Discover Oncology

Research

RNA methylated long non-coding RNAs as potential biomarkers for prognosis prediction in patients with lung adenocarcinoma: development of a risk assessment model

Xin Lin[1] · Jialin Cui[1] · Yangyang Cheng[1] · Huimin Xu[1] · Wanlin Xie[1] · Jingya Zeng[1] · Yihua Sun[1,2]

Received: 5 November 2024 / Accepted: 13 May 2025
Published online: 29 May 2025

Abstract

Introduction Lung adenocarcinoma (LUAD) is the most prevalent form of lung cancer worldwide. Long non-coding RNA (lncRNAs) are non-protein coding RNAs that are involved in lung cancer. This study aimed to develop a lncRNA-based risk assessment model based on RNA methylation to evaluate the prognosis of patients with LUAD.

Method The TCGA-LUAD dataset consisted of 524 primary tumor samples and 59 normal samples, and the validation set (GSE3121011), which included 246 patients with LUAD, was used for this analysis.Pearson's correlation analysis was used to identify lncRNAs associated with RNA methylation in LUAD. Univariate, least absolute shrinkage and selection operators, and multivariate Cox analyses were used to construct the prognostic model. Gene oncology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were used to identify enriched biological processes. Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were used to visualize the dataset and gene set variation analyses. Kaplan–Meier and decision curve analyses were used to assess the accuracy. qRT-PCR was used to verify the expression of lncRNAs. Invasion and cell scratch assays were conducted to evaluate migration capacity, and colony formation experiments were performed to assess proliferation ability.

Result Ten RNA methylation-associated lncRNAs were identified to construct risk features. According to the risk model, the patients were categorized into low- and high-risk groups, with the latter exhibiting a less favorable prognosis. The expression levels of the lncRNAs exceeded those in lung epithelial cells. After siRNA transfection, the proliferation and migration abilities of the tumor cells were significantly reduced. The risk-scoring model may be a potential indicator for predicting the sensitivity of patients with LUAD to immunotherapy.

Conclusion The model constructed in this study can accurately predict the prognosis of patients with LUAD, and holds promise for future immunotherapies.

Keywords RNA modification · long non-coding RNA · Lung adenocarcinoma · Risk scoring · Bioinformatics

Xin Lin and Jialin Cui have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02693-y.

✉ Yihua Sun, 600611@hrbmu.edu.cn | [1]Harbin Medical University Cancer Hospital, Harbin 150000, China. [2]Department of Clinical Laboratory, Harbin Medical University Cancer Hospital, Baojian Rd, Harbin 150081, China.

1 Introduction

Lung cancer, originating from the bronchial mucosal epithelium and alveoli, has the highest mortality rate worldwide. An overall 5-year survival rate of approximately 16% poses a serious threat to human health [1]. Lung cancer is broadly categorized into two types, viz., small and non-small cell lung cancer. Among these, non-small cell lung cancer accounts for approximately 85% of all cases, with adenocarcinoma accounting for 60%, squamous cell carcinoma for 30%, and other subtypes constituting the remaining 10%. Lung adenocarcinoma (LUAD) has the highest incidence among all lung cancers, and its onset and progression are influenced by numerous factors.

Gene expression is closely related to various mammalian developmental processes, including stem cell self-renewal and differentiation, embryonic development, and cancer susceptibility. In recent years, substantial progress has been made in the study of epigenetic mechanisms such as DNA methylation, histone covalent modification, chromatin remodeling, non-coding RNA activity, and RNA chemical modification. These mechanisms are associated with different forms of tumorigenesis, malignancies, and resistance to therapeutic drugs [2, 3]. To date, more than 170 RNA modifications have been identified, including N6 adenylate methylation (m6A), N7 methylguanosine (m7G), N1 adenylate methylation (m1A), and cytosine hydroxylation (m5C) [4–6]. Abnormal RNA modification has emerged as a crucial research avenue for regulating the growth of various malignant tumors; for example, the m5C- and m6A-modified long non-coding RNA (lncRNA) NKILA accelerates the progression of cholangiocarcinoma through the miR-582-3p-YAP1 axis [7]. In gastric cancer, lncRNA DIAPH2-AS1 enhances the m5C modification of NTN1 by stabilizing NSUN2, thus promoting neural invasion in gastric cancer [8]. These findings offer new insights into cancer from the perspective of RNA modifications and the exploration of novel treatment methods.

LncRNAs represent a class of single-stranded RNA that lack protein-coding functionality. They exert regulatory control over diverse biological processes, including histone processing and transcriptional regulation, thereby promoting cancer cell proliferation, metastasis, and invasion [9, 10]. Numerous studies have demonstrated a close association between m5C- and m6A-related lncRNAs and cancer progression and patient prognosis across various cancer types, including breast, liver, and prostate cancer, as well as glioma. However, the RNA modifications associated with LUAD prognosis and their associated lncRNAs remain unclear.

Therefore, the aim of this study is to identify RNA-modifying genes associated with LUAD prognosis and construct a risk assessment model based on lncRNAs associated with key RNA-modifying genes. Additionally, it also intends to investigate their potential underlying mechanisms and predictive abilities.

2 Materials and methods

2.1 Data acquisition

The gene expression profile data related to LUAD was downloaded from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) in FPKM-UQ format. The TCGA-LUAD dataset contains transcriptome data from 585 samples, comprising 524 primary tumor samples (01A) and 59 normal samples (11A), which were used in this analysis. Additionally, the clinical data (phenotype) of the patients from the dataset was obtained from the UCSC Xena database (Supplementary Table 1). The clinical data included age, survival status [including overall survival (OS), progression-free survival (PFS), disease-specific survival (DSS), and disease-free survival (DFS)], follow-up time, and cancer stage. The gene expression profile data of patients with LUAD were matched with their clinical information, using 511 patients as the training set for the prognosis model. The LUAD-related dataset, GSE31210 [11], was obtained from the GEO database to serve as a validation set for the prognosis model. The GSE31210 dataset, derived from the GPL570 sequencing platform, included 246 patients with LUAD (species source: *Homo Sapiens*).

Here, 81 RNA modification-related genes [6] from the literature, including those related to m1A, m5C, m6A, and m7G modifications were extracted (Supplementary Table 2). The intersection of these genes with the gene expression profile data of patients with LUAD resulted in the identification of 75 RNA modification-related genes.

2.2 Identification of RNA-modified genes

To analyze the impact of the expression levels of RNA modification-related genes on the occurrence and development of LUAD, the R package limma [12] was used to identify RNA modified differentially expressed genes (DEGs) between LUAD tumors and control tissues. A p-value < 0.05 was considered significantly different.

Subsequently, to evaluate the prognostic value of RNA-modified genes in LUAD, the 511 patients with OS information were divided into high- and low-expression groups based on the best cutoff value for RNA-modified gene expression levels. Key RNA-modifying genes associated with LUAD were identified by determining the intersection between the RNA-modifying DEGs and those related to LUAD prognosis. The R package maftools [13] was used to depict the mutation landscape of the RNA-modified genes in patients with LUAD.

2.3 Identification and construction of lncRNA models

Using the R package Hmisc, a Pearson's correlation analysis was used to identify lncRNAs related to key RNA-modifying genes in the entire dataset. The criteria used for selection were |Pearson R |> 0.3 and $p < 0.05$.

To analyze the influence of the expression levels of key RNA-modifying gene-related lncRNAs on LUAD prognosis, a prognostic scoring system was established. Univariate Cox regression analysis was used to screen candidate genes from the lncRNAs. Subsequently, the least absolute shrinkage and selection operator (LASSO-Cox) method was used to further screen candidate genes obtained from the univariate Cox regression. This was accomplished using the R package glmnet [14], with the optimal lambda value selected. Following regression analyses, only ten characteristic genes were retained and a prognostic model based on these selected genes was constructed. An optimal prognosis model was constructed using ten-fold cross-validation screening. A risk-scoring formula was established for each normalized gene expression value weighted by the penalty coefficient of the characteristic gene to predict the OS rate of patients with LUAD.

$$\text{riskScore} = \sum_{i} \text{Coefficient}\left(\text{gene}_i\right) * \text{mRNA Expression}\left(\text{gene}_i\right)$$

Kaplan–Meier analysis was conducted using the R packages survival [15] and surviver [16], with a log-rank test to evaluate the prognostic value of the risk score. The expression data of characteristic genes in the GSE31210 dataset and their corresponding prognostic data were used for verification.

2.4 Identification of biological characteristics related to risk score

The R package limma [12] was used to screen DEGs associated with the risk score. These genes were defined by an absolute value of log2 fold change (log2FC) > 1 and a threshold value of $p < 0.05$. Genes with FC > 1.5 and $p < 0.05$ were upregulated, whereas those with FC < – 1.5 and $p < 0.05$ were downregulated. For the gene oncology (GO) function [17] annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway [18] enrichment analysis, the R package clusterProfiler [19, 20] was used to identify significantly enriched biological processes, with statistical significance set at $p < 0.05$. Based on the gene expression profile dataset of LUAD samples, the reference gene sets "c5.go.v7.4. entrez. gmt" and "c2.cp.kegg.v7.4. entrez.gmt" were obtained from the MSigDB database [21]. Enrichment analysis and dataset visualization were conducted using the gene set enrichment analysis (GSEA) [22] method in the R package clusterProfiler. Statistical significance was set at an adjusted p-value < 0.05. Gene set variation analysis (GSVA) was performed using the R package "GSVA" [23]. Fifty landmark pathways were downloaded from the reference gene set "h.all.v7.4. symbols.gmt" in the MSigDB database [21]. For each sample in the dataset, the enrichment score for each hallmark was calculated and Kaplan–Meier analysis was performed to assess the prognostic value of these carcinogenic marker pathways. Statistical significance was set at $p < 0.05$.

2.5 Protein–protein interaction (PPI)

From the STRING database [24], a comprehensive score of 700 was selected as the cutoff value for constructing the PPI network. Subsequently, Cytoscape [25] software was used for further analysis. The MCODE [26] and Cytohubba [27] plug-ins were used to identify hub genes. Hub nodes related to LUAD risk score were determined by considering the intersection of these results.

2.6 Mutation analysis

To explore somatic mutations in LUAD samples related to risk scores, "Masked Somatic Mutation" data was retrieved from TCGA GDC official website (https://portal.gdc.cancer.gov/). Somatic mutation data from TCGA-LUAD were downloaded

and the R package maftools [13] was used to visualize the mutation landscape of patients with both high- and low-risk scores.

2.7 Evaluation of immune infiltration

The ssGSEA algorithm was used to calculate 28 reported immune cell types [28], and the immune cell compositions of patients in the high- and low-scoring LUAD groups were visualized using a box plot. The differences in immune cell proportions were determined using the Wilcoxon test and the correlation between immune cell infiltration levels were plotted using the R package corrplot [29]. Furthermore, Spearman's correlation analysis helped establish an association between the risk score and immune cells. To assess the prognostic value of these immune cells, a Kaplan–Meier survival analysis was conducted. Finally, the difference in immune scores between patients with LUAD in the high- and low-score groups were estimated using the R package ESTIMATE [30]. Statistical significance was set at $p < 0.05$.

