Open Access Full Text Article

ORIGINAL RESEARCH

MiRNA and Potential Prognostic Value in Non-Smoking Females with Lung Adenocarcinoma by High-Throughput Sequencing

Hao Wang[1], Lijun Wang[2], Gengyun Sun[1]

[1]Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Anhui Medical University, Hefei, People's Republic of China; [2]Department of Respiratory Disease, Tongling People's Hospital, Tongling, People's Republic of China

Correspondence: Gengyun Sun, Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Anhui Medical University, No. 218, Jixi Road, Hefei, Anhui, 230022, People's Republic of China, Email sungengy@126.com

Background: Non-smoking females with lung adenocarcinoma (LUAD) account for a unique disease entity and miRNA play critical roles in cancer development and progression. The purpose of this study is to explore prognosis-related differentially expressed miRNA (DEmiRNA) and establish a prognostic model for non-smoking females with LUAD.

Methods: Eight specimens were collected from thoracic surgery of non-smoking females with LUAD and implemented the miRNA sequencing. The intersection of our miRNA sequencing data and TCGA database were identified as common DEmiRNA. Then, we predicted the target genes of the common DEmiRNAs (DETGs) and explored the functional enrichment and prognosis of DETGs. A risk model by overall survival (OS)-related DEmiRNA was constructed based on multivariate Cox regression analyses.

Results: A total of 34 overlapping DEmiRNA were obtained. The DETGs were enriched in pathways including "Cell cycle" and "miRNAs in cancer". The DETGs (*KPNA2*, *CEP55*, *TRIP13*, *MYBL2*) were risk factors, significantly related to OS, progression-free survival (PFS), and were also hub genes. ScRNA-seq data also validated the expression of the four DETGs. Hsa-mir-200a, hsa-mir-21, and hsa-mir-584 were significantly associated with OS. The prognostic prediction model constructed by the 3 DEmiRNA could effectively predict OS and can be used as an independent prognostic factor of non-smoking females with LUAD.

Conclusion: Hsa-mir-200a, hsa-mir-21, and hsa-mir-584 can serve as potential prognostic predictors in non-smoking females with LUAD. A novel prognostic model based on the three DEmiRNAs was also constructed to predict the survival of non-smoking females with LUAD and showed good performance. The result of our paper can be helpful for treatment and prognosis prediction for non-smoking females with LUAD.

Keywords: women, lung adenocarcinoma, miRNA, prognosis, bioinformatics

Introduction

Lung cancer is an issue of increasing global concern, with the most incidence and mortality worldwide.[1] About 80% of the lung cancers are non-small cell lung cancer (NSCLC) and lung adenocarcinoma (LUAD) is the most common subtype. Smoking is a major risk factor for lung carcinogenesis[2]; however, about a quarter of lung cancer cases are not related to smoking, and it is reported that more than half of female patients never smoke.[3] Furthermore, the number of LUAD in female never smokers is gradually increasing.[4,5] Non-smoking females with LUAD account for a unique disease entity, showing different pathogenesis, epidemiological characteristics and course of disease.[6] Therefore, it is crucial to explore the molecular mechanism that drives LUAD in female never smokers. Exploring prognostic factors and models can be helpful for treatment and prognosis prediction for non-smoking females with LUAD.

miRNAs are a highly conserved class of approximately 19–24 nucleotide-long non-protein coding RNA, which monitor cell proliferation, apoptosis, tumor growth and metastasis by negatively regulating gene expression.[7,8] Aberrant expressions of miRNAs have widely been reported to regulate NSCLC progression,[9] including cell growth, stemness, drug resistance, and immune evasion.[10] miRNAs could also act as biomarkers for NSCLC diagnosis, prediction of
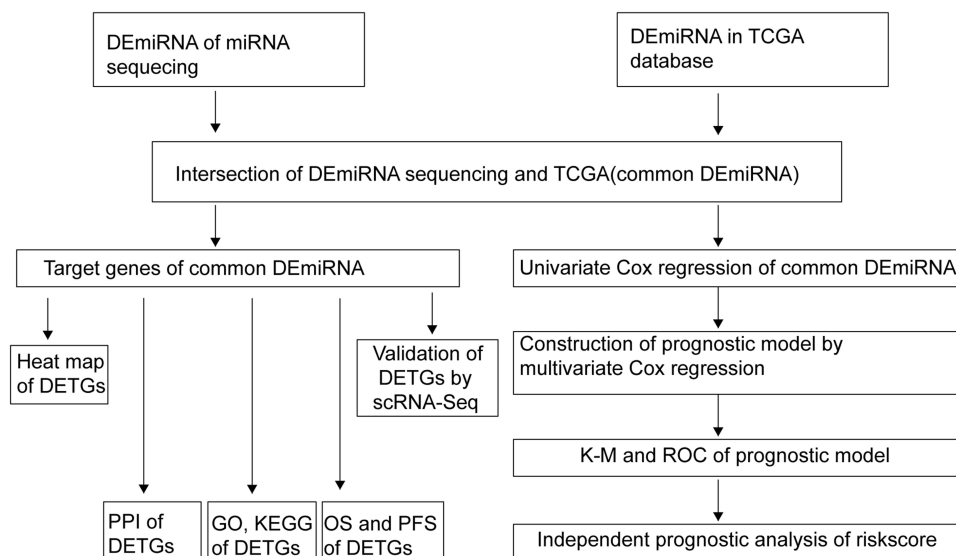
Figure 1 Flow chart of the analysis process in our study.
Abbreviations: DEmiRNA, differentially expressed miRNA; DETGs, differentially expressed target genes; PPI, protein–protein interaction; OS, overall survival; PFS, progression-free survival; ROC, receiver operating characteristic curves; DETGs, target genes of the common differentially expressed miRNA.

