scientific reports

Check for updates

OPEN

The role of miR-21 in early-stage lung adenocarcinoma: clinical and bioinformatics insights

Tiantian Qin[1,2] & Guojun Zhang[1]✉

Lung adenocarcinoma (LUAD) is one of the most common types of lung cancer and a leading cause of cancer-related mortality worldwide. This study aimed to identify key microRNAs (miRNAs) and genes involved in LUAD pathogenesis. We enrolled 50 patients diagnosed with early-stage LUAD between October 2021 and September 2023 as the observation group, along with 50 healthy individuals as the control group. Peripheral blood samples (5 mL) were collected from each participant for exosome isolation and quantitative analysis of miR-21. The results demonstrated that miR-21 expression was significantly higher in the observation group compared to the control group ($P < 0.05$). To further explore the molecular mechanisms underlying LUAD, publicly available RNA sequencing (RNA-seq) and single-cell RNA sequencing (scRNA-seq) datasets were analyzed using bioinformatics approaches. Differential expression analysis of The Cancer Genome Atlas (TCGA)-LUAD and Gene Expression Omnibus (GEO) datasets identified 296 differentially expressed miRNAs (192 upregulated, 104 downregulated) and 2942 differentially expressed genes (1691 upregulated, 1251 downregulated). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses indicated that miR-21 target genes are involved in critical biological processes, including cell proliferation, apoptosis, and signal transduction. A protein–protein interaction (PPI) network analysis revealed key protein modules associated with LUAD progression.Immunohistochemistry (IHC) confirmed the high expression of miR-21 target genes IL12A and TGFBI in LUAD tissues. Additionally, the interaction between miR-21 and long non-coding RNAs (lncRNAs) was explored, revealing complex regulatory networks. Gene co-expression analysis of the miR-21 target gene ZNF367 identified highly correlated genes associated with mitosis and DNA repair. Immune infiltration analysis using CIBERSORT demonstrated significant differences in immune cell composition between LUAD patients and healthy controls. Survival analysis further indicated that high miR-21 expression is associated with poorer prognosis, underscoring its potential as a prognostic biomarker.Furthermore, scRNA-seq provided insights into the dynamic molecular changes occurring during the transition from normal to malignant states. Collectively, these findings highlight the critical role of miR-21 in LUAD pathogenesis and suggest potential avenues for targeted therapies and personalized treatment strategies.

Keywords Lung adenocarcinoma, Single-cell RNA-sequencing, Networking analysis, miR-21

Abbreviations

NSCLC	Non-small cell lung cancer
scRNA-seq	Single-cell RNA sequencing
GEO	Gene expression omnibus
DEGs	Differentially expressed genes
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
PPI	Protein–protein interaction
ROC	Receiver operating characteristics
LUAD	Lung adenocarcinoma
STRING	Search tool for the retrieval of interacting genes
AUC	Area under curve
logFC	Log of fold change

[1]Department of Respiratory and Critical Medicine, The First Affiliated Hospital of Zhengzhou University, 1 Jianshedong Road, Zhengzhou, Henan, China. [2]Department of Respiratory and Critical Medicine, Zhengzhou Central Hospital Affiliated to Zhengzhou University, 16 Tongbaibei Road, Zhengzhou, Henan, China. ✉email: gjzhangzzu@126.com

Cancer, as a severe disease that endangers human health, has become a major global health issue[1]. Among various types of cancer, lung cancer stands out as one of the most prevalent, with lung adenocarcinoma being the primary subtype. Its incidence and mortality rates remain alarmingly high, imposing a significant burden on society and families[2–4]. According to global cancer statistics, lung cancer is one of the most common cancers worldwide, causing approximately 184,000 deaths in the United States alone in 2018, making it one of the leading causes of cancer-related deaths[5].

In China, lung cancer is also a leading cause of mortality. Data from the Chinese Center for Disease Control and Prevention indicate that the incidence and mortality rates of lung cancer are continually rising, making it one of the most common cancers in the country. This trend imposes a heavy burden on the national healthcare system and patients' families. As social and economic development progresses, lifestyle changes and environmental pollution exacerbate this issue, highlighting the critical need for research and treatment of lung adenocarcinoma.

Lung adenocarcinoma, a common subtype of lung cancer, is closely associated with various factors, including smoking, environmental influences, and genetic variations[2]. Despite significant research advancements, the etiology and development mechanisms of lung adenocarcinoma remain insufficiently understood. Therefore, in-depth studies of the molecular mechanisms underlying lung adenocarcinoma are essential to better comprehend the disease's pathogenesis, ultimately aiding in the prevention, diagnosis, and treatment of this malignancy[4].

MicroRNAs (miRNAs) are a class of short non-coding RNAs that have been shown to play crucial regulatory roles in tumorigenesis and cancer progression[6]. By modulating the expression levels of target genes, miRNAs can influence biological processes such as tumor cell proliferation, metastasis, and apoptosis[6]. Specifically, miR-21 plays a significant role in lung adenocarcinoma; its overexpression promotes tumor cell proliferation, inhibits apoptosis, and enhances invasion and metastasis[7,8]. Moreover, miR-21 impacts tumorigenesis and progression by regulating gene expression within the tumor microenvironment[7,8]. Understanding these interactions and identifying related markers is vital for developing diagnostic and therapeutic strategies.

In this study, we investigate the molecular mechanisms driving lung adenocarcinoma (LUAD), with a focus on miR-21. We analyzed blood samples from 50 early-stage LUAD patients and 50 healthy controls to assess miR-21 expression. Through gene expression analysis, bioinformatics predictions, and single-cell sequencing, we identified key pathways and interactions involved in LUAD onset and progression. Therefore, this study aimed to investigate the role of serum exosomal miR-21 in early-stage lung adenocarcinoma and validate its clinical and biological significance through bioinformatics analyses.

Materials and methods

All methods were performed in accordance with the relevant guidelines and regulations. The study protocol was approved by the Research Ethical Committee of Zhengzhou Central Hospital Affiliated to Zhengzhou University, and informed consent was obtained from all participants.

Clinical data

Patients diagnosed with early-stage lung adenocarcinoma (LUAD) and admitted between October 2021 and September 2023 were included as the observation group, while healthy individuals served as the control group. This study strictly screened participants who met the inclusion criteria for both early-stage LUAD and the control group. A 5 mL peripheral blood sample were collected from each participant using anticoagulant tubes containing a coagulation activator. The samples were centrifuged at 3,000 × g for 10 min at room temperature to obtain serum, which was aliquoted and stored at − 80 °C until exosome isolation and subsequent quantitative analysis of miR-21. All blood collection and processing steps were conducted at room temperature following standardized operating procedures to ensure sample quality and experimental reproducibility.The inclusion criteria for the observation group were as follows: confirmed diagnosis of early-stage LUAD (stage I–II, according to the American Joint Committee on Cancer TNM classification), age between 18 and 75 years, and no prior chemotherapy or radiotherapy. The exclusion criteria include a history of other cancer types or major systemic diseases such as cardiovascular or renal diseases, recent (within the past three months) lung surgery or biological therapy, and pregnancy or lactation.For the control group, participants must have no history of cancer or chronic diseases and be matched to the observation group by age and sex. Each group included 50 participants, totaling 100 cases, ensuring an adequate sample size for statistically meaningful comparisons while maintaining sample representativeness and diversity. Serum samples were chosen because blood collection is minimally invasive, readily accessible, and suitable for non-invasive early detection of LUAD. To complement this, tissue RNA-seq data from TCGA and GEO databases were analyzed to validate our findings and provide cross-sample confirmation.The study protocol was approved by the Research Ethical Committee of Zhengzhou Central Hospital Affiliated to Zhengzhou University, and informed consent was obtained from all participants.

Main reagents

The miR-21 primers were synthesized by Invitrogen (Guangzhou, China). TRIzol reagent was purchased from Invitrogen (USA). The reverse transcription and real-time quantitative PCR kits were obtained from Tiangen Biotech (Beijing, China).

Quantification of miR-21 expression by qRT-PCR

Total RNA was extracted from the blood samples of lung adenocarcinoma patients using the TRIzol method, and RNA concentration and purity were assessed with a NanoDrop 2000c spectrophotometer. cDNA was synthesized from total RNA according to the instructions of the reverse transcription kit. The primers for miR-21 and U6 were as follows: miR-21 forward 5′-ATGGGCTGTCTGACAAGCTA-3′, reverse5′-TCTCGCATAGT

ACGCTAGCT-3′;U6forward5′-ATTGGAACGATACAGAGAAGATT-3′, reverse 5′-GGAACGCTTCACGAATTTG-3′, all of which were synthesized by Sangon Biotech (Shanghai, China). The qRT-PCR reaction mixture consisted of 2 μL cDNA, 10 μL Real-Time Master Mix, 1 μL forward primer, 1 μL reverse primer, and RNase-Free ddH_2O to a final volume of 20 μL. The amplification conditions were 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s. miR-21 expression was normalized to U6, and the relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. U6 was used as the internal reference gene for normalization, as it has been widely adopted in circulating miRNA studies[9].

Bioinformatics data sources

We utilized miRNA expression data, RNA-seq expression data, and clinical data from the TCGA-LUAD (The Cancer Genome Atlas—Lung Adenocarcinoma) database. Additionally, single-cell RNA sequencing (scRNA-seq) data for lung adenocarcinoma (LUAD) were obtained from the Gene Expression Omnibus (GEO) database. Specifically, we used the single-cell samples of adenocarcinoma in situ (AIS) GSM5699778 and invasive adenocarcinoma (IAC) GSM5699784 from the GSE189357 dataset. We applied strict inclusion criteria for data selection: (i) ensuring the data included information from both case and control groups, (ii) selecting only miRNA and RNA microarray data, (iii) ensuring the samples were from peripheral blood or tissue, and (iv) using human samples. All datasets used met these inclusion criteria to ensure data reliability and comparability, allowing for accurate and comprehensive results to better understand the role of miRNA in lung adenocarcinoma[10–12].

Differential miRNA and gene expression analysis

Differential expression analyses of miRNA and RNA-seq data from TCGA-LUAD were conducted using the R package DESeq2 (v1.42.0). Differentially expressed miRNAs (DEmiRNAs) were defined as $|\log2FC| \geq 1$ and adjusted $P < 0.05$ (Benjamini–Hochberg correction). Differentially expressed genes (DEGs) were identified using the R packages edgeR (v3.42.0) and limma (v3.56.0) based on the Wilcoxon rank-sum test.[11,13].

Target gene prediction and functional analysis (GO and KEGG)

Differentially expressed miRNAs were identified using "DESeq2", resulting in 296 miRNAs. miR-21 was identified among the top 100 based on log2FC. DEGs (n = 2942) were identified using "edgeR" and "limma" through wilcox analysis. Target genes of miR-21 were predicted using "TargetScan" and "miRDB", with intersections determined using "VENNY", resulting in 172 target genes. Functional enrichment analyses of predicted target genes and DEGs were performed using the R package "clusterProfiler" for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. Enrichment significance was determined with a P-value < 0.05, covering cellular components, molecular functions, and biological processes. The top 4 GO biological processes and significant KEGG pathways were selected based on ascending P-values to explore gene-disease associations[14–16].

Construction of protein–protein interaction (PPI) network

Predicted target genes were used to construct a PPI network using the STRING database (http://string-db.org, version 12.0), with a confidence score > 0.4[17].

Protein expression and validation of target genes

Immunohistochemistry (IHC) images of LUAD tissues were obtained from the Human Protein Atlas (HPA) database (https://www.proteinatlas.org) and compared to matched normal tissues to evaluate differential protein expression of miR-21 target genes.

Interaction network of miR-21 and lncRNAs (ENCORI)

Interactions between miR-21 and lncRNAs were analyzed using the ENCORI software (https://rnasysu.com/encori/) to create an interaction network.

Gene co-expression analysis

Co-expression analysis of miR-21 target genes was performed using the R packages "limma", "ggExtra", "future.apply", "ggplot2", and "ggpubr". Genes with an absolute correlation coefficient $|cor| > 0.8$ and $P < 0.05$ were visualized in correlation plots.

Immune infiltration analysis

Immune infiltration analysis of TCGA-LUAD gene expression data was conducted using CIBERSORT, with visualization performed using the R package "ggplot2". Statistical significance was defined as $P < 0.05$.

Survival analysis

Survival analysis was conducted using TCGA-LUAD expression and clinical data with the R packages "survminer" and "survival". Kaplan–Meier analysis was used to assess the prognostic relevance of miR-21 expression levels in LUAD patients. Time-dependent ROC curves were generated using "survivalROC", "survival", "survminer", "timeROC", and "tidyverse" to evaluate the predictive accuracy of miR-21.

Prognostic model construction and validation (univariate Cox + Lasso + multivariate Cox)

TCGA-LUAD data were randomly divided into training (80%) and validation (20%) sets. Univariate Cox regression, Lasso regression, and multivariate Cox regression analyses were performed using the R packages "tidyverse", "glmnet", "survival", "caret", "limma", "survivalROC", "survminer", and "timeROC" to construct and validate a survival model.

Group	Male (percentage)	Female (percentage)	Average age of males (years)	Average age of females (years)	Overall average age (years)	miR-21
Observation Group	29(58.000%)	21(42.000%)	64.172 ± 3.752	63.333 ± 3.526	63.820 ± 3.646	2.139 ± 0.246
Control Group	25(50.000%)	25(50.000%)	62.480 ± 3.405	63.600 ± 2.799	63.040 ± 3.136	0.783 ± 0.054
t-value	0.797		− 1.725	0.286	− 1.147	− 38.072
P-value	0.427		0.091	0.776	0.254	0.000

Table 1. Comparison between the observation group and the control group.

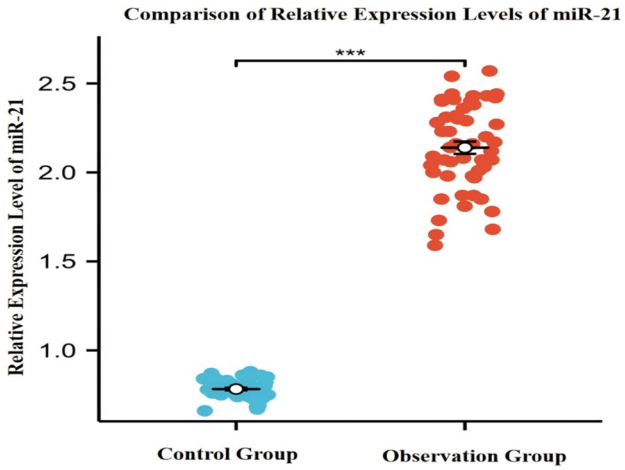

Fig. 1. Relative expression of miR-21 in serum samples from the observation (early-stage LUAD) and control groups. miR-21 expression was significantly higher in the observation group. ***$P < 0.001$.

TIDE analysis (immune therapy efficacy prediction)
TIDE (Tumor Immune Dysfunction and Exclusion) analysis was conducted using the TIDE software (http://tide.dfci.harvard.edu/) on TCGA-LUAD data to predict responses to immune checkpoint blockade (including anti-PD1 and anti-CTLA4 therapies), treatment efficacy, and patient survival, aiding in the identification of tumor responses to immunotherapy for strategic treatment selection.

Single-cell analysis
Single-cell RNA sequencing data for LUAD were obtained from GEO, specifically GSE189357 (AIS sample GSM5699778 and IAC sample GSM5699784). Single-cell analysis and pseudotime trajectory analysis were performed using the R packages Seurat (v4.3.0) and Monocle2 (v2.22.0) to investigate dynamic changes in miR-21 and its target genes at the single-cell level[18,19].