2.8 Construction of nomogram model

By incorporating the risk score and clinicopathological features into the model, a clinical predictive line map (nomogram) was constructed using the R-packet rms [31]. To evaluate the consistency between the predicted values and actual observations, calibration curves were used. Additionally, the R package ggDCA [32] was used to analyze the decision curve and draw clinical impact curves to evaluate whether model-based decision-making benefits patient prognosis.

2.9 Prediction of treatment response

The Wilcoxon test was used to explore the correlation between risk score and potential immunotherapy markers, including immune checkpoint blockade (ICB)-related genes [33–35] and interferon-gamma (IFN-γ) pathway markers [36].

The LUAD cell-drug action dataset was obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database [37]. The expression data of patients with LUAD were analyzed using the R-packet oncoPredict [38], enabling a comparison of drug sensitivity between high- and low-risk patients.

2.10 Cell culture

A549(ATCC CCL-185™),NCI-H1299(ATCC CRL-5803),PC9(SCSP-5085), NCI-H1915(ATCC CRL-5904) and HBE(ATCC CRL-2741 ™) cell lines were provided by the Shanghai Cell Bank of the Chinese Academy of Sciences. The cells were cultured in RPMI 1640 medium(Gibco BRL,Grand Island,NY,USA) supplemented with 10% FBS (Gibco BRL,Grand Island,NY,USA) and incubated at 37 ℃ in a 5% CO_2 incubator. The cells were then seeded in a 6-well culture plate at a density of 4×10^5 cells per well and cultured overnight for subsequent experiments.

2.11 qRT-PCR

RNA was prepared using TRIzol reagent (Seven Biotech,Beijing,China). For reverse transcription, RNA (1 μg) was used with a Transcriptor First Strand cDNA Synthesis Kit (Roche Diagnostics, Basel, Switzerland) according to the manufacturer's instructions. qRT-PCR was performed using a LightCycler 480 instrument (Roche Diagnostics) and SYBR Premix Ex Taq II kit (Takara, Shiga, Japan). The primer sequences used for PCR are listed in Supplementary Table 3. Housekeeping genes, GAPDH and U6, were used for normalization, and RNA expression was normalized against GAPDH using the $2^{-\Delta\Delta Ct}$ method.

2.12 Colony formation

In six-well plates, 1000 LUAD cells were cultured for two weeks. Cell colonies were fixed with 4% formaldehyde and stained with 1% crystal violet (Sigma-Aldrich, St. Louis, MO, USA). The number of visible colonies on each plate was counted.

2.13 Cell scratch assay

Cells (5×10^5) were cultured in 6-well plates. After being cultured overnight, cell monolayers were scraped using a 100 µL pipette tip. The wounds were photographed at 0 and 24 h, and the percentage of migration was calculated using Image J software.

2.14 Invasion assay

Cells (1×10^5) were suspended in FBS-free medium and plated in the upper trans-well chamber (Millipore, Braunschweig, Germany) pre-coated with Matrigel (BD Bioscience, San Jose, CA, USA). The lower chamber was filled with 500 µL of culture medium (20% FBS). The number of invading cells were counted after staining with 1% crystal violet.

2.15 Statistical analysis

Statistical analysis was performed using R software (version 4.1.1). The independent Student's t-test was used to compare normally distributed continuous variables, whereas differences between non-normally distributed continuous variables were assessed using the Mann–Whitney U test (Wilcoxon's rank-sum test). Categorical variables were analyzed using either the chi-squared or Fisher's exact test. Survival analysis was conducted using the R survival package, with the Kaplan–Meier survival curve illustrating survival differences. The log-rank test was used to evaluate the significance of these differences in survival time between the two groups. Both univariate and multivariate Cox analyses were based on survival R packages, and the LASSO analysis was based on the R package glmnet. For all analyses, statistical significance was set at $p < 0.05$.

3 Results

3.1 Identification of key RNA modification genes

Sixty-eight RNA-modifying DEGs were identified in the RNA-seq dataset from patients with LUAD (Fig. 1A). Thirty-nine RNA-modifying genes related to prognosis were identified based on prognostic data (Supplementary Fig. 2). It was determined that RNA-modifying genes exhibited a higher mutation frequency in patients (Fig. 1B–E), particularly m5C- and m6A-related genes. The somatic mutations in patients with LUAD (Fig. 1F) were also analyzed. The most common mutation type was missense mutations, characterized by frequent single nucleotide mutations and C > A substitutions.

3.2 Shared genes of OS and LUAD

Thirty-five key RNA-modifying genes associated with LUAD were identified through the intersection of DEGs and prognosis-related genes. One thousand five hundred and eighty-nine lncRNAs associated with key RNA modification genes through Pearson correlation analysis ($|R| > 0.3$ and $p < 0.05$, Supplementary Table 4) was determined. These were included in univariate Cox regression analysis, and 493 lncRNAs were identified to significantly affect the prognosis of patients with LUAD. Subsequently, ten lncRNAs and their penalty coefficients were obtained through LASSO Cox regression analysis (AC092168.2, SATB2-AS1, PP2672, FOCAD-AS1, DSG2-AS1, HIF1A-AS1, OGFRP1, WWC2-AS2, TMPO-AS1, and LINC01117). These lncRNAs were included in the risk-scoring model to predict the prognosis of patients with LUAD. Analysis of the expression levels of the ten prognostic genes demonstrated a significant positive correlation ($p < 0.05$; Fig. 2A). A correlation map of the risk factors was developed based on the risk scores of patients with LUAD, which revealed significantly higher mortality rates and shorter survival times in the high-risk group than that of the low-risk group (Fig. 2B). Kaplan–Meier analysis also showed significant differences between the low- and high-risk group in terms of OS ($p < 0.0001$; Fig. 2C), PFI ($p < 0.0001$; Fig. 2D), DSS ($p < 0.0001$; Fig. 2E), and DFI ($p < 0.0001$; Fig. 2F). Subsequently, the GSE31210 dataset was used to further evaluate and validate the impact of risk scores on patients with LUAD and

Fig. 1 Identification of RNA-modified genes. **A** Heatmap of the expression levels of m1A, m5C, m6A, and m7G related genes in LUAD (sky blue) and normal tissues (pink), with the horizontal and vertical axes representing genes and samples, respectively. Red and blue represent high and low gene expressions, respectively. Mutation waterfall plot of **B** m1A, **C** m7G, **D** m6A, and **E** m5C related genes in LUAD. **F** Overview of gene mutations in LUAD samples

determine their prognosis. The results indicated significant differences in OS and DSS between the low- and high-risk groups ($p < 0.05$; Fig. 2G, H).

3.3 Potential molecular mechanisms of risk scoring

Differential analysis of the gene expression profiles of the two patient groups resulted in the identification of 586 differentially expressed coding genes. GO functional annotations revealed their involvement in various biological processes. These processes were primarily associated with chromosome segregation, nuclear division, mitotic nuclear division, and organelle fission ($p < 0.05$; Fig. 3A). Additionally, they were enriched in cellular components such as chromosomal regions, centromeric regions, condensed chromosomes, and kinetochores ($p < 0.05$, Fig. 3B). Molecular functions, such as DNA helicase activity, microtubule binding, single-stranded DNA helicase activity, and ATPase activity, were also enriched ($p < 0.05$; Fig. 3C). Furthermore, these genes were enriched in KEGG pathways such as the cell cycle, DNA replication, oocyte meiosis, Fanconi anemia pathway, and p53 signaling pathway ($p < 0.05$, Fig. 3D).

Subsequently, the GSEA analysis revealed that biological processes such as GO ATPase activity and catalytic activity acting on DNA were upregulated (Fig. 3E), and immune-related biological processes such as GO lamellar body and leukotriene biosynthetic process involved in ventricular cardiac muscle cell action potential repolarization were downregulated in the high-risk group (Supplementary Table 5). Pathway activity analysis indicated that the KEGG pathways related to alpha-linolenic acid metabolism and asthma were inhibited in the high-risk patient group (Fig. 3F). Conversely, KEGG pathways associated with cell cycle, DNA replication, spliceosomes, p53 signaling, and homologous recombination were upregulated (Supplementary Table 6).

3.4 PPI network analysis

The PPI network related to DEGs between the low- and high-risk samples includes 5263 interactions and 346 DEGs. Key genes were screened using the CytoHuba and MCODE plugins. Sixty and seventy-six CytoHuba- and MCODE-related hub nodes were identified, respectively. This intersection was then used to identify 49 hub genes (Fig. 4A). Then, GO semantic similarity analysis of these hub genes were calculated using the R package "GOSemSim" [33] which revealed functional associations with several genes, including *BIRC5, PLK4, NDC80, ASPM*, and *CDC6* (Fig. 4B). Further enrichment analysis indicated their involvement in processes such as mitotic nuclear division, chromosome segregation, nuclear division, spindle formation, chromosomal regions, kinetochores, DNA replication origin binding, microtubule binding, and tubulin binding (Fig. 4C–E). These genes also affected cell cycle, oocyte meiosis, progesterone-mediated oocyte maturation, DNA replication, and the p53 signaling pathway (Fig. 4F).

The GSVA results showed significant enrichment differences between the low- and high-risk samples in most cancer marker pathways ($p < 0.05$, Fig. 5A). These included hallmark DNA repair, E2F targets, MYC targets v1, and MYC targets v2, which were significantly upregulated in the high-risk group and associated with well-known oncogenic pathways. In addition, patients in the low-risk group showed a stronger correlation between cancer marker pathway GSVA scores and hub gene expression levels (Fig. 5B, C). Moreover, correlation analysis of the ten characteristic lncRNAs identified TMPO AS1 as being significantly associated with several well-known oncogenic pathways (Fig. 5D). It exhibited a significant positive correlation with hallmark G2/M checkpoint ($r = 0.68$, $p < 0.05$) and a significant negative correlation with hallmark KRAS signaling DN ($r = -0.21$, $p < 0.05$).