clinical outcomes.[11] Risk score model constructed by miRNAs can well predict lung cancer prognosis.[12] However, miRNAs in non-smoking females with LUAD have not been well explored.

In the present study, we investigated the differentially expressed miRNAs (DEmiRNAs) and explored the GO, KEGG, OS and PFS of their target genes. Additionally, we evaluated the prognostic value of the DEmiRNAs, and constructed three miRNA (hsa-mir-200a, hsa-mir-21 and hsa-mir-584) based risk model that could effectively predict overall survival (Figure 1).

Materials and Methods

Patients and Sample Collection

All tissue samples were collected from non-smoking females with LUAD who were treated by thoracic surgery in Anhui Provincial Cancer Hospital from June 2020 to June 2021. Patients treated with preoperative targeted therapy, chemotherapy, immunotherapy, and radiotherapy were excluded. The pathological type of lung adenocarcinoma was confirmed by pathologists. Tumor specimens and matched normal tissues were quickly frozen at −80 °C for further high-throughput sequencing. The present study was approved by The Human Investigation and Ethical Committees of Anhui Provincial Cancer Hospital (Ethics Number-41/2020) and performed according to the Declaration of Helsinki. All participants provided written informed consent.

Small RNA Sequencing and DEmiRNA Analysis

Total RNA was extracted from fresh tissues using TRIzol according to the manufacturer's instructions (Thermo Fisher Scientific, Inc.). The purity and concentration of RNA were examined using NanoDrop 2000. The integrity and quantity of RNA were measured using the Agilent 2100/4200 system. Three micrograms of total RNA per sample was used to generate the small RNA library. Small RNA sequencing was performed using Illumina NextSeq500sequencing (BG, Beijing, China). Low quality reads were removed. Adaptor sequences and small RNAs which perfectly match in the human genome database (ftp://ncbi.nlm.nih.gov/genomes/Homo_sapiens) were considered valid for subsequent analysis. The "edgeR" R package was utilized to find the differentially expressed miRNA (DEmiRNA) with the threshold set at absolute $\log_2$Fold Change >1 and P value <0.05.

Validation of DEmiRNA by TCGA

The miRNA expression data, mRNA expression data, single nucleotide polymorphism (SNP) data and corresponding clinical features of non-smoking females with lung adenocarcinoma were also downloaded from The Cancer Genome Atlas (TCGA). The DEmiRNA between tumor samples and normal tissues were determined using "edgeR" R package. The $\log_2$ Fold Change and P value cutoff were set as 1 and 0.05. We applied "VennDiagram" R package to identify the common DEmiRNA in our miRNA-seq data and TCGA miRNA data. The common DEmiRNA was used for the following analysis. Tumor mutation burden (TMB) is a measure of the total number of somatic non-synonymous mutations per base pair. SNP data from TCGA was utilized to calculate TMB.

Common DEmiRNA and Their Target Genes

The common DEmiRNAs were predicted their target mRNA using StarBase (http://starbase.sysu.edu.cn/), miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/), miRDB (http://www.mirdb.org/), and TargetScan (http://www.targetscan.org/) data-sets. The target genes were submitted to STRING v11.0 (https://www.string-db.org/) database, for the exploration of protein–protein interaction (PPI) network. Cytoscape software (version 3.8.0) was used to perform visualization of PPI network. Cytohubba, a plug-in of Cytoscape, was applied to find the key genes in the PPI network. Target genes were also inputted into DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/) online database, to present Gene Ontology (GO) analysis.

