with a gamut of processes related to cancer, including enhanced cell growth, proliferation, differentiation, and tumorigenesis, as well as increased resistance to chemotherapy and other drugs [11–14]. Studies have found that the m6A demethylase FTO promotes lung cancer cell proliferation by modulating the m6A levels of USP7 mRNA [15]. According to Liu et al. [16], m5C RNA regulatory factors serve a critical function

in governing the tumor microenvironment (TME) in LUAD and hold the potential for prognostic predictions of LUAD individuals. The importance of m1A regulatory factors in both the treatment and prognosis of LUAD was explored by Bao and colleagues [17]. They observed that individuals with decreased m1A levels exhibited higher mortality rates, as well as increased tumor mutation burden (TMB) and genomic mutation frequency. In summary, RNA methylation modifications feature in the prognosis of LUAD, and they have a certain influence on the occurrence and development of tumors. Despite recent advances in the study of m6A/m5C/m1A modifications, there remains a notable gap in our understanding of their prognostic value, particularly in LUAD. As such, the development of a comprehensive m6A/m5C/m1A-related prognostic model for foretelling the outcomes of LUAD individuals would hold significant practical significance. Such a model could enable more accurate prognostic assessments and potentially guide personalized treatment strategies for individuals with this deadly disease.

In this work, the subtypes of differentially expressed m6A/m5C/m1A regulatory factors in LUAD were determined, and a prognostic model was devised based on the differentially expressed genes (DEGs) across the identified clusters. Additionally, we conducted drug sensitivity prediction and virtual drug screening, and mined new therapeutic targets based on the relationship between drugs and targets. Finally, we performed a single-cell analysis of the feature genes in the model and interpreted their distribution in different immune cell types. At the same time, the relationship between key gene expression and drug sensitivity was detected by cell experiments. This will provide a reference for prognostic prediction and drug selection for LUAD individuals.

## Results

### Expression and mutation analysis of m6A/m5C/m1A related differential genes

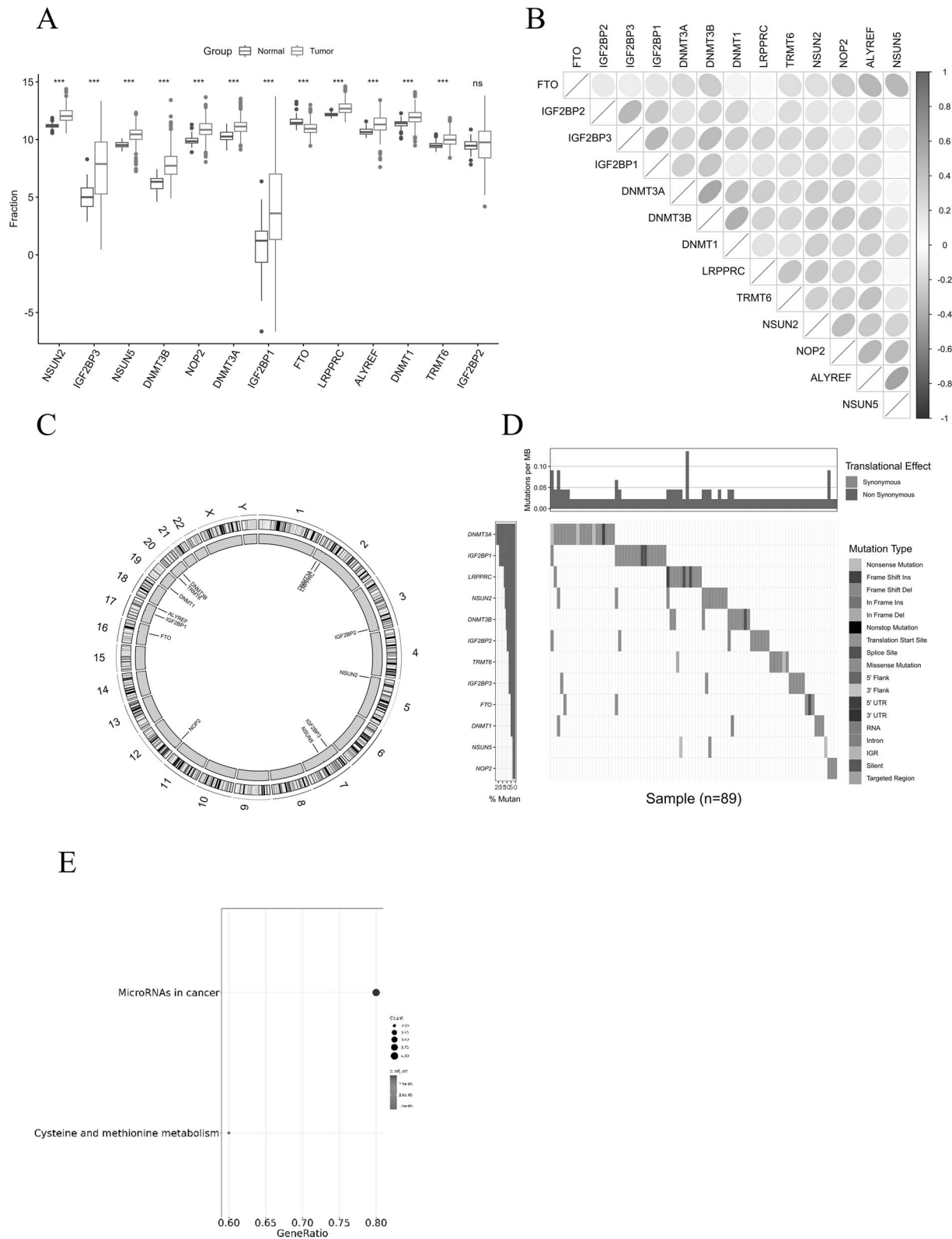
By analyzing the differential genes between tumor and normal groups in The Cancer Genome Atlas (TCGA) database, 8694 LUAD-related DEGs were obtained. By referring to the literature, 50 m6A/m5C/m1A related genes were acquired, and the intersection of these two sets of genes yielded 13 m6A/m5C/m1A genes differentially expressed in LUAD. We analyzed the expression differences of these 13 genes between tumor and normal groups, and ascertained the significant upregulation of NSUN2, IGF2BP3, NSUN5, DNMT3B, NOP2, DNMT3A, IGF2BP1, LRPPRC, ALYREF, DNMT1, and TRMT6 ( $P < 0.001$ ), but downregulation of FTO in tumor tissues ( $P < 0.001$ ) (Fig. 1A). Correlation heatmap analysis revealed strong positive correlations among these genes (Fig. 1B). Next, we visualized the specific locations of the differential m6A/m5C/m1A genes with significant copy number variation (CNV) differences in tumors on human chromosomes (Fig. 1C). We then analyzed the mutation status of these 13 genes in tumor groups and found that the top 5 mutated genes were DNMT3A, IGF2BP1, LRPPRC, NSUN2, and DNMT3B according to the TMB waterfall plot (Fig. 1D). The above results suggest that there are 13 DEGs related to m6A/m5C/m1A in LUAD, and these genes have mutated. The 13 m6A/m5C/m1A-related genes were enriched in MicroRNAs in Cancer, Cysteine, and methionine metabolism signaling pathways (Fig. 1E) (Kyoto Encyclopedia of Genes and Genomes (KEGG), approved by Kanehisa laboratories, Kyoto, Japan).

### Tumor subtype identification on the ground of m6A/m5C/m1A-regulated genes

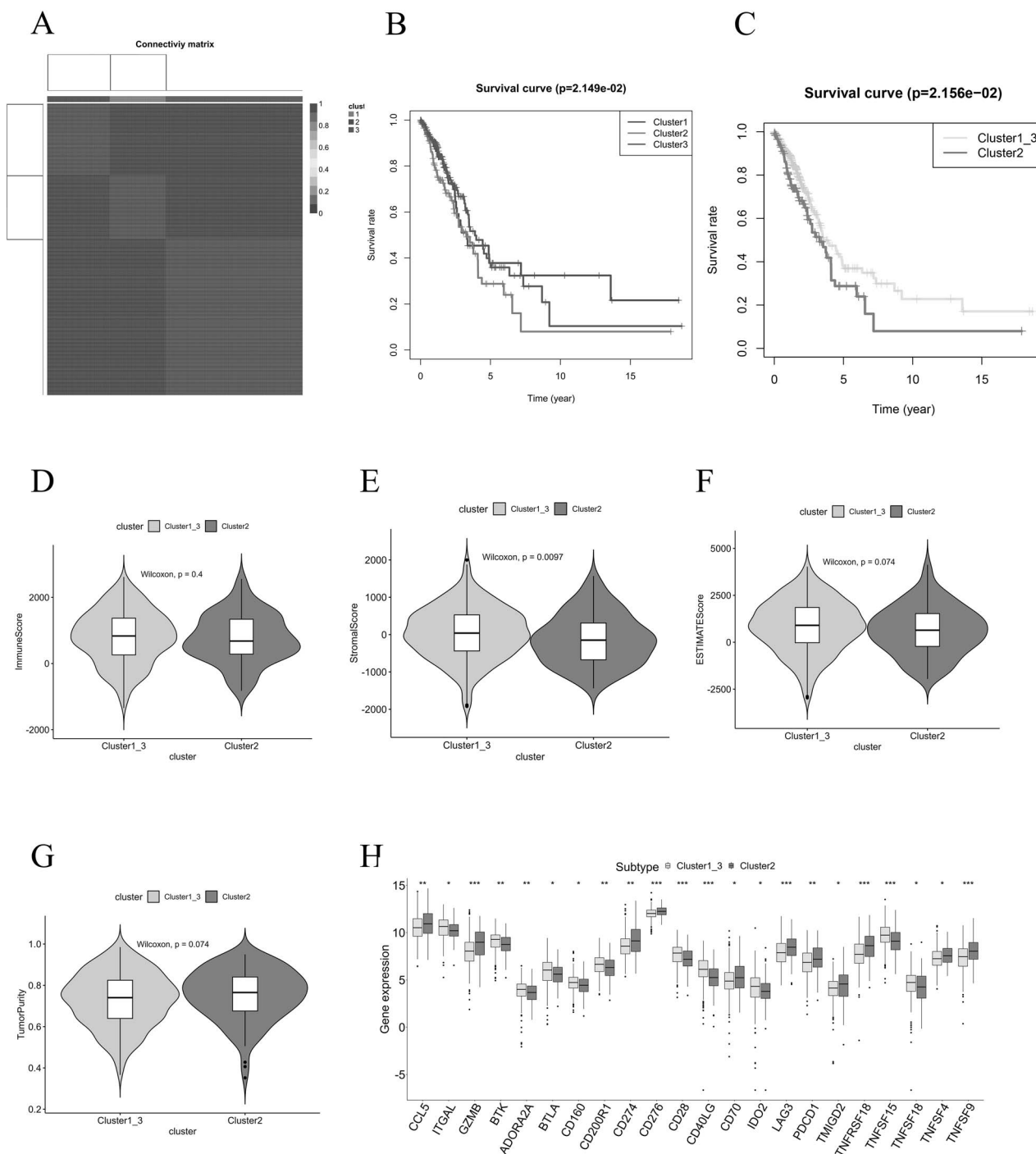
Based on 13 m6A/m5C/m1A-regulated genes, the samples were stratified by the nonnegative matrix factorization (NMF) clustering method, and rank = 3 was selected as the best number of clusters (Fig. 2A). We then conducted survival and immune analyses on the three clusters, showing that there was a significant difference in the overall survival rate between clusters 1, 2, and 3, with a lower survival rate in cluster 2 ( $P = 1.149 \times 10^{-2}$ ). There was no significant difference between cluster 1 vs. cluster 2 ( $P = 1.007 \times 10^{-1}$ ) and cluster 1 vs. cluster 3 ( $P = 9.213 \times 10^{-1}$ ), but there was a significant difference between cluster 2 and cluster 3 ( $P = 2.827 \times 10^{-2}$ ) (Fig. 2B). The survival rate of cluster 1\_3 was significantly higher than that of cluster 2 ( $P = 1.256 \times 10^{-2}$ ) (Fig. 2C). Additionally, there was a significant difference in stromal score between cluster 2 and cluster 1\_3 ( $P < 0.05$ ). The immune score and ESTIMATE score of cluster 2 were lower than those of cluster 1\_3, but not statistically significant (Fig. 2D–F). Immune score was used to evaluate the degree of tumor immune cell infiltration, and the stromal score was used to evaluate the degree of tumor interstitial cell infiltration. The higher the score, the better the prognosis. ESTIMATE score is a comprehensive evaluation index of tumor microenvironment. It is the sum of the immune score and stromal score, which is used to infer tumor purity. The higher the score, the lower the tumor purity. The tumor purity score of cluster 2 was also higher than that of cluster 1\_3, but not statistically significant (Fig. 2G). The box plot of immune checkpoints revealed that the expression levels of ITGAL, BTK, ADORA2A, BTLA, CD160, CD200R1, CD28, CD40LG, IDO2, TNFSF18 and TNFSF18 were tellingly lower ( $P < 0.05$ ) in cluster 2 than in cluster 1\_3 (Fig. 2H). The above results show that there are differences in survival time, immune cell infiltration, and immune checkpoints among the three clusters, indicating a better stratification effect on LUAD patients based on 13 m6A/m5C/m1A-regulated genes. Since the survival rate of cluster 1\_3 was significantly different from that of cluster 2, we chose to combine clusters 1 and 3 for analysis with cluster 2.