Statistical analysis
Clinical data were analyzed using SPSS software (version 17.0, IBM, USA). Continuous variables were expressed as mean ± standard deviation and compared using Student's t-test or one-way ANOVA, while categorical variables were analyzed using the chi-square test or Fisher's exact test, as appropriate. Bioinformatics analyses were performed using R software (version 4.3.1). Differential expression of miRNAs was analyzed using the DESeq2 package, survival analysis was conducted with the survival and survminer packages, and visualization was carried out using ggplot2. For single-cell RNA-seq data, preprocessing, normalization, and clustering were performed with the Seurat package. A P-value < 0.05 was considered statistically significant.

Results

3.1 The comparison of the relative expression levels of miR-21 in the blood between the observation group and the healthy control group, as well as the comparison of basic information between the two groups[20], is shown. The relative expression level of miR-21 in the observation group was significantly higher than that in the healthy control group ($P < 0.05$), as shown in Table 1 and Fig. 1.

Differential miRNAs and differentially expressed genes

Expression data for LUAD patients and normal controls were obtained from the TCGA database (Tables 2, 3). Using the criteria |log2FC|>1 and $P < 0.05$, we identified 296 differentially expressed miRNAs, with 192 upregulated and 104 downregulated (Fig. 2). Additionally, we identified 2942 differentially expressed genes, with 1691 upregulated and 1251 downregulated (Fig. 3). miR-21 was significantly overexpressed in lung adenocarcinoma (Fig. 4), suggesting its potential role in the pathophysiology of LUAD.

Data	Normal control	LUAD patients	Upregulated miRNAs	Downregulated miRNAs
TCGA-miRNAs	45	510	192	104

Table 2. Summary of miRNA expression data in LUAD patients and normal control (TCGA-miRNAs).

Data	Normal control	LUAD patients	Upregulated genes	Downregulated genes
TCGA-RNAseq	58	513	1691	1251

Table 3. Summary of gene expression data in LUAD patients and normal control (TCGA-RNAseq).

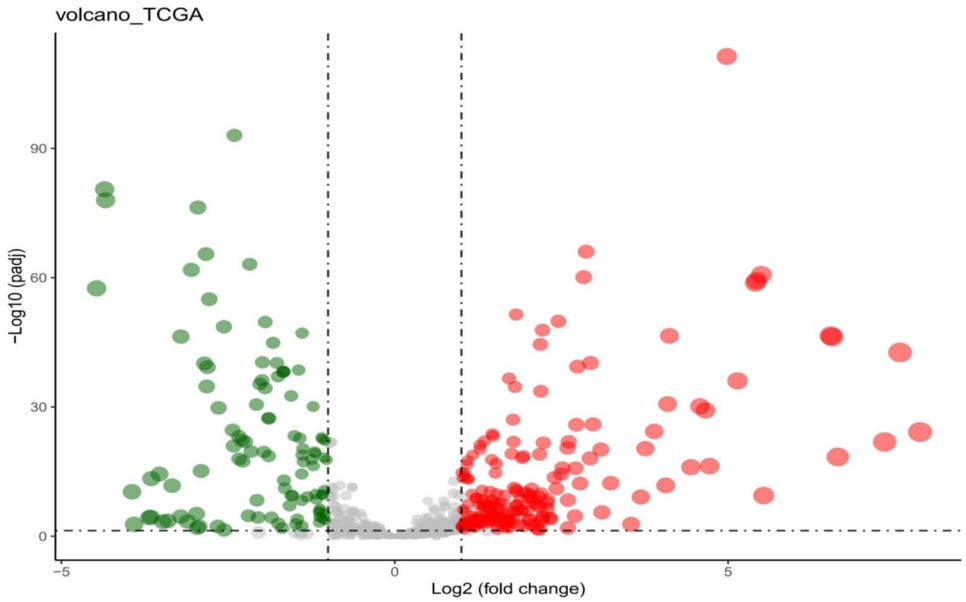

Fig. 2. Volcano plot of differentially expressed miRNAs. Red dots indicate upregulated miRNAs, and green dots indicate downregulated miRNAs. $^{*}P < 0.05$.

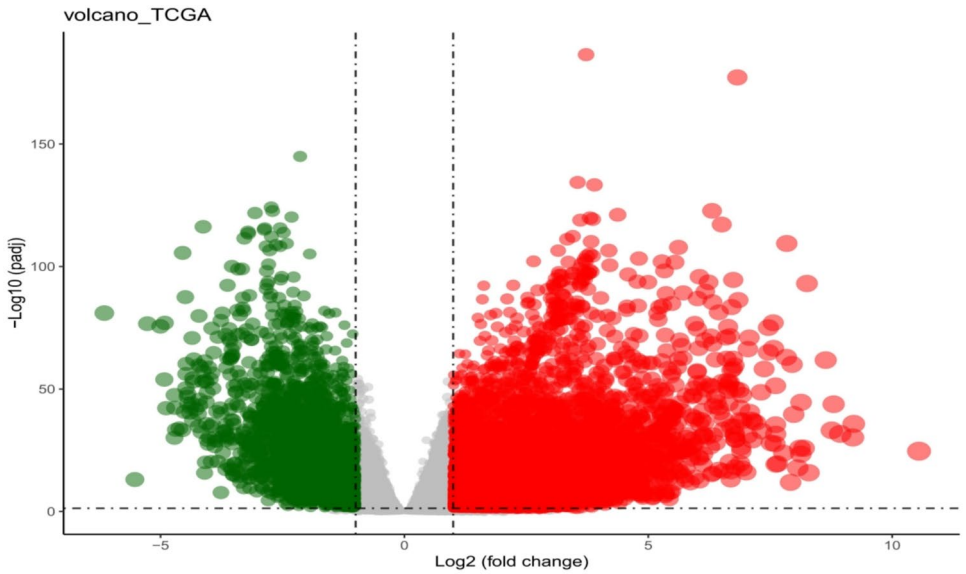

Fig. 3. Volcano plot of differentially expressed genes. Red represents upregulated genes, and green represents downregulated genes. $^{*}P < 0.05$.

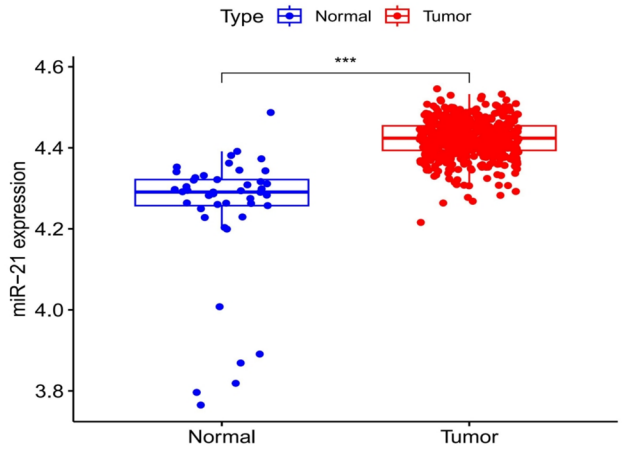

Fig. 4. Boxplot of miR-21 expression levels in tumor (red) and normal (blue) groups. miR-21 expression was significantly higher in the tumor group. $^{***}P < 0.001$.

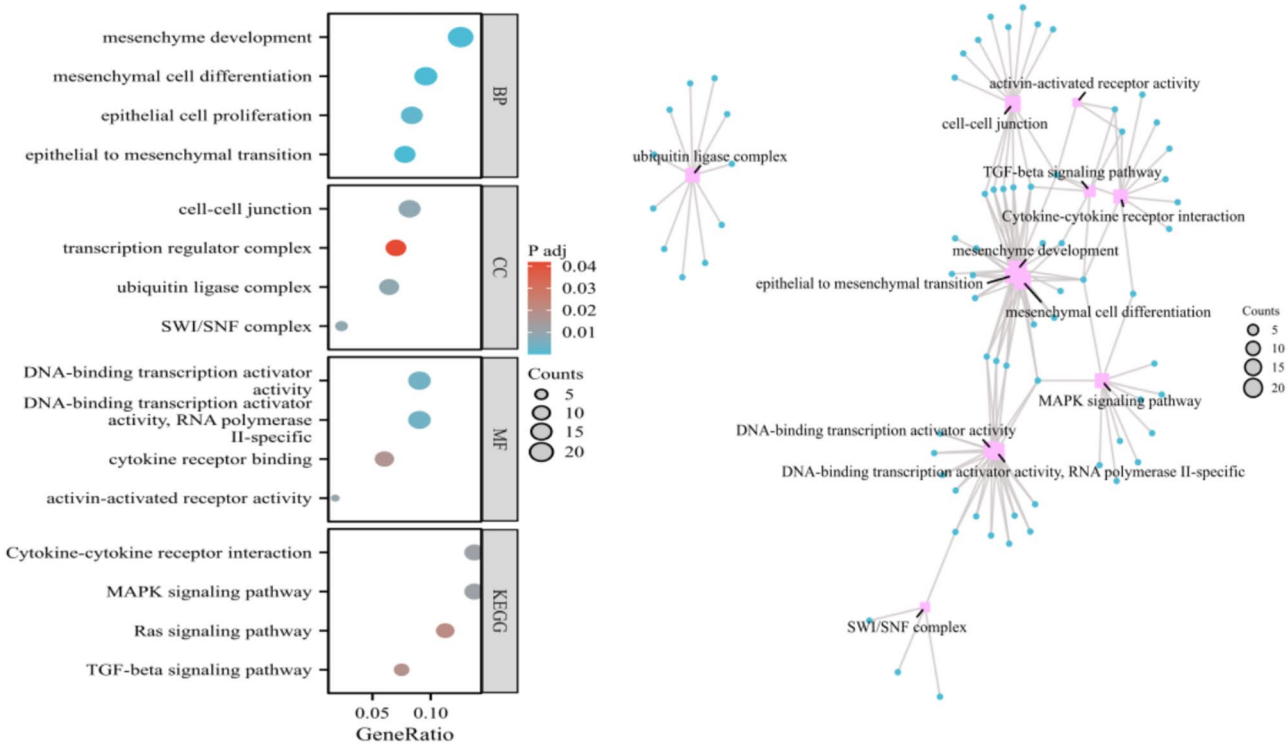

Fig. 5. Bubble chart showing Gene Ontology (GO) and KEGG pathway enrichment analysis of miR-21 target genes. Pathway diagrams were obtained from KEGG (www.kegg.jp) with permission from Kanehisa Laboratories.

Target gene prediction and functional analysis

We identified 2942 differentially expressed genes from RNA-seq data. Using "TargetScan" and "miRDB," we predicted 172 target genes for miR-21 by intersecting the predictions with "VENNY". GO enrichment analysis of miR-21 target genes indicated significant enrichment in processes such as mesenchyme development, mesenchymal cell differentiation, epithelial-mesenchymal transition, cell junction, ubiquitin ligase complex, SWI/SNF complex, and activated receptor activity, highlighting their roles in cell growth, proliferation, apoptosis, cell cycle regulation, DNA repair, signal transduction, migration, and differentiation. KEGG pathway analysis revealed significant enrichment in cytokine-cytokine receptor interaction, MAPK signaling pathway, and TGF-beta signaling pathway, implicating their roles in physiological processes including cell growth, proliferation, differentiation, apoptosis, inflammatory response, embryonic development, immune regulation, tissue repair, and tumorigenesis (Fig. 5).

GO enrichment analysis of DEGs revealed significant enrichment in tissue migration, epithelial cell migration, tissue homeostasis, ERK1 and ERK2 cascade, cell–cell junction, chromosomal region, cell cortex, DNA replication preinitiation complex, cell adhesion mediator activity, growth factor binding, transmembrane receptor protein kinase activity, and CXCR chemokine receptor binding. KEGG pathway analysis indicated significant enrichment in TNF signaling pathway, PI3K-Akt signaling pathway, and transcriptional misregulation in cancer, highlighting their regulatory roles in cell survival, growth, apoptosis, immune response, inflammation, and tumorigenesis (Fig. 6).

miR-21 target gene protein–protein interaction (PPI) network

We constructed a PPI network for miR-21 target genes using the STRING database (http://stringdb.org), with a confidence score > 0.4 (Fig. 7).

Protein expression validation of miR-21 target genes

We selected IL12A and TGFBI, miR-21 target genes, and obtained immunohistochemistry (IHC) images from the HPA database (https://www.proteinatlas.org). Comparing normal lung tissue with LUAD tissue, both IL12A and TGFBI showed high expression in LUAD (Fig. 8).

miR-21 and lncRNA interaction network (ENCORI)

Using ENCORI (https://rnasysu.com/encori/), we analyzed the interaction between miR-21 and lncRNAs, illustrating the interaction network (Fig. 9). The interaction network showed that miR-21 could form complexes with various lncRNAs, influencing cellular biological processes. miR-21-lncRNA interactions can regulate lncRNA expression levels, act as sponges to miR-21, and compete for common target genes, impacting tumorigenesis, cardiovascular diseases, and neurological disorders.

Gene co-expression analysis

We conducted a co-expression analysis of miR-21 target gene ZNF367 using the R packages "limma," "ggExtra," "future.apply," "ggplot2," and "ggpubr". Genes with an absolute correlation coefficient $|cor| > 0.8$ and $P < 0.05$ were visualized (Fig. 10). ZNF367 showed high correlation with genes such as WDR76, TTK, STIL, SMC2, PLK4, KNL1, KIF15, KIF11, FANCI, FAM11B, DTL, BRIP1, BRCA1, CLSPN, and CKAP2L.

Immune infiltration analysis

Using CIBERSORT, we analyzed immune infiltration in TCGA-LUAD gene expression data for normal controls and LUAD patients, visualized with "ggplot2" (Fig. 11). Significant differences were found ($P < 0.05$). LUAD patients had higher levels of naive B cells, plasma cells, CD8 + T cells, and dendritic cells compared to controls, indicating diverse immune cell involvement in LUAD immune infiltration.

Survival analysis

We performed survival analysis using TCGA-LUAD expression and clinical data with "survminer" and "survival". Kaplan–Meier analysis indicated poorer overall survival for patients with high miR-21 expression (Fig. 12). Time-dependent ROC and AUC curves evaluated the predictive accuracy of miR-21, demonstrating its robust predictive performance over time.

Prognostic model construction and validation

TCGA-LUAD data were randomly divided into training (80%) and validation (20%) sets. Univariate Cox regression identified miR-21 as a significant risk factor for LUAD prognosis ($P = 0.037$) (Fig. 13). LASSO regression (lambda.min = 0.02761091, log lambda = −3.589544) (Fig. 14) and multivariate Cox regression confirmed miR-21 as an independent prognostic factor ($P = 0.016$) (Fig. 15). A prognostic model was developed (Fig. 15), and expression heatmaps (Fig. 16), survival curves (Fig. 17), nomogram (Fig. 18), and ROC, DCA, and calibration curves validated the model's accuracy and clinical utility (Figs. 19, 20, 21).

TIDE analysis for immunotherapy efficacy prediction

Using TIDE (http://tide.dfci.harvard.edu/), we analyzed TCGA-LUAD data to predict responses to immune checkpoint blockade (anti-PD1 and anti-CTLA4). TIDE scores showed no significant difference between high-risk and low-risk groups ($P = 0.92$), nor in immune dysfunction ($P = 0.19$) and immune exclusion ($P = 0.31$) (Fig. 22).

Single-cell analysis

We obtained scRNA-seq data for LUAD from GEO, specifically GSE189357 (AIS GSM5699778 and IAC GSM5699784). Using "Seurat" and "Monocle2" we conducted quality control, clustering, and annotation of epithelial cells (epi), myeloids, T cells, B cells, mast cells, and endothelial cells (endo) (Figs. 23, 24, 25). InferCNV identified malignant cells among epithelial cells (Fig. 26), confirming c0, c2, and c5 as malignant[21,22] (Fig. 27). Pseudotime analysis (Figs. 28, 29) revealed the transition of epithelial to malignant cells, with miR-21 target genes showing significant changes (Figs. 30, 31).