Prognosis-Related Common DEmiRNA and Target Genes

Target genes' expression level and corresponding clinical characteristics were obtained from TCGA dataset. We employed univariate Cox proportional-hazards regression analyses to evaluate the progression-free survival (PFS) and overall survival (OS) of common DEmiRNA and target genes. P value less than 0.05 was considered statistically significant.
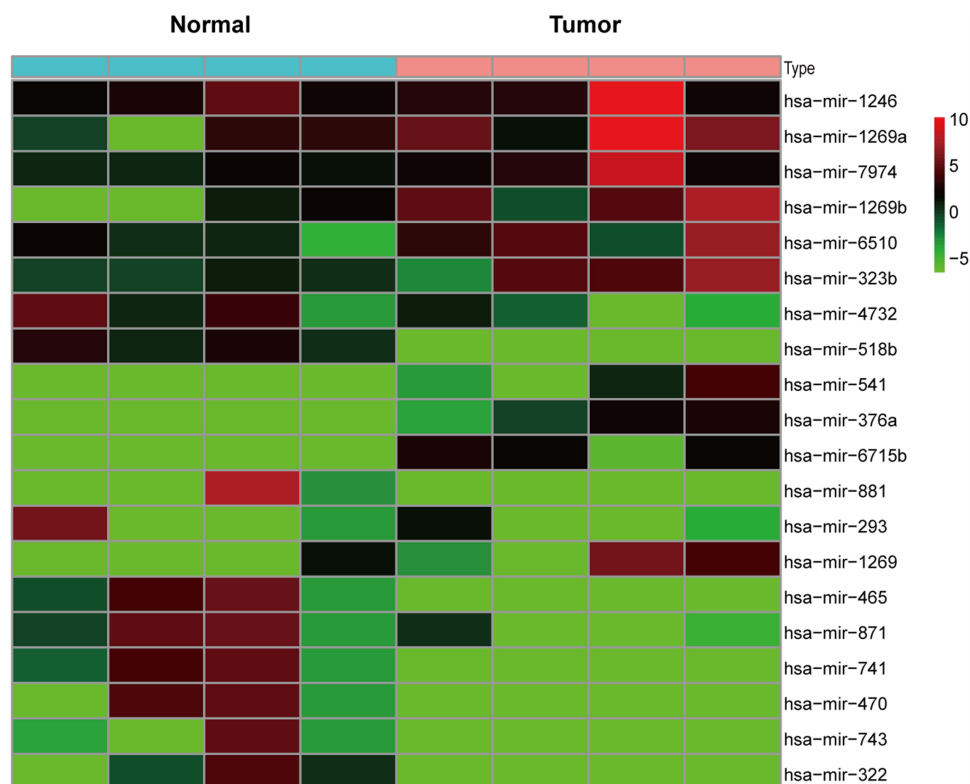

Figure 2 The top 20 differentially expressed miRNA between tumor and corresponding normal tissues in sequencing data.

The Expression of DETGs Was Validated by scRNA-Seq

The scRNA-seq data of 19 non-smoking women with lung adenocarcinoma patients (13 tumor tissues and 6 normal tissues) were also downloaded from the GEO database (GEO 131907). The package of Seurat was applied to process the scRNA-seq data. The data of the samples were batch corrected by the package of harmony. The FindVariableFeatures was used to find the top 2000 highly variant genes and performed principal component analysis. The first 30 principal components and resolution 0.6 were used to find clusters function to generate clusters. The FindAllMarkers were used to identify marker genes in each cluster.

Construction of the Risk Score by DEmiRNA

The univariate Cox proportional hazards regression analyses were employed to evaluate the association between OS and DEmiRNA. DEmiRNA with a p-value less than 0.05 was selected for construction the risk prognostic model. Then, we employed multivariate Cox proportional hazards regression analysis to produce coefficients which were acted as weights in the prognostic model. The risk scores equal to the sum of the product of each DEmiRNA and its weight. Based on the prognostic model, we divided patients into high-risk or low-risk groups. In order to test the robustness of the risk prognostic model, we randomly divided all samples of non-smoking females with LUAD into train-dataset and test-dataset. We conducted "survival" package of R to draw Kaplan–Meier (KM) curve, which was used to compare the OS between the high-risk group and low-risk group. We also applied "timeROC" package of R to draw the receiver operating characteristic (ROC) curves, which were used to validate the accuracy of the risk model in predicting the survival prognosis of the non-smoking females with lung adenocarcinoma.

Independent Prognostic Analysis

We conducted univariate and multivariate Cox proportional hazard regression analyses to assess whether the prognostic model we constructed, independent of the clinical characteristics (age, primary tumor, regional lymph node, stage, tumor mutation burden) in predicting OS of non-smoking females with lung adenocarcinoma.
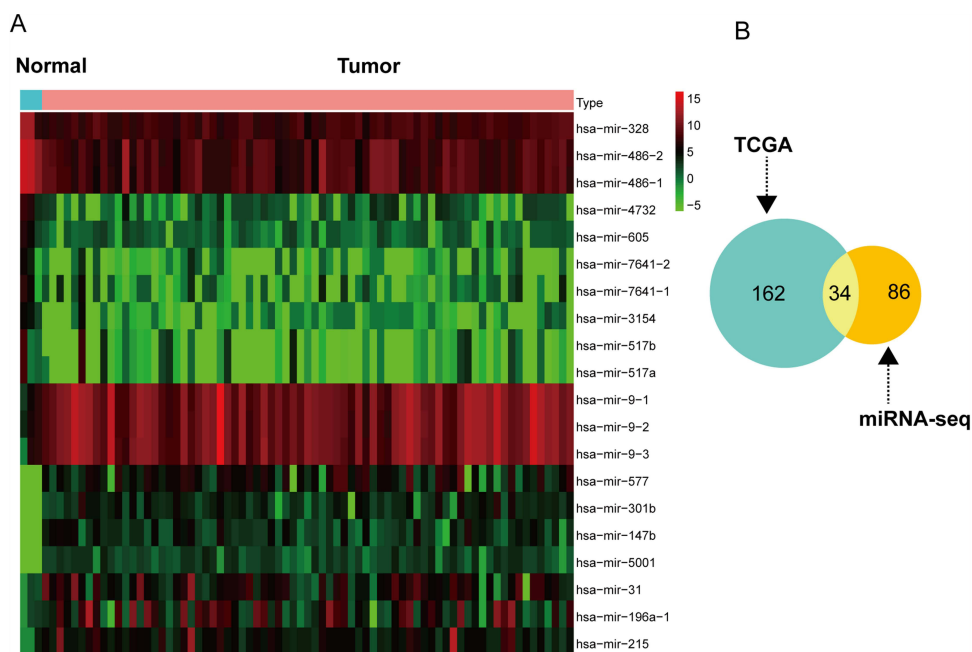