3.5 Genomic differences in distinct risk groups

The overall mutation status in the LUAD dataset as well as the mutation status in high- and low-risk patients were analyzed. The top five genes with the highest mutation frequencies were RYR2 (36%), MUC16 (40%), TTN (46%), LRP1B (32%), and CSMD3 (37%). Although the top five genes with the highest mutation frequencies were consistent between high- and low-risk groups, the mutation types and frequencies of these genes differed. For instance, *TP53* exhibited a 63% mutation frequency in the high-risk group (Fig. 6A) compared to that of the low-risk group, which was 44% (Fig. 6B),

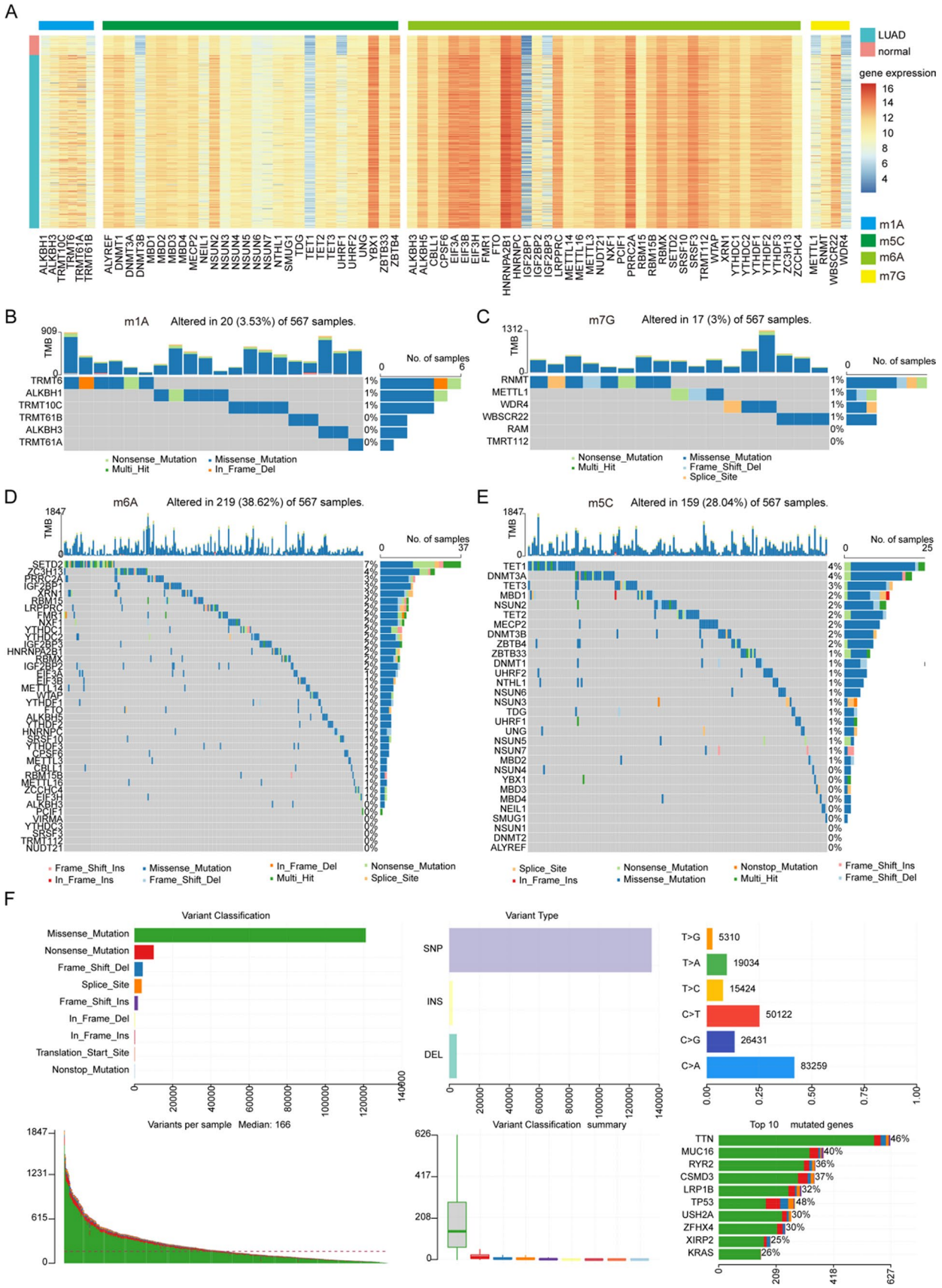

A
LUAD
normal
gene expression
16
14
12
10
8
6
4
m1A
m5C
m6A
m7G
ALKBH1 ALKBH3 TRMT10C TRMT6 TRMT61A TRMT61B
ALYREF DNMT1 DNMT3A DNMT3B MBD1 MBD2 MBD3 MBD4 MECP2 NEIL1 NSUN2 NSUN3 NSUN4 NSUN5 NSUN6 NSUN7 NTHL1 SMUG1 TDG TET1 TET2 TET3 UHRF1 UHRF2 UNG YBX1 ZBTB33 ZBTB4
ALKBH3 ALKBH5 CBLL1 CPSF6 EIF3A EIF3B EIF3H FMR1 FTO HNRNPA2B1 HNRNPC IGF2BP1 IGF2BP2 IGF2BP3 LRPPRC METTL14 METTL16 METTL3 NUDT21 NXF1 PCIF1 PRRC2A RBM15 RBM15B RBMX SETD2 SRSF10 SRSF3 TRMT112 WTAP XRN1 YTHDC1 YTHDF1 YTHDF2 YTHDF3 ZC3H13 ZCCHC4
METTL1 RNMT WBSCR22 WDR4
B
m1A
Altered in 20 (3.53%) of 567 samples.
TMB
909
0
No. of samples
0
6
TRMT6
ALKBH1
TRMT10C
TRMT61B
ALKBH3
TRMT61A
1%
1%
1%
0%
0%
0%
Nonsense_Mutation
Missense_Mutation
Multi_Hit
In_Frame_Del
C
m7G
Altered in 17 (3%) of 567 samples.
TMB
1312
0
No. of samples
0
RNMT
METTL1
WDR4
WBSCR22
RAM
TMRT112
1%
1%
1%
1%
0%
0%
Nonsense_Mutation
Missense_Mutation
Multi_Hit
Frame_Shift_Del
Splice_Site
D
m6A
Altered in 219 (38.62%) of 567 samples.
TMB
1847
0
No. of samples
0
37
SETD2
ZC3H13
PRRC2A
IGF2BP1
XRN1
RBM15
LRPPRC
FMR1
NXF1
YTHDC1
YTHDF2
IGF2BP3
HNRNPA2B1
RBMX
IGF2BP2
EIF3A
EIF3B
METTL14
WTAP
YTHDF1
FTO
ALKBH5
YTHDF2
HNRNPC
SRSF10
YTHDF3
CPSF6
METTL3
CBLL1
RBM15B
METTL16
ZCCHC4
EIF3H
ALKBH3
PCIF1
VIRMA
YTHDC2
SRSF3
TRMT112
NUDT21
7%
5%
3%
3%
2%
2%
2%
2%
2%
2%
2%
2%
2%
2%
2%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
0%
0%
0%
0%
0%
0%
0%
Frame_Shift_Ins
In_Frame_Ins
Missense_Mutation
Frame_Shift_Del
In_Frame_Del
Multi_Hit
Nonsense_Mutation
Splice_Site
E
m5C
Altered in 159 (28.04%) of 567 samples.
TMB
1847
0
No. of samples
0
25
TET1
DNMT3A
TET3
MBD1
NSUN2
TET2
MECP2
DNMT3B
ZBTB4
ZBTB33
DNMT1
UHRF2
NTHL1
NSUN6
NSUN3
TDG
UHRF1
UNG
NSUN5
NSUN7
MBD2
NSUN4
YBX1
MBD3
MBD4
NEIL1
SMUG1
NSUN1
DNMT2
ALYREF
4%
4%
3%
2%
2%
2%
2%
2%
2%
2%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
1%
0%
0%
0%
0%
0%
0%
0%
0%
Splice_Site
In_Frame_Ins
Nonsense_Mutation
Missense_Mutation
Nonstop_Mutation
Multi_Hit
Frame_Shift_Ins
Frame_Shift_Del
F
Variant Classification
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Splice_Site
Frame_Shift_Ins
In_Frame_Del
In_Frame_Ins
Translation_Start_Site
Nonstop_Mutation
0
20000
40000
60000
80000
100000
120000
140000
Variant Type
SNP
INS
DEL
0
20000
40000
60000
80000
100000
120000
140000
T>G 5310
T>A 19034
T>C 15424
C>T 50122
C>G 26431
C>A 83259
0.00
0.25
0.50
0.75
1.00
Variants per sample Median: 166
1847
1231
615
0
Variant Classification summary
626
417
208
0
Top 10 mutated genes
TTN 46%
MUC16 40%
RYR2 36%
CSMD3 37%
LRP1B 32%
TP53 48%
USH2A 30%
ZFHX4 30%
XIRP2 25%
KRAS 26%
0
209
418
627

Fig. 2 Construction of the risk score. **A** Correlation analysis of the expression levels of characteristic genes, with numerical values indicating the magnitude of the correlation. **B** Risk factor correlation chart for risk scoring; yellow: high-risk score; horizontal axis: patient risk score ranking; vertical axis: patient risk score (top); correlation analysis between survival time, survival status of patients, and risk scores. The yellow nodes represent deceased patients, horizontal axis represents the ranking of patient risk scores, and vertical axis represents the survival time of patients (middle). The heat map of characteristic gene expression for risk scoring, with the horizontal axis representing patients and the vertical axis representing characteristic genes, with blue indicating high gene expression (bottom). **C**–**F** Survival curves for the training set risk scores (OS, PFI, DFI, and DSS). **G**, **H** Survival curves for the validation set of risk scores (OS and DSS). The horizontal axis represents survival time and the vertical axis represents survival rate. Blue and yellow lines represent patients with low- and high-risk scores, respectively

whereas fewer mutation types were observed in the high-risk group. Mutation association in the high-risk group was significantly lower than that in the low-risk group. In the high-risk group, *TP53* was significantly co-mutated with X*IRP2, ZNF804A, NAV3, ZNF536, MUC17, LRRC7*, and *USH2A* ($p < 0.05$; Fig. 6C). In the low-risk group, TP53 exhibited co-mutation relationships with multiple genes including *TTN, MUC16, CSMD3, RYR2, LRP1B, USH2A, ZFHX4* and *PCLO* ($p < 0.05$; Fig. 6D). Additionally, *COL5A2, RP1L1*, and *COL22A1* exhibited higher mutation rates in the low-risk group ($p < 0.05$), whereas *BZRAP1, ACSM2B, KIF19*, and *SP140* exhibited higher mutation rates in the high-risk group ($p < 0.05$; Fig. 6E). At the same time, we found that the high-risk group had a higher tumor mutation burden (TMB), fewer immune cells, and higher levels of activated CD4 + T cells ($p < 0.05$; Fig. 6F).