Discussion

Lung adenocarcinoma (LUAD) is one of the most prevalent malignant tumors worldwide, with high incidence and mortality[23–26]. Despite advances in diagnosis and treatment, the molecular mechanisms underlying LUAD pathogenesis remain incompletely understood. In this study, we combined analyses of serum-derived exosomal miR-21 with bioinformatics evaluations of TCGA and GEO tissue datasets to investigate its role in early-stage

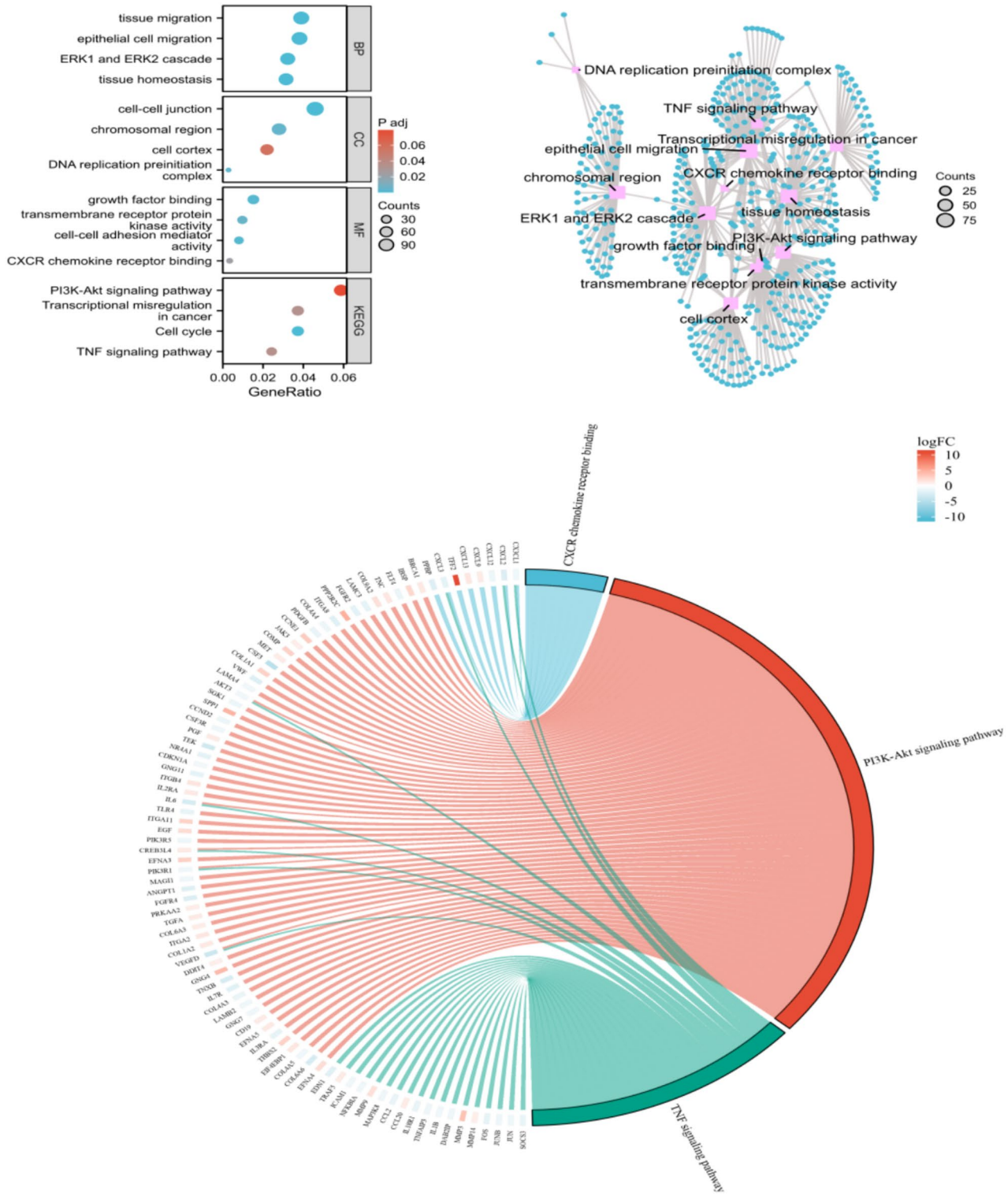

Fig. 6. Bubble chart, network diagram, and chord diagram of Gene Ontology (GO) and KEGG pathway enrichment analyses for differentially expressed genes identified by RNA-seq.

LUAD. Consistent with previous reports, miR-21 was significantly overexpressed in LUAD[27–30], supporting its pivotal role in tumor progression. The use of serum samples provides a minimally invasive approach suitable for early detection and disease monitoring. Integrating tissue RNA-seq data allowed us to validate miR-21 expression across sample types, enhancing the robustness and clinical relevance of our findings. Compared with prior tissue-based studies, our results underscore the potential of circulating miR-21 as a non-invasive biomarker in early-stage LUAD.

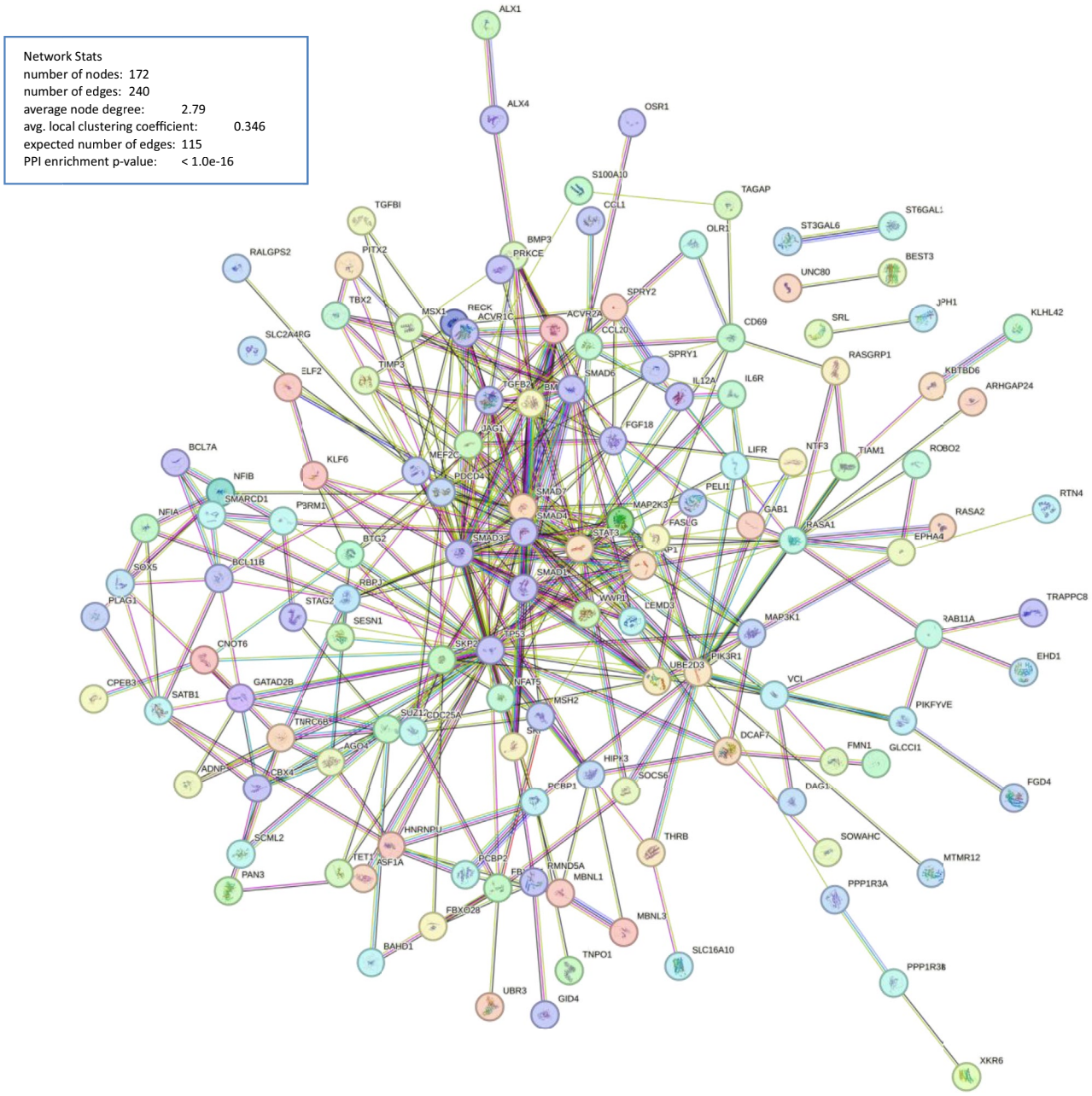

Fig. 7. Protein–protein interaction (PPI) network of miR-21 target genes.

Functional enrichment and target gene analyses suggested that miR-21 may regulate multiple oncogenic processes, including cell proliferation, apoptosis, cell cycle progression, DNA repair, and invasion. Pathway analyses implicated PI3K/AKT, MAPK, TGF-beta signaling, and epithelial-mesenchymal transition (EMT) as potential mediators of miR-21-driven LUAD progression. Furthermore, immune infiltration analysis revealed significant alterations in tumor-infiltrating lymphocytes, including elevated CD8 + T cells, plasma cells, and dendritic cells, suggesting that miR-21 may modulate tumor immune escape mechanisms. These mechanistic insights are consistent with previous literature, indicating that miR-21 can promote malignancy via both intrinsic tumor pathways and modulation of the tumor microenvironment[28–32]. Additionally, our analysis of miR-21-lncRNA interactions revealed complex regulatory networks, implying that miR-21 may exert its oncogenic effects through ceRNA mechanisms, further emphasizing its multifaceted role in LUAD pathogenesis.

Clinically, our findings underscore the utility of miR-21 as a prognostic biomarker[33–38]. High miR-21 expression correlated with poorer overall survival, and time-dependent ROC analyses confirmed its predictive accuracy. The prognostic model constructed using TCGA-LUAD data further validated miR-21 as an independent risk factor, offering a framework for personalized risk stratification. From a translational perspective, circulating

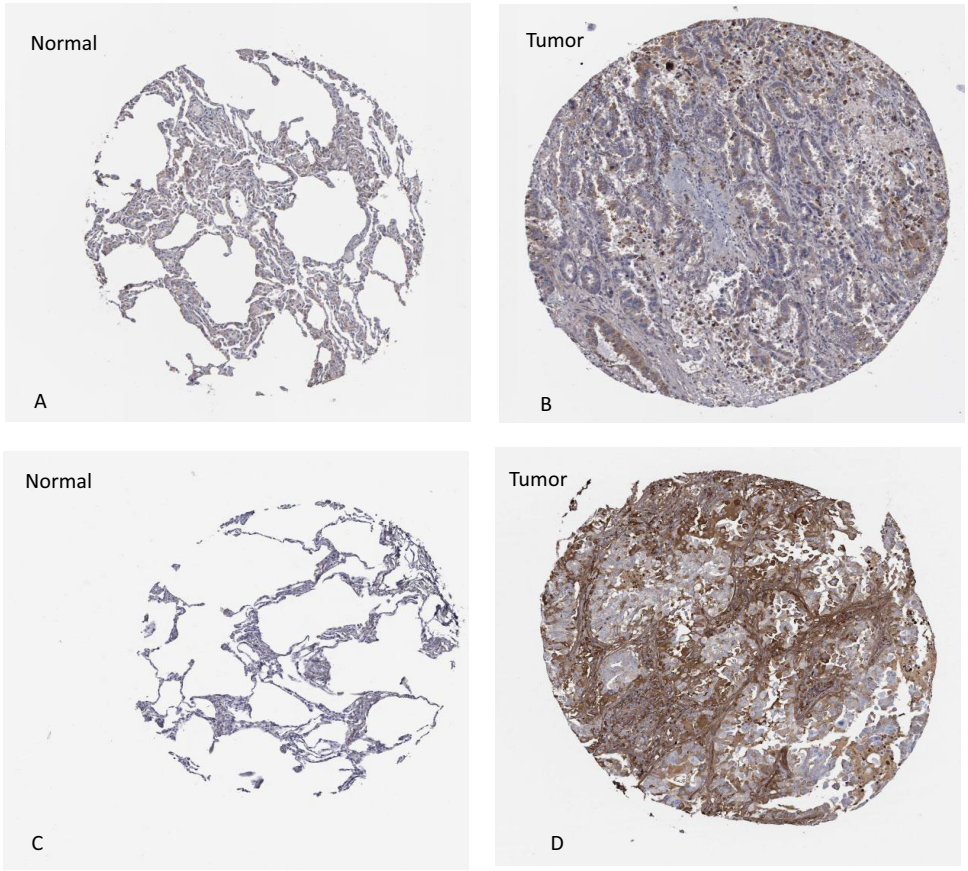

Fig. 8. Differential protein expression of miR-21 target genes IL12A and TGFBI. (**A**,**B**) IL12A expression in normal lung tissue and lung adenocarcinoma, respectively. (**C**,**D**) TGFBI expression in normal lung tissue and lung adenocarcinoma, respectively. Both IL12A and TGFBI show higher expression in lung adenocarcinoma tissues.

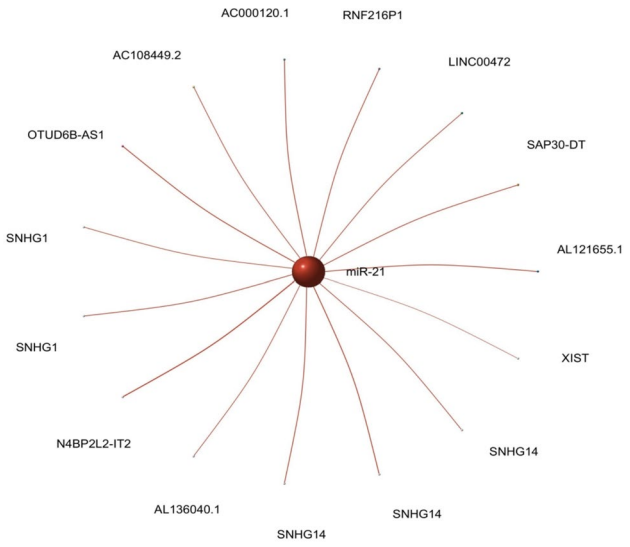

Fig. 9. Interaction network between miR-21 and long non-coding RNAs (lncRNAs).