### Construction of the 6-gene prognostic model for LUAD

From the results of survival and immune analysis among different clusters, we conducted a differential analysis of genes between cluster 2 and cluster 1\_3 (control group) with  $|\log FC| > 1$  and FDR  $< 0.05$ , resulting in 1091 DEGs (Fig. 3A). After combining with clinical information, univariate regression analysis was done, and 43 candidate genes significantly linked with survival were screened based on a  $P$  value  $< 0.001$  (Table S1). Subsequently, LASSO analysis was employed to obtain 19 feature genes (Fig. 3B and C, Table S2) to mitigate multicollinearity and reduce model complexity by selecting the optimal penalty parameter Lambda. Finally, we performed multiple regression analysis on these 19 feature genes and selected 6 feature genes based on the Lambda value to construct the prognostic model:  $\text{Risk score} = 0.0934 \times \text{NTSR1} - 0.0358 \times \text{ABCC12} + 0.0448 \times \text{KRT76} - 0.1064 \times \text{RIC3} - 0.1278 \times \text{ATP8A2} + 0.0468 \times \text{KRT6C}$  (Fig. 3D). Among them, the  $P$  value of the coefficient is less than 0.05, indicating that this factor has a significant impact on survival. For example, the  $P$  value of NTSR1 is less than 0.001, indicating that it has a very significant impact on survival. The  $P$  value of the model is  $2.2931 \times 10^{-12}$  and the C index was 0.71.



**Figure 1.** Data analysis of differentially regulated genes of m6A/m5C/m1A in LUAD. (A) Boxplot of differentially expressed m6A/m5C/m1A genes in normal and tumor groups. (B) Correlation heatmap of differential m6A/m5C/m1A genes. (C) CNV chromosome displays a circle diagram of differentially expressed m6A/m5C/m1A genes. (D) TMB waterfall plot of differentially expressed m6A/m5C/m1A genes in tumors. (E) KEGG analysis of 13 m6A/m5C/m1A genes enriched signaling pathways. (<https://www.genome.jp/kegg/>) [65–67]. Alt-text (A) The box plot showed that NSUN2, IGF2BP3, NSUN5, DNMT3B, NOP2, DNMT3A, IGF2BP1, LRPPRC, ALYREF, DNMT1, and TRMT6 were highly expressed in LUAD tumor tissues, FTO was lowly expressed, and the expression of IGF2BP2 did not change significantly; (B) The correlation heat map showed that there was a positive correlation between genes except FTO, and there was a negative correlation between FTO and other genes; (C) The circle shows the specific location of these genes in chromosome copy number variation; (D) The waterfall diagram analyzes the tumor mutation burden and shows the genes with a greater degree of mutation that may affect the tumor; (E) The bubble diagram showed the functions involved in the differential m6A/m5C/m1A genes.

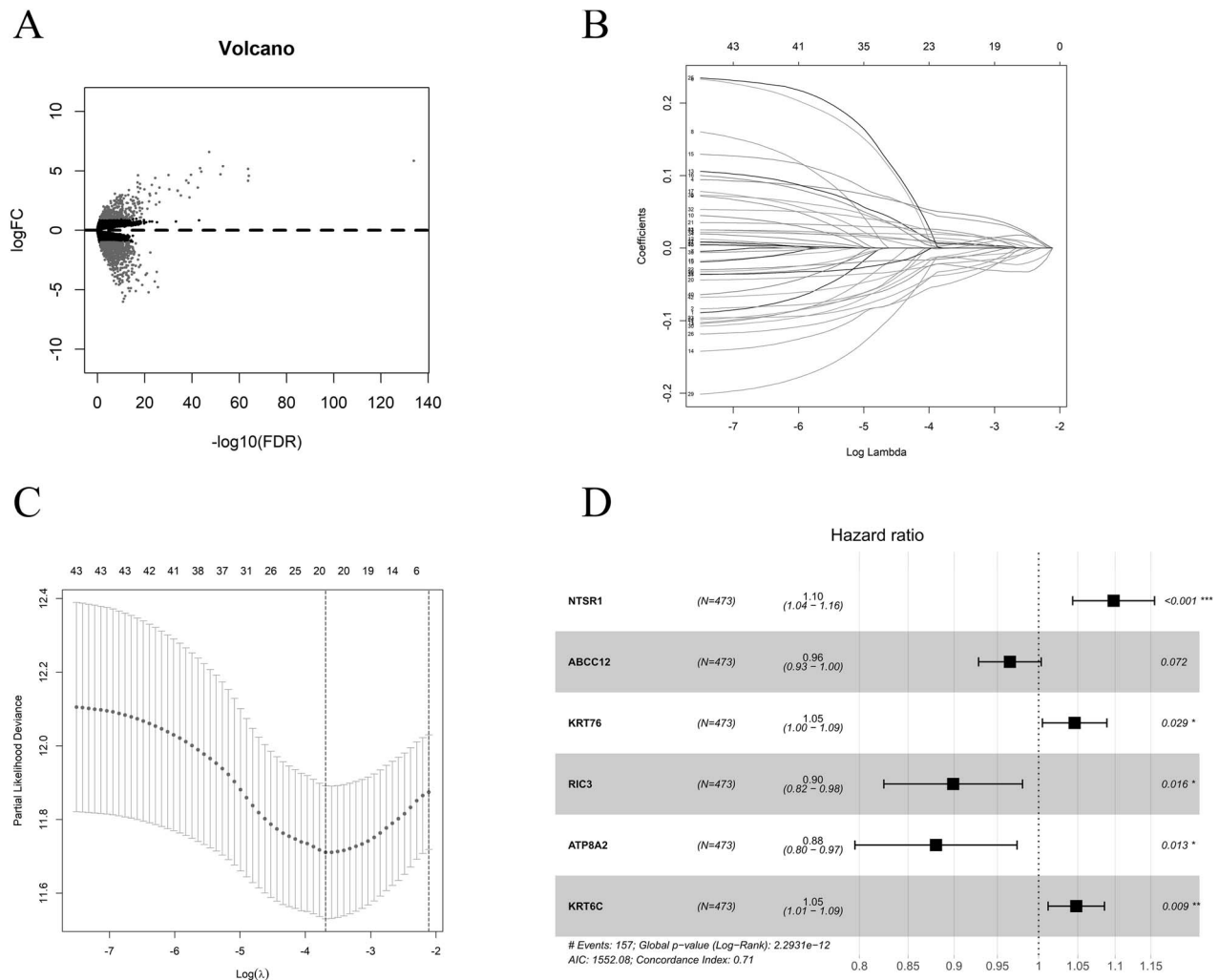


**Figure 2.** Identification and analysis of tumor subtypes based on m6A/m5C/m1A regulatory genes. (A) NMF clustering diagram. (B) Cluster survival analysis. (C) Cluster 1\_3 and cluster 2 survival analysis. (D) Immune score. (E) Stromal score. (F) ESTIMATE score. (G) Tumor purity score. (H) Boxplot of immune checkpoint expression. Alt-text (A) Heat map display non-negative matrix algorithm based on m6A/m5C/m1A gene expression patients were divided into 3 clusters; (B) K-M curve showed that cluster1 had a higher survival rate than cluster2 and cluster3; (C) The K-M curve showed that the survival rate of patients with combined cluster1\_3 was higher than that of cluster2; (D-F) The violin plot showed that the immune score, stromal score, and ESTIMATE score of cluster2 were relatively lower than those of cluster1\_3; (G) Violin diagram showed that there was no significant difference between cluster2 and cluster1\_3 in tumor purity; (H) The box plot showed that there were differences in the expression of immune checkpoint genes between cluster2 and cluster1\_3.

## Validation of the 6-gene prognosis model for LUAD

In order to evaluate the reliability of the model, we calculated riskScores for clinical samples in the training set (Fig. 4A) and grouped them into high- and low-risk groups based on the median riskScore, which disclosed a positive correlation between

riskScores and the number of patient deaths (Fig. 4B). In the high-risk group, the expression levels of ATP8A2, ABCC12, and RIC3 genes were low, while the expression levels of KRT76, NTSR1, and KRT6C were high (Fig. 4C). The area under the curve (AUC) values of the 1-, 3-, and 5-year receiver operating characteristic (ROC) curves were 0.78, 0.71, and 0.69, respectively, indicating



**Figure 3.** Construction of m6A/m5C/m1A genes-related prognostic model of LUAD. (A) Volcano plot for differential analysis. (B) Coefficient distribution plot was generated for  $\log(\lambda)$  sequences in the Lasso model. (C) Lasso coefficient spectrum for Lasso Cox analysis. (D) Forest plot for multiple factor regression analysis. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , indicating that the gene has a significant effect on the survival status. Alt-text (A) Volcano map shows the differential genes between cluster2 and cluster1\_3; (B and C) Twenty genes were obtained by LASSO analysis; (D) The forest plot showed that six characteristic genes were obtained after multivariate Cox regression analysis, and the influence of these genes on the survival rate of patients was demonstrated by  $P$  value.

good predictive performance of the prognosis model we built (Fig. 4D). The survival analysis indicated that the group with higher riskScores exhibited significantly lower survival rates ( $P = 1.495e-07$ ) (Fig. 4E).

Next, we validated the model using the Gene Expression Omnibus (GEO) dataset GSE72094. We computed the riskScore values for the validation set samples using the built model (Fig. 4F) and put the samples into high- and low-risk groups according to the median riskScore. The survival status (Fig. 4G) and prognostic gene expression levels (Fig. 4H) of the validation set samples showed the same results as those of the training set samples. The AUC values of the 1-, 3-, and 5-year ROC curves for this set of samples were 0.67, 0.71, and 0.85, respectively (Fig. 4I). Survival curve results also disclosed tellingly lower survival rates for the high-risk group ( $P = 1.161e-04$ ) (Fig. 4J). Overall, the prognosis model we built here exhibited good predictive performance.

We also compared the 1-, 3-, and 5-year ROC values of the prognostic model based on the training set with the other 14 published prognostic features [18–31] to test the prognostic performance of our model. The results showed that our prognostic model had

better performance than any other model in TCGA-LUAD cohort (Fig. 5).

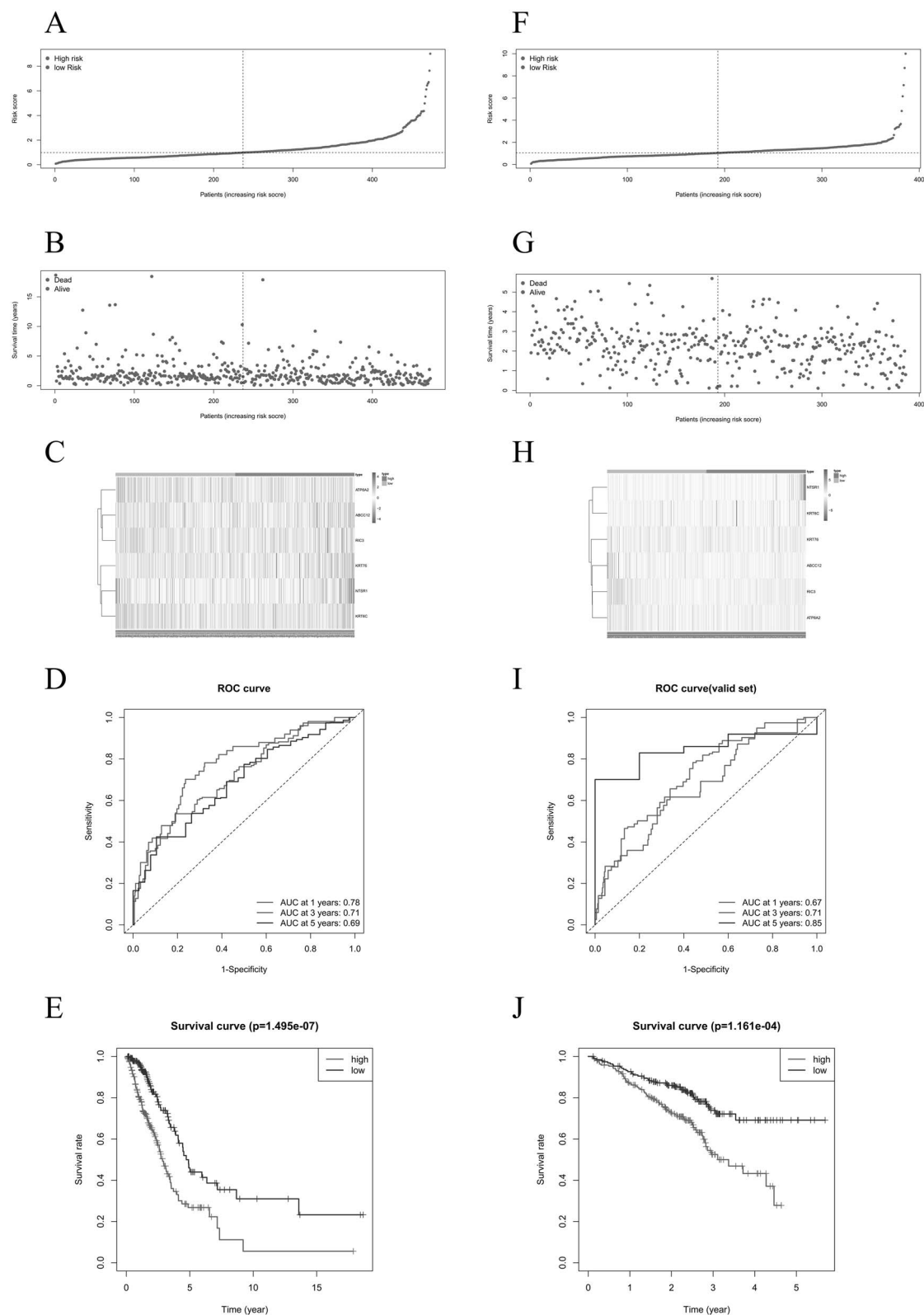
### Enrichment pathway analysis of significantly enriched pathways between two risk groups

We conducted KEGG pathway enrichment analysis using gene set enrichment analysis (GSEA) software to identify significantly enriched pathways between the two groups based on the gene sets, with outcomes showing that the genes in the high-risk group were significantly enriched in several pathways, including PENTOSE PHOSPHATE PATHWAY, PATHOGENIC ESCHERICHIA COLI INFECTION, CELL CYCLE, DNA REPLICATION, P53 SIGNALING PATHWAY, and HOMOLOGOUS RECOMBINATION (Fig. 6). These pathways are involved in cancer progression, indicating that the feature genes we obtained may affect LUAD by participating in multiple signaling pathways that affect cancer.