3.6 Analysis of differences in immune cell infiltration levels

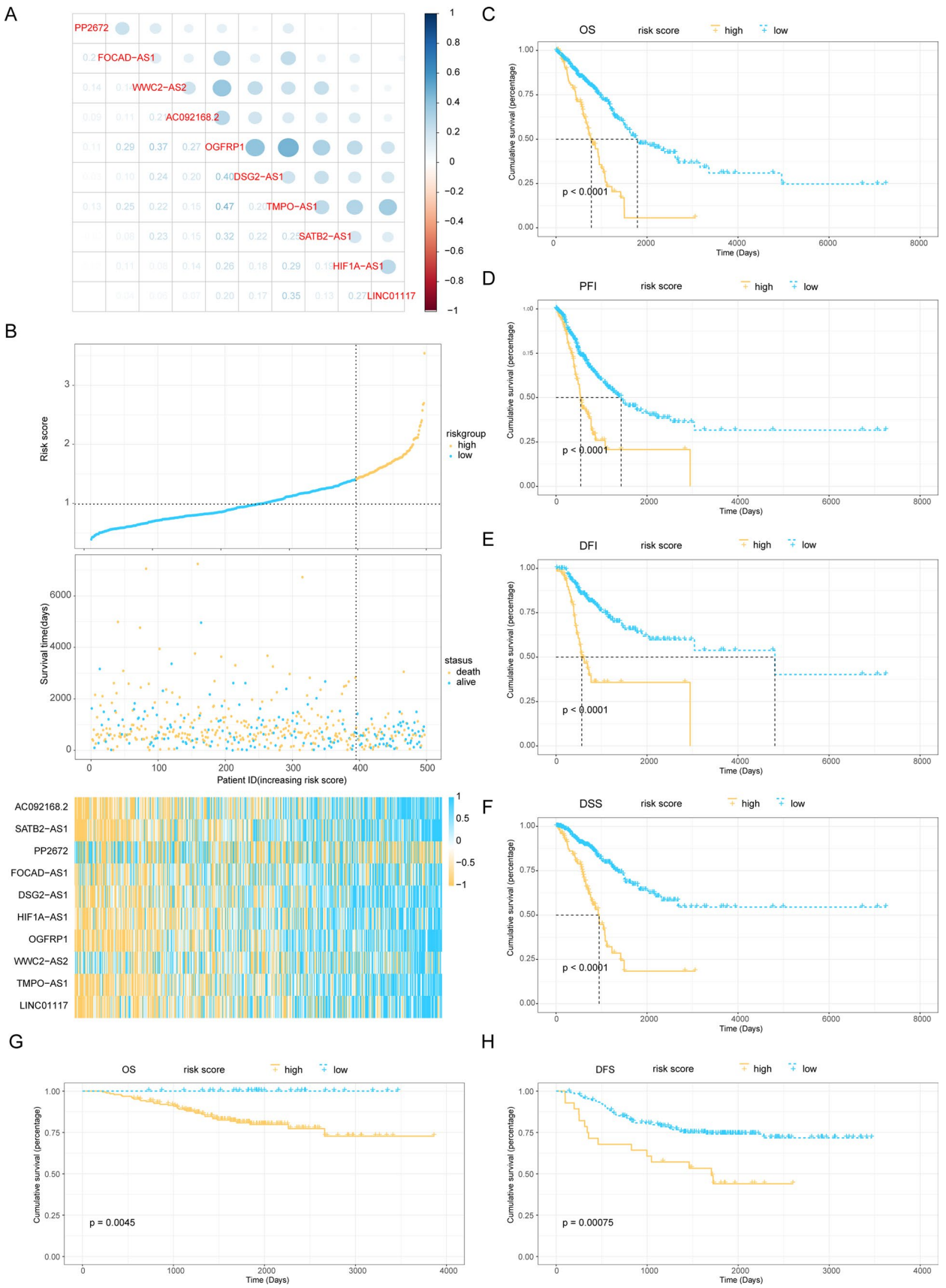
Subsequently, ssGSEA was performed on the two sample groups. The results revealed significant differences in the infiltration levels of 14 immune cells, including activated dendritic cells, eosinophils, immature B cells, and Myeloid-derived suppressor cells (MDSCs), between the low- and high-risk samples ($p < 0.05$; Fig. 7A). Except for activated $CD4^+$ T cells, all other immune cells were highly abundant in the low-risk samples. There was a positive correlation between various immune cells in the high- (Fig. 7B) and low-risk samples (Fig. 7C), but there were significant differences in the clustering modules between immune cells. In addition, there was a significant negative correlation between the risk score and immune cells, indicating that patients in the high-risk group had fewer immune cells ($r < 0$, $p < 0.05$; Fig. 7D). The association between OS and TME cell infiltration in patients with LUAD were also analyzed. Survival differences between the high- and low-infiltration groups were associated with 19 types of TME cells ($p < 0.05$; Supplementary Fig. 3). Analysis of the correlation between TME cells and hub genes in the low- and high-risk samples revealed that activated $CD4^+$ T cells were significantly positively correlated with multiple hub genes in the high-risk samples ($p < 0.05$; Fig. 7E). In contrast, activated B cells were significantly negatively correlated with multiple hub genes in low-risk samples ($p < 0.05$; Fig. 7F). By taking the intersection of differentially enriched immune cells, immune cells significantly related to risk scores, and immune cells that affected the prognosis of patients with LUAD, 11 key TME cells were obtained (Fig. 7G). ESTIMATE analysis revealed that high-risk samples had lower ESTIMATE scores ($p < 0.05$, Fig. 7H), lower immune scores ($p < 0.05$, Fig. 7I), higher tumor purity ($p < 0.05$, Fig. 7J), and lower stromal scores ($p < 0.05$, Fig. 7K).

3.7 Nomogram model

Univariate and multivariate analyses indicated that the risk score prediction model and Tumor stage were prognostic factors for LUAD, with $HR < 1$ and $p < 0.05$ for a low-risk score, respectively (Fig. 8A). A nomogram model was constructed based on the patient risk scores and staging to predict the prognosis of patients with LUAD (Fig. 8B). After calibration of the model, it was found to predict patient survival in general agreement with the actual patient survival (Fig. 8C). At the same time, the three-year DCA curve of the nomogram model was plotted, and the results showed that the nomogram model based on the patient's risk score and tumor stage had a better effect on the prognosis of LUAD patients (Fig. 8D). Moreover, the time-ROC results also showed that the one- and three-year survival predictive efficacy of the prognostic model was greater than 70% (Fig. 8E).

3.8 Expression of lncRNA in lung epithelial and LUAD cells

To verify the aforementioned findings, the expression levels of ten lncRNAs (i.e., AC092168.2, SATB2-AS1, PP2672, FOCAD-AS1, DSG2-AS1, HIF1A-AS1, OGFRP1, WWC2-AS2, TMPO-AS1, and LINC01117) were evaluated in HBE and LUAD cells (A549, NCI-H1299, PC9, and NCI-H1915). The results indicated that the expression of nine lncRNAs in LUAD cells, except


A
PP2672
FOCAD−AS1
WWC2−AS2
AC092168.2
OGFRP1
DSG2−AS1
TMPO−AS1
SATB2−AS1
HIF1A−AS1
LINC01117
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
B
Risk score
riskgroup
high
low
Survival time(days)
stasus
death
alive
Patient ID(increasing risk score)
AC092168.2
SATB2−AS1
PP2672
FOCAD−AS1
DSG2−AS1
HIF1A−AS1
OGFRP1
WWC2−AS2
TMPO−AS1
LINC01117
C
OS
risk score
high
low
Cumulative survival (percentage)
p < 0.0001
Time (Days)
D
PFI
risk score
high
low
Cumulative survival (percentage)
p < 0.0001
Time (Days)
E
DFI
risk score
high
low
Cumulative survival (percentage)
p < 0.0001
Time (Days)
F
DSS
risk score
high
low
Cumulative survival (percentage)
p < 0.0001
Time (Days)
G
OS
risk score
high
low
Cumulative survival (percentage)
p = 0.0045
Time (Days)
H
DFS
risk score
high
low
Cumulative survival (percentage)
p = 0.00075
Time (Days)

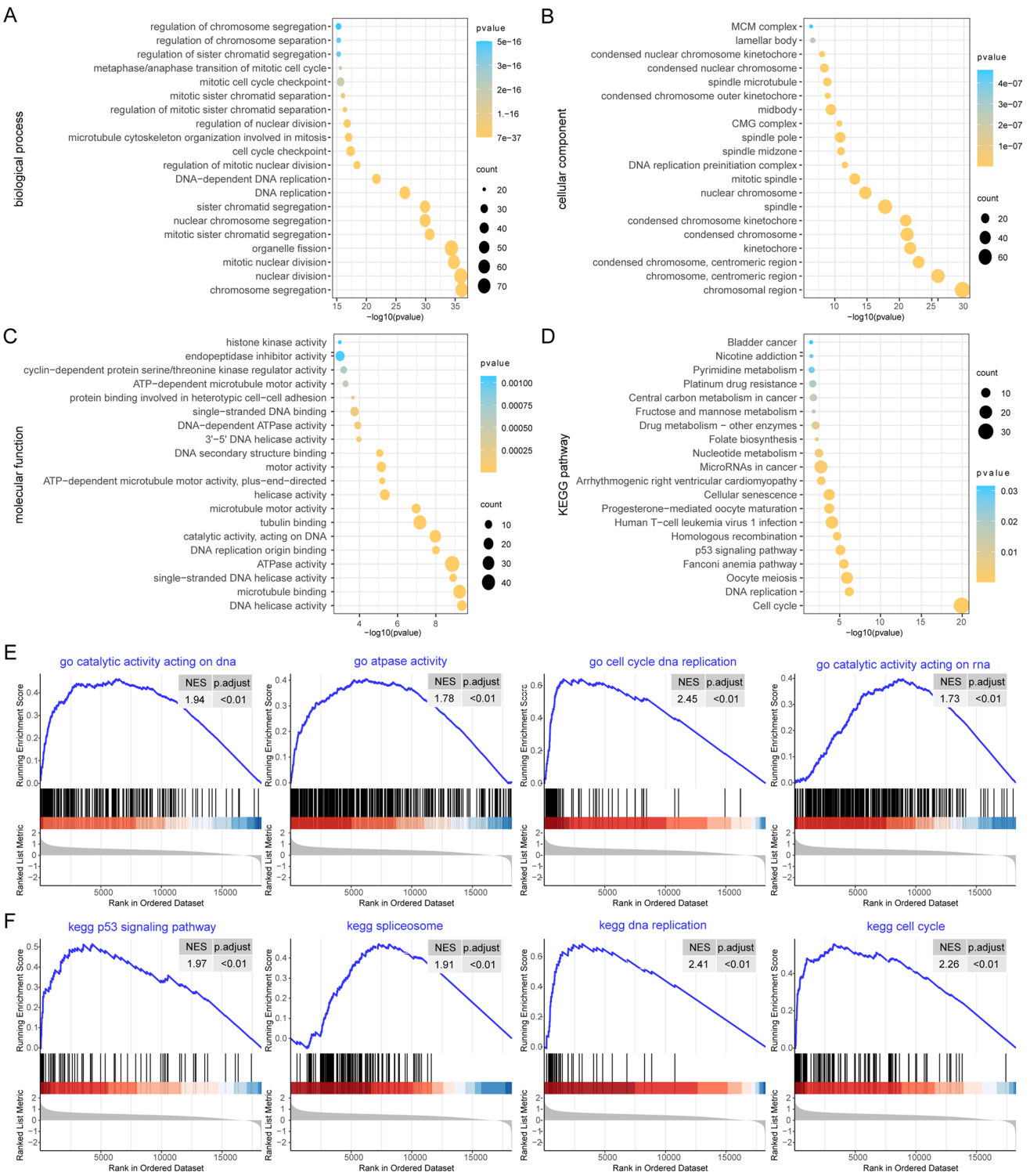


Fig. 3 Differences in biological processes between low- and high-risk samples. **A**–**D** The BP, CC, MF GO term analysis, and KEGG enrichment analysis between low- and high-risk samples, where the node color corresponds to the significance of enrichment results, and the node size represents the number of enriched genes. The horizontal and vertical axes represent − log10 (p-value) and biological process, respectively. **E** The top four most significantly enriched GO terms in GSEA-GO analysis. **F** Top four most significantly enriched KEGG pathways are displayed