Figure 3 The top 20 differentially expressed miRNA in TCGA and common DEmiRNA. (**A**) The top 20 differentially expressed miRNA (DEmiRNA) in TCGA. (**B**) Venn diagram shown the Common DEmiRNA.

Statistical Analysis

All statistical analyses were performed by statistical package R (version 4.1.0, https://www.r-project.org/). Fisher's exact tests were used to evaluate the target genes in Gene Ontology. Univariate and multivariate Cox regression analyses were implied to assess whether the fact was related to prognosis. P value less than 0.05 was considered statistically significant.

Results

Differentially Expressed miRNA in Sequencing Data

We collect eight samples from thoracic surgery specimens (Supplementary Data 1). After completing sequencing and data quality control, we applied "edgeR" R package to identify differentially expressed miRNA between tumor and normal tissues in non-smoking females with LUAD. Finally, 120 differentially expressed miRNAs (DEmiRNA) were obtained, in which 66 DEmiRNAs upregulated, and 54 DEmiRNAs downregulated. The top 10 up- and downregulated DEmiRNAs were visualized by heat map, as shown in Figure 2.

Identification of DEmiRNA in TCGA and Common DEmiRNA

Through screening of TCGA clinical data, we finally obtained 73 non-smoking females with lung adenocarcinoma and 3 normal samples. For the purpose of obtaining DEmiRNA, we conducted "edgeR" R package. At last, we got 196 DEmiRNA (70 upregulated, 126 downregulated) in TCGA data. We performed the top 10 up and down regulated DEmiRNAs by heat map, as shown in Figure 3A. In order to increase the credibility of DEmiRNA, we used the Venn diagram to select the intersection part of DEmiRNA in TCGA and our miRNA sequencing data, a total of 34 overlapping DEmiNRAs were obtained, as shown in Figure 3B.
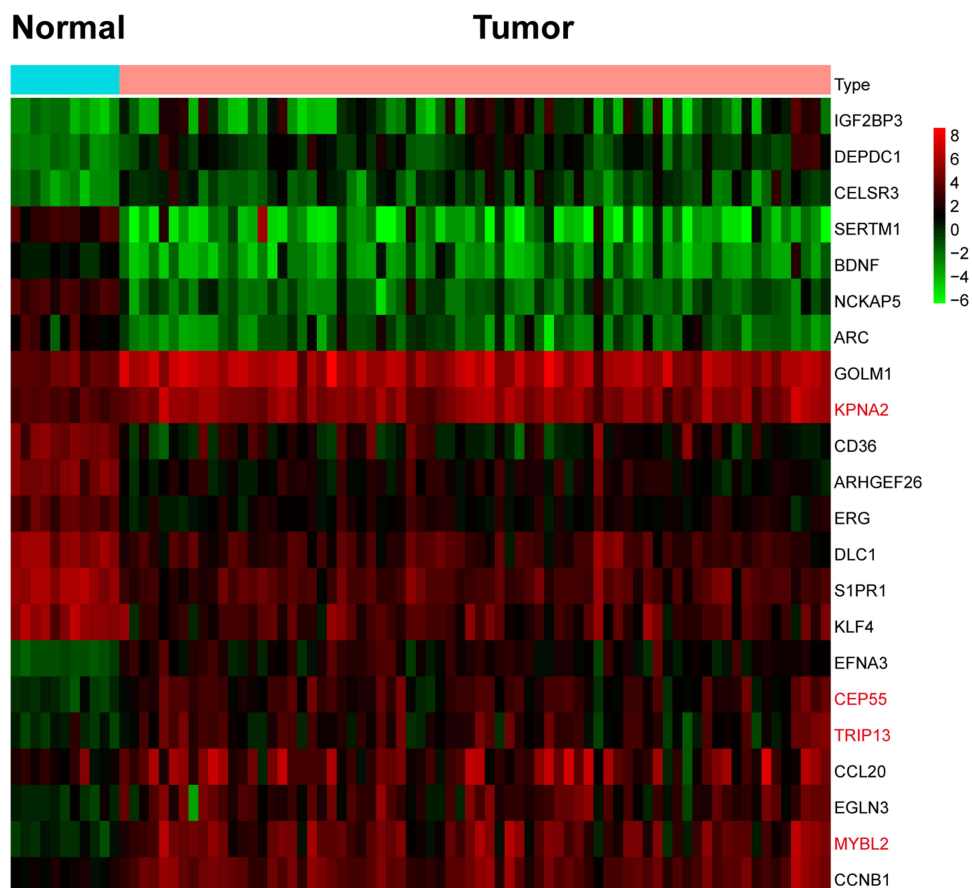

Figure 4 The heat map of the top 20 DETGs and 4 hub genes. The four hub genes (*KPNA2*, *CEP55*, *TRIP13*, *MYBL2*) were marked in red. *CEP55* and *MYBL2* are not only in the top 20 genes but also hub genes.