### The prognostic model has an independent prognostic ability

We calculated the riskScore for different genders, T, N, and stage classifications to evaluate their correlations and plotted violin

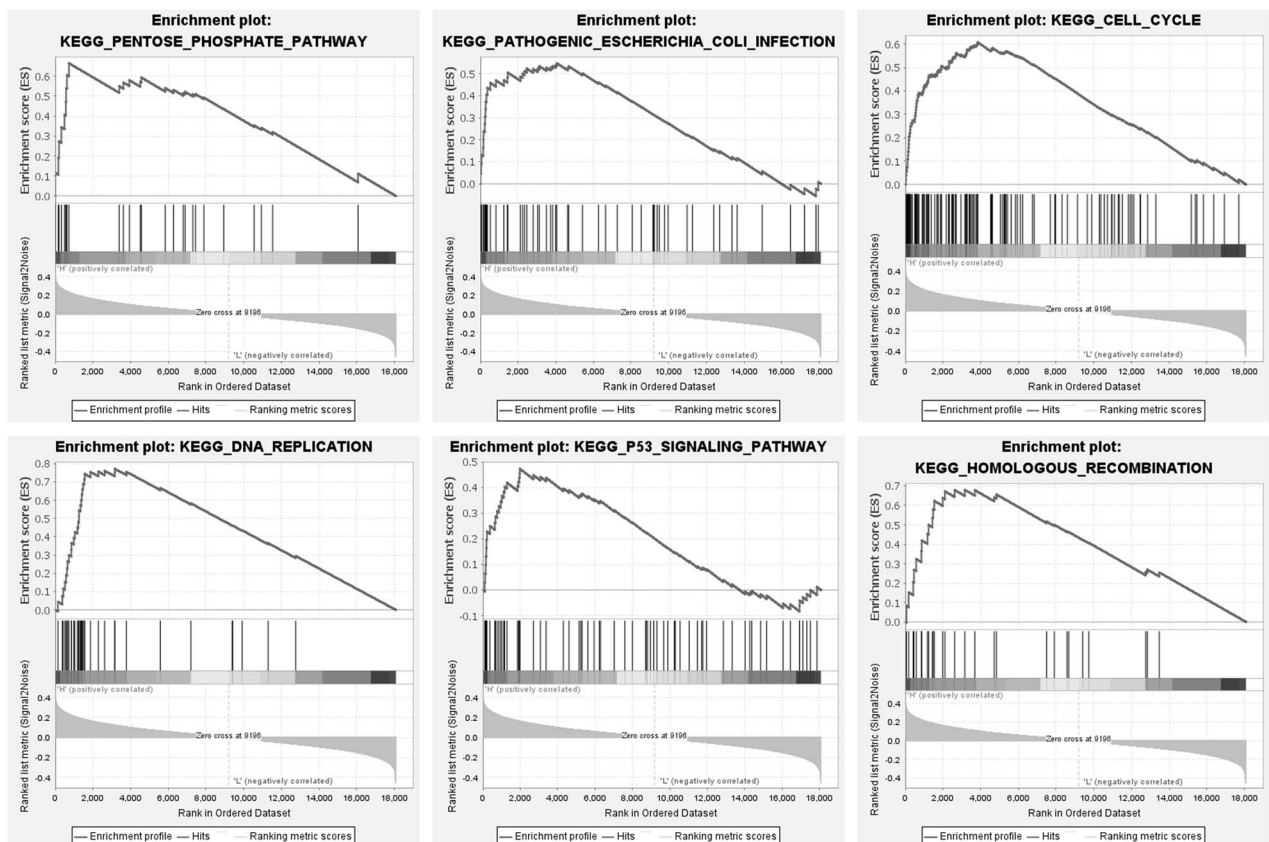




**Figure 4.** Validation of the prognostic model. (A) Distribution of riskScore for TCGA training set. (B) Distribution of survival status for TCGA training set. (C) Heatmap of expression levels for TCGA training set. (D) ROC curve for TCGA training set. (E) Survival analysis for high- and low-risk groups in TCGA training set. (F) Distribution of risk group scores in the GEO validation set. (G) Distribution of survival status for GEO validation set. (H) Heatmap of expression levels for GEO validation set. (I) ROC curve for GEO validation set. (J) Survival analysis for high- and low-risk groups in the GEO validation set. Alt-text (A and B) The scatter plot shows the riskscore values of each sample and ranks them by size. At the same time, it shows the survival status of patients; (C) Heat map showed that in the high-risk group, ATP8A2, ABCC12, and RIC3 were lowly expressed, and KRT76, NTSR1, and KRT6C were highly expressed; (D and E) ROC curve and survival curve showed that the AUC value was greater than 0.6, and the survival rate of the high-risk group was lower than that of the low-risk group; (F and G) In the validation set, the scatter plot showed the riskscore value of each sample and sorted by size. At the same time, it showed the survival status of each patient, and the patients who died were mostly distributed in the high-risk group; (H) In the validation set, the heat map showed low expression of ATP8A2, ABCC12, and RIC3, and high expression of KRT76, NTSR1, and KRT6C in the high-risk group; (I and J) In the validation set, the ROC curve and survival curve showed that the AUC value was greater than 0.6, and the survival rate of the high-risk group was lower than that of the low-risk group.



**Figure 5.** Comparison between m6A/m5C/m1A gene-related prognostic model and other models. The 1-, 3-, and 5-year AUC values of the prognostic model and other models developed in TCGA-LUAD cohort. Alt-text The bar chart shows that compared with other models, the AUC values of the model in this study are higher in 1 year, 3 years and 5 years.



**Figure 6.** GSEA pathway enrichment analysis of high- and low-risk groups. The line chart is called the enrichment score (ES) line chart. The peak represents the ES of the gene set, and the gene on the left side of the peak represents the core gene of the gene set. The upper ordinate corresponds to running ES; the lower abscissa, the distribution of rank values of all genes after sorting; barcode, also known as hits plot, marks the location of the gene in the gene ordination list; the left bar graph indicates the high-risk group genes, and the right bar graph indicates the low-risk group genes. Alt-text GSEA enrichment analysis showed that high-risk group genes were enriched in PENTOSE PHOSPHATE PATHWAY, PATHOGENIC ESCHERICHIA COLI INFECTION, CELL CYCLE, DNA REPLICATION, P53 SIGNALING PATHWAY, and HOMOLOGOUS RECOMBINATION signaling pathways.

plots. The results showed significant differences among them, with male individuals ( $P = 0.0076$ ) (Fig. 7A), T3 + T4 stage ( $P = 0.019$ ) (Fig. 7B), N1 + N2 + N3 stage ( $P = 3.9 \times 10^{-6}$ ) (Fig. 7C), and stage 3–4

( $P = 0.00054$ ) (Fig. 7D) having higher riskScore values and disease risks. Then, we performed univariate and multivariate regression analysis combining riskScore with clinical information, and found

that stage, T, N, and riskScore were all significant in the univariate regression ( $P < 0.001$ ), with riskScore being the most significant in the multivariate regression analysis (stage,  $P = 0.022$ ; N,  $P = 0.023$ , riskScore,  $P < 0.001$ ) (Fig. 7E and F). Subsequently, we used the riskScore and combined it with clinical information to draw a nomogram of the model to foretell the survival rate of LUAD individuals (Fig. 7G). The calibration curves for 1-year, 3-year, and 5-year survival rates demonstrated a strong agreement between the predicted and actual survival rates of the model (Fig. 7H–J). Therefore, the 6-gene prognostic model has independent prognostic performance and can accurately predict patient prognosis.

### Low-risk patient group may benefit more from immunotherapy

Next, we performed an immune-related analysis on the two groups. Firstly, using the immune cell score obtained by single-sample gene set enrichment analysis (ssGSEA), the scores of aDCs, iDCs, Mast\_cells, and Neutrophils were found significantly higher in the low-risk group, while the NK\_cells score was significantly lower in the low-risk group ( $P < 0.001$ ) (Fig. 8A). Regarding immune function scoring, the low-risk group exhibited significantly higher scores ( $P < 0.05$ ) for HLA and Type\_II\_IFN\_Response but lower scores ( $P < 0.01$ ) for APC\_co\_inhibition, Inflammation-promoting, MHC\_class\_I, and Parainflammation in contrast to the high-risk group (Fig. 8B). Further analysis of infiltrating immune cells using CIBERSORT showed that T cells CD4 memory activated, Macrophages M0, Macrophages M1, and Neutrophils were significantly higher in the high-risk group, while B cells naive, T cells CD4 memory resting, Monocytes, and Dendritic cells activated were significantly higher in the low-risk group ( $P < 0.05$ ) (Fig. 8C). The HLA results indicated that the expression level of HLA was significantly higher in the low-risk group than in the other group ( $P < 0.05$ ) (Fig. 8D). Furthermore, we found that the expression levels of immune checkpoint genes were significantly different between the high- and low-risk groups. Compared with the low-risk group, the expression of most immune checkpoints in the high-risk group was low ( $P < 0.05$ ) (Fig. 8E). Additionally, we found that the low-risk group had a lower tumor immune dysfunction and exclusion (TIDE) score ( $P = 2.3e-08$ ) (Fig. 8F) and a higher immunophenoscore (IPS) ( $P < 0.001$ ) (Fig. 8G–H). Drawing on the results of the immunological analysis, it could be inferred that the low-risk group in LUAD might derive greater benefit from immunotherapy.

### Correlation analysis between feature genes and single-cell characteristics

To explore the expression patterns of model feature genes across various immune cell types, we conducted an analysis of single-cell sequencing data from GSE99254 utilizing the Tumor Immune Single-cell Hub (TISCH) database. The Uniform Manifold Approximation and Projection (UMAP) diagram showed that there were mainly 6 cell clusters, of which the most abundant cells were CD4 Tconv, CD8 Tex, CD8 T cells, and Treg cells (Fig. 9A). Next, we evaluated the expression of feature genes and found that RIC3 was expressed in these cell clusters, but was mainly expressed in CD4 Tconv cells and CD8 Tex cells (blue dots were densely distributed in these two cell clusters) (Fig. 9B), while other feature genes were less expressed in these cell clusters (blue dots were sparsely distributed in the six cell clusters) (Figs S1–S5). Additionally, we further quantitatively analyzed the expression of feature genes in these cell clusters. Similar to the results of Fig. 9B, compared with the expression of other feature genes, RIC3 was expressed in the six cell clusters, and its expression level was

relatively high. In addition, these genes were highly expressed in monocytes/macrophages, which may be due to the small number of monocytes/macrophages (Fig. 9C). The above results suggest that the expression levels of feature genes are different in six types of immune cells.

### High-risk group has a higher TMB

To understand the mutation status of genes in different tumor samples, we utilized TCGA mutation data to assay the TMB of two group samples, with results showing that the TMB of the high-risk group was notably higher ( $P = 1.3e-05$ ) (Fig. 10A). Patients with high TMB may achieve better immunotherapy results due to higher antigen counts. The detailed gene mutation spectrum of patients in the high- and low-risk groups was further explored through the waterfall diagram. Specific mutation analysis of the top 30 genes in the two groups revealed that TP53 ranked first in the high-risk group (Fig. 10B) and second in the low-risk group (Fig. 10C), with a higher mutation frequency in the high-risk group. Different expression patterns may lead to different clinical outcomes of immune response [21].