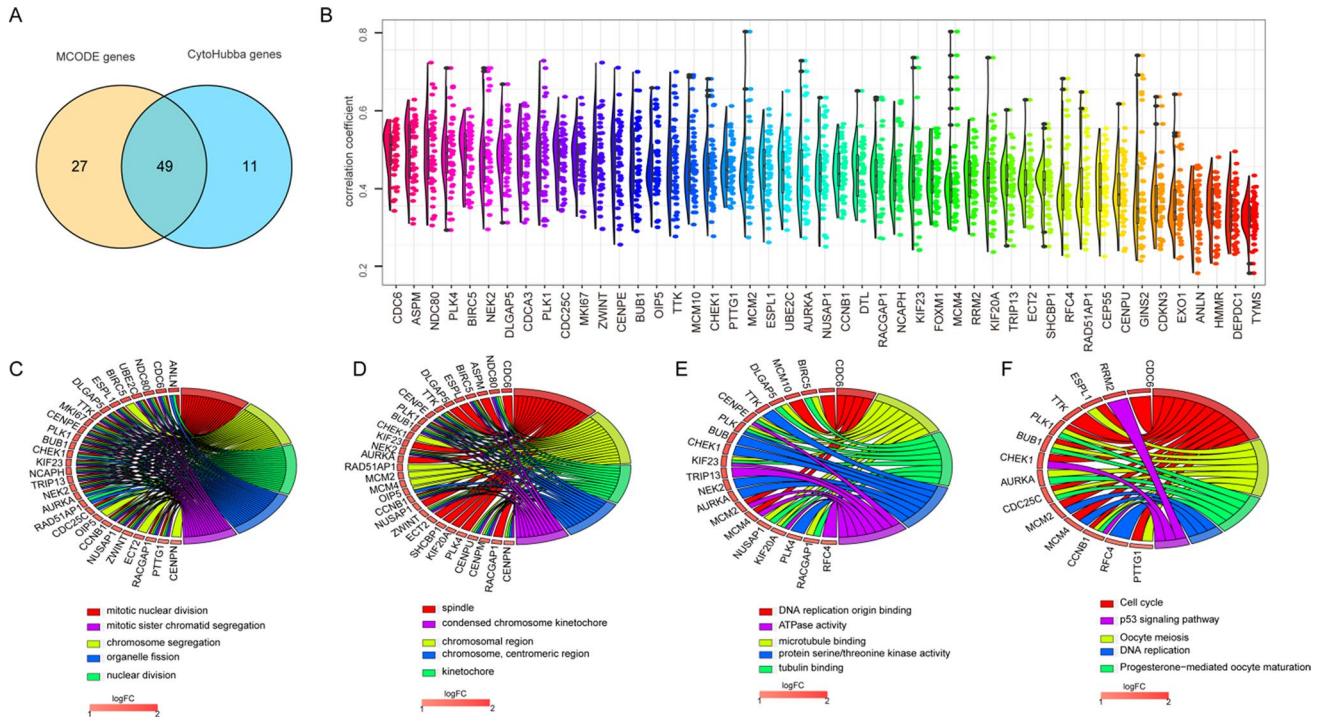


Fig. 4 PPI network analysis. **A** Wayne diagram of hub nodes related to CytoHuba and MCODE, where yellow and blue represent hub nodes related to MCODE and CytoHuba, respectively. **B** GO semantic similarity scores of hub genes, with the horizontal axis representing the similarity level and the vertical axis representing genes. **C**–**F** BP, CC, and MF GO term analysis, and KEGG enrichment analysis of hub genes, with line colors indicating biological functions

for AC092168.2, was higher than that in HBE cells. The expression of lncRNAs was highest in NCI-H1299 and A549 cells, with TMPO-AS1 and HIF1A-AS1 being the most highly expressed among the ten lncRNAs (Fig. 9).

3.9 Si-HIF1A-AS1 and si-TMPO-AS1 reduces the metastatic ability of LUAD cells

si-TMPO-AS1 and si-HIF1A-AS1 were subsequently transfected into H1299 and A549 cells to assess LUAD cell proliferation and migration ability. The results revealed that the silencing of lncRNAs TMPO-AS1 and HIF1A-AS1 significantly reduced cell migration (Fig. 10A, B) and proliferation (Fig. 10C). No noticeable changes were observed in the si-NC group, suggesting that the lncRNAs TMPO-AS1 and HIF1A-AS1 may influence LUAD metastasis and prognosis.

4 Discussion

Early diagnosis and treatment are critical for managing malignant tumors. Advances in medical technology have extended the survival of patients with LUAD. However, the complexity of the mechanisms underlying LUAD onset makes reliable prognosis prediction challenging using traditional factors such as disease stages and differentiation levels. Hence, investigating the genetic and epigenetic factors involved in LUAD is crucial for identifying novel therapeutic targets and diagnostic biomarkers.

The epigenetic transcriptome is complex, with over 170 different types of chemical modifications affecting both coding and ncRNAs. Currently, research on these modifications are emerging, revealing their significant impact on human pathology. m6A is the most abundant and characteristic modification of internal mRNA. It regulates the self-renewal of embryonic stem cells and cancer cells. In addition to its role in mRNA, m6A modification occurs in ncRNAs, such as microRNAs (miRNAs), lncRNAs, and circRNAs. RNA modifications influence numerous finely regulated molecular processes, including RNA metabolism, decay, splicing, translation, localization, stability, turnover, and interactions with RNA-binding

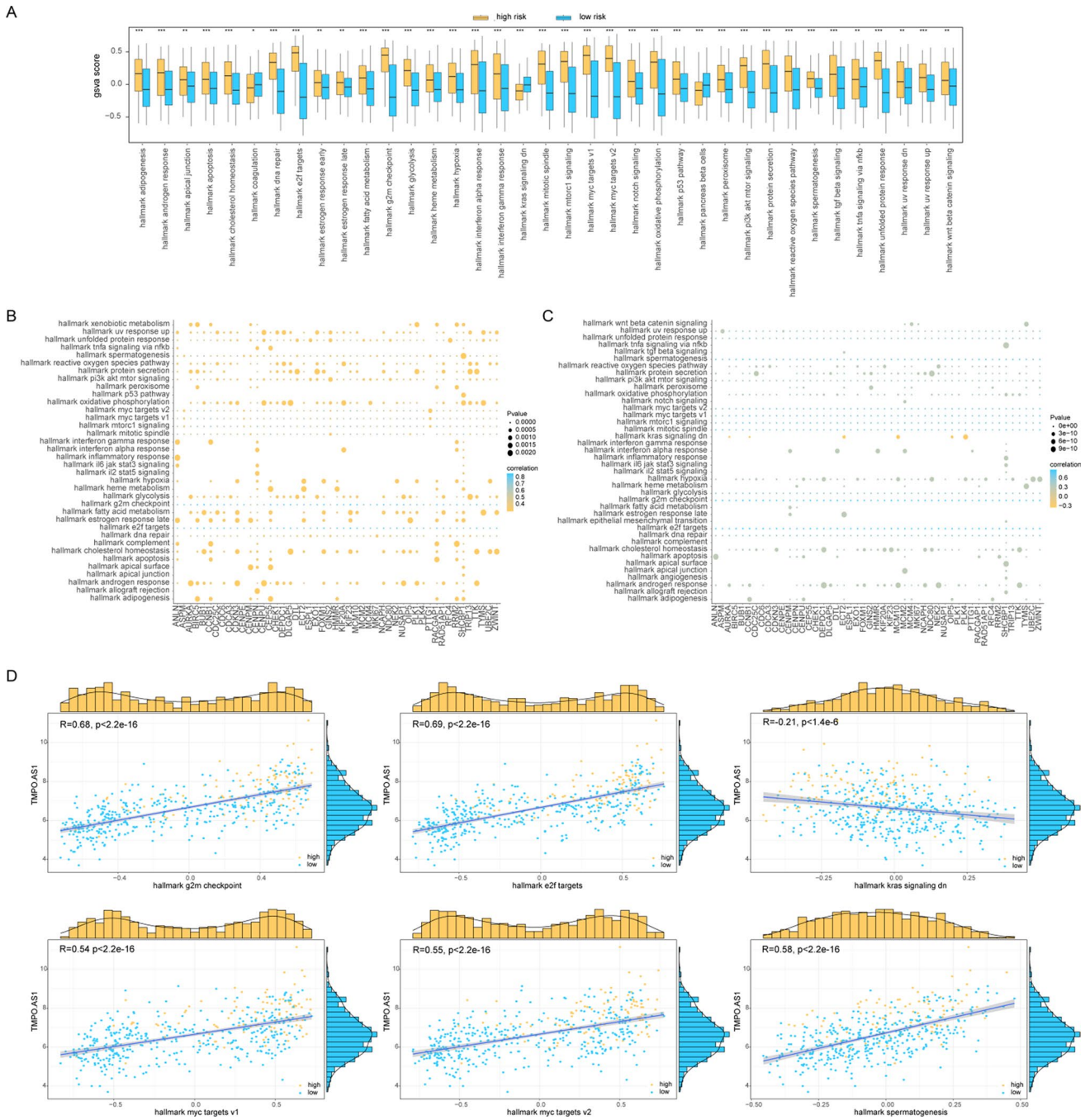


Fig. 5 GSVA analysis. **A** Differences in GSVA scores between high- and low-risk samples represented by hallmark coordinates. The vertical axis represents the GSVA scores, and yellow indicates high-risk scoring samples (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$). Correlation analysis between the GSVA scores of hallmarks and hub gene expression levels in **B** high- and **C** low-risk samples. The horizontal axis represents hub genes and the vertical axis represents hallmarks. The node size corresponds to the significance level and the node color represents the correlation level. **D** Correlation map of the most relevant relationships between hallmarks and hub genes, with metabolic hallmarks on the horizontal axis, hub genes on the vertical axis, and high-risk scoring samples indicated by yellow dots

proteins or other RNAs, thereby diversifying the genetic information. Recent discoveries include N6, m6Am, m5C, hm5C, pseudouridine, and m1A modifications of mRNA [39].