PPI Network and Functional Enrichment of Differentially Expressed Target Genes (DETGs)

By predicting the target genes of 34 common miRNAs, 863 target genes were obtained. We used the "limma" package of R to explore the differentially expressed target genes (DETGs). Then, 123 DETGs (60 upregulated, 63 downregulated) in TCGA data set were acquired. This study performed the top 10 up and down regulated DETGs by heat map, as shown in Figure 4. We also explored the protein–protein interaction (PPI) network of DETGs. The PPI network was visualized by Cytoscape. There were 141 edges, 80 nodes, and 14 hub genes in the PPI network, as shown in Figure 5. In order to explore the influence of DEmiRNA on the pathogenesis in non-smoking females with lung adenocarcinoma, GO analysis on DETGs were performed. The 123 DETGs were grouped into BP (biological process), MF (molecular function), and CC (cellular component) categories. The most enriched GO terms in the biological process(BP) category was "signal transduction", "positive regulation of RNA polymerase II promoter". The most enriched GO terms in the cellular component (CC) category were "cytoplasm", 'nucleus'. The most enriched GO terms in the molecular function (MF) category were "protein binding" and "transcription factor activity, sequence −specific DNA binding" (Figure 6).

DETGs and Common DEmiRNA Associated with Prognosis

After merging overall survival (OS), progression-free survival (PFS) and the expression of differentially expressed target genes (DETGs), we found 16 DETGs were associated with overall survival, using univariate Cox regression. AIFM3,
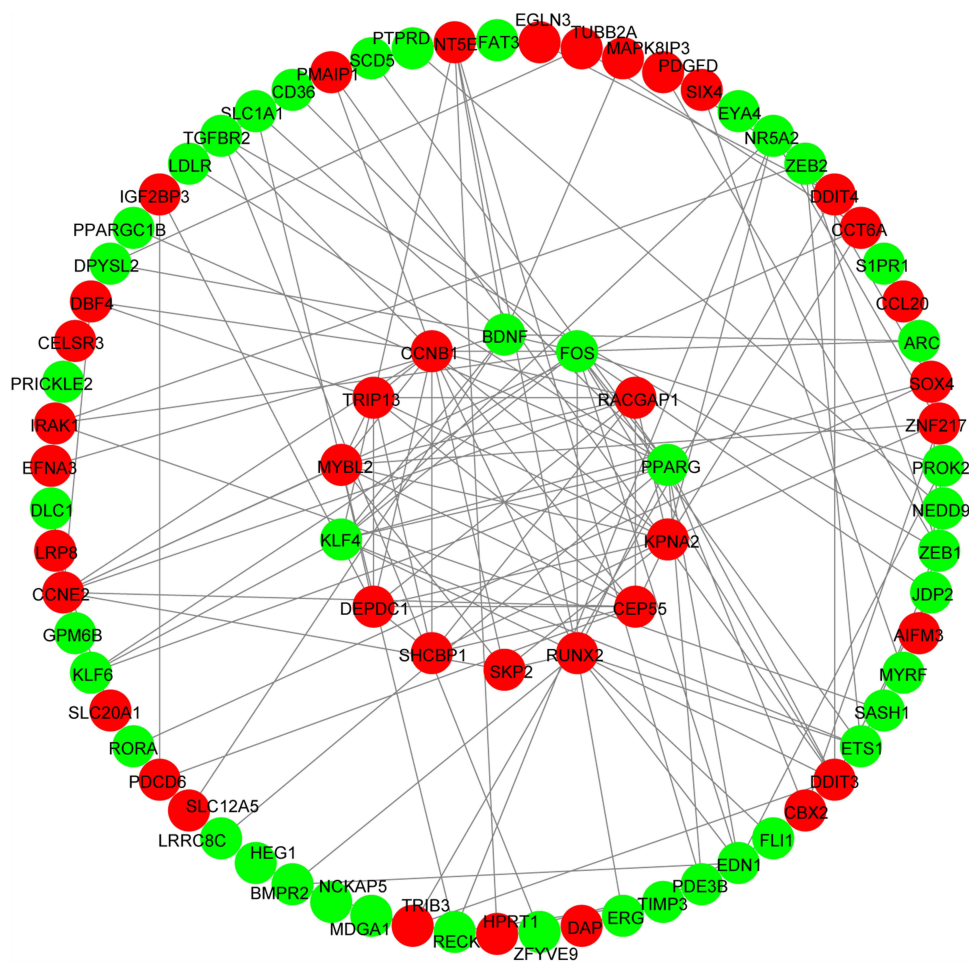

Figure 5 The protein–protein interaction network of DETGs. The small circle represents hub-genes. The red color represents up-regulated genes. The green color represents down-regulated genes.
Abbreviation: DETGs, target genes of the common differentially expressed miRNA.