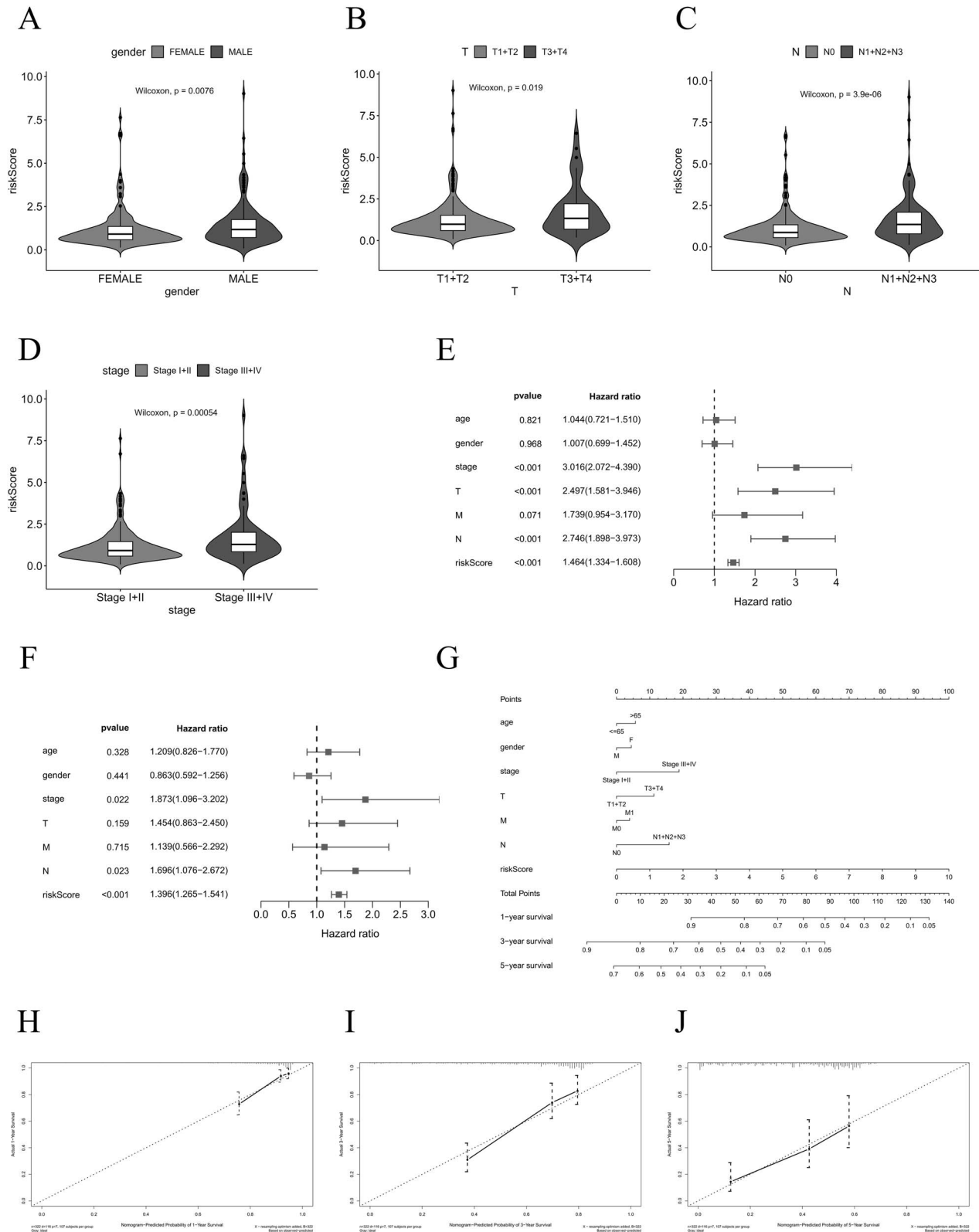
### Drug sensitivity analysis

In order to further explore the potential application of the prognostic model in the treatment of LUAD, the possible chemotherapeutic drugs were predicted by analyzing the correlation between the expression of feature genes and the IC<sub>50</sub> value of conventional chemotherapeutic drugs. At the same time, the difference in chemotherapy sensitivity between the high- and low-risk groups was evaluated by analyzing the relationship between the two risk groups and the IC<sub>50</sub> value. The results showed significant positive correlations between KRT6C and Alectinib, Estramustine, and Brigatinib (correlation coefficients of 0.710, 0.546, and 0.525, respectively), as well as a significant positive correlation between KRT76 and Elesclomol (correlation coefficient of 0.646) (Fig. 11A). It is suggested that the low expression of KRT6C and KRT76 in LUAD may be highly sensitive to Alectinib, Estramustine, Brigatinib, and Elesclomol. Among several chemotherapeutic drugs (Elesclomol, Pyrimethamine, Saracatinib, and Temsirolimus), there are differences in drug sensitivity between high- and low-risk groups, and the high-risk group is relatively more sensitive ( $P < 0.05$ ) (Fig. 11B). It is suggested that these four chemotherapy drugs may be potential drugs for the treatment of LUAD patients in the high-risk group. These results provide a reference for the stratified treatment of LUAD.

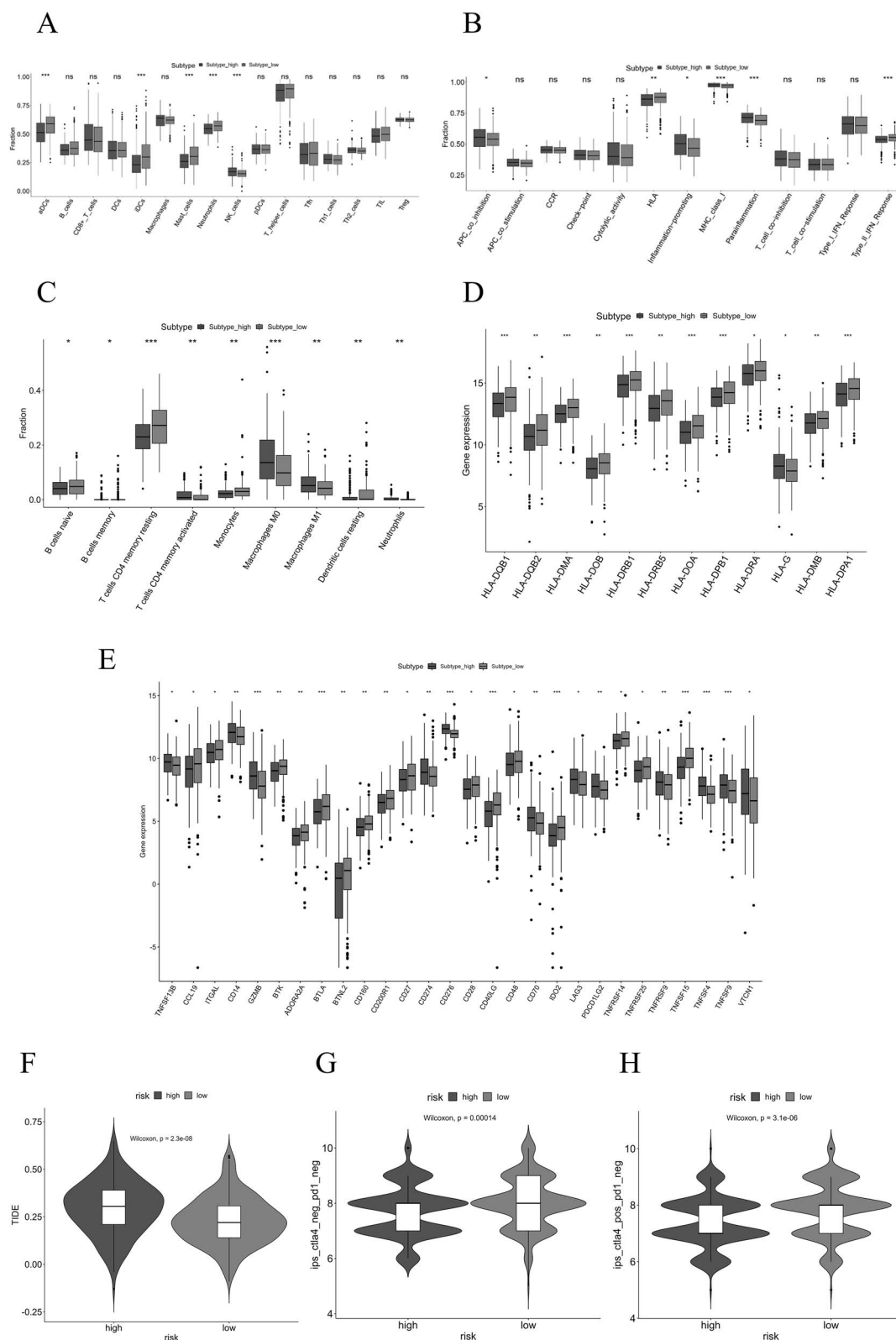
### Virtual screening of target drugs for feature genes

We further conducted a virtual screening of drugs targeting the three feature genes upregulated in the prognostic model: NTSR1, KRT76, and KRT6C. The three-dimensional structures of the proteins corresponding to these genes were obtained from the PDB database, and 2115 small molecule compounds approved by the Food and Drug Administration (FDA) were screened virtually from the ZINC database, with outcomes showing the top two small molecules and proteins in terms of affinity score prediction (Table 1). The binding mode of compounds and proteins was plotted, and the interaction between compounds and binding sites and surrounding amino acid residues was observed. NTSR1 had affinity scores of  $-10.8$  (Fig. 12A) and  $-10.4$  kcal/mol (Fig. 12B) for Irinotecan and Dihydroergotamine, respectively. KRT76 had affinity scores of  $-8.9$  (Fig. 12C) and  $-8.6$  kcal/mol (Fig. 12D) for Dihydroergotamine and Ergotamine, respectively. KRT6C had