To explore gene mutations in LUAD, RNA modifications associated with the disease were screened and it was determined that m5C- and m6A-related genes were more prone to mutations. The most frequent mutations in somatic cells are missense mutations, single-nucleotide mutations, and C > A substitutions.

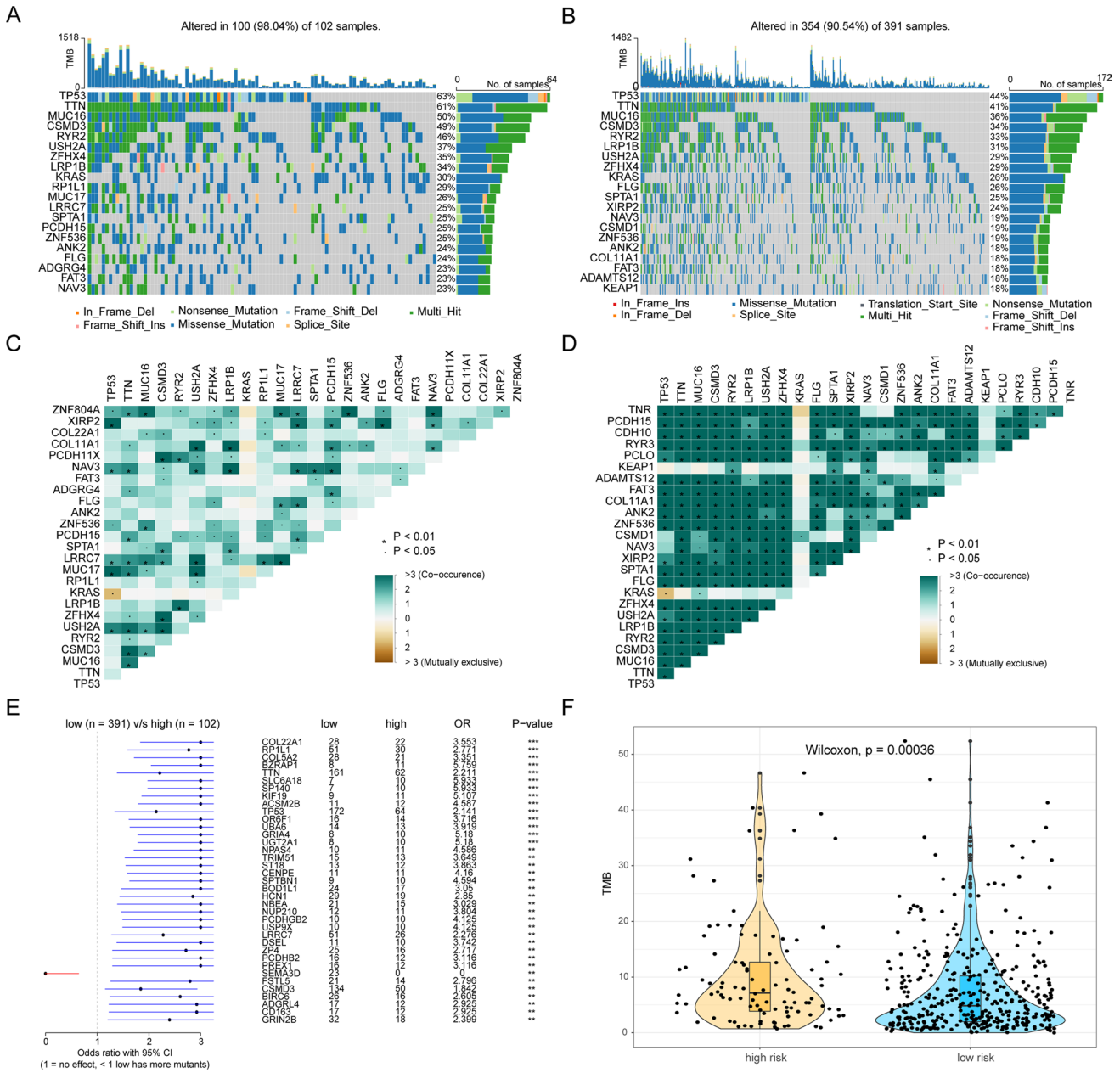


Fig. 6 Analysis of mutation mechanism. Mutation waterfall plot for **A** high- and **B** low-risk scoring samples. The horizontal and vertical axes represents the sample and mutated gene, respectively, and the rectangular color represents the type of mutation. Analysis of co-mutations and excluded mutations in **C** high- and **D** low-risk scoring samples, with yellow rectangles indicating rejected mutations (*$p < 0.05$, **$p < 0.01$). **E** Genes with significant differences in mutation status between high- and low-risk scoring samples (*$p < 0.05$, **$p < 0.01$). **F** TMB box chart with the horizontal axis representing high- and low-risk scoring samples and the vertical axis representing TMB

LncRNAs are transcripts of over 200 nucleotides and are predominantly located in the nucleus [40]. Although lncRNAs do not encode proteins, their tissue-specific expression across different tissues and developmental stages suggests important biological functions. Thousands of lncRNAs have been discovered using sequencing technology; however, their biological functions remain unclear. Recent studies have highlighted the close association between lncRNAs and RNA modifications. For example, H19 lncRNA can inhibit S-adenosyl-L-homocysteine hydrolase (SAHH). Knockdown of H19 activates SAHH, leading to increased methylation of the lncRNA-encoding gene *Nctc1*. Genome-wide methylation mapping revealed that numerous methylation changes at gene loci were related to H19, suggesting that certain lncRNAs can cause extensive epigenetic changes [40]. This regulatory mechanism may be the foundation of methylation dynamics and may play a pivotal role in developmental and pathological processes.

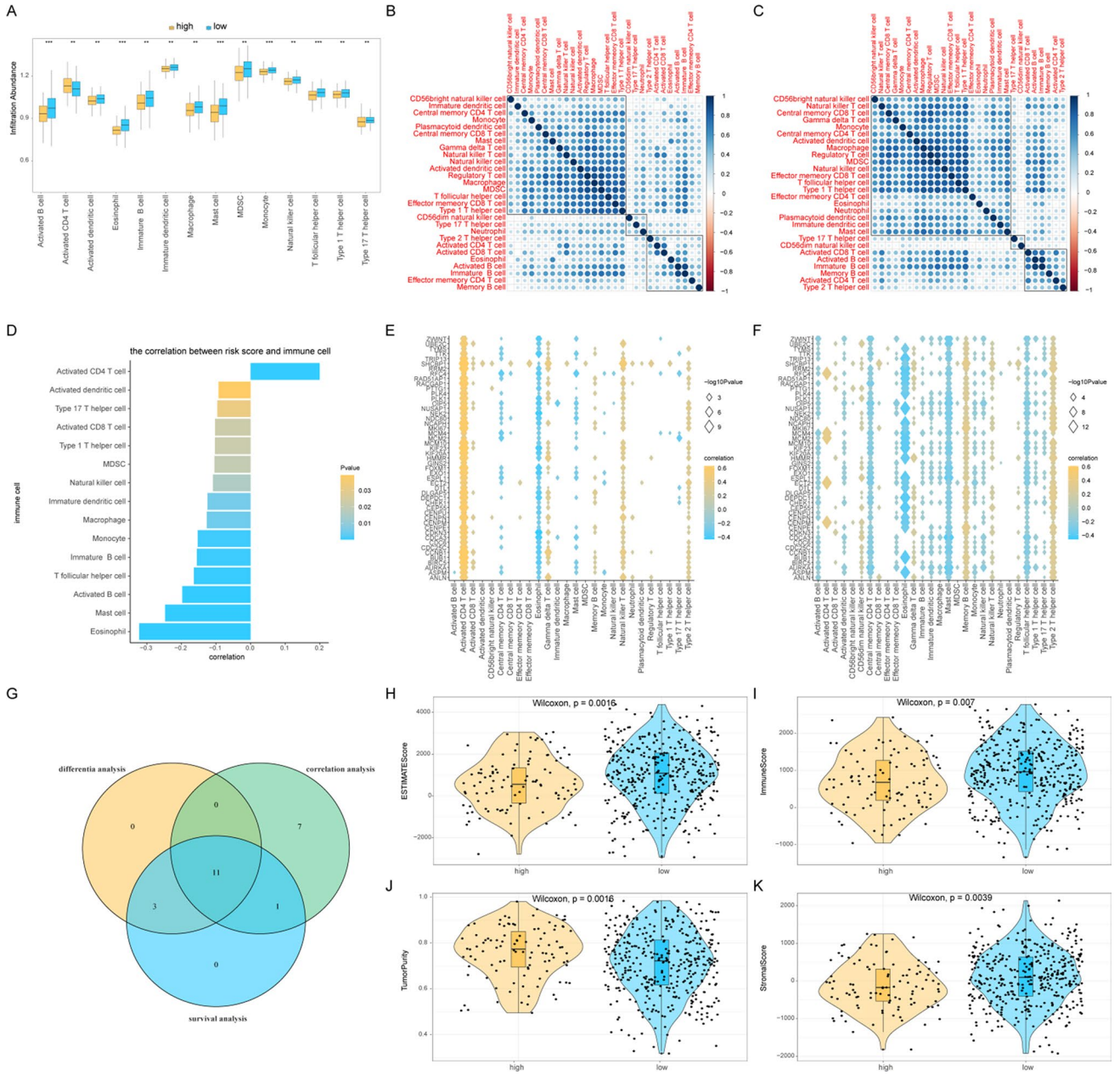


Fig. 7 Immuno-infiltration analysis. **A** Box chart of infiltration scores for TME cells, with yellow representing high-risk scoring samples. The horizontal axis represents the TME cells and the vertical axis represents the cell content. Correlation analysis between the TME cell content in **B** high- and **C** low-risk scoring samples, with red indicating a negative correlation and blue indicating a positive correlation. **D** Correlation between TME cell infiltration and risk scores, with the horizontal axis representing the magnitude of the correlation and the vertical axis representing TME cells. The bar color represents the significance level of the correlation. Correlation analysis between TME cells and hub gene expression levels in **E** high- and **F** low-risk scoring samples, with node size indicating significance and node color indicating a correlation. The horizontal axis represents TME cells, and the vertical axis represents characteristic genes. **G** Venn plots of differential TME cells (yellow), TME cells associated with risk scores (green), and TME cells affecting patient prognosis (blue). **H**–**K** Immune score, tumor purity, stromal score, and ESTIMATE score box plots in high-(yellow) and low-risk (blue) scoring samples

In this study, RNA-modifying genes and lncRNAs associated with LUAD prognosis were screened. Several lncRNAs, including AC092168.2, SATB2-AS1, PP2672, FOCAD-AS1, DSG2-AS1, HIF1A-AS1, OGFRP1, WWC2-AS2, TMPO-AS1, and LINC01117, were strongly associated with patient prognosis. Based on this, a risk-scoring model was constructed to predict the prognosis and potential molecular mechanisms of LUAD. The risk score significantly affected the prognosis of patients with LUAD, with significant differences in OS, PFI, DSS, and DFI between the low- and high-risk groups. Additionally, the stratified analysis demonstrated that the risk score provided valuable predictive information across