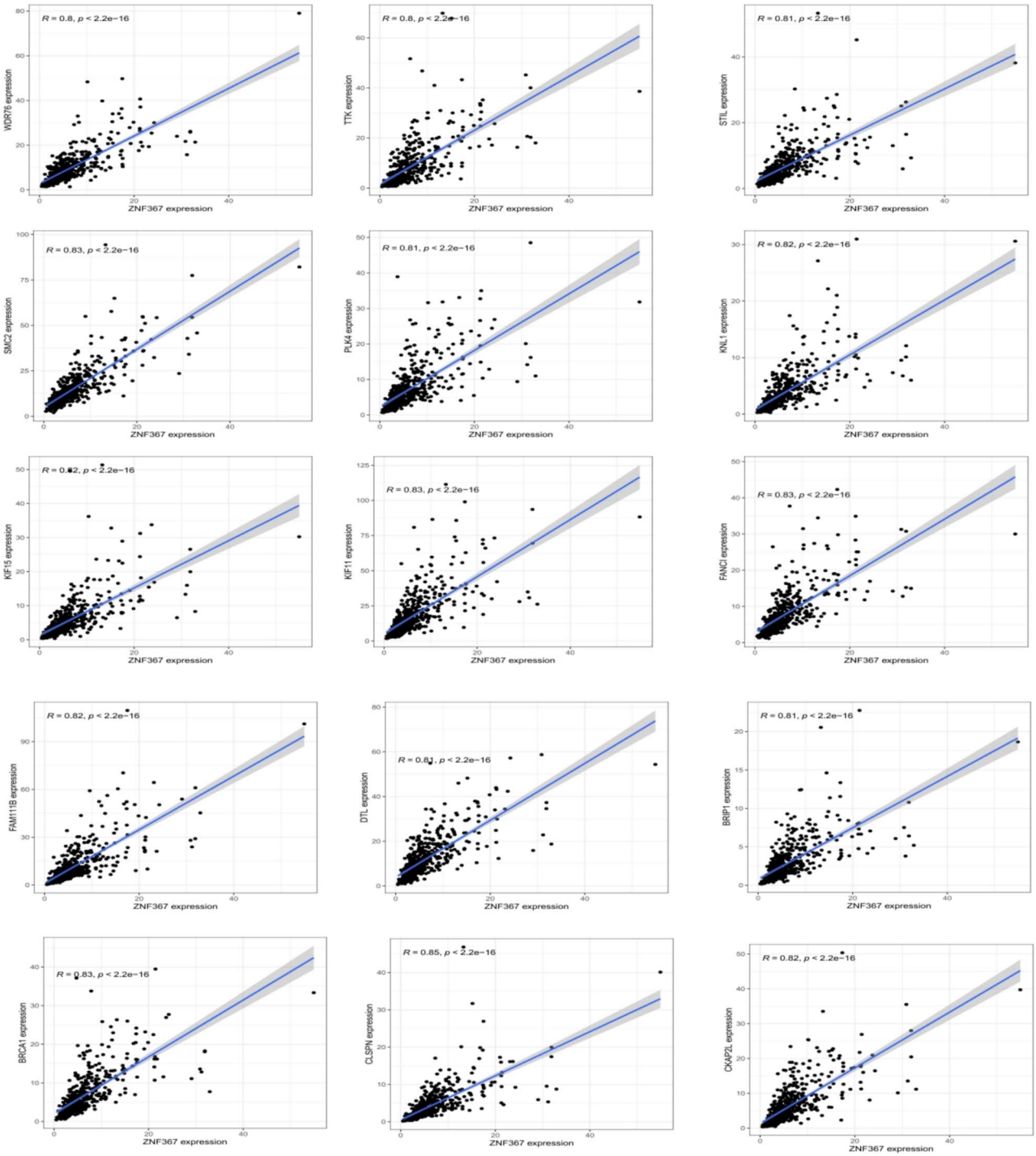

Fig. 10. Gene co-expression network of the miR-21 target gene ZNF367.

miR-21 could facilitate early detection, guide patient monitoring, and potentially serve as a therapeutic target, aligning with ongoing efforts to develop miRNA-based interventions in oncology.

A notable limitation of this study is the use of U6 as the sole reference gene for qRT-PCR normalization. Although U6 is commonly applied in miRNA studies[9], its expression may fluctuate under certain pathological conditions, potentially affecting normalization accuracy. MIQE guidelines recommend multiple validated reference genes, and future work should incorporate this approach. Other limitations include the single-center design, relatively small sample size, and the lack of in vitro or in vivo experimental validation of bioinformatics findings. In addition, information on smoking status and environmental exposures was not available for all patients, which may influence circulating miRNA levels and represents a potential source of confounding.

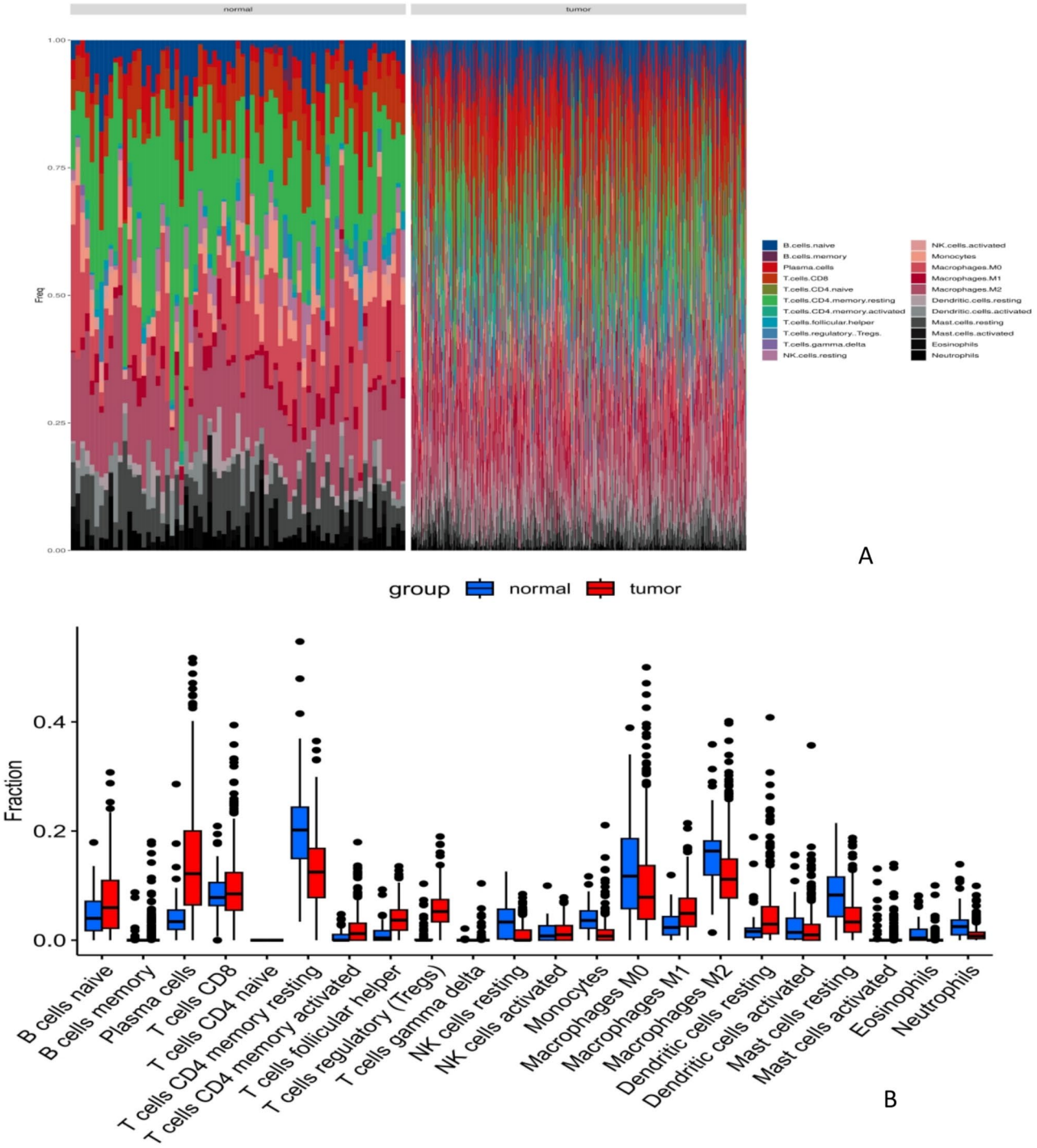

Fig. 11. Immune infiltration analysis in lung adenocarcinoma (LUAD) and normal control tissues. (**A**) Comparison of immune cell abundance between LUAD and normal tissues shown as a bar graph. (**B**) Comparison of immune cell abundance between LUAD and normal tissues shown as a boxplot.

Despite these constraints, the integration of clinical and computational analyses provides a robust framework for understanding miR-21 biology in LUAD.

Future studies should aim to validate these findings in larger, multicenter cohorts, collect comprehensive clinical and environmental data to control for potential confounders, and use multiple reference genes for qRT-PCR normalization. Mechanistic investigations, including functional studies of miR-21 in LUAD cell lines and animal models, are warranted to delineate its role in tumor proliferation, metastasis, and immune regulation. Additionally, exploring the interplay between miR-21, lncRNAs, and immune components may uncover

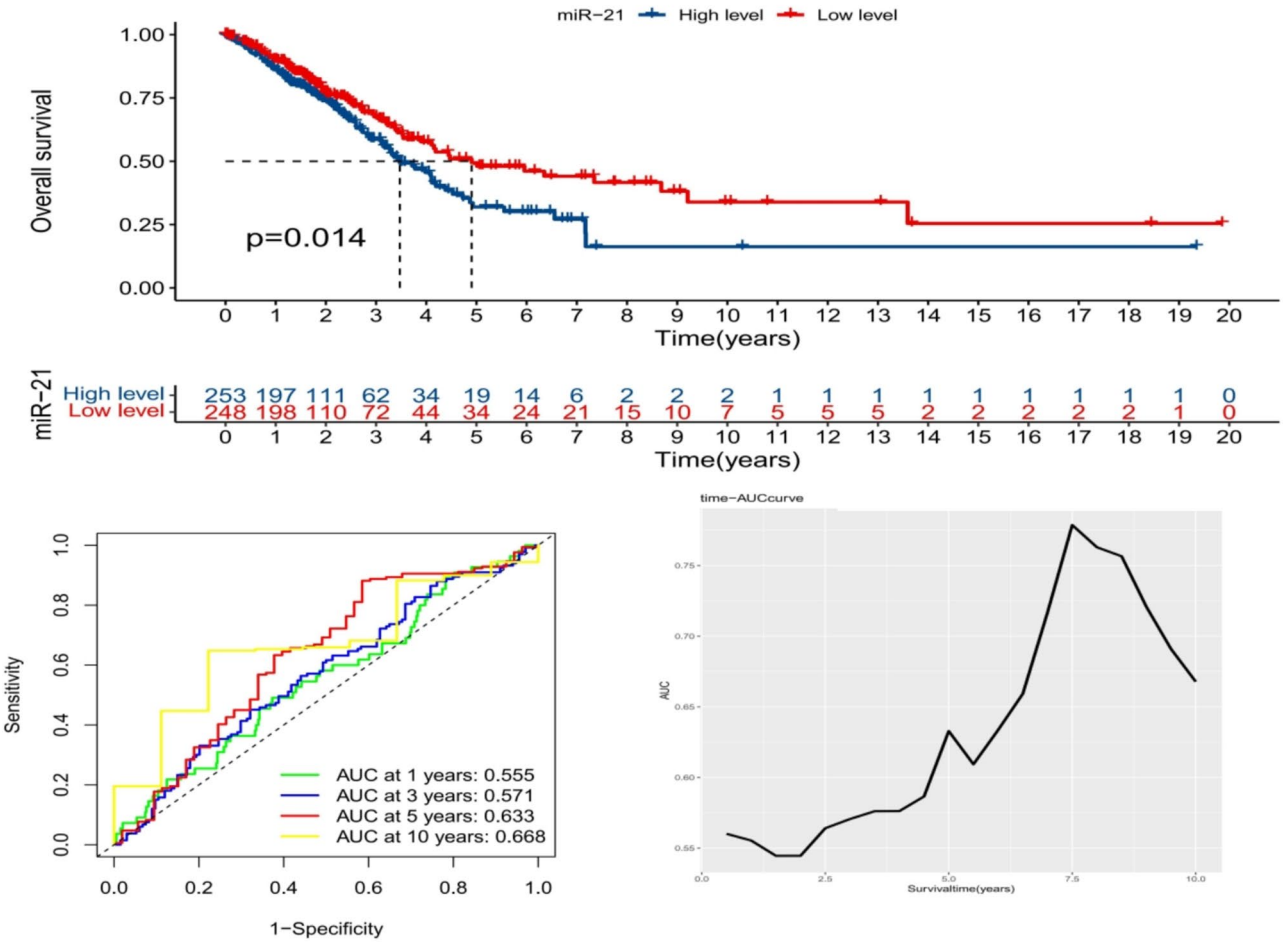

Fig. 12. Kaplan–Meier survival curves, time-dependent ROC curves, and time-dependent AUC curves for LUAD patients.

novel therapeutic targets[39–43]. Such comprehensive studies will advance the clinical translation of miR-21 as a biomarker and therapeutic target in LUAD.

In summary, our study demonstrates that miR-21 is significantly upregulated in early-stage LUAD and regulates multiple oncogenic pathways and immune processes. High miR-21 expression correlates with poor prognosis, highlighting its potential for early detection, risk stratification, and therapeutic targeting. These findings provide a foundation for future research into the clinical and biological implications of miR-21 in LUAD, paving the way for improved patient management and precision medicine strategies.

Conclusion

This study demonstrates that miR-21 is significantly upregulated in early-stage lung adenocarcinoma (LUAD) and regulates multiple oncogenic pathways and immune processes. High miR-21 expression was associated with poorer patient survival, supporting its potential as a prognostic biomarker and therapeutic target. These findings highlight the clinical relevance of miR-21 for early detection, risk stratification, and personalized treatment strategies in LUAD. Future studies should validate these results in larger, multicenter cohorts, incorporate multiple reference genes for qRT-PCR normalization, and perform mechanistic experiments to further elucidate miR-21's role in LUAD pathogenesis and tumor-immune interactions.

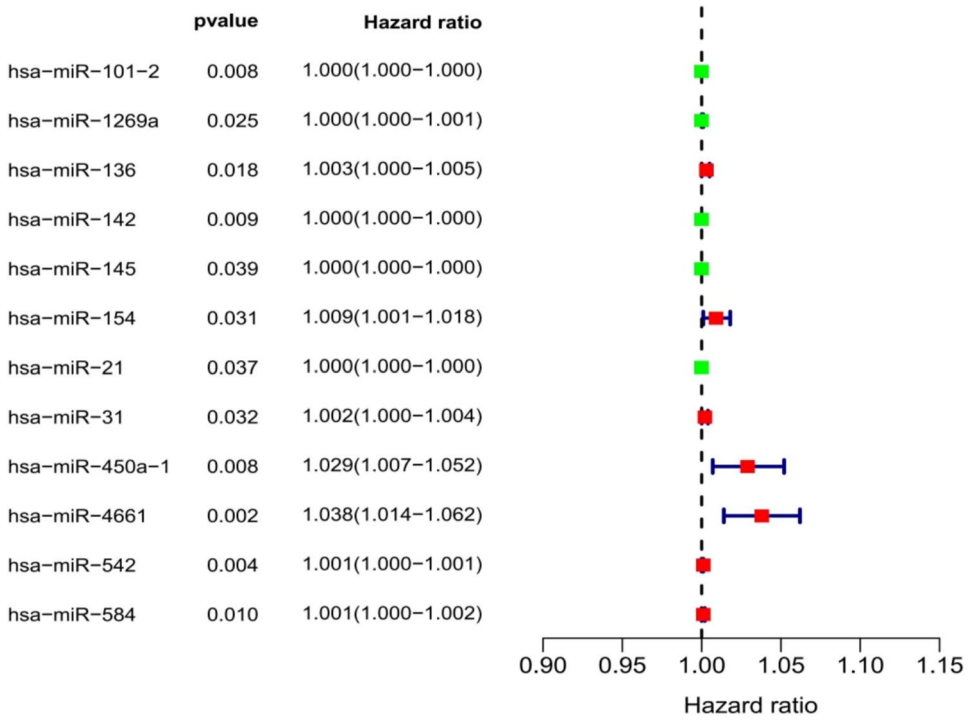

Fig. 13. Univariate Cox regression analysis of prognostic factors in LUAD patients.