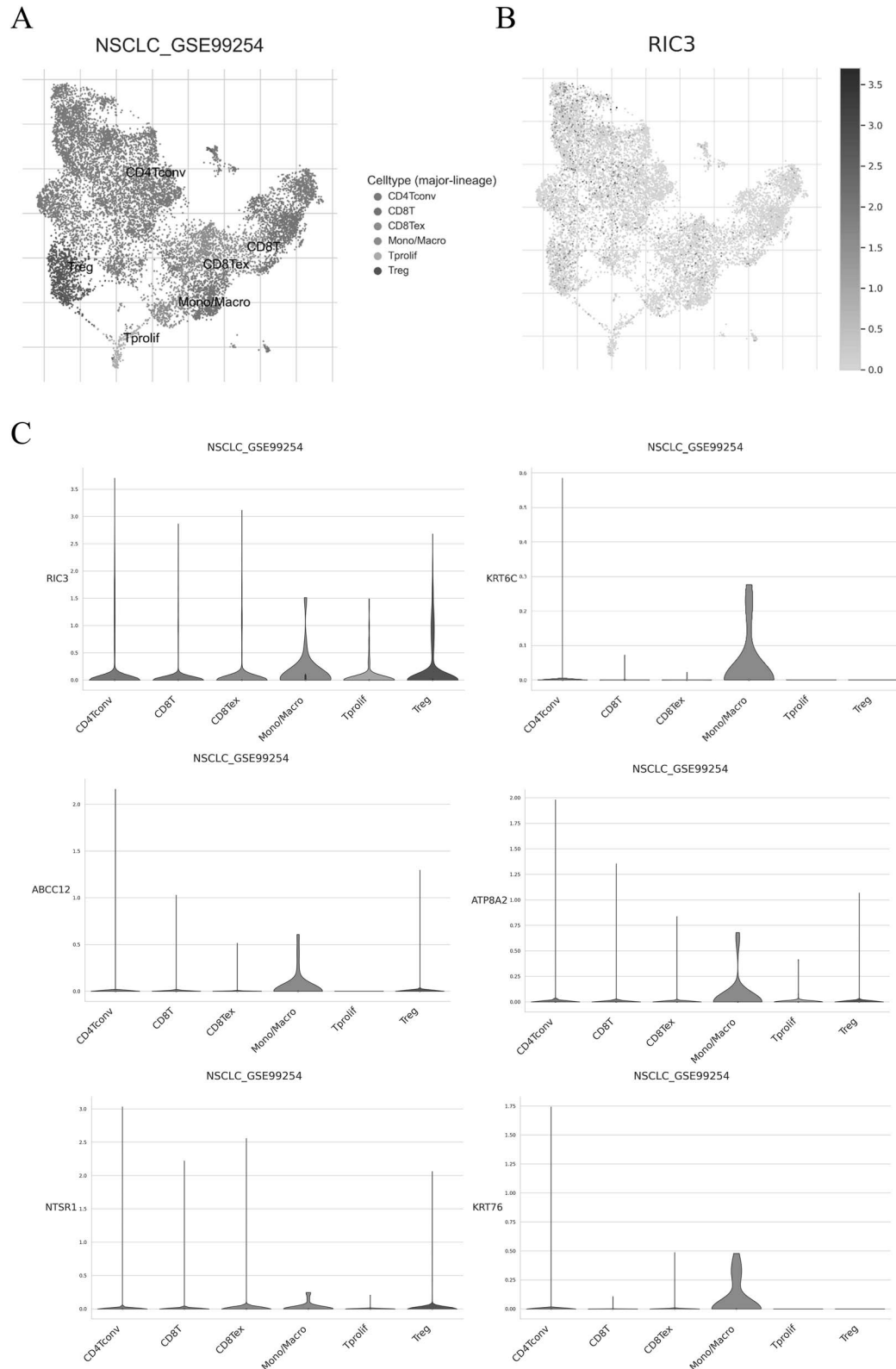




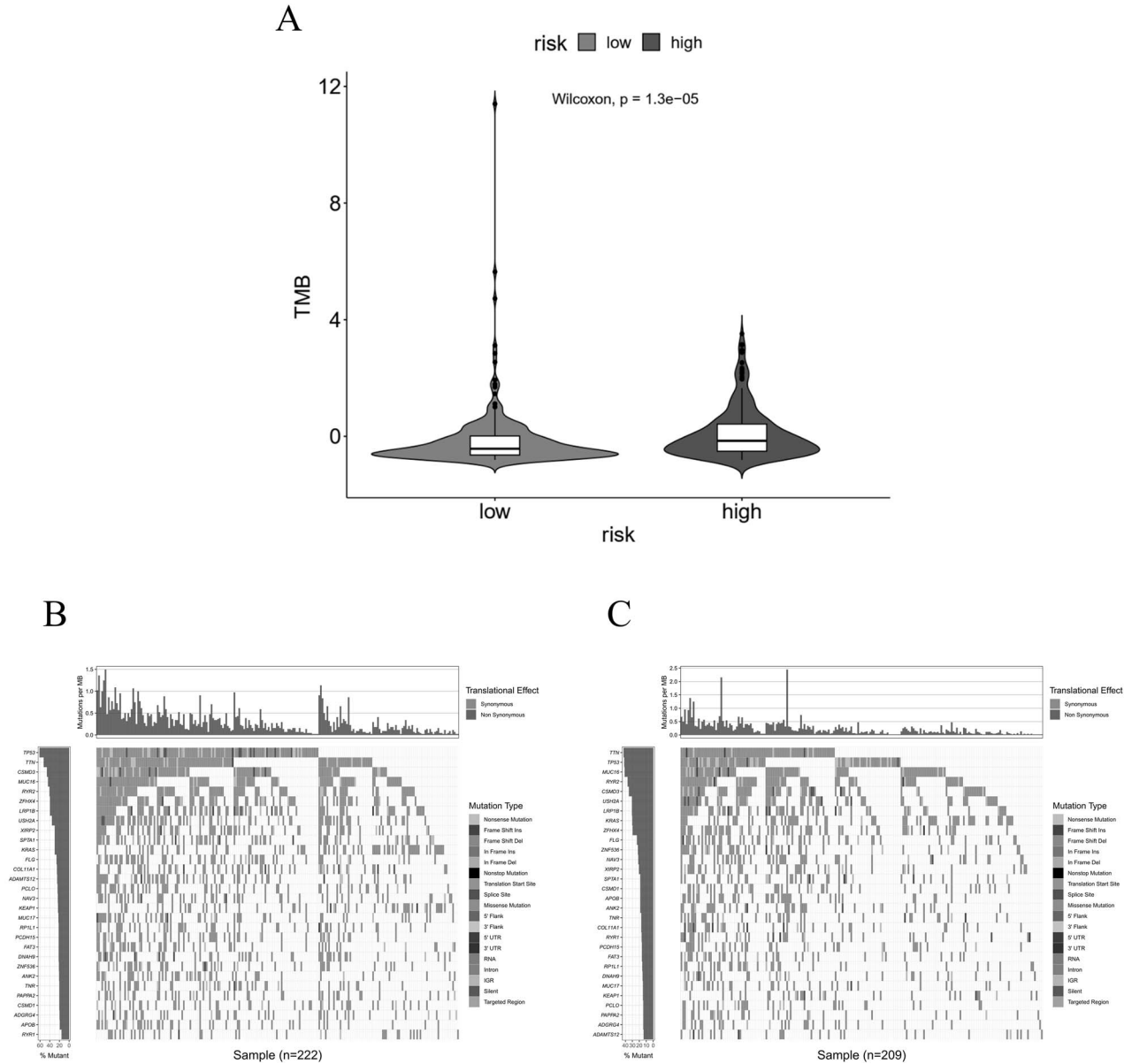
**Figure 7.** Independent prognostic ability analysis for riskScore and clinical features. (A) Relationship between age and riskScore. (B) Relationship between tumor T stage and riskScore. (C) Relationship between tumor N stage and riskScore. (D) Relationship between tumor grade and riskScore. (E) Forest plot for univariate regression analysis. (F) Forest plot for multiple factor regression analysis. (G) Nomogram of prognostic model riskScore combined with clinical information. (H) Calibration curve for 1-year risk prediction. (I) Calibration curve for 3-year risk prediction. (J) Calibration curve for 5-year risk prediction. Alt-text (A–D) The violin plot showed that the risk scores of male group, T3 + T4 group, N1 + N2 + N3 group and stage III + IV group were higher; (E and F) the Forest plot showed that the P value of riskScore was less than 0.001 in univariate and multivariate Cox regression analysis; (G) RiskScore combined with clinical information to draw nomogram. (H–J) The calibration curves for 1-year, 3-year, and 5-year survival rates showed strong consistency between the predicted and actual survival rates of the model.



**Figure 8.** Immune analysis of high- and low-risk groups. (A) ssGSEA algorithm analysis of immune cells. (B) ssGSEA algorithm analysis of immune function. (C) CIBERSORT algorithm analysis of immune cell infiltration level. (D) Box plots of HLA expression levels. (E) Box plots of immune checkpoint expression levels. (F) Box plots of TIDE scores. (G) Box plots of IPS. (H) Box plots of IPS. Alt-text (A-C) The box plot showed that there were differences in the infiltration of some immune cells between the high and low risk groups; (D) Box plot showed that most of the HLA-related genes were highly expressed in the low-risk group; (E) Box plot showed that most of the immune checkpoint genes were highly expressed in the low-risk group; (F) The violin plot showed that the scores of tumor immune dysfunction and exclusion in the low-risk group were lower than those in the high-risk group; (G and H) The violin plot showed that the immune cell proportion score of the low-risk group was higher than that of the high-risk group.



**Figure 9.** Expression of feature genes in different immune cell types. (A) UMAP plot of 6 major cell clusters in TME. (B) Distribution of RIC3 in cells. (C) Expression of feature genes in different immune cell types. Alt-text (A) Uniform manifold approximation and projection display GSE99254 was divided into 6 cell clusters; (B) The points in the figure represent RIC3, showing the expression and distribution of RIC3 in different cell clusters; (C) The violin plot showed that the characteristic genes RIC3, KRT6C, ABCC12, ATP8A2, NTSR1 and KRT76 were expressed in 6 cell clusters, but the expression levels were different.



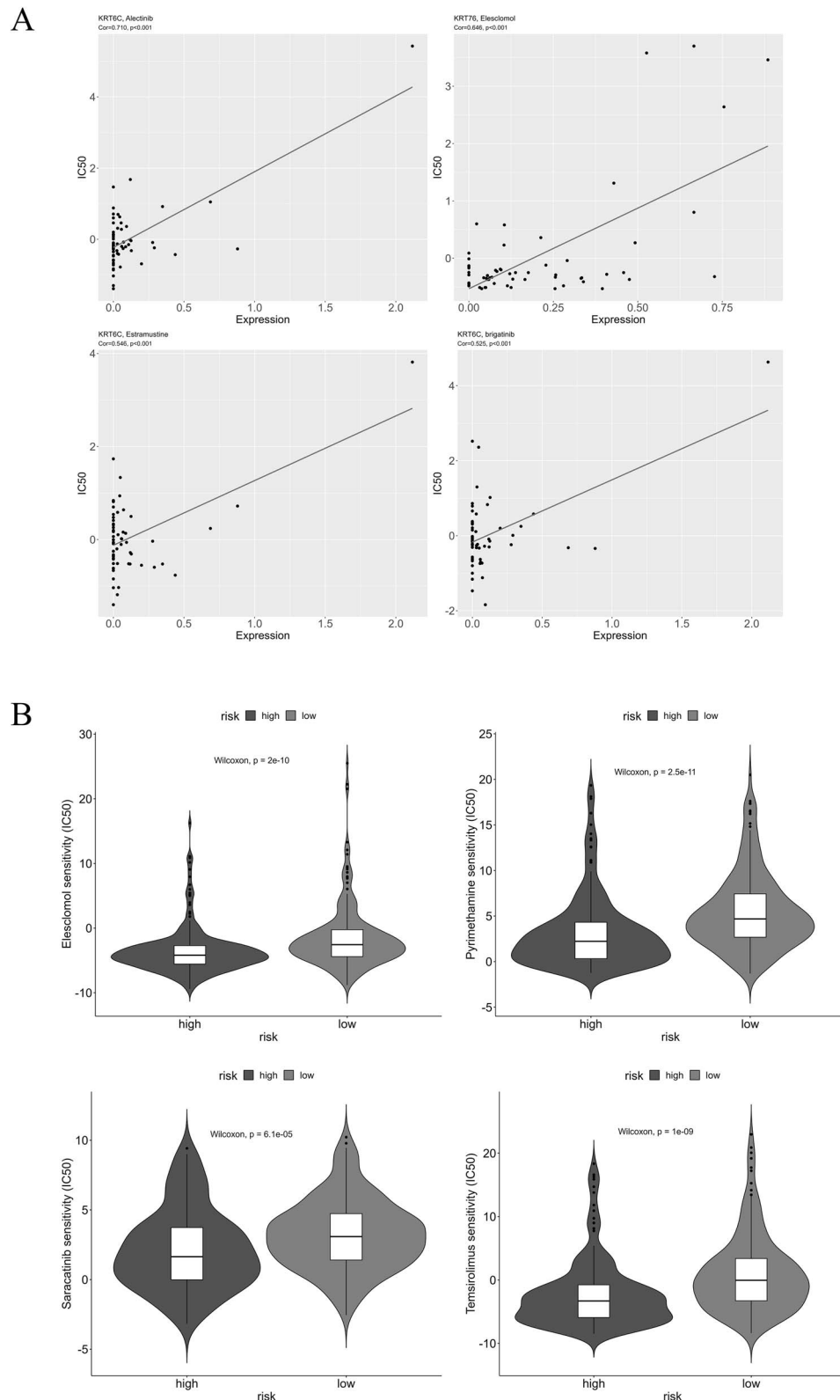
**Figure 10.** Tumor mutation analysis of high- and low-risk groups. (A) Violin plot of TMB in high- and low-risk groups. (B) Waterfall plot of top 30 mutated genes in the high-risk group. (C) Waterfall plot of top 30 mutated genes in the low-risk group. Alt-text (A) The violin plot showed that the tumor mutation burden in the high-risk group was higher than that in the low-risk group; (B and C) The waterfall diagram was used to analyze the tumor mutation burden in the high and low risk groups, and to show the genes with a greater degree of mutation that may affect the tumor.

affinity scores of  $-7.7$  (Fig. 12E) and  $-7.1$  kcal/mol (Fig. 12F) for Differin and Nilotinib, respectively. The three protein targets can form hydrophobic interactions and van der Waals forces with the compounds and bind stably. Dihydroergotamine (with NTSR1 and KRT76) formed hydrogen bonds on Glu165, Glu276, and Asn422 residues. Ergotamine (with KRT76) formed hydrogen bonds on five residues such as Glu267 and Lys271. Nilotinib (with KRT6C) forms hydrogen bonds on the Tyr253 residue, which makes the binding stronger. This indicates that there are differences in the sensitivity of high- and low-risk groups to some chemotherapeutic drugs, which suggests that attention should be paid to differences in the use of chemotherapeutic drugs in different groups of patients.

**The expression of key gene NTSR1 affects the sensitivity of A549 cells to irinotecan**  
Because NTSR1 had the highest binding energy with irinotecan, we used cell experiments to verify the relationship between

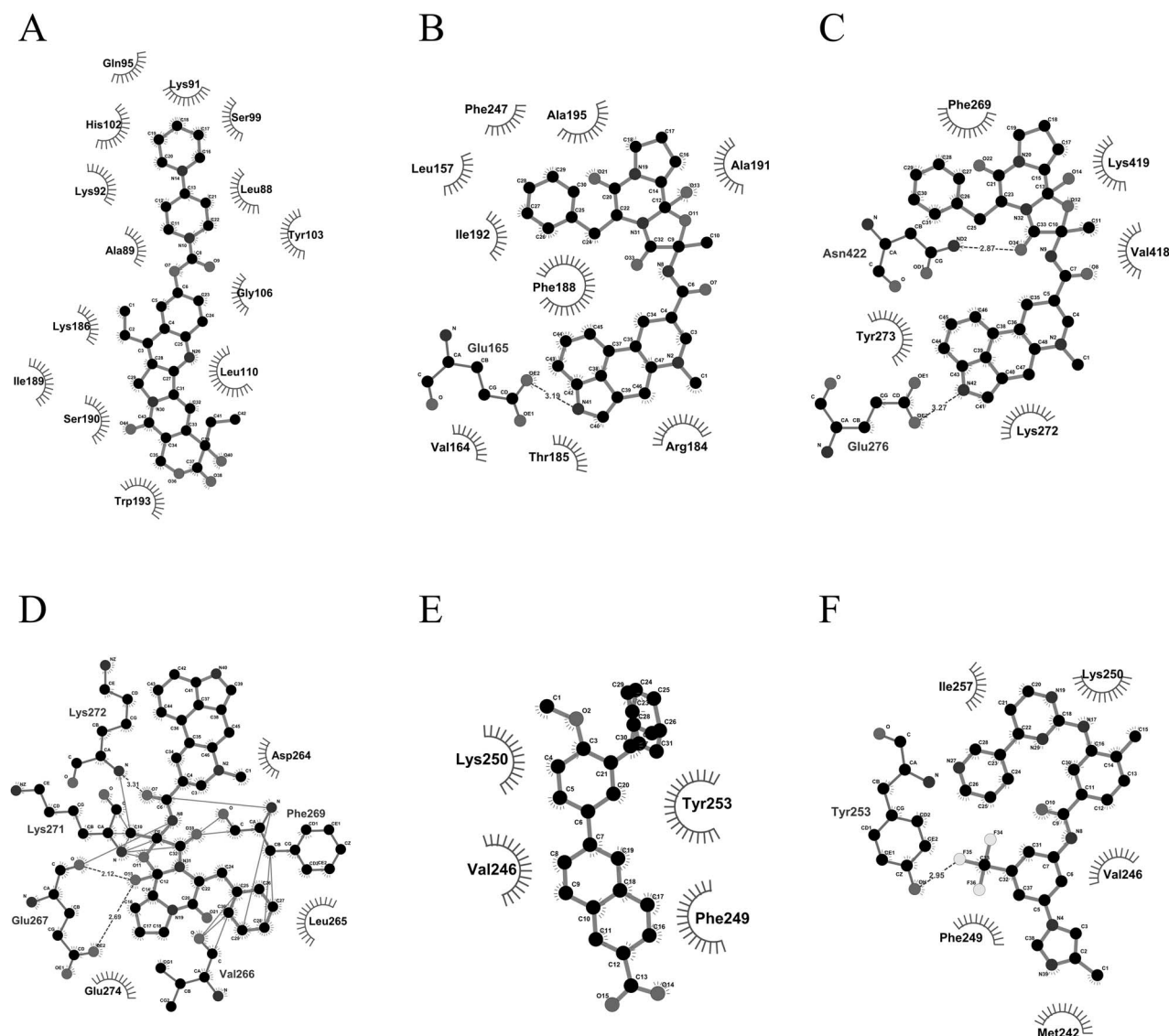
Table 1. Virtual screening affinity scores.		
Target	Molecular	Affinity (kcal/mol)
NTSR1	Irinotecan	$-10.8$
NTSR1	Dihydroergotamine	$-10.4$
KRT76	Dihydroergotamine	$-8.9$
KRT76	Ergotamine	$-8.6$
KRT6C	Differin	$-7.7$
KRT6C	Nilotinib	$-7.1$

NTSR1 expression and irinotecan. Firstly, based on TCGA-LUAD dataset, the expression of NTSR1 in LUAD was analyzed. The results showed that NTSR1 was highly expressed in LUAD compared with normal tissues ( $P < 0.05$ ) (Fig. 13A), which was also verified in cell experiments (Fig. 13B). Therefore, we silenced the



**Figure 11.** Drug sensitivity analysis. (A) Plot of gene expression and IC<sub>50</sub>. (B) Violin plots of the sensitivity to the four drugs in high- and low-risk groups. Alt-text (A) Pearson correlation analysis showed that the expression of the characteristic gene KRT76C was positively correlated with the IC<sub>50</sub> values of Alectinib, Estramustine, and Brigatinib, and the same was true between KRT76 and Eleclomol; (B) The violin plot showed that the IC<sub>50</sub> values of Eleclomol, Pyrimethamine, Saracatinib, and Temsirolimus in the low-risk group were higher than those in the high-risk group.