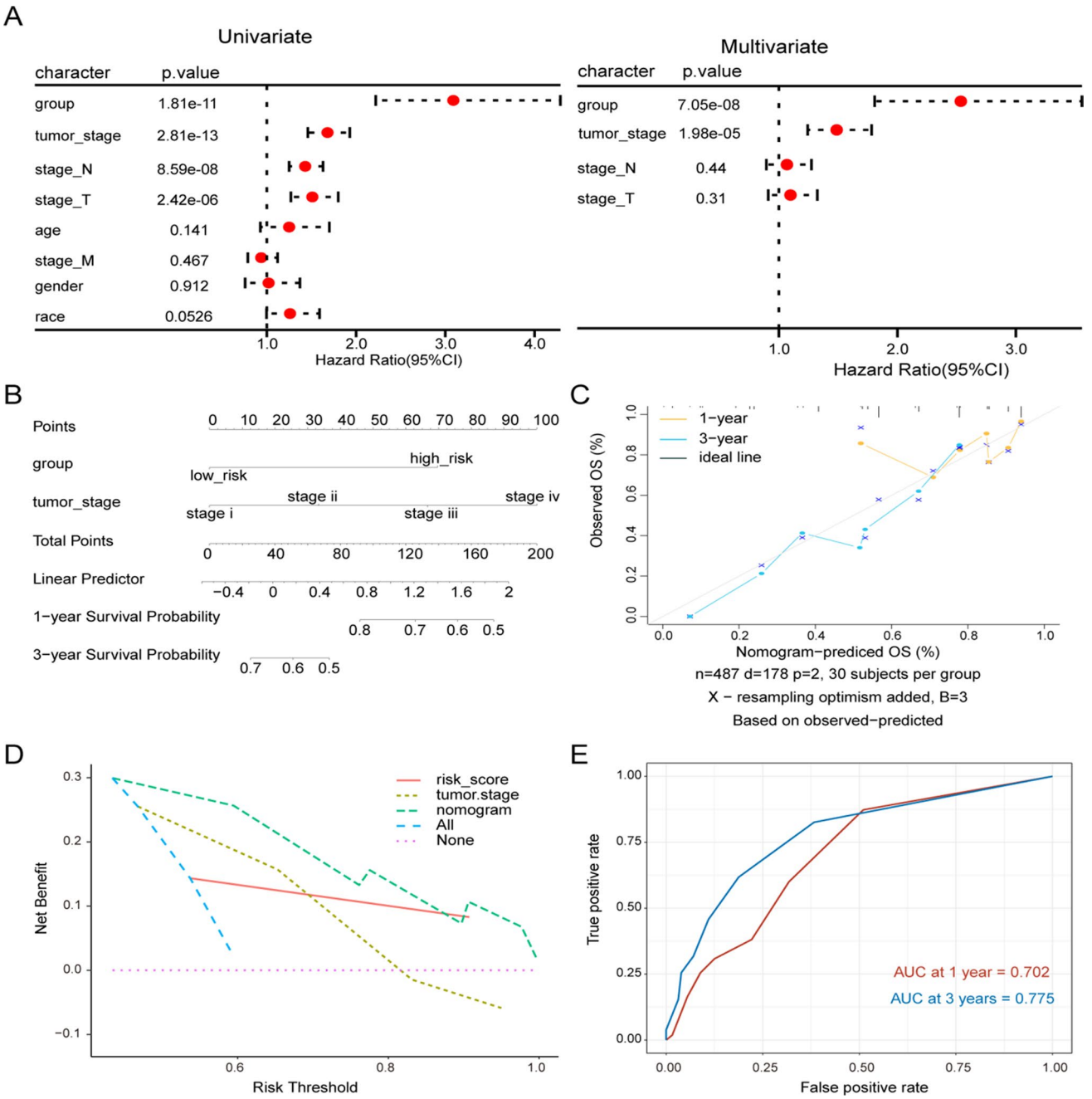


Fig. 8 Construction and evaluation of the risk score nomogram. **A** Univariate and multivariate analysis of risk scores and prognostic indicators for LUAD. **B** Prognostic nomogram for patients with LUAD. **C** Calibration curve of the nomogram model. **D** Nomogram model DCA curve. **E** Time ROC curve of one- and three-year survival prediction for patients with LUAD

subgroups classified by age, sex, race, stage, and grade ($p < 0.05$; Supplementary Fig. 4). These findings underscore the reliability of the risk score as a prognostic feature. In vitro experiments have also shown that the migratory ability of LUAD cells is significantly weakened after silencing key lncRNAs.This is similar to the results of Wang et al. who found that LncRNA-TMPO-AS1 Upregulates STRIP2 Expression and Predicts Poor Prognosis in LUAD [41].

Additionally, research on the differences in drug sensitivity between different risk groups were also conducted and it was determined that ARRY-520_474, THZ-2–102-1_346, and gemcitabine _135 were highly sensitive in the high-risk group of patients.It is noteworthy that the IFN-γ pathway genes are significantly upregulated in the high-risk group($p < 0.05$, Supplementary Fig. 5A, B). While IFN-γ signaling can activate anti-tumor immunity, it may also promote immune escape by inducing PD-L1 expression. This paradoxical phenomenon could be associated with

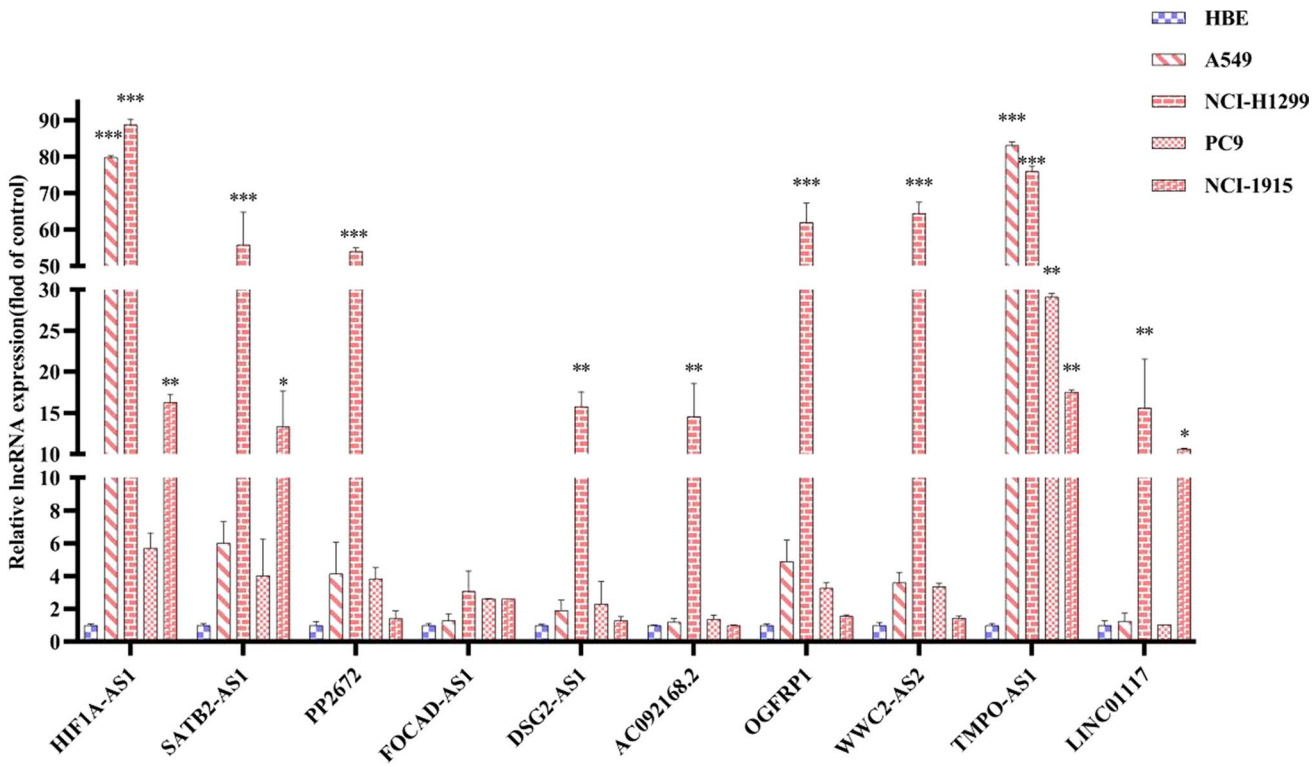


Fig. 9 Expression of lncRNA in lung epithelial and LUAD cells

the dynamic balance of immune cell subsets within the tumor microenvironment (TME).Xu et al. also found that the LncRNA HIF1A-AS1 enhances glycolysis and promotes gemcitabine resistance in pancreatic cancer by regulating the AKT/YB1/HIF1alpha pathway [42]. This is consistent with our findings,suggesting that metabolic reprogramming may be an important mechanism by which risk scores predict response to chemotherapy.

At the same time, we found that the high-risk group exhibited a significantly elevated tumor mutation burden (TMB, $p < 0.05$), along with a unique phenotype characterized by reduced immune cell infiltration but an increased proportion of activated CD4 + T cells.This phenomenon may reflect the presence of increased neoantigen production due to somatic mutations in patients in the high-risk group, which activates CD4 + T cells via MHC-II molecular delivery, creating a potentially immunogenic microenvironment. Notably, despite a reduction in the total number of immune cells in the tumor microenvironment (TME), sustained activation of CD4 + T cells may partially counteract the immunosuppressive state by facilitating antigen cross-presentation or enhancing CD8 + T-cell function, thereby enhancing patients' sensitivity to immune checkpoint blockade (ICB) therapy. In contrast, patients in the low-risk group have a relatively stable immune cell composition in TME and may be more dependent on the direct antitumor effects of conventional chemotherapy or targeted therapy.According to previous reports, the expression levels of ICB-related genes are correlated with the therapeutic response to immune checkpoint inhibitors and targeting checkpoints with ICB has become a promising strategy for treating cancer. This study showed that the expression of ICB-related genes was significantly correlated with risk scores.And the expression levels of immune phenotype related genes CXCL1, CXCL10, and CD274 were also significantly positively correlated with risk scores ($r > 0$, $p < 0.05$, Supplementary Fig. 5C, D).The upregulation of CXCL10 may reflect the sustained activation of the IFN-γ signaling pathway in the TME of the high-risk group, while the synchronous up regulation of CD274 may form the mechanism of "adaptive immune resistance", which provides a theoretical basis for patients in high-risk group to benefit more from PD-1/PD-L1 inhibitors. These findings indicate that risk scoring has potential in assessing the response of patients with LUAD to immunotherapy.In the future, the correlation between risk score and the spatial distribution and functional status of immune cells in TME can be analyzed by polychromatic immunohistochemistry or single cell sequencing technology.