DDIT3, LIMCH1, METTL7A were protective factors, and the other 12 DETGs were risk factors, as shown in Figure 7A. Meanwhile, 13 DETGs were associated with progression-free survival. CD36, METTL7A were protective factors, and the other 11 DETGs were risk factors, as shown in Figure 7B. *KPNA2*, *CEP55*, *TRIP13*, *MYBL2*, which were risk factors, were significantly related with OS, PFS and were also hub genes.

The data of overall survival and the expression of common differentially expressed miRNA (DEmiRNA) were merged. Three common DEmiRNA (hsa-mir-200a, hsa-mir-21, hsa-mir-584) were significantly associated with overall survival. Hsa-mir-200a, hsa-mir-21 were protective factors, and Hsa-mir-584 was risk factor, as shown in Figure 8.
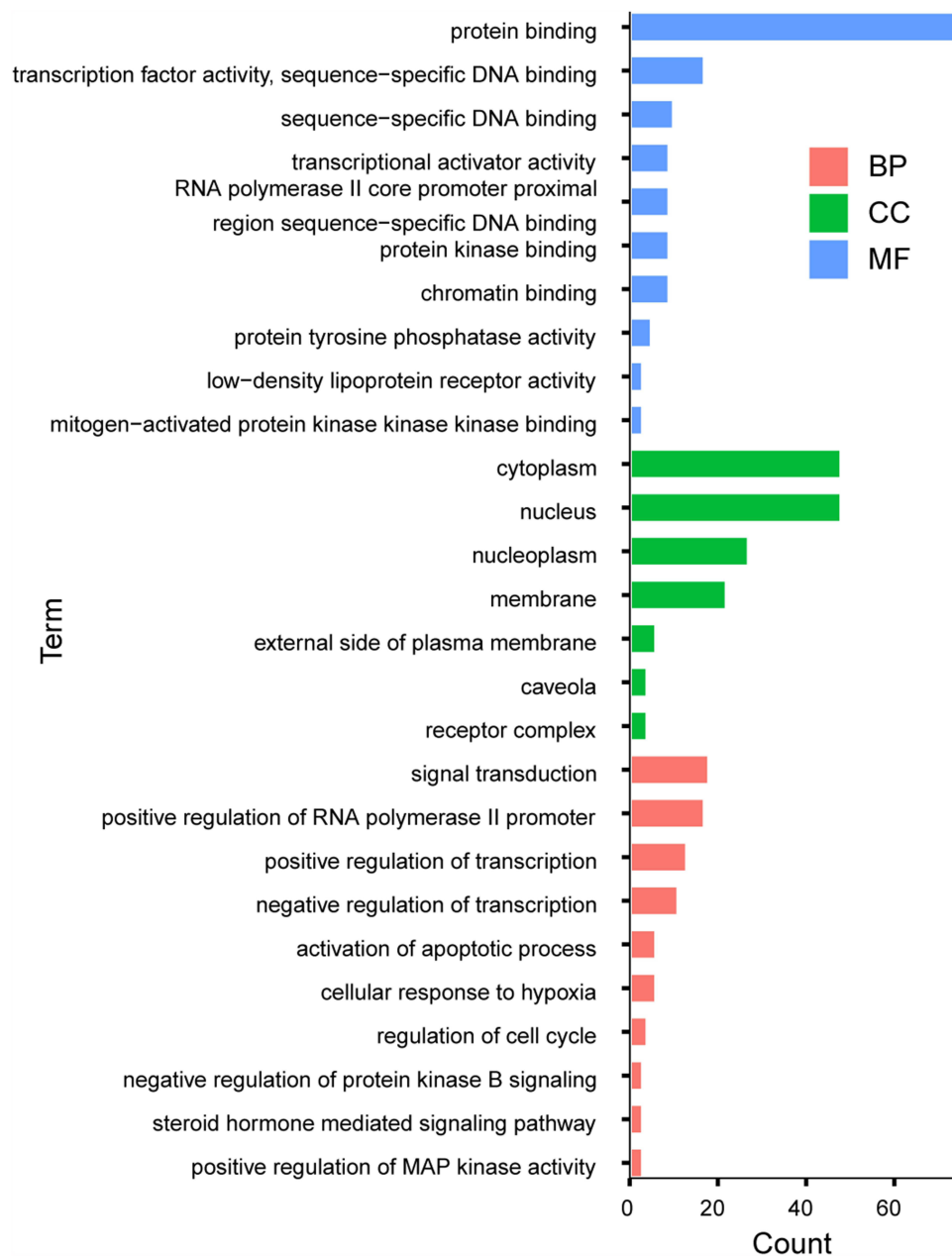