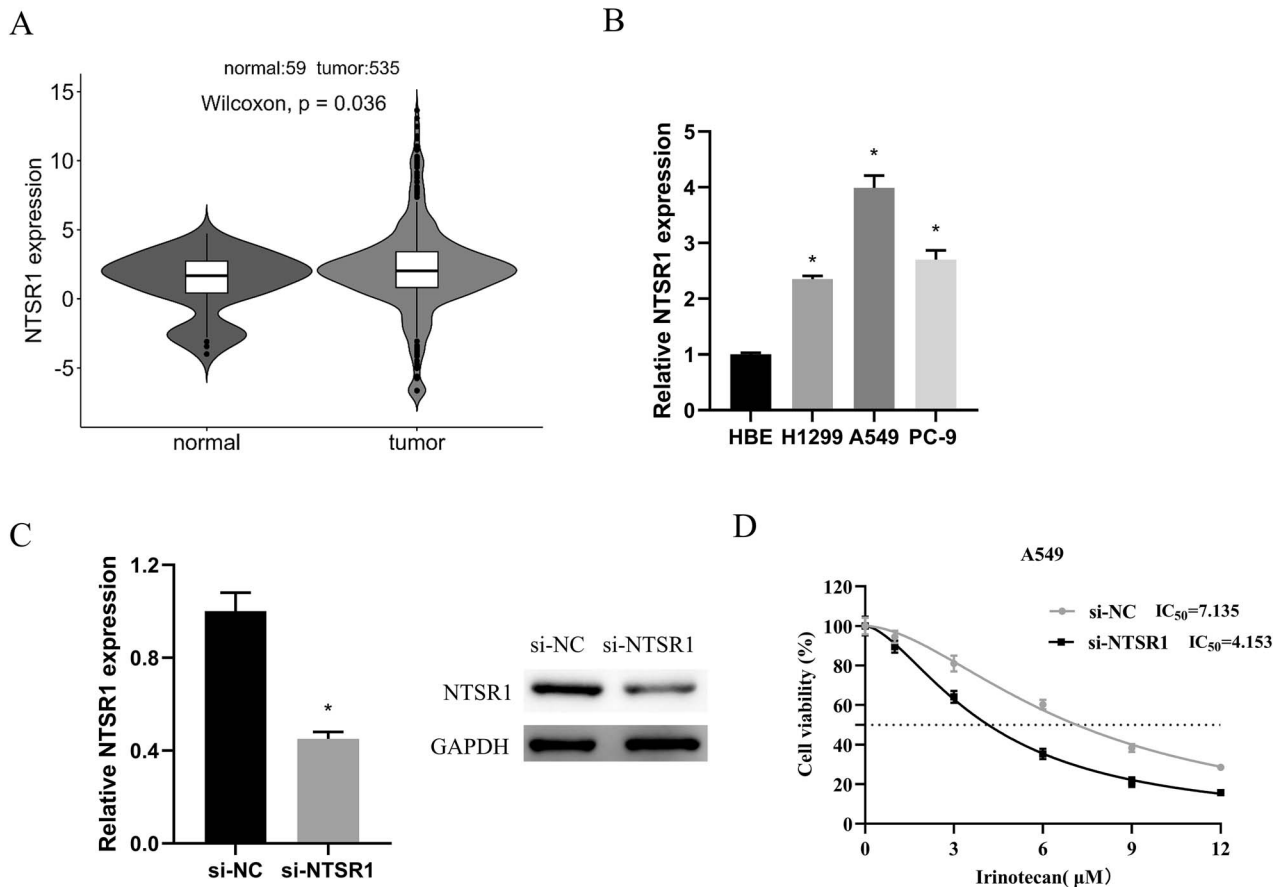
**Figure 12.** Virtual screening of drugs. (A) Two-dimensional display of the docking of NTSR1 with Irinotecan. (B) Two-dimensional display of the docking of NTSR1 with Dihydroergotamine. (C) Two-dimensional display of the docking of KRT76 with Dihydroergotamine. (D) Two-dimensional display of the docking of KRT76 with ergotamine. (E) Two-dimensional display of the docking of KRT6C with Differin. (F) Two-dimensional display of the docking of KRT6C with Nilotinib. Semicircle: Hydrophobic interaction. Alt-text (A–F) The binding modes of compounds to proteins showed the interaction modes between NTSR1 and irinotecan and dihydroergotamine, KRT76 and dihydroergotamine and ergotamine, KRT6C and Differin and Nilotinib, including hydrophobic interaction, van der Waals force and hydrogen bond.

expression of NTSR1 (Fig. 13C) and detected the change of  $IC_{50}$  value after irinotecan treatment of A549 cells with low expression of NTSR1. The results showed that the low expression of NTSR1 significantly reduced the  $IC_{50}$  value ( $IC_{50} = 4.153$ ) (Fig. 13D). Low expression of NTSR1 increased the sensitivity of A549 cells to irinotecan.

## Discussion

This study found that the prognosis of cluster 2 was worse, indicating a certain correlation between m6A/m5C/m1A and LUAD prognosis. We conducted a regression analysis of DEGs related to cluster 2 and ultimately constructed a 6-gene LUAD prognostic risk scoring model. The AUC values of the ROC curve for the performance evaluation of our model at 1 year, 3 years, and 5 years were 0.78, 0.71 and 0.69, respectively. The AUC values of the prognostic model constructed by Luo et al. [31] based

on hypoxia-related genes were 0.649 (1 year), 0.667 (3 years), and 0.575 (5 years). The 3-year and 5-year AUC values of the autophagy-related prognostic model constructed by Fei et al. [23] were 0.699 and 0.673, respectively. Compared with these models, the AUC value of our model is relatively high, and the AUC value in the validation set (1 year = 0.67, 3 years = 0.71, 5 years = 0.85) is relatively stable, indicating that the performance of our model is relatively good. Of course, the comparison with other LUAD prognostic models needs to be further explored. Additionally, the high-risk group exhibited higher expression levels of feature genes such as KRT76, NTSR1, and KRT6C, whereas the low-risk group showed higher expression levels of ATP8A2, ABCC12, and RIC3 ( $P < 0.05$ ). Some of these genes have been found to have vital functions in LUAD, but some genes have not been deeply probed into. For instance, NTSR1 is the neurotensin receptor 1. Ou-Yang et al. [32] found that knocking down NTSR1 can weaken the invasion of glioblastoma cells, and this anti-invasion effect is produced via the Jun/miR-494/SOCS6 axis. Huang et al. [33] also found in their



**Figure 13.** The effect of the expression of key gene NTSR1 on the sensitivity of LUAD cells to irinotecan. (A) Based on the TCGA database, the expression of NTSR1 in LUAD and normal tissues; (B) The expression of NTSR1 in human bronchial epithelial cell line and LUAD cell lines detected by qRT-PCR; (C) The expression of NTSR1 after transfection of si-NC and si-NTSR1 detected by qRT-PCR and western blot; (D) CCK-8 detection of the IC<sub>50</sub> value of irinotecan after LUAD cells were transfected with si-NC and si-NTSR1. \* $P < 0.05$ , indicating statistical significance. Alt-text (A) Violin diagram showed that NTSR1 was highly expressed in lung adenocarcinoma tumor tissues; (B) The expression of NTSR1 in lung adenocarcinoma cell lines (H1299, A549, PC-9) was higher than that in normal cells (HBE) by qRT-PCR; (C) The expression of NTSR1 was decreased after transfection of NTSR1 silencing plasmid by qRT-PCR and western blot; (D) CCK-8 method was used to detect the cells treated with different concentrations of Irinotecan with or without silencing NTSR1. The IC<sub>50</sub> value of the cells after silencing NTSR1 was lower than that of the non-silencing group.

study on LUAD that the synergistic effect of MAP3K19 and NTSR1 has clinical significance for the prognosis and immunotherapy of LUAD individuals. Although these outcomes are consistent with the conclusions of this study, they still need prospective studies to verify. KRT6C is one of the differentiation-specific keratin genes [34]. Hu et al. [35] ascertained the high expression of KRT6C in LUAD tissues and cell lines. After its expression was inhibited, the invasion, proliferation, and migration of LUAD cancer cells were also inhibited through the regulation of epithelial-mesenchymal transition, and this particular gene was recognized as a promising candidate for therapeutic intervention in LUAD. KRT76 is a type II intermediate filament. Mice with KRT76 defects showed increased susceptibility to tumors [36]. The ATP8A2 gene is a member of the P4 ATPase protein family and has been found to have abnormal methylation in various cancer tissues [37, 38]. RIC3 is a methylation-driven gene that can be a prognostic biomarker for pancreatic cancer and pulmonary fibrosis. In our study, RIC3 is also used as a prognostic marker for LUAD, indicating that RIC3 has the potential to be a target in cancer.

Through enrichment analysis, we found significant enrichment of high-risk group genes in the PENTOSE PHOSPHATE PATHWAY, CELL CYCLE, and P53 SIGNALING PATHWAY, etc. ( $P < 0.05$ ). The methylation level of genes in cancer cells is related to glucose

conversion to the pentose phosphate pathway. Tumor cells can extensively utilize the pentose phosphate pathway to synthesize nucleotides [39, 40]. The methyltransferase METTL3 has been reported to regulate gene methylation in pancreatic cancer cells to further regulate the cell cycle and promote cancer progression [41]. Xu et al. [42] found that high methylation of NPTX2 can foster cancer cell proliferation and inhibit cancer cell apoptosis by inhibiting the p53 signaling pathway. These pathways play specific roles in the high-risk group, resulting in lower survival rates in high-risk populations.

Immunotherapy aims to enhance the natural defense to exclude malignant cells, which is a huge breakthrough in cancer treatment. Understanding the degree of immune infiltration in the TME is key to improving the immune response rate and developing new cancer treatment strategies with immunotherapy [43]. In this study, it was found that the immune cells, immune function score, and immune infiltration level in the low-risk group were generally higher than those in the high-risk group ( $P < 0.05$ ). Besides, the low-risk group also had higher HLA expression levels, lower TIDE scores, and higher IPS, which all indicate that cancer individuals in the low-risk group may be more likely to benefit from immunotherapy ( $P < 0.05$ ) [44]. Because some researchers have found that the TIDE score can better evaluate the efficacy

of anti-PD1 and anti-PD-L1 immunotherapy, and by comparing with other immune checkpoint inhibitors, TIDE score is the best predictor of immune checkpoint inhibitor treatment, with stable predictive function [45]. Furthermore, we found that the prognostic feature gene RIC3 (methylation-driven gene) is mainly expressed in CD4<sup>+</sup>T cells and CD8<sup>+</sup>T cells. This is an interesting finding that may be further explored in the future.

Through immune analysis, we found that high-risk patient populations might have a harder time benefiting from immunotherapy. Therefore, this study predicted chemotherapy drugs that could treat high-risk patient populations. The results identified four drugs with higher sensitivity in the high-risk group: Elesclomol, Pyrimethamine, Saracatinib, and Temsirolimus ( $P < 0.05$ ). Elesclomol is an anticancer drug that targets mitochondrial metabolism and can promote the degradation of ATP7A to slow the proliferation of colon cancer cells [46, 47]. Pyrimethamine can inhibit oncogenic proteins in different types of cancer and can synergize with other anticancer agents, such as Temozolomide, to drive tumor cell apoptosis [48]. Saracatinib is an oral tyrosine kinase inhibitor that, in combination with other agents, induces autophagy and apoptosis of lung cancer cells [49]. Temsirolimus is an mTOR inhibitor with significant antitumor effects on many cancer cells, such as lung cancer cells [50], clear cell ovarian cancer [51], and colon cancer [52]. Additionally, in drug virtual screening, we also identified some therapeutic drugs targeting prognostic markers for high-risk groups, which had been reported to have strong anticancer effects in cancer. For example, Estramustine is an effective anticancer cell proliferation drug targeting prostate cancer [53]. Brigatinib is a tyrosine kinase receptor inhibitor and anticancer drug mainly used for the treatment of advanced non-small cell lung cancer [54]. Nilotinib is a tyrosine kinase inhibitor that inhibits the growth and migration of cancer cells in thyroid cancer and breast cancer and promotes cancer cell apoptosis [55, 56]. Irinotecan is a versatile chemotherapy drug that has been used for many years to treat various malignancies in combination with other anticancer drugs [57]. These findings lent additional support to the clinical use of these drugs for the treatment of LUAD. In addition, we found that there were differences in drug sensitivity between high- and low-risk groups. This may be related to the differences in the immune microenvironment and gene mutations in the high- and low-risk groups. In the immune microenvironment, immune cells can inhibit the role of immune cells in a variety of ways, thereby reducing the effect of chemotherapy. It has been reported that the expression of PD-L1 increased significantly after 1 day and 2 days of chemotherapy treatment of breast cancer cells [58], which is obviously unfavorable for the treatment of patients. In our study, there were significant differences in the expression of PD-L1 between the high- and low-risk groups ( $P < 0.01$ ). In addition, some gene mutations can also reduce the sensitivity of tumor cells to chemotherapy. For example, P53 mutation can lead to chemoresistance in glioblastoma [59]. Similarly, there were differences in gene mutations between the high- and low-risk groups in our study ( $P < 0.05$ ), and TP53 mutation ranked first in the high-risk group.