There are some limitations in this study, we only selected two key LncRNAs for simple validation, we did not use clinical patient samples to re-validate the predictive efficiency of the risk assessment model, and we did not explore the specific

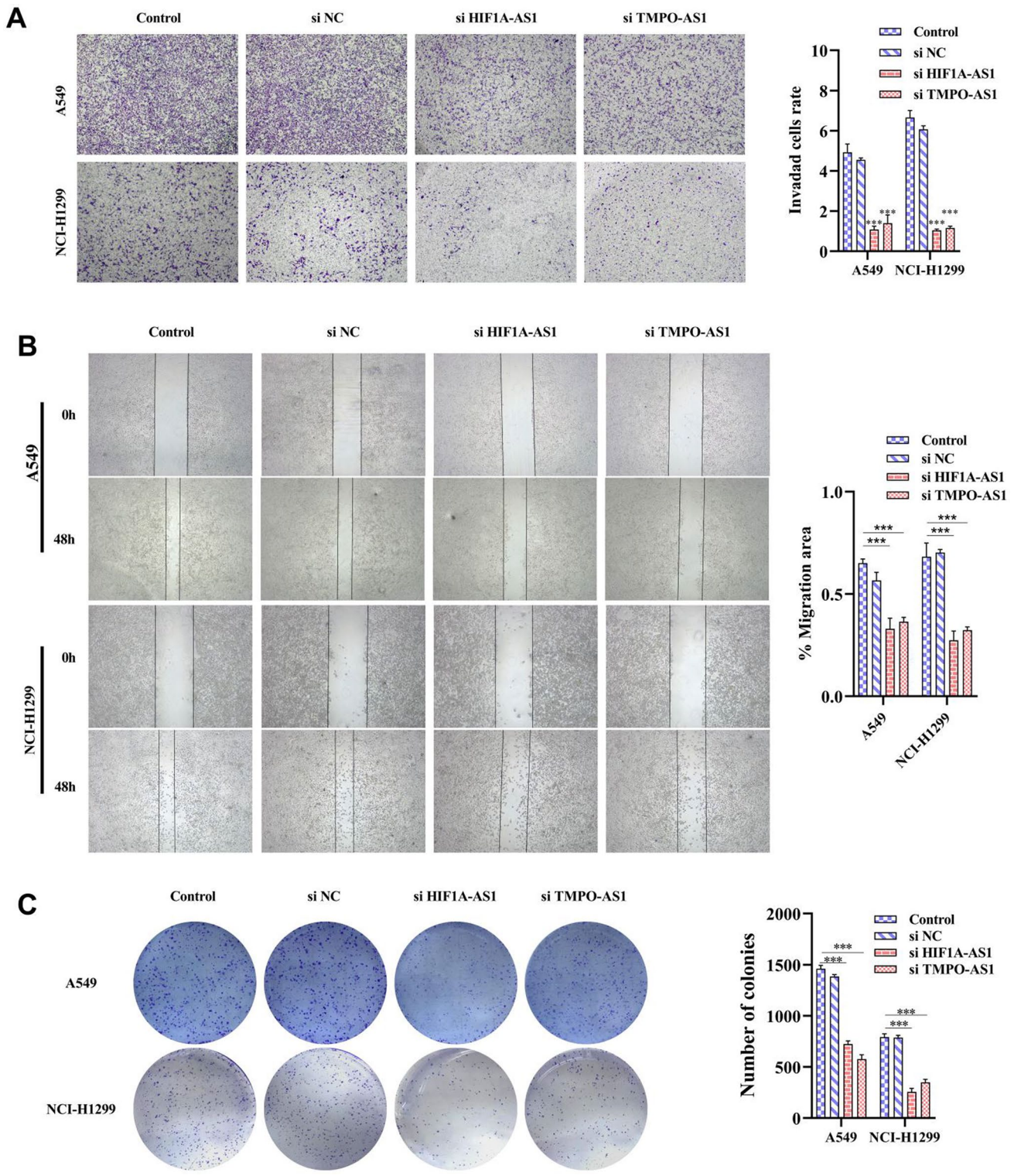


Fig. 10 Si-HIF1A-AS1 and si-TMPO-AS1 reduces the metastatic ability of LUAD cells. **A** The invasion of cells was evaluated by trans-well assay. **B** The migration of cells was evaluated by wound healing assay. **C** Cell proliferation was evaluated by colony-formation assay.*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$

mechanism of its role in depth, which will be the focus of the next step of our research, with a view to providing a more reliable and portable means of prognostic assessment of lung cancer patients.

5 Conclusions

This study provided evidence that lncRNAs related to RNA modification can influence LUAD proliferation, metastasis, and patient prognosis. Our risk assessment model holds promise as a novel approach for evaluating the prognosis of patients with LUAD.

Acknowledgements Not applicable.

Author contributions Xin Lin and Jialin Cui designed the study. Yangyang Cheng and Huimin Xu carried out the experiments and performed the analysis. Wanlin Xie and Jingya Zeng interpreted the results. Xin Lin wrote the manuscript. All co-authors reviewed the manuscript and provided feedback.Yihua Sun provided the funding for all the experiments and is the corresponding author of this article. All authors have read and agreed to the published version of the manuscript.

Funding The work in this study was funded by the Natural Science Foundation of Heilongjiang Province, China (LH2021H082).

Data availability Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, Ameh S, Sandel D, Liang XS, Mazzilli S, et al. Commensal Microbiota Promote Lung Cancer Development via gammadelta T Cells. Cell. 2019;176(5):998–1013.
2. Wu ZY, Shi ZY. The prognostic value and immune landscapes of m1A/m5C/m6A-associated lncRNA signature in osteosarcoma. Eur Rev Med Pharmacol Sci. 2022;26(16):5868–83.
3. Wu Y, Wang Z, Han L, Guo Z, Yan B, Guo L, Zhao H, Wei M, Hou N, Ye J, et al. PRMT5 regulates RNA m6A demethylation for doxorubicin sensitivity in breast cancer. Mol Ther. 2022;30(7):2603–17.
4. Barbieri I, Kouzarides T. Role of RNA modifications in cancer. Nat Rev Cancer. 2020;20(6):303–22.
5. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. RNA. 2017;23(12):1754–69.
6. Zhao LY, Song J, Liu Y, Song CX, Yi C. Mapping the epigenetic modifications of DNA and RNA. Protein Cell. 2020;11(11):792–808.
7. Zheng H, Zhu M, Li W, Zhou Z, Wan X. m(5) C and m(6) A modification of long noncoding NKILA accelerates cholangiocarcinoma progression via the miR-582-3p-YAP1 axis. Liver Int. 2022;42(5):1144–57.
8. Li Y, Xia Y, Jiang T, Chen Z, Shen Y, Lin J, Xie L, Gu C, Lv J, Lu C, et al. Long noncoding RNA DIAPH2-AS1 promotes neural invasion of gastric cancer via stabilizing NSUN2 to enhance the m5C modification of NTN1. Cell Death Dis. 2023;14(4):260.
9. Qian X, Zhao J, Yeung PY, Zhang QC, Kwok CK. Revealing lncRNA structures and interactions by sequencing-based approaches. Trends Biochem Sci. 2019;44(1):33–52.
10. Hong JH, Jin EH, Kang H, Chang IA, Lee SI, Sung JK. Correlations between genetic polymorphisms in long non-coding RNA PRNCR1 and gastric cancer risk in a Korean population. Int J Mol Sci. 2019;20(13):3355.
11. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, Furuta K, Tsuta K, Shibata T, Yamamoto S, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res. 2012;72(1):100–11.

12. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7): e47.
13. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747–56.
14. Friedman J et al. Package 'glmnet'. Journal of Statistical Software. 2010a, 2021. 33(1).
15. Therneau TM, Lumley T. Package 'survival.' R Top Doc. 2015;128(10):28–33.
16. Kassambara A et al. Package 'survminer'. Drawing survival curves using 'ggplot2'(R package version 03 1). 2017;3.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
18. MaSG K. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
19. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7.
20. Wu T, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (N Y). 2021;2(3):100141.
21. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–25.
22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50.
23. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinform. 2013;14:7.
24. von Mering C, et al. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003;31(1):258–61.
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
26. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. 2003;4:2.
27. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8(Suppl 4):S11.
28. Barbie DA, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108–12.
29. Wei T, et al. Package 'corrplot.' Statistician. 2017;56(316): e24.
30. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
31. Nunez E, Steyerberg EW, Nunez J. Regression modeling strategies. Rev Esp Cardiol. 2011;64(6):501–7.
32. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565–74.
33. Wu XN, Su D, Mei YD, Xu MQ, Zhang H, Wang ZY, Li LL, Peng L, Jiang JY, Yang JY, et al. Identified lung adenocarcinoma metabolic phenotypes and their association with tumor immune microenvironment. Cancer Immunol Immunother. 2021;70(10):2835–50.
34. Wu J, Li L, Zhang H, Zhao Y, Zhang H, Wu S, Xu B. A risk model developed based on tumor microenvironment predicts overall survival and associates with tumor immunity of patients with lung adenocarcinoma. Oncogene. 2021;40(26):4413–24.
35. Liu Z, Lu T, Wang L, Liu L, Li L, Han X. Comprehensive molecular analyses of a novel mutational signature classification system with regard to prognosis, genomic alterations, and immune landscape in glioma. Front Mol Biosci. 2021;8: 682084.
36. Xu F, Chen JX, Yang XB, Hong XB, Li ZX, Lin L, Chen YS. Analysis of lung adenocarcinoma subtypes based on immune signatures identifies clinical implications for cancer therapy. Mol Ther Oncolytics. 2020;17:241–9.
37. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41(Database issue):955–61.
38. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Brief Bioinform. 2021;22(6):bbab260. https://doi.org/10.1093/bib/bbab260
39. Li X, Xiong X, Yi C. Epitranscriptome sequencing technologies: decoding RNA modifications. Nat Methods. 2016;14(1):23–31.
40. Zhou J, Yang L, Zhong T, Mueller M, Men Y, Zhang N, Xie J, Giang K, Chung H, Sun X, et al. H19 lncRNA alters DNA methylation genome wide by regulating S-adenosylhomocysteine hydrolase. Nat Commun. 2015;6:10221.
41. Wang J, Yuan YX, Tang L, Zhai HQ, Zhang DH, Duan LC, Jiang XL, Li C. Long non-coding RNA-TMPO-AS1 as ceRNA binding to let-7c-5p upregulates STRIP2 expression and predicts poor prognosis in lung adenocarcinoma. Front Oncol. 2022;12:921200.
42. Xu FY, Huang MQ, Chen QY, Niu Y, Hu YH, Hu P, Chen D, He C, Huang K, Zeng Z, et al. LncRNA HIF1A-AS1 promotes gemcitabine resistance of pancreatic cancer by enhancing glycolysis through modulating the AKT/YB1/HIF1α pathway. Can Res. 2021;81(22):5678–91.