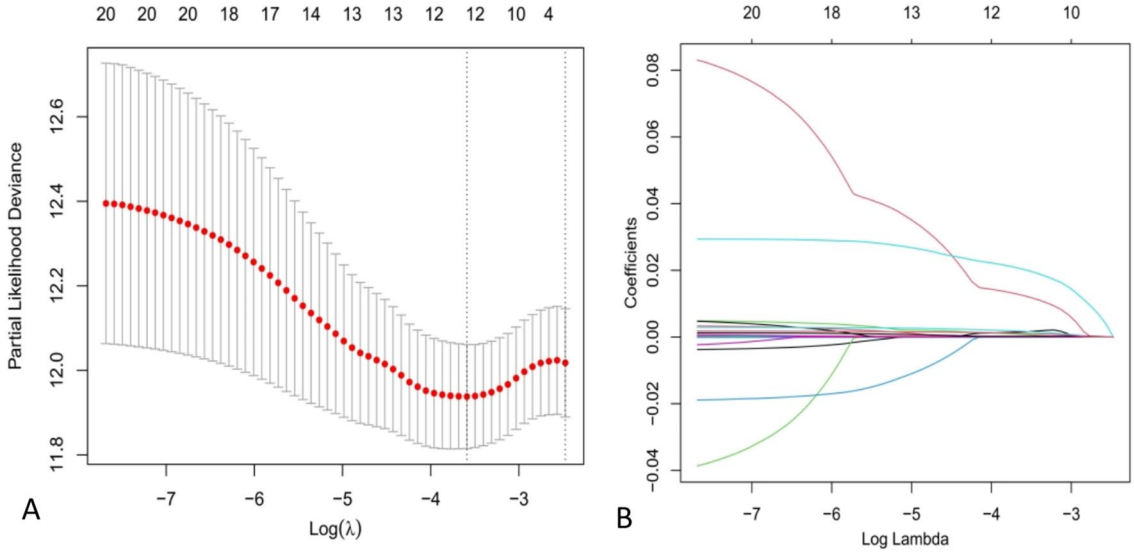

Fig. 14. LASSO-Cox regression analysis. (**A**) Partial likelihood deviance plot. (**B**) LASSO coefficient profile plot. Lambda.min = 0.02761091 (log λ = − 3.589544) was selected for model construction.

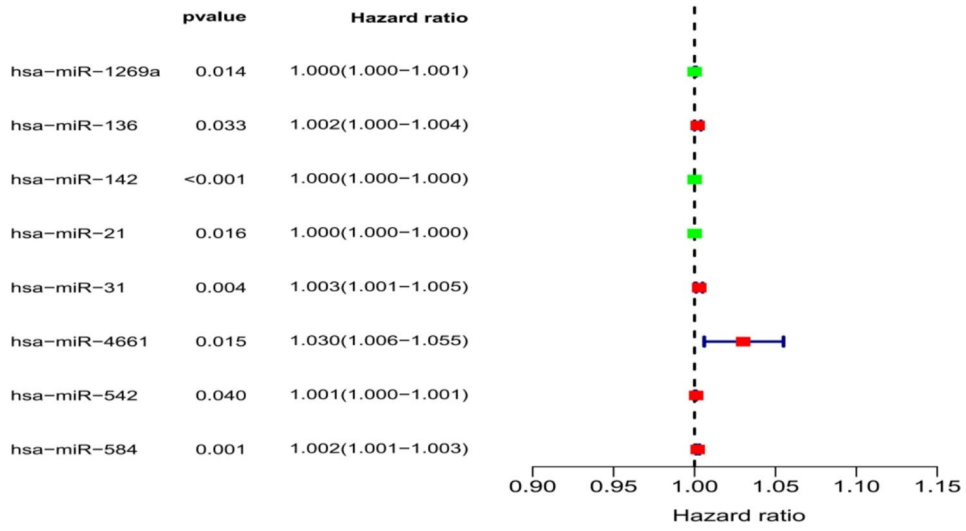

Fig. 15. Multivariate Cox regression analysis and prognostic scoring model for LUAD patient survival. The prognostic score was calculated as follows: Score = 3.562e-04*hsa-mir-1269a + 2.327e-03*hsa-mir-136- 1.308e-04*hsa-mir-142 + 1.729e-06*hsa-mir-21 + 2.927e-03*hsa-mir-31 + 2.976e-02*hsa-mir-4661 + 5.628e-04*hsa-mir-542 + 1.720e-03*hsa-mir-584.

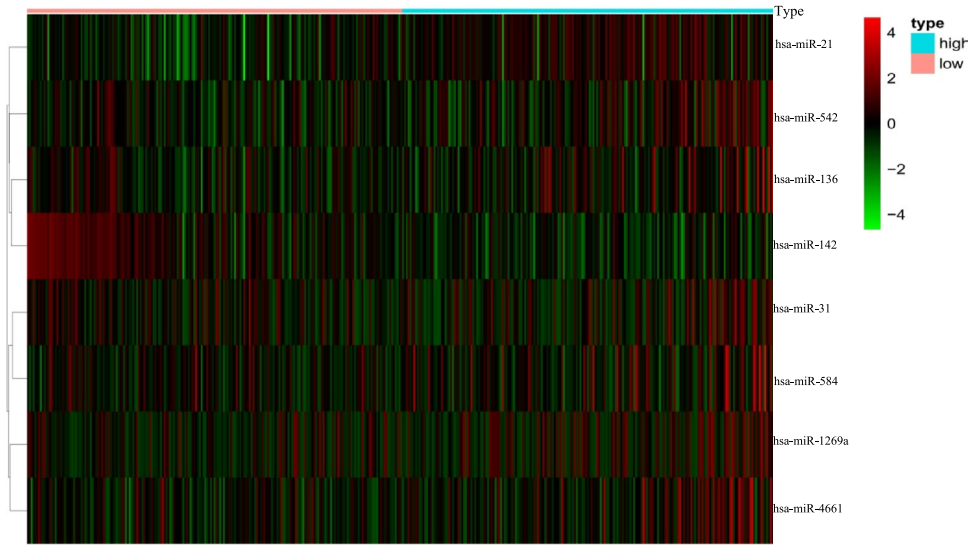

Fig. 16. Heatmap showing the expression of the 8 selected risk miRNAs in high-risk and low-risk LUAD patient groups.

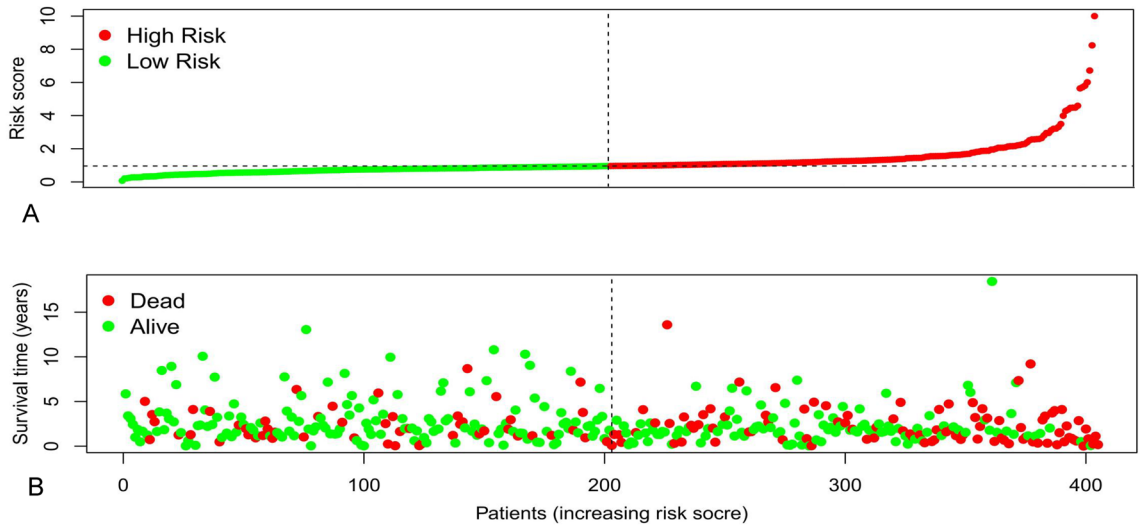

Fig. 17. Risk stratification of LUAD patients based on the prognostic score. (**A**) Division of patients into high-risk and low-risk groups. (**B**) Comparison of survival outcomes between high-risk and low-risk groups.

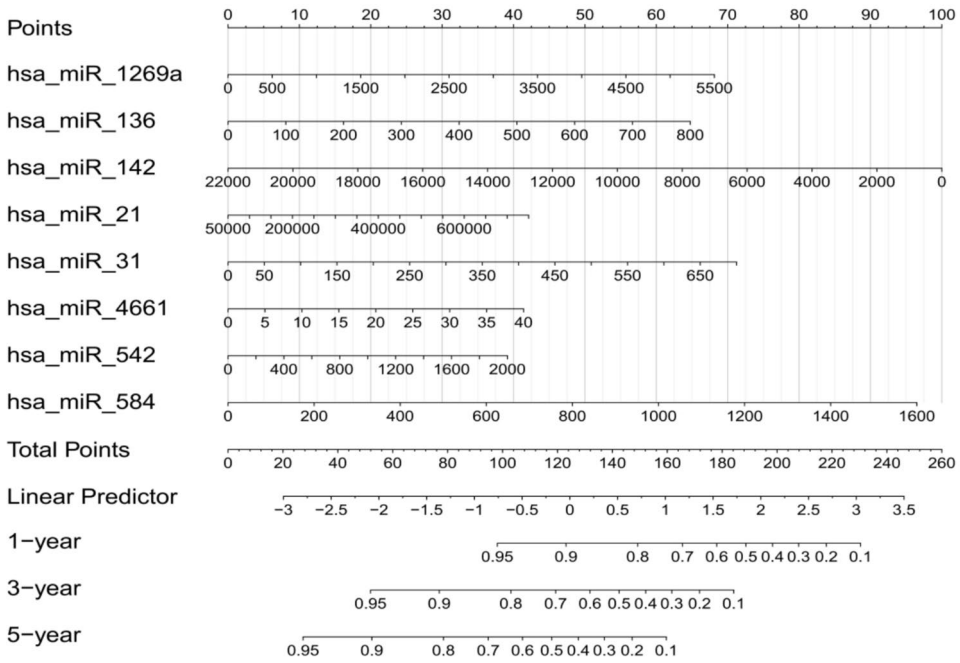

Fig. 18. Nomogram for predicting 1-year, 3-year, and 5-year survival probabilities in LUAD patients.

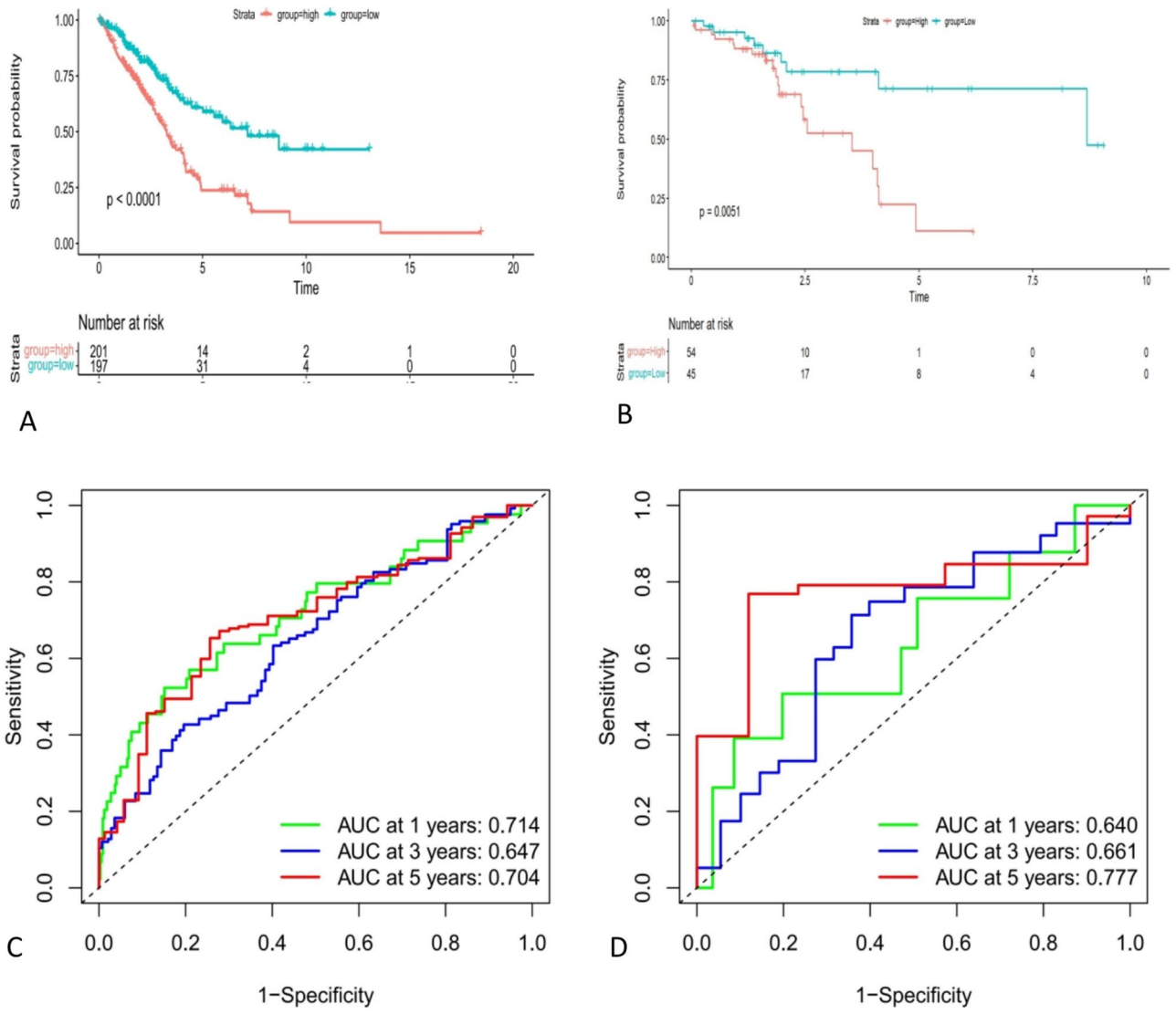

Fig. 19. Kaplan–Meier (KM) survival curves and ROC curves for training and validation sets of LUAD patients. (**A**,**C**) KM survival curves and ROC curves for the training set, respectively. (**B**,**D**) KM survival curves and ROC curves for the validation set, respectively.