We built a 6-gene risk prognosis model with strong predictive performance that can be an independent prognostic factor for predicting the prognosis of individuals with LUAD. Low-risk group individuals based on this model had higher levels of immune infiltration, which might be more suitable for immunotherapy. For high-risk group individuals who benefited less from immunotherapy, we conducted a series of anti-cancer drug predictions, and most of the predicted drugs had strong anti-tumor effects in

cancer, providing more choices for clinical usage in LUAD individuals. However, some shortcomings still existed, such as the need for clinical validation of the constructed model in the future, and further molecular and cellular experimental research on the molecular mechanisms of the 6 prognostic markers in tumors. In addition, the specific time for patients in the high-risk group to be suitable for immunotherapy is unclear. Overall, our study will offer some help for the prognosis analysis of LUAD individuals.

## Materials and methods

### Data download

As the training dataset, we obtained mRNA expression data (normal: 51, tumor: 647), mutation, copy number variation, and clinical information for LUAD from TCGA (<https://portal.gdc.cancer.gov/>). We got chip data GSE72094 (platform: GPL15048) as the validation set from GEO (<https://www.ncbi.nlm.nih.gov/geo/>). We obtained 50 feature genes (Table S3) for m6A/m5C/m1A from the reference articles [60, 61].

### Differential analysis of m6A/m5C/m1A genes

We utilized the edgeR package [62] to do differential expression analysis on genes in the normal and tumor groups of LUAD. The criteria for screening were set as  $|\log FC| > 0.585$  and  $FDR < 0.05$ . We then obtained the intersection of the DEGs and m6A/m5C/m1A genes to obtain differentially expressed m6A/m5C/m1A genes. We compared the expression of differentially expressed m6A/m5C/m1A genes in the normal and tumor groups, and plotted box plots and correlation heatmaps of these genes. The CNV data from TCGA was utilized, and chi-square tests were conducted to compare the CNVs between the normal and tumor groups. We identified differential m6A/m5C/m1A genes with significant CNV differences and used the RCircos package to plot the specific locations of these genes on chromosomes [63]. We also used the TMB data from TCGA to count the TMB status of differentially expressed m6A/m5C/m1A genes in the tumor group and plotted a waterfall chart using the GenVisR package [64]. KEGG enrichment analysis was performed on the differentially expressed m6A/m5C/m1A genes [65–67].

### Subtype identification of m6A/m5C/m1A genes

To identify m6A/m5C/m1A-related tumor subtypes, we used the NMF algorithm to cluster LUAD tumor samples grounded in the expression profile matrix of differentially expressed m6A/m5C/m1A genes. We utilized the survival package [68] to do survival analysis on clustered samples to identify differences in survival between different subtypes. We then used GSVA [69] to perform ssGSEA analysis on the clustered samples and plotted violin plots of immune scores, stromal scores, ESTIMATE scores, and tumor purity using estimate package [70]. Finally, the expression levels of immune checkpoint genes in the clustered samples were quantified and presented as box plots.

### Construction of a prognostic model by screening prognostic-relevant features

Based on the difference in survival rate between clustered samples, we selected samples from cluster 2 and cluster 1\_3 (The samples of cluster 1 and cluster 3 were combined as the control group) for differential analysis ( $|\log FC| > 1.0$ ,  $FDR < 0.05$ ) to obtain DEGs. Afterward, data for DEGs and clinical information were read in R language. The expression data of DEGs were sorted out, and the samples with expression level less than 0 were deleted. Then, the clinical information and survival information were sorted out and

merged, and the samples with incomplete survival information or survival time of less than 30 days were excluded. The expression matrix of DEGs was matched with clinical information to obtain sample data with both clinical data and expression data of DEGs. Subsequently, We did univariate regression analysis on the DEGs using the “survival” package to identify genes that significantly affect survival as candidate genes (selected according to P-value,  $P < 0.05$ , considered to have a significant effect on survival). In order to avoid overfitting of univariate Cox regression analysis and reduce the complexity of the model (that is, reduce the number of genes), we used the “glmnet” package [71] to fulfill LASSO regression analysis on the candidate genes, and used cross-validation to remove genes with strong correlation between genes during the analysis. Finally, multivariate Cox regression was performed on the genes screened by LASSO analysis using the “survival” package, and the feature genes significantly related to survival were obtained. Based on these genes, a prognostic model was constructed.

In order to evaluate model performance, we initially computed the riskScores of individuals in the training set on the ground of the expression levels and risk coefficients of the feature genes in the model, and divided the samples into high- and low-risk groups by applying the median riskScore. Subsequently, we generated heatmaps to display the distribution of riskScores, survival status, and expression levels for the high- and low-risk groups, and utilized the “timeROC” package [71] to plot a ROC curve and calculate the AUC values at 1-, 3-, and 5-year time points, in order to evaluate the performance of our model. Using the “survival” package, we generated survival curves for both high- and low-risk groups. Furthermore, we validated the model utilizing the GEO validation set, dividing the validation set samples into high- and low-risk groups on the ground of the median riskScore and using the same methods to depict the riskScore distribution plot, survival status distribution plot, expression level heatmap, ROC curve, and survival curve.

### Gene set enrichment analysis (GSEA)

To understand the pathways enriched in two risk groups, we used GSEA (version: 4.1.0) software [72] to perform pathway enrichment analysis. GSEA is a statistical method that identifies biological pathways or functions associated with changes in gene expression by assessing the correlation between predefined gene sets and gene expression data.

### Independence verification of the prognostic model

We performed statistical analysis on the riskScores of each clinical group and presented them in the form of violin plots. Then, we conducted univariate and multivariate regression analyses on the samples according to clinical information and riskScores and made corresponding forest plots to determine if the model can serve as an independent prognostic factor. We then used the “rms” package [73] to create nomograms and predict the 1-year, 3-year, and 5-year survival rates of individuals. We also depicted corresponding calibration curves to validate the predictive effect of the nomograms.

### Immune analysis

We compared the immune microenvironment between the two groups by performing ssGSEA based on the expression levels of immune cell-related marker genes using GSVA and estimate packages. The resulting immune cell and immune function scores were used to create box plots that compared the scores between

the two groups. CIBERSORT algorithm [74] was utilized to calculate the levels of immune infiltration in the two groups and plotted box plots accordingly. We also statistically assayed the levels of immune checkpoint and HLA genes and drew box plots. Furthermore, we calculated the TIDE scores for the two groups using the TIDE (<http://tide.dfci.harvard.edu/login/>) database [75] and plotted violin plots for TIDE scores. We downloaded the IPS for LUAD from TCIA (<https://tcia.at>) database [76] and plotted violin plots of IPS for the two groups.

### Single-cell sequencing analysis

The TISCH database (<http://tisch.comp-genomics.org/>) [77] involves 79 high-quality single-cell transcriptome datasets for 27 types of tumors, primarily from GEO and ArrayExpress databases (<https://ngdc.cncb.ac.cn/databasecommons/database/id/155>), along with corresponding clinical information. The TISCH database provides detailed cell type annotations at the single-cell level and has advantages like comprehensive data, simple operation, user-friendly interface, and data visualization. Using the TISCH database (<http://tisch.comp-genomics.org/home/>), we employed UMAP plots to visualize the distribution and expression patterns of the feature genes in the GSE99254 dataset. Furthermore, we compared the gene expression levels among different immune cell types.

### TMB analysis

Using the mutation data for LUAD, we computed the TMB scores for each sample, performed Wilcoxon tests on the TMB values for the two groups, and plotted violin plots. We also organized and summarized the mutation data for the top 30 genes and used the “GenVisR” package [64] to draw waterfall plots. TMB = Total number of somatic mutations (including non-synonymous point mutations, insertions, and deletions in the exon coding region)/Target area size.

### Drug sensitivity prediction

CellMiner database (<https://discover.nci.nih.gov/cellminer/>) [78] was utilized to screen for anti-tumor drugs significantly correlated with prognostic genes. The “pRRophetic” package [79] in R version (4.1.2) was utilized to predict the half-maximal inhibitory concentration ( $IC_{50}$ ) of different drugs in high-risk and low-risk groups [80]. The lower the  $IC_{50}$  value of a drug, the more effective it was in treating tumors.

### Virtual screening

Virtual screening was employed by utilizing AutoDock Vina 1.5.7 [81] to predict the best candidate small molecule for docking. Differential upregulated feature genes were selected for virtual screening, and the 3D structures predicted by AlphaFold in the Protein Data Bank (<https://www.rcsb.org/>) were selected. 2115 FDA-approved small molecule drugs were extracted from the ZINC database (<https://zinc.docking.org/>). LigPlus+2.2.5 software was used to visualize the interactions between the protein and the small molecules.

### Cell culture and transfection

Human LUAD cell lines (H1299, A549 and PC-9) were purchased from BeNa Culture Collection (China), and human bronchial epithelial cell lines were purchased from ATCC (USA). These cells were cultured in RPMI-1640 medium containing 10% fetal bovine serum (FBS) and placed in a humidified cell incubator containing 5%  $CO_2$ .



si-NTSR1 or si-NC was purchased from GenePharma (China). The cells were added to a 6-well plate and the purchased si-NTSR1 or si-NC was transfected into the LUAD cell line using Lipofectamine 3000 (Thermo Fisher, USA). After 48 h, the cells were collected for subsequent analysis. The transfected LUAD cells were treated with different concentrations (0, 3, 6, 9, 12  $\mu$ M) of Irinotecan (Sigma, USA), and the IC<sub>50</sub> values were calculated. The concentration of Irinotecan used was based on previous studies [82].

### qRT-PCR

The transfected cells were collected. Total RNA was isolated using Trizol (Beyotime, China). Total RNA was reverse transcribed into cDNA with the help of ReverTra-Ace qPCR RT Master Mix kit (Toyobo, Japan) and NTSR1 primers. Finally, RT-PCR was performed using SYBR® Green Real-time PCR Master Mix (Toyobo, Japan). The fold change was measured using the  $2^{-\Delta\Delta CT}$  method and used to calculate the relative RNA level.

### Western blot analysis

The transfected cells were collected and lysed using RIPA lysis buffer (Beyotime, China). Protein concentration was determined by protein quantification kit. The protein was then separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a polyvinylidene fluoride (PVDF) membrane. The membrane was incubated with the primary antibody overnight at 4°C, and then incubated with the secondary antibody at room temperature for 2 h. Finally, according to the instructions, the protein bands were observed using the enhanced chemiluminescence (ECL) substrate kit (Yeast, China). Antibodies included anti-NTSR1 (ab183088, Abcam, UK) and GAPDH (ab9485, Abcam, UK).

### CCK-8 analysis

The transfected cells were seeded in 96-well plates at a concentration of  $3 \times 10^4$  cells/mL. After 24 h of culture, the medium was replaced with an equal volume of medium containing different concentrations of irinotecan. After 24 h, 10  $\mu$ L of CCK-8 solution (Bimake, China) was added to each well and cultured in an incubator for 2 h. The optical density (OD) value was detected by a microplate reader. The IC<sub>50</sub> value of the drug was calculated by statistical analysis software.

### Statistical analysis

Graphpad Prism8.0 was used to analyze experimental data. The experimental data was presented as mean  $\pm$  SD. T-test was used to analyze the significance of differences between two groups.  $P < 0.05$  indicated that the difference was statistically significant.

### Acknowledgements

Not applicable.

### Authors contributions

Hao Ding contributed to the study design. Yuanyuan Teng conducted the literature search. Ping Gao and Qi Zhang acquired the data. Mengdi Wang wrote the article. Yi Yu performed data analysis. Yueping Fan and Li Zhu revised the article and gave the final approval of the version to be submitted. All authors read and approved the final manuscript.

### Supplementary data

Supplementary data is available at HMG Journal online.

### Conflict of interest statement

The authors declare that they have no potential conflicts of interest.

### Funding

This work was supported by 2020 Jiangsu Province High Level Health Talents “Six One Project” Top Talent Project (LGY2020063); 2021 Zhenjiang City Science and Technology Innovation Fund (Key R&D Plan - Social Development) Project (SH2021049); and 2021 Zhenjiang City “Jinshan Talent” High Level Leading Talent Training Plan (2021-169DT-20).

### Data availability

The data and materials in the current study are available from the corresponding author on reasonable request.

### References

1. Sung H, Ferlay J, Siegel RL. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**:209–49.
2. Hirsch FR, Scagliotti GV, Mulshine JL. et al. Lung cancer: current therapies and new targeted treatments. *Lancet* 2017;**389**:299–311.
3. Zhu M, Cui Y, Mo Q. et al. Characterization of m(6)A RNA methylation regulators predicts survival and immunotherapy in lung adenocarcinoma. *Front Immunol* 2021;**12**:782551.
4. Oerum S, Dégut C, Barraud P. et al. m1A post-transcriptional modification in tRNAs. *Biomol Ther* 2017;**7**:20.
5. Liu J, Jia G. Methylation modifications in eukaryotic messenger RNA. *J Genet Genomics* 2014;**41**:21–33.
6. Boccaletto P, Machnicka MA, Purta E. et al. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* 2018;**46**:D303–d307.
7. Boccaletto P, Stefaniak F, Ray A. et al. MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res* 2022;**50**:D231–5.
8. Chen YS, Yang WL, Zhao YL. et al. Dynamic transcriptomic m(5)C and its regulatory role in RNA processing. *Wiley Interdiscip Rev RNA* 2021;**12**:e1639.
9. Teng C, Kong F, Mo J. et al. The roles of RNA N(6)-methyladenosine in esophageal cancer. *Heliyon* 2022;**8**:e11430.
10. Zhang X, Su H, Chen H. et al. RNA modifications in gastrointestinal cancer: current status and future perspectives. *Biomedicine* 2022;**10**:1918.
11. Cui Q, Shi H, Ye P. et al. m(6)A RNA methylation regulates the self-renewal and tumorigenesis of glioblastoma stem cells. *Cell Rep* 2017;**18**:2622–34.
12. Su R, Dong L, Li C. et al. R-2HG exhibits anti-tumor activity by targeting FTO/m(6)A/MYC/CEBPA Signaling. *Cell* 2018;**172**:90–105.e23.
13. Weng H, Huang H, Wu H. et al. METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes Leukemogenesis via mRNA m(6)A modification. *Cell Stem Cell* 2018;**22**:191–205.e9.