Figure 6 GO analysis on DETGs. The most enriched GO terms in the biological process (BP) category were "signal transduction", "positive regulation of RNA polymerase II promoter". The most enriched GO terms in the cellular component (CC) category were "cytoplasm", "nucleus". The most enriched GO terms in the molecular function (MF) category were "protein binding" and "transcription factor activity, sequence–specific DNA binding".
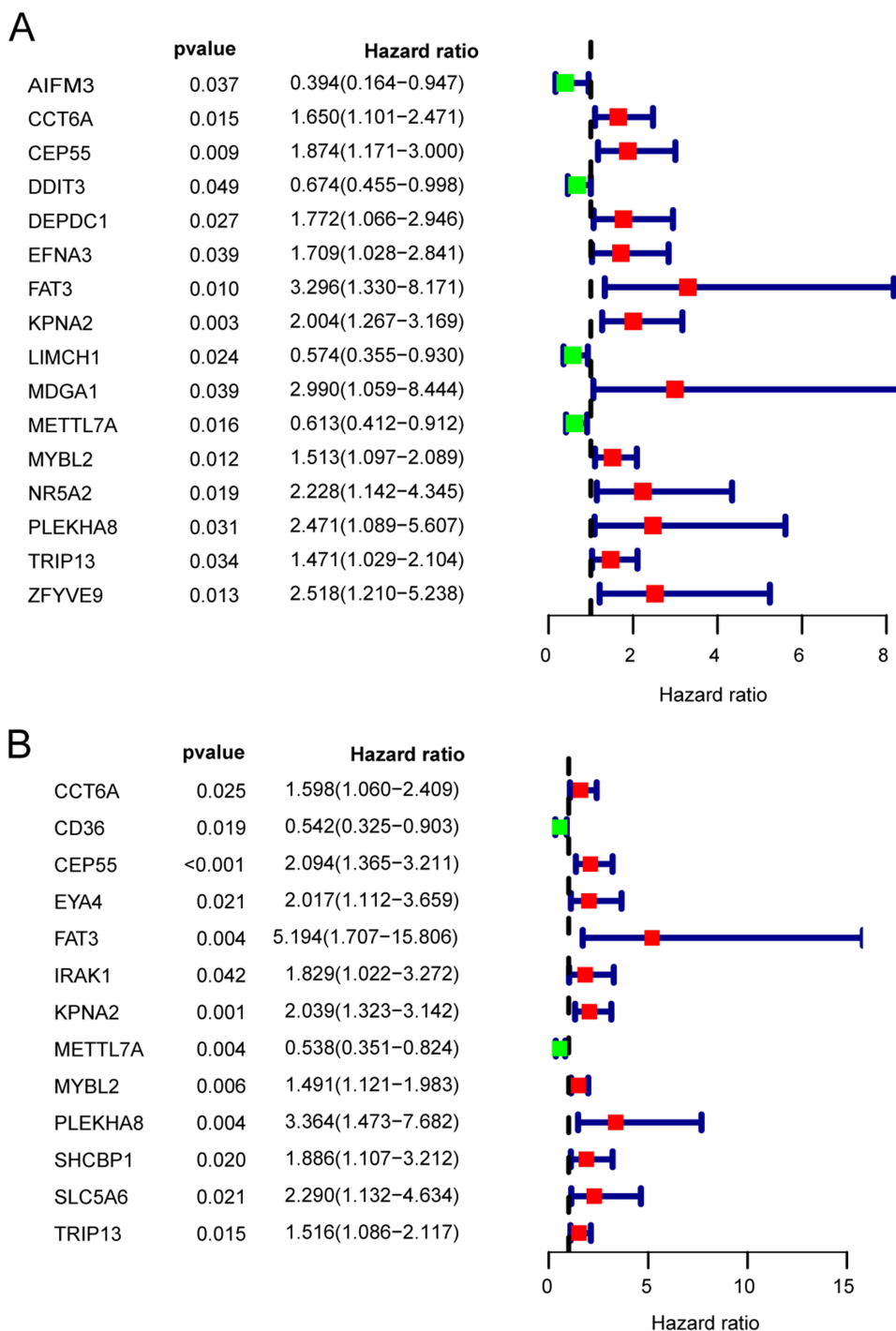**Abbreviation**: DETGs, target genes of the common differentially expressed miRNA.

Figure 7 Prognosis-related DETGs. Univariate Cox proportional-hazards regression analyses were used to evaluate the progression-free survival (**A**) and overall survival (**B**) of differentially expressed target genes (DETGs).
Abbreviation: DETGs, target genes of the common differentially expressed miRNA.