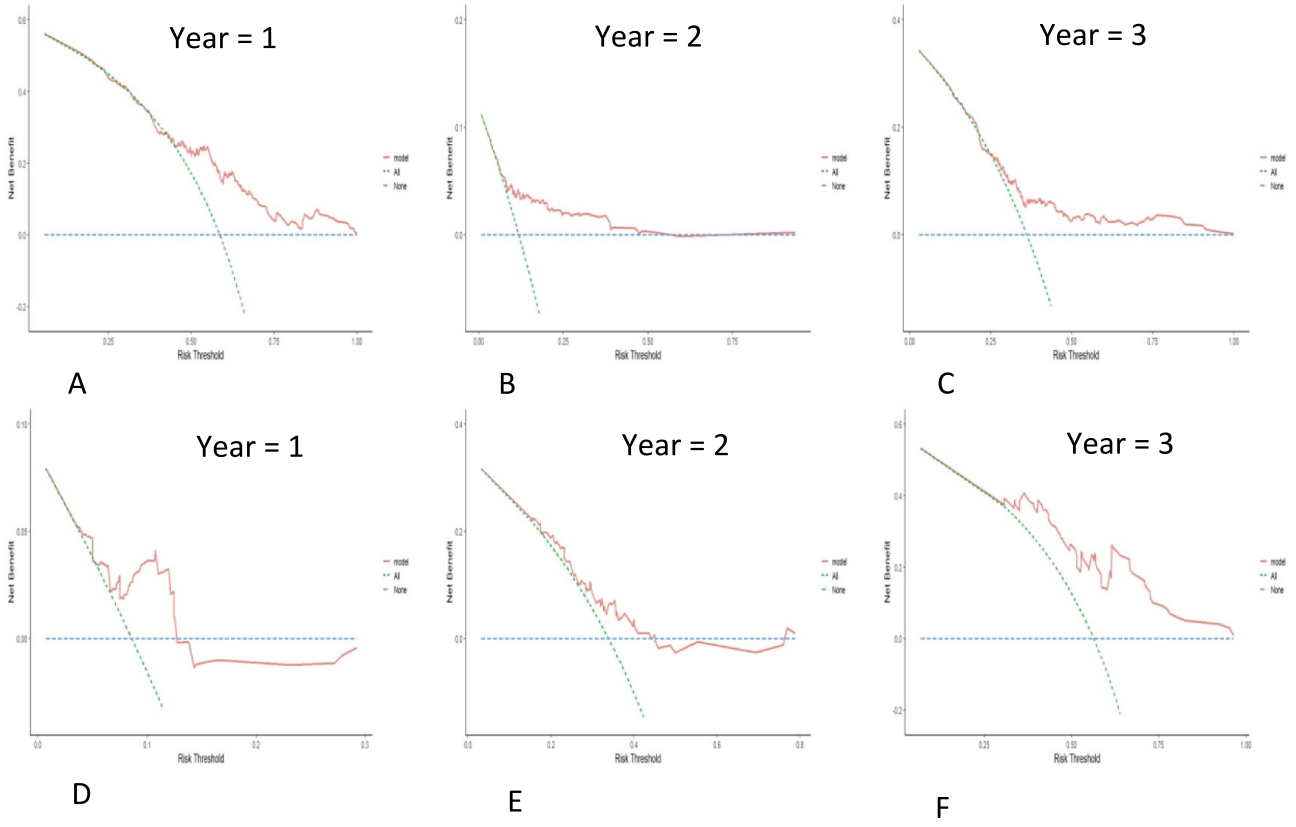

Fig. 20. Decision curve analysis (DCA) evaluating the clinical utility of the LUAD survival prediction model at 1-, 3-, and 5-year time points. (**A**–**C**) DCA for the training set. (**D**–**F**) DCA for the validation set.

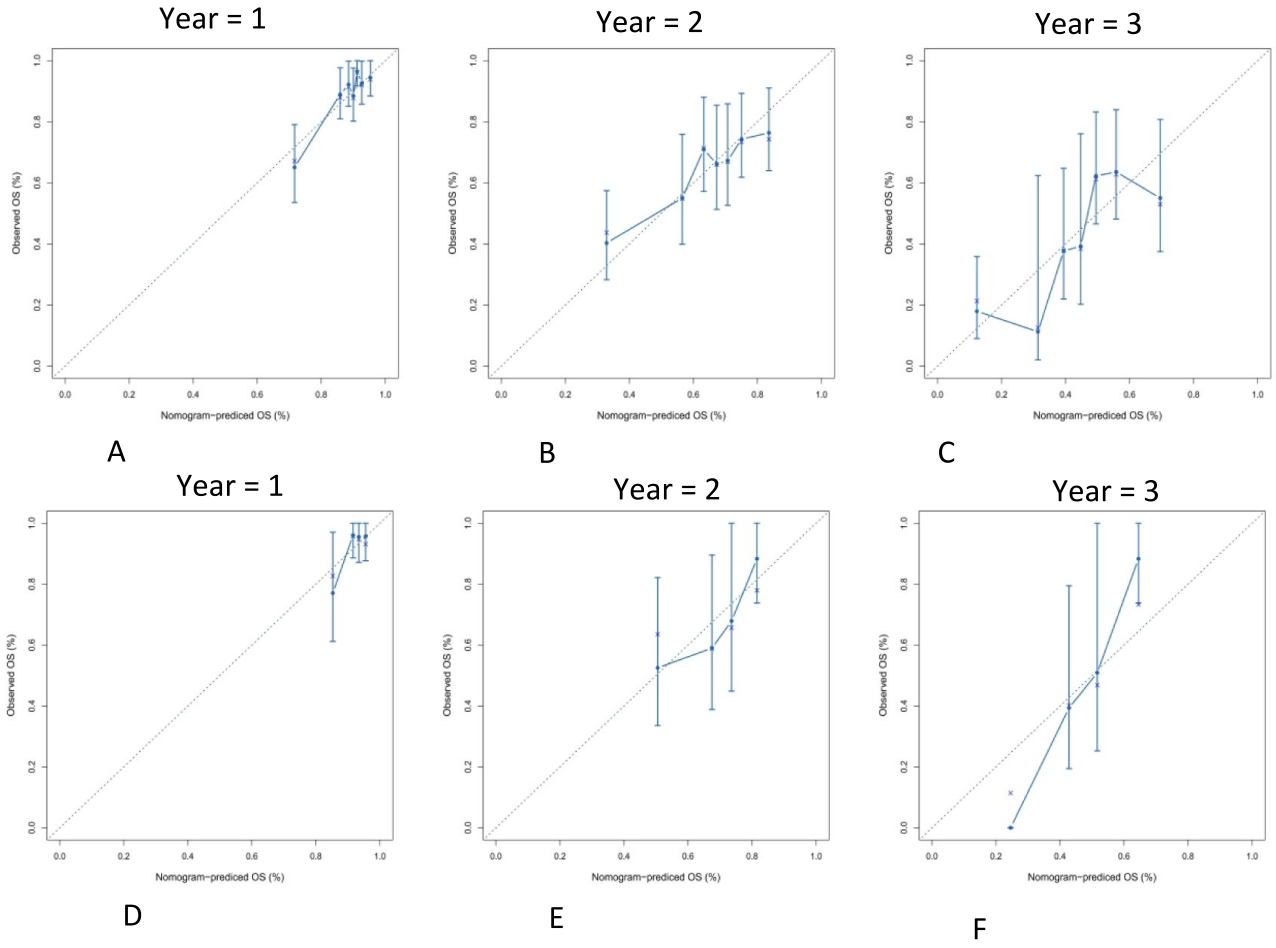

Fig. 21. Calibration plots of the LUAD nomogram for predicting 1-, 3-, and 5-year survival. (**A**–**C**) Calibration plots for the training set. (**D**–**F**) Calibration plots for the validation set.

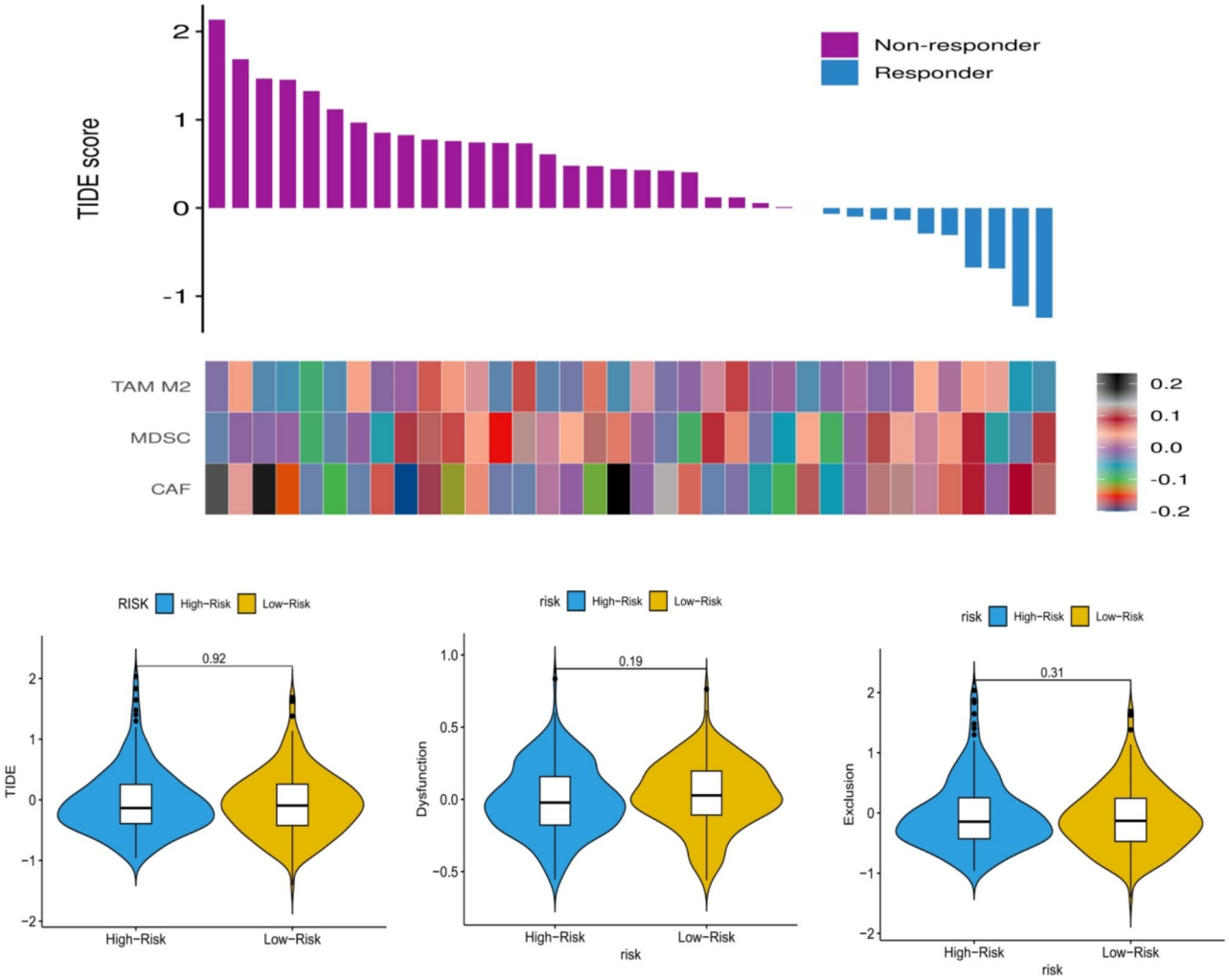

Fig. 22. TIDE (Tumor Immune Dysfunction and Exclusion) analysis in LUAD patients. TAM M2: tumor-associated macrophages (M2 subtype); MDSC: myeloid-derived suppressor cells; CAF: cancer-associated fibroblasts.

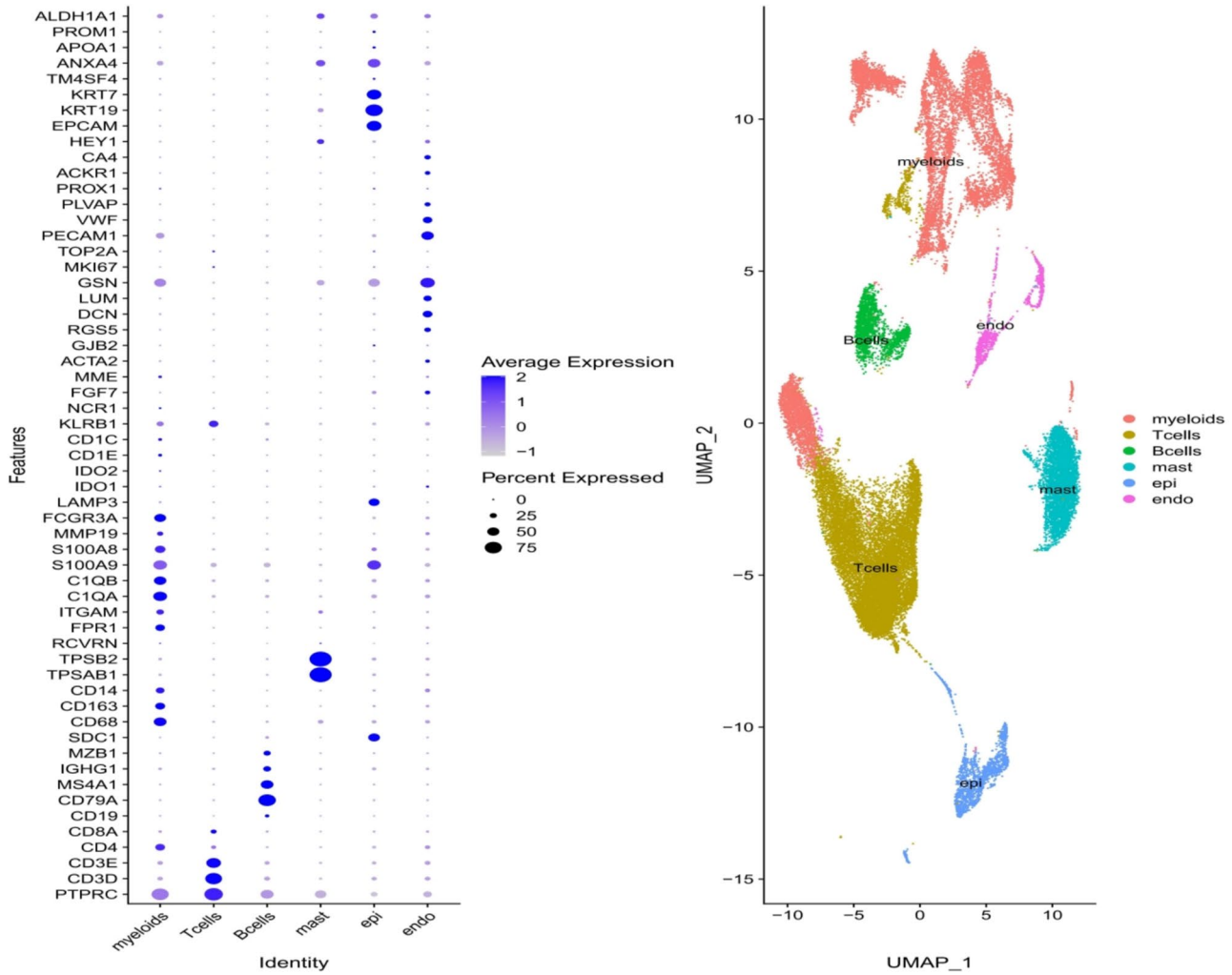

Fig. 23. Expression of marker genes across cell clusters and the UMAP visualization of identified clusters.

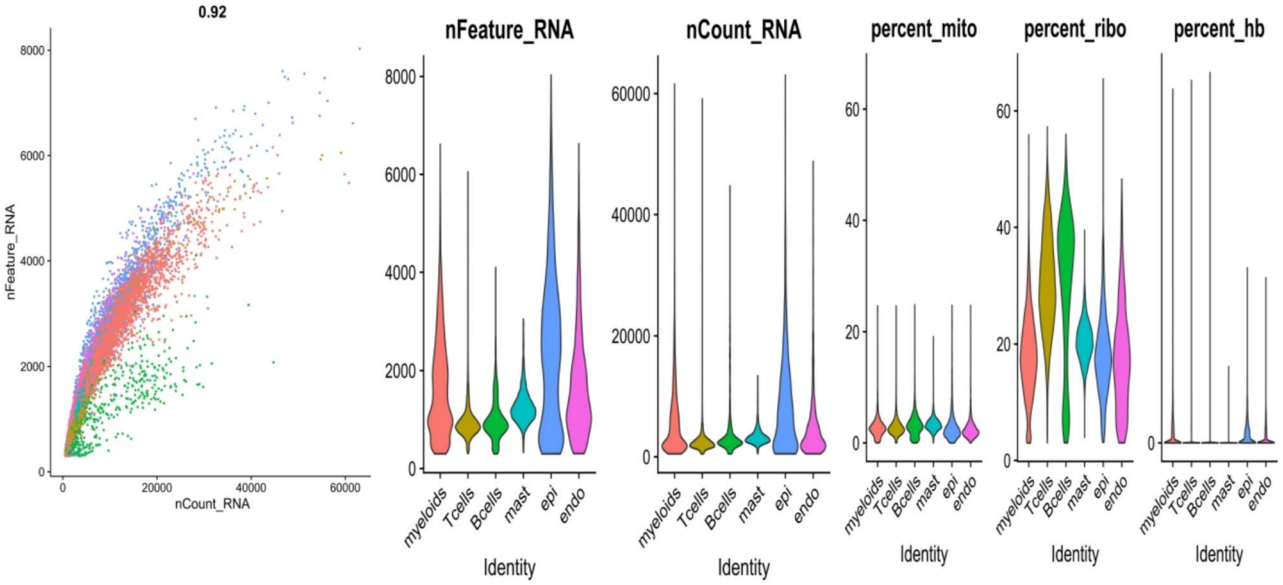

Fig. 24. Quality control plot of single-cell RNA sequencing (scRNA-seq) data. nFeature_RNA: number of genes; nCount_RNA: number of RNA molecules; percent_mito: percentage of mitochondrial genes; percent_ribo: percentage of ribosomal genes; percent_hb: percentage of hemoglobin genes.