14. Yang S, Wei J, Cui YH. et al. m(6)A mRNA demethylase FTO regulates melanoma tumorigenicity and response to anti-PD-1 blockade. *Nat Commun* 2019;**10**:2782.
15. Li J, Han Y, Zhang H. et al. The m6A demethylase FTO promotes the growth of lung cancer cells by regulating the m6A level of USP7 mRNA. *Biochem Biophys Res Commun* 2019;**512**:479–85.
16. Liu T, Hu X, Lin C. et al. 5-methylcytosine RNA methylation regulators affect prognosis and tumor microenvironment in lung adenocarcinoma. *Ann Transl Med* 2022;**10**:259.
17. Bao G, Li T, Guan X. et al. Comprehensive analysis of the function, immune profiles, and clinical implication of m1A regulators in lung adenocarcinoma. *Front Oncol* 2022;**12**:882292.
18. Zhang W, Wan S, Qu Z. et al. Establishment of a prognostic signature for lung adenocarcinoma by integration of 7 pyroptosis-related genes and cross-validation between the TCGA and GEO cohorts: a comprehensive bioinformatics analysis. *Medicine (Baltimore)* 2022;**101**:e29710.
19. Zhang H, Shi Y, Yi Q. et al. A novel defined cuproptosis-related gene signature for predicting the prognosis of lung adenocarcinoma. *Front Genet* 2022;**13**:975185.
20. Wang Y, Xu J, Fang Y. et al. Comprehensive analysis of a novel signature incorporating lipid metabolism and immune-related genes for assessing prognosis and immune landscape in lung adenocarcinoma. *Front Immunol* 2022;**13**:950001.
21. Liu J, Shen H, Gu W. et al. Prediction of prognosis, immunogenicity and efficacy of immunotherapy based on glutamine metabolism in lung adenocarcinoma. *Front Immunol* 2022;**13**:960738.
22. Gu X, Cai L, Luo Z. et al. Identification and validation of a muscle failure index to predict prognosis and immunotherapy in lung adenocarcinoma through integrated analysis of bulk and single-cell RNA sequencing data. *Front Immunol* 2022;**13**:1057088.
23. Fei Y, Xu J, Ge L. et al. Establishment and validation of individualized clinical prognostic markers for LUAD patients based on autophagy-related genes. *Aging (Albany NY)* 2022;**14**:7328–47.
24. Yang Y, Huang T, Fan Y. et al. Significance of spliceosome-related genes in the prediction of prognosis and treatment strategies for lung adenocarcinoma. *Biomed Res Int* 2022;**2022**:1753563.
25. Yu H, Zhang W, Xu XR. et al. Drug resistance related genes in lung adenocarcinoma predict patient prognosis and influence the tumor microenvironment. *Sci Rep* 2023;**13**:9682.
26. Yu S, Tang L, Zhang Q. et al. A cuproptosis-related lncRNA signature for predicting prognosis and immunotherapy response of lung adenocarcinoma. *Hereditas* 2023;**160**:31.
27. Ren Q, Zhang P, Lin H. et al. A novel signature predicts prognosis and immunotherapy in lung adenocarcinoma based on cancer-associated fibroblasts. *Front Immunol* 2023;**14**:1201573.
28. Yang F, Zhao Y, Huang X. et al. A cell differentiation trajectory-related signature for predicting the prognosis of lung adenocarcinoma. *Genet Res (Camb)* 2022;**2022**:3483498.
29. Diao X, Guo C, Li S. Identification of a novel anoikis-related gene signature to predict prognosis and tumor microenvironment in lung adenocarcinoma. *Thorac Cancer* 2023;**14**:320–30.
30. Ye W, Huang Y, Li X. Cuproptosis-related gene signatures for predicting prognosis of lung adenocarcinoma. *Medicine (Baltimore)* 2022;**101**:e30446.
31. Luo J, Du X. A promising prognostic signature for lung adenocarcinoma (LUAD) patients basing on 6 hypoxia-related genes. *Medicine (Baltimore)* 2021;**100**:e28237.
32. Ou-Yang Q, He X, Yang A. et al. Interference with NTSR1 expression exerts an anti-invasion effect via the Jun/miR-494/SOCS6 Axis of glioblastoma cells. *Cell Physiol Biochem* 2018;**49**:2382–95.
33. Huang X, Lu Z, He M. et al. A prognostic risk model of a novel oxidative stress-related signature predicts clinical prognosis and demonstrates immune relevancy in lung adenocarcinoma. *Oxidative Med Cell Longev* 2022;**2022**:2262014.
34. Forrest CE, Casey G, Mordaunt DA. et al. Pachyonychia Congenita: a Spectrum of KRT6a mutations in Australian patients. *Pediatr Dermatol* 2016;**33**:337–42.
35. Hu HB, Yang XP, Zhou PX. et al. High expression of keratin 6C is associated with poor prognosis and accelerates cancer proliferation and migration by modulating epithelial-mesenchymal transition in lung adenocarcinoma. *Genes Genomics* 2020;**42**:179–88.
36. Sequeira I, Watt FM. The role of keratins in modulating carcinogenesis via communication with cells of the immune system. *Cell Stress* 2019;**3**:136–8.
37. Yan H, Guan Q, He J. et al. Individualized analysis reveals CpG sites with methylation aberrations in almost all lung adenocarcinoma tissues. *J Transl Med* 2017;**15**:26.
38. Ohara K, Arai E, Takahashi Y. et al. Genes involved in development and differentiation are commonly methylated in cancers derived from multiple organs: a single-institutional methylome analysis using 1007 tissue specimens. *Carcinogenesis* 2017;**38**:bgw209–51.
39. Yamamoto T, Takano N, Ishiwata K. et al. Reduced methylation of PFKFB3 in cancer cells shunts glucose towards the pentose phosphate pathway. *Nat Commun* 2014;**5**:3480.
40. Thomas AA, Le Huerou Y, De Meese J. et al. Synthesis, in vitro and in vivo activity of thiamine antagonist transketolase inhibitors. *Bioorg Med Chem Lett* 2008;**18**:2206–10.
41. Tatekawa S, Tamari K, Chijimatsu R. et al. N(6)-methyladenosine methylation-regulated polo-like kinase 1 cell cycle homeostasis as a potential target of radiotherapy in pancreatic adenocarcinoma. *Sci Rep* 2022;**12**:11074.
42. Xu G, Fan L, Zhao S. et al. Neuronal pentraxin II (NPTX2) hypermethylation promotes cell proliferation but inhibits cell cycle arrest and apoptosis in gastric cancer cells by suppressing the p53 signaling pathway. *Bioengineered* 2021;**12**:1311–23.
43. Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell Mol Immunol* 2020;**17**:807–21.
44. Jiang P, Gu S, Pan D. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 2018;**24**:1550–8.
45. Marcu A, Bichmann L, Kuchenbecker L. et al. HLA ligand atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J Immunother Cancer* 2021;**9**:e002071.
46. Zheng P, Zhou C, Lu L. et al. Elesclomol: a copper ionophore targeting mitochondrial metabolism for cancer therapy. *J Exp Clin Cancer Res* 2022;**41**:271.
47. Gao W, Huang Z, Duan J. et al. Elesclomol induces copper-dependent ferroptosis in colorectal cancer cells via degradation of ATP7A. *Mol Oncol* 2021;**15**:3527–44.
48. Ramchandani S, Mohan CD, Mistry JR. et al. The multifaceted antineoplastic role of pyrimethamine against human malignancies. *IUBMB Life* 2022;**74**:198–212.
49. Chiu LY, Hsin IL, Tsai JN. et al. Combination treatment of Src inhibitor Saracatinib with GMI, a Ganoderma microsporium immunomodulatory protein, induce synthetic lethality via autophagy and apoptosis in lung cancer cells. *J Cell Physiol* 2021;**236**:1148–57.

50. Kim SY, Jeong EH, Lee TG. et al. The combination of trametinib, a MEK inhibitor, and temsirolimus, an mTOR inhibitor, Radiosensitizes lung cancer cells. *Anticancer Res* 2021;**41**:2885–94.
51. Farley JH, Brady WE, O'Malley D. et al. A phase II evaluation of temsirolimus with carboplatin and paclitaxel followed by temsirolimus consolidation in clear cell ovarian cancer: an NRG oncology trial. *Gynecol Oncol* 2022;**167**:423–8.
52. Wang HW, Yang SH, Huang GD. et al. Temsirolimus enhances the efficacy of cetuximab in colon cancer through a CIP2A-dependent mechanism. *J Cancer Res Clin Oncol* 2014;**140**:561–71.
53. Karmakar S, Kostrohova H, Ctvrtlikova T. et al. Platinum(IV)-Estramustine multiaction prodrugs are effective antiproliferative agents against prostate cancer cells. *J Med Chem* 2020;**63**:13861–77.
54. LiverTox: Clinical and Research Information on Drug-Induced Liver Injury. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases, 2012, in press.
55. Wang S, Xie Y, Bao A. et al. Nilotinib, a discoidin domain receptor 1 (DDR1) inhibitor, induces apoptosis and inhibits migration in breast cancer. *Neoplasma* 2021;**68**:975–82.
56. Meng L, Zhao P, Hu Z. et al. Nilotinib, a tyrosine kinase inhibitor, suppresses the cell growth and triggers autophagy in papillary thyroid cancer. *Anti Cancer Agents Med Chem* 2022;**22**:596–602.
57. Bailly C. Irinotecan: 25 years of cancer treatment. *Pharmacol Res* 2019;**148**:104398.
58. Majidi M, Safaee S, Amini M. et al. The effects of chemotherapeutic drugs on PD-L1 gene expression in breast cancer cell lines. *Med Oncol* 2021;**38**:147.
59. Afjei R, Sadeghipour N, Kumar SU. et al. A new Nrf2 inhibitor enhances chemotherapeutic effects in glioblastoma cells carrying p53 mutations. *Cancers (Basel)* 2022;**14**:6120.
60. Li D, Li K, Zhang W. et al. The m6A/m5C/m1A regulated gene signature predicts the prognosis and correlates with the immune status of hepatocellular carcinoma. *Front Immunol* 2022;**13**:918140.
61. Wang E, Li Y, Ming R. et al. The prognostic value and immune landscapes of a m(6)A/m(5)C/m(1)A-related lncRNAs signature in head and neck squamous cell carcinoma. *Front Cell Dev Biol* 2021;**9**:718974.
62. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40.
63. Naquin D, d'Aubenton-Carafa Y, Thermes C. et al. CIRCUS: a package for Circos display of structural genome variations from paired-end and mate-pair sequencing data. *BMC Bioinformatics* 2014;**15**:198.
64. Skidmore ZL, Wagner AH, Lesurf R. et al. GenVisR: genomic visualizations in R. *Bioinformatics* 2016;**32**:3012–4.
65. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019;**28**:1947–51.
66. Kanehisa M, Furumichi M, Sato Y. et al. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 2023;**51**:D587–92.
67. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30.
68. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. *Springer-Verlag* 2000;**20**:2053–4.
69. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;**14**:7.
70. Yoshihara K, Shahmoradgoli M, Martínez E. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;**4**:2612.
71. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22.
72. Subramanian A, Tamayo P, Mootha VK. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;**102**:15545–50.
73. Huang C, Liu Z, Xiao L. et al. Clinical significance of serum CA125, CA19-9, CA72-4, and fibrinogen-to-lymphocyte ratio in gastric cancer with peritoneal dissemination. *Front Oncol* 2019;**9**:1159.
74. Newman AM, Steen CB, Liu CL. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;**37**:773–82.
75. Fu J, Li K, Zhang W. et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med* 2020;**12**:21.
76. Guo JN, Chen D, Deng SH. et al. Identification and quantification of immune infiltration landscape on therapy and prognosis in left- and right-sided colon cancer. *Cancer Immunol Immunother* 2022;**71**:1313–30.
77. Sun D, Wang J, Han Y. et al. TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res* 2021;**49**:D1420–30.
78. Wang Z, Dai Z, Zheng L. et al. Ferroptosis activation scoring model assists in chemotherapeutic Agents' selection and mediates cross-talk with Immunocytes in malignant glioblastoma. *Front Immunol* 2021;**12**:747408.
79. Gleeleher 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.
80. Qu X, Zhao X, Lin K. et al. M2-like tumor-associated macrophage-related biomarkers to construct a novel prognostic signature, reveal the immune landscape, and screen drugs in hepatocellular carcinoma. *Front Immunol* 2022;**13**:994019.
81. Nguyen PTV, Huynh HA, Truong DV. et al. Exploring Aurone derivatives as potential human pancreatic lipase inhibitors through molecular docking and molecular dynamics simulations. *Molecules* 2020;**25**:4657.
82. Park IH, Kim JY, Choi JY. et al. Simvastatin enhances irinotecan-induced apoptosis in human non-small cell lung cancer cells by inhibition of proteasome activity. *Investig New Drugs* 2011;**29**:883–90.