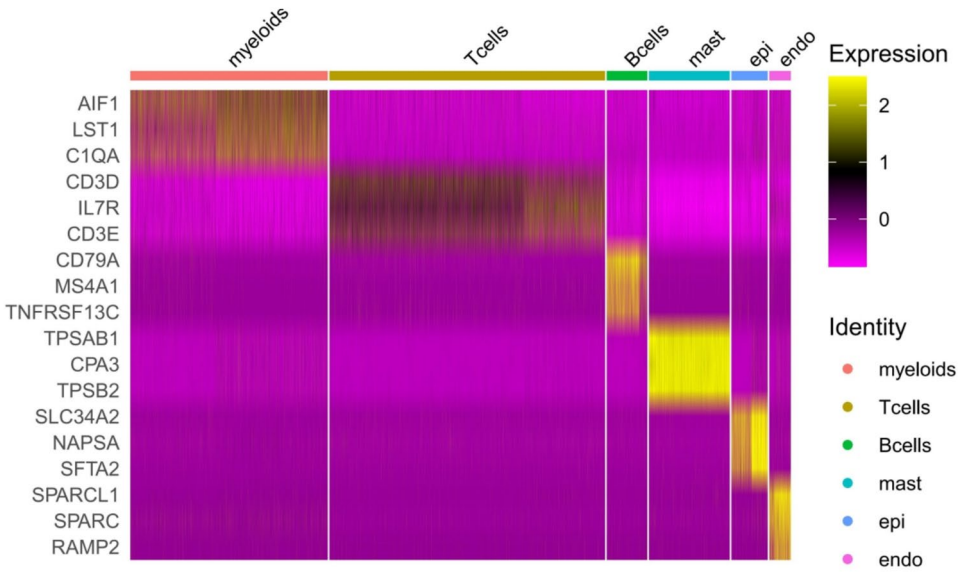

Fig. 25. Heatmap showing the top three marker genes expressed in each cell cluster.

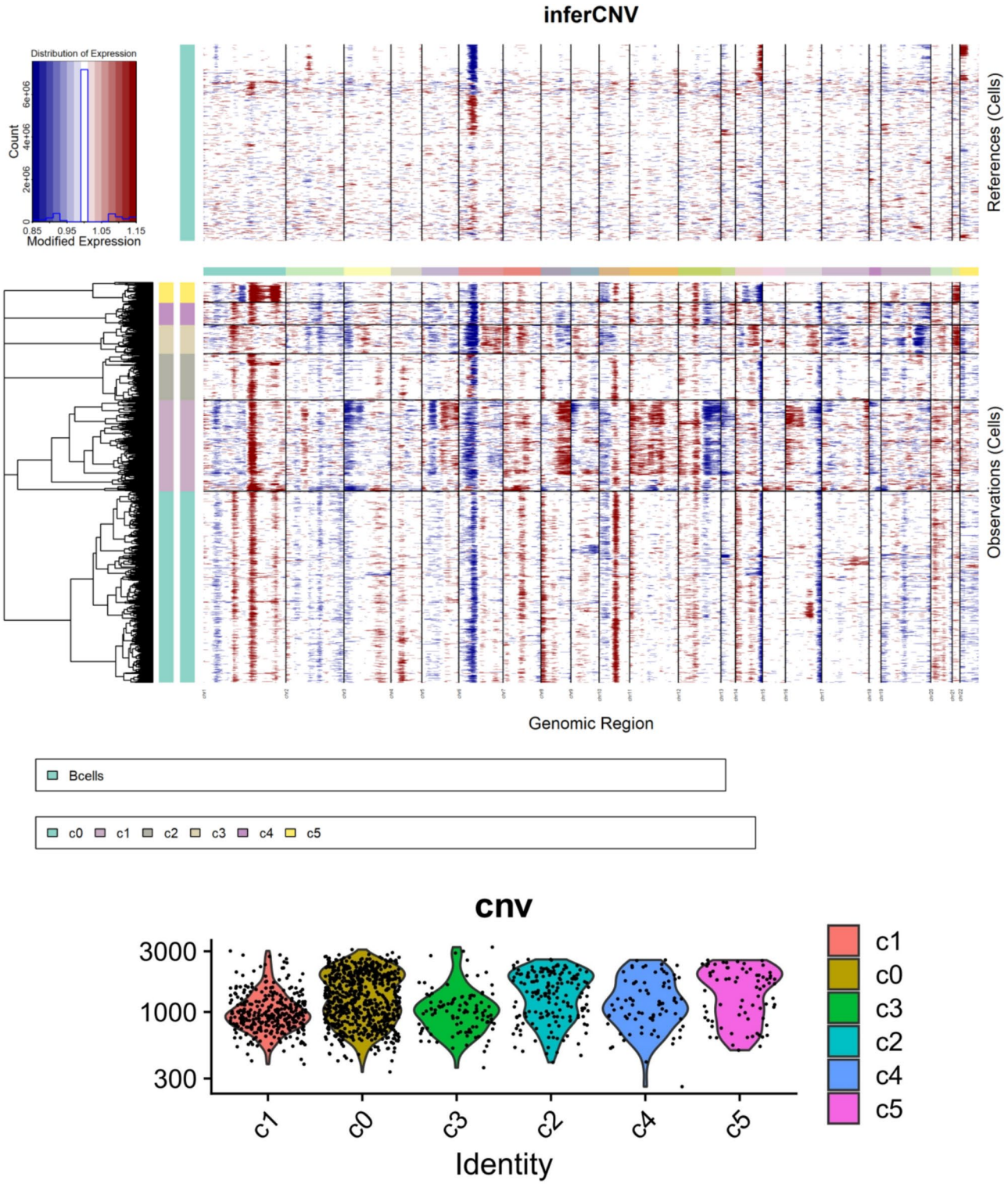

Fig. 26. Results of inferCNV analysis in LUAD single-cell data.

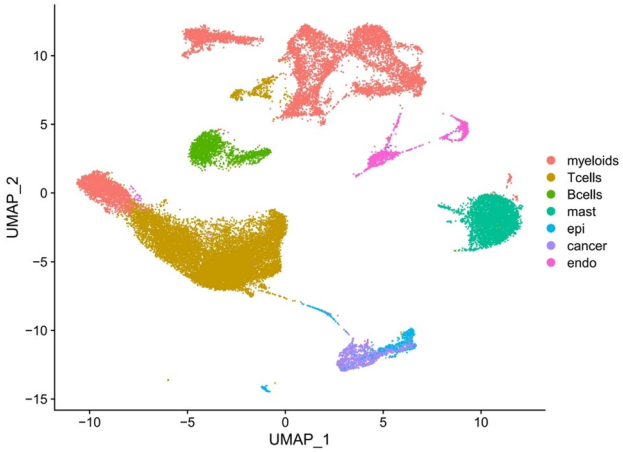

Fig. 27. UMAP visualization of isolated tumor cells in LUAD single-cell data.

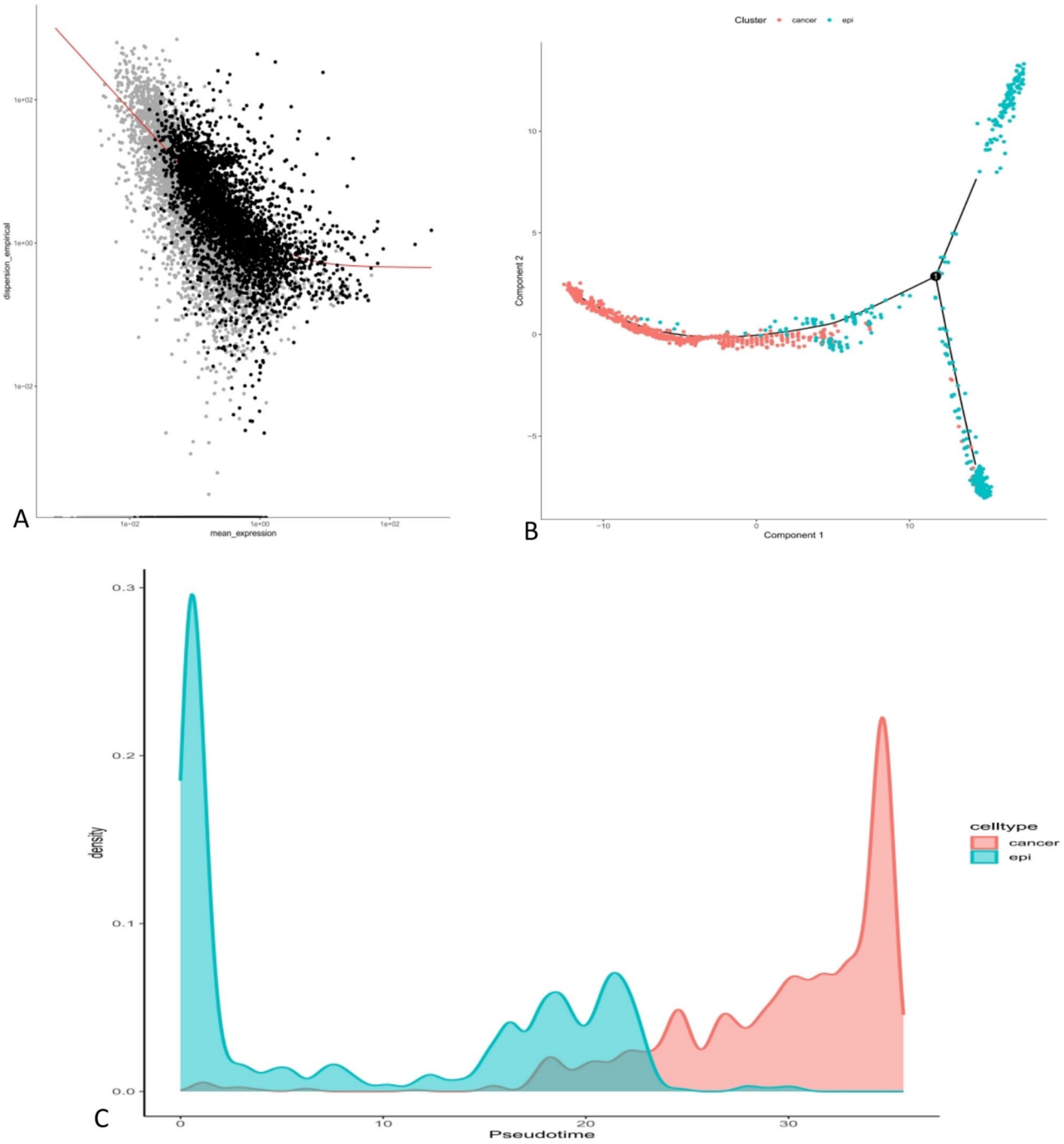

Fig. 28. Single-cell analysis of LUAD. (**A**) Gene scatter plot of single cells. (**B**) Pseudotime trajectory of epithelial (epi) and malignant (cancer) cells. (**C**) Time density plot of epithelial (epi) and malignant (cancer) cells.

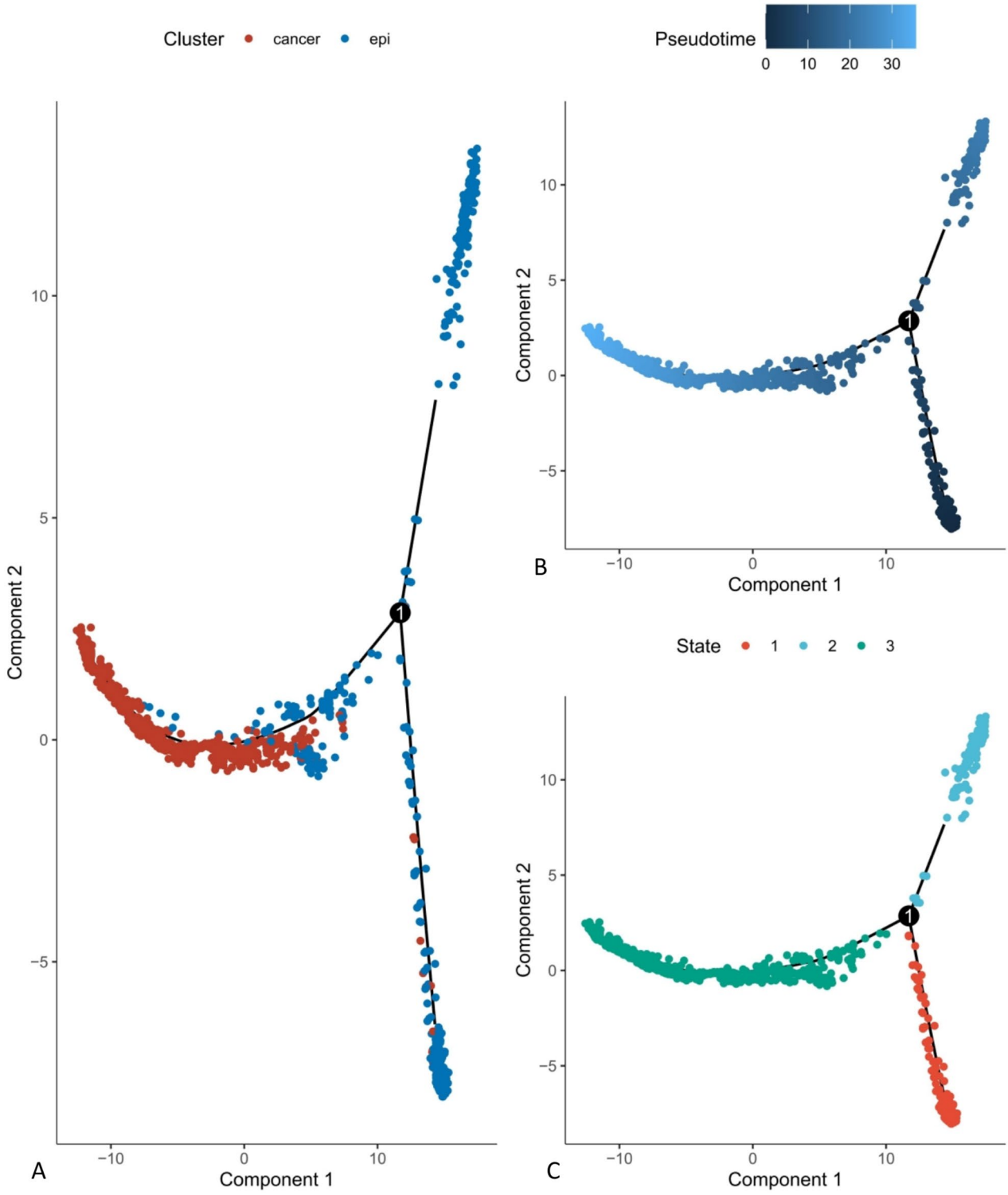

Fig. 29. Single-cell trajectory analysis of LUAD. (**A**) Cell type trajectory plot. (**B**) Pseudotime trajectory plot. (**C**) Cell state trajectory plot.

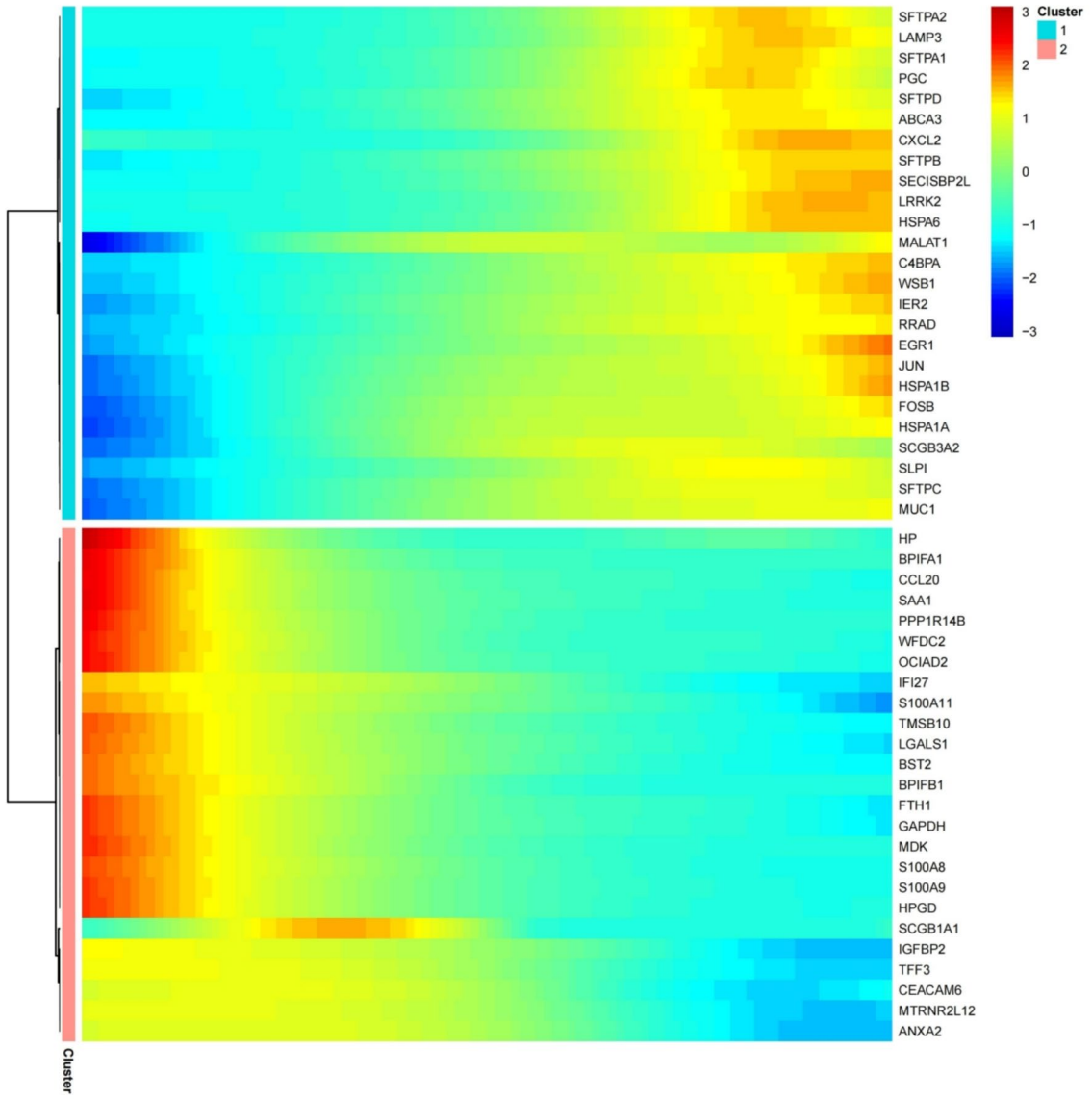

Fig. 30. Heatmap of the top 50 genes expressed in epithelial (epi) and malignant (cancer) cells in LUAD single-cell data.

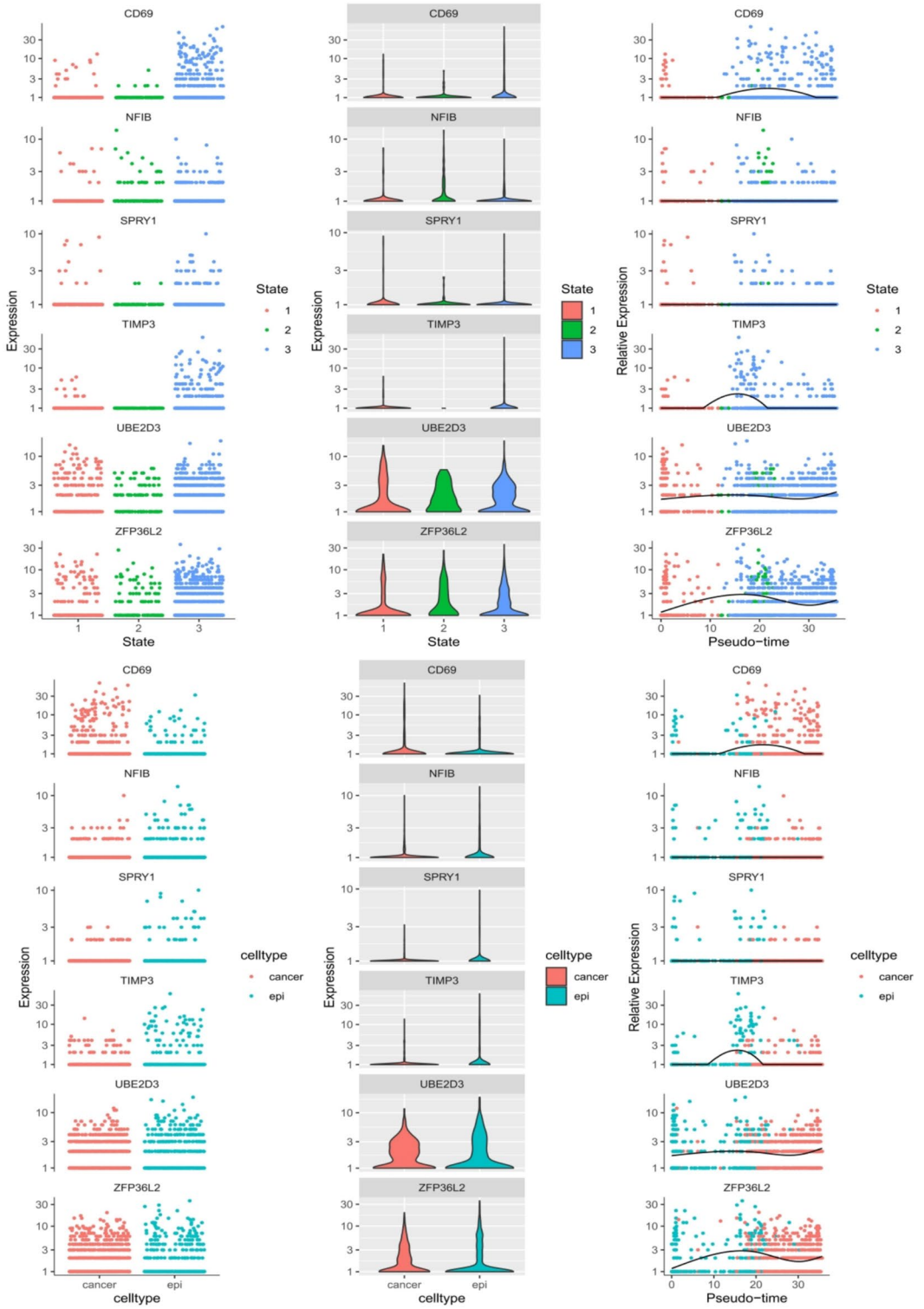

Fig. 31. Expression of miR-21 target genes across different cell states and cell types in LUAD single-cell data.

Data availability

The scRNA-seq data of LUAD were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) with accession number GSE189357. miRNA expression data, RNA-seq expression data, and clinical data were obtained from the TCGA-LUAD (The Cancer Genome Atlas—Lung Adenocarcinoma) database(https://www.cancer.gov/ccg/research/genome-sequencing/tcga).

Received: 11 May 2025; Accepted: 23 September 2025
Published online: 29 October 2025

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**(3), 209–249 (2021).
2. Chen, Z. et al. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat. Rev. Cancer* **14**(8), 535–546 (2014).
3. Hirsch, F. R. et al. The prognostic and predictive role of histology in advanced non-small cell lung cancer: A literature review. *J. Thorac. Oncol.* **3**(12), 1468–1481 (2008).
4. Ettinger, D. S. et al. Non–small cell lung cancer. *J. Natl. Compr. Canc. Netw.* **10**(10), 1236–1271 (2012).
5. American Cancer Society. Lung Cancer Survival Rates 2019. Available from: https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html.
6. Kalinina, T. S. et al. Expression of AhR-regulated miRNAs in non-small cell lung cancer in smokers and never smokers. *Biomed. Khim.* **70**(1), 52–60 (2024).
7. He, M. et al. Systematic analysis to identify the MIR99AHG-has-miR-21-5p-EHD1 CeRNA regulatory network as potential biomarkers in lung cancer. *J. Cancer* **15**(8), 2391 (2024).
8. Zhong, J. et al. miR-21-5p promotes lung adenocarcinoma progression partially through targeting SET/TAF-Iα. *Life Sci.* **231**, 116539 (2019).
9. Wu, S. et al. microRNA-21 aggravates lipopolysaccharide-induced inflammation in MH7A cells through targeting SNF5. *Inflammation* **43**(2), 441–454 (2020).
10. Xu, J. et al. Using machine learning modeling to explore new immune-related prognostic markers in non-small cell lung cancer. *Front. Oncol.* **10**, 550002 (2020).
11. Song, F. et al. Identification of key microRNAs and hub genes in non-small-cell lung cancer using integrative bioinformatics and functional analyses. *J. Cell. Biochem.* **121**(3), 2690–2703 (2020).
12. Wu, Q., Zhang, B. & Sun, Y., et al. Identification of novel biomarkers and candidate small molecule drugs in non-small-cell lung cancer by integrated microarray analysis. *OncoTargets Therapy* 3545–3563 (2019).
13. Wang, Y. et al. Bioinformatics analyses of the role of vascular endothelial growth factor in patients with non-small cell lung cancer. *PLoS ONE* **10**(9), e0139285 (2015).
14. Hsu, S. D. et al. miRTarBase: A database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* **39**(suppl_1), D163–D169 (2011).
15. Gene Ontology Consortium. The gene ontology (GO) project in 2006. *Nucleic Acids Res.* **34**(suppl_1): D322–D326 (2006).
16. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**(1), 27–30 (2000).
17. Szklarczyk, D. et al. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**(D1), D607–D613 (2019).
18. Chen, X., Teichmann, S. A. & Meyer, K. B. From tissues to cell types and back: Single-cell gene expression analysis of tissue architecture. *Annu. Rev. Biomed. Data Sci.* **1**, 29–51 (2018).
19. Olsen, T. K. & Baryawno, N. Introduction to single-cell RNA sequencing. *Curr. Protoc. Mol. Biol.* **122**(1), e57 (2018).
20. Sun, Q. et al. Differential expression and bioinformatics analysis of circRNA in non-small cell lung cancer. *Front. Genet.* **11**, 586814 (2020).
21. Cochain, C. et al. Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis. *Circ. Res.* **122**(12), 1661–1674 (2018).
22. Min, J. W. et al. Identification of distinct tumor subpopulations in lung adenocarcinoma via single-cell RNA-seq. *PLoS ONE* **10**(8), e0135817 (2015).
23. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**(6190), 1396–1401 (2014).
24. Gong, K. et al. Identification and integrate analysis of key biomarkers for diagnosis and prognosis of non-small cell lung cancer based on bioinformatics analysis. *Technol. Cancer Res. Treat.* **20**, 15330338211060202 (2021).
25. Kim, J., Xu, Z. & Marignani, P. A. Single-cell RNA sequencing for the identification of early-stage lung cancer biomarkers from circulating blood. *NPJ Genom. Med.* **6**(1), 87 (2021).
26. Warren, G. W. & Cummings, K. M. Tobacco and lung cancer: risks, trends, and outcomes in patients with cancer. *Am. Soc. Clin. Oncol. Educ. Book* **33**(1), 359–364 (2013).
27. Chan, B. A. & Hughes, B. G. M. Targeted therapy for non-small cell lung cancer: Current standards and the promise of the future. *Transl. Lung Cancer Res.* **4**(1), 36 (2015).
28. Hayes, J., Peruzzi, P. P. & Lawler, S. MicroRNAs in cancer: Biomarkers, functions and therapy. *Trends Mol. Med.* **20**(8), 460–469 (2014).
29. Cho, W. C. S. OncomiRs: The discovery and progress of microRNAs in cancers. *Mol. Cancer* **6**, 1–7 (2007).
30. Chen, W. et al. Identification of key candidate tumor biomarkers in non-small-cell lung cancer by in silico analysis. *Oncol. Lett.* **19**(1), 1008–1016 (2020).
31. Shen, Y., Pan, X. & Yang, J. Gene regulation and prognostic indicators of lung squamous cell carcinoma: TCGA-derived miRNA/mRNA sequencing and DNA methylation data. *J. Cell. Physiol.* **234**(12), 22896–22910 (2019).
32. Tang, L. X. et al. Long non-coding RNA OGFRP1 regulates LYPD3 expression by sponging miR-124-3p and promotes non-small cell lung cancer progression. *Biochem. Biophys. Res. Commun.* **505**(2), 578–585 (2018).
33. Zhang, Y. et al. LncRNA DSCAM-AS1 interacts with YBX1 to promote cancer progression by forming a positive feedback loop that activates FOXA1 transcription network. *Theranostics* **10**(23), 10823 (2020).
34. Qin, R. X. et al. Transcriptomic analysis reveals the potential biological mechanism of AIS and lung adenocarcinoma. *Front. Neurol.* **14**, 1119160 (2023).
35. Välk, K. et al. Gene expression profiles of non-small cell lung cancer: survival prediction and new biomarkers. *Oncology* **79**(3–4), 283–292 (2011).
36. Puzone, R. et al. Glyceraldehyde-3-phosphate dehydrogenase gene over expression correlates with poor prognosis in non small cell lung cancer patients. *Mol. Cancer* **12**, 1–8 (2013).
37. Wang, K., Chen, M. & Wu, W. Analysis of microRNA (miRNA) expression profiles reveals 11 key biomarkers associated with non-small cell lung cancer. *World J. Surg. Oncol.* **15**, 1–10 (2017).

38. Behera, M. et al. Lung adenocarcinoma staging using the 2011 IASLC/ATS/ERS classification: A pooled analysis of adenocarcinoma in situ and minimally invasive adenocarcinoma. *Clin. Lung Cancer* **17**(5), e57–e64 (2016).
39. Zhang, L. et al. Identification of key genes in non-small cell lung cancer by bioinformatics analysis. *PeerJ* **7**, e8215 (2019).
40. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**(10), 1550–1558 (2018).
41. Fu, J. et al. Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med.* **12**, 1–8 (2020).
42. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**(4), 381–386 (2014).
43. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**(10), 979–982 (2017).

Author contributions
QTT:Data curation, Formal analysis, Writing – original draft, Writing – review & editing. ZGJ:Conceptualization, Supervision, Writing – review & editing.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
This study was in strict agreement with international guidelines for the care and use of laboratory and also approved by the Research Ethical Committee of Zhengzhou Central Hospital Affiliated to Zhengzhou University, and informed consent was obtained from all participants. All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication
All the authors have provided consent for the publication.

Additional information
Correspondence and requests for materials should be addressed to G.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.