Prognosis-Related DETGs Validated by scRNA-Seq

To elucidate the DETGs expression in tumor Epithelial cells and normal Epithelial cells, we analyzed the GSE 131907. We cataloged the cells into nine distinct cell lineages annotated with canonical marker gene expression: Epithelial cells (EPCAM, KRT19, KRT18, CDH1), Fibroblast (DCN, THY1, COL1A1, COL1A2), Endothelial cells (PECAM1, CLDN5, FLT1, RAMP2), T lymphocyte (CD3D, CD3E, CD3G, TRAC), NK cells (NKG7, GNLY, NCAM1, KLRD1),
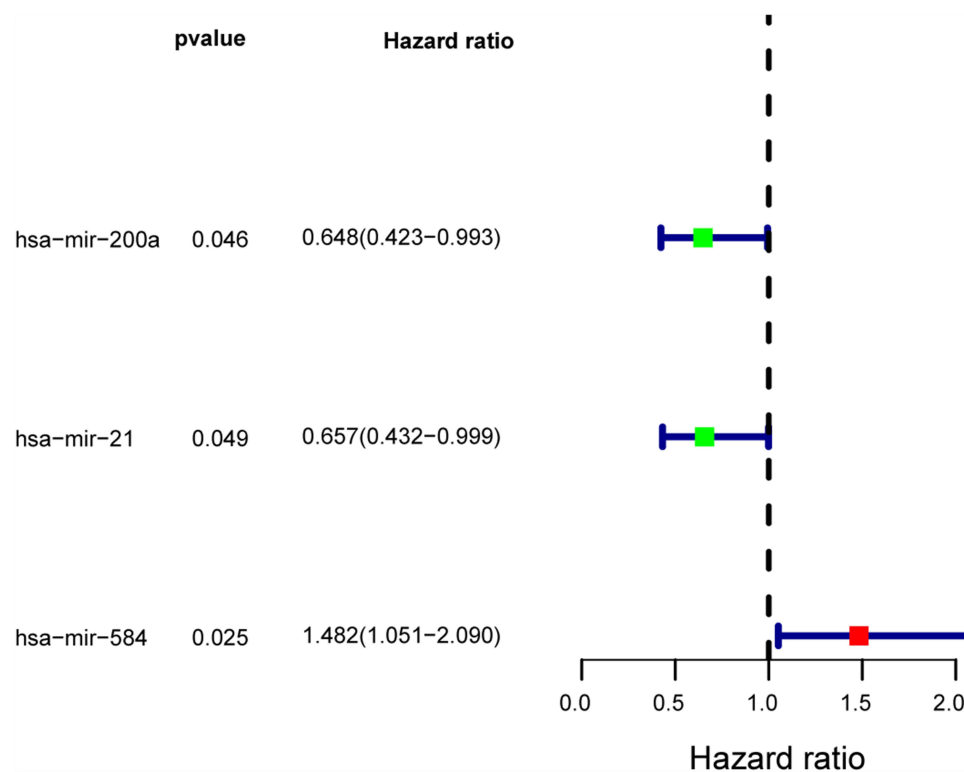
Figure 8 Prognosis-related DEmiRNA. Hsa-mir-200a, hsa-mir-21 were protective factors, and Hsa-mir-584 was risk factor.
Abbreviation: DETGs, target genes of the common differentially expressed miRNA.

B cells (CG79A, IGHM, IGHG3, IGHA2), Myeloid (LYZ, MARCO, CD68, FCGR3A), MAST (KIT, MS4A2, GATA2), and Oligodendrocytes (OLIG1, OLIG2, MOG, CLDN11), as shown in Figure 9A and B. Epithelial cells in tumor tissues were defined as tumor-epithelial cells, and Epithelial cells in normal tissues were defined as Normal-epithelial cells (Figure 9C). Consistent with our previous study, KPNA2, CEP55, and MYBL2 were up-regulated in tumor epithelial cells, while TRIP13 was up-regulated in normal epithelial cells (Figure 9D).

Construction of a Prognostic Model by DEmiRNA

The prognostic prediction model including 3 DEmiRNA. The risk Score = (0.325544* hsa-mir-584 expression level) – (0.46053* hsa-mir-21 expression level) –(0.34102* hsa-mir-200a expression level). The risk score of each sample was calculated according to the constructed model. These samples were divided into high- and low-risk groups using the median value as the cut-off point. Then, Kaplan–Meier survival curve analysis was used to examine the significant differences in survival time between the two sample groups. Figure 10A shows that patients in the high-risk group had significantly lower overall survival than that in the low-risk group. The predictive performance of the risk score for OS was evaluated by time-dependent receiver operating characteristic curves (ROC), and the area under the curve (AUC) reached 0.714 at 1 year, 0.816 at 3 years, and 0.635 at 5 years (Figure 10B). Both in train-dataset (Figure 10C) and test-dataset (Figure 10D), patients in the lower-risk group had significantly longer overall survival than that in the higher-risk group. The univariate and multivariate Cox regression analyses were performed to explore whether risk score is an independent factor from traditional clinical parameters. As shown in univariate Cox regression analysis (Figure 11A), stage and risk score were independent factors, but in multivariate Cox regression analysis (Figure 11B), only risk score was independent factors.

Discussion

In this study, through bioinformatics analysis of eight samples by high-throughput sequencing and TCGA database, We found The DETGs (*KPNA2*, *CEP55*, *TRIP13*, *MYBL2*) were risk factors, at the same time, were significantly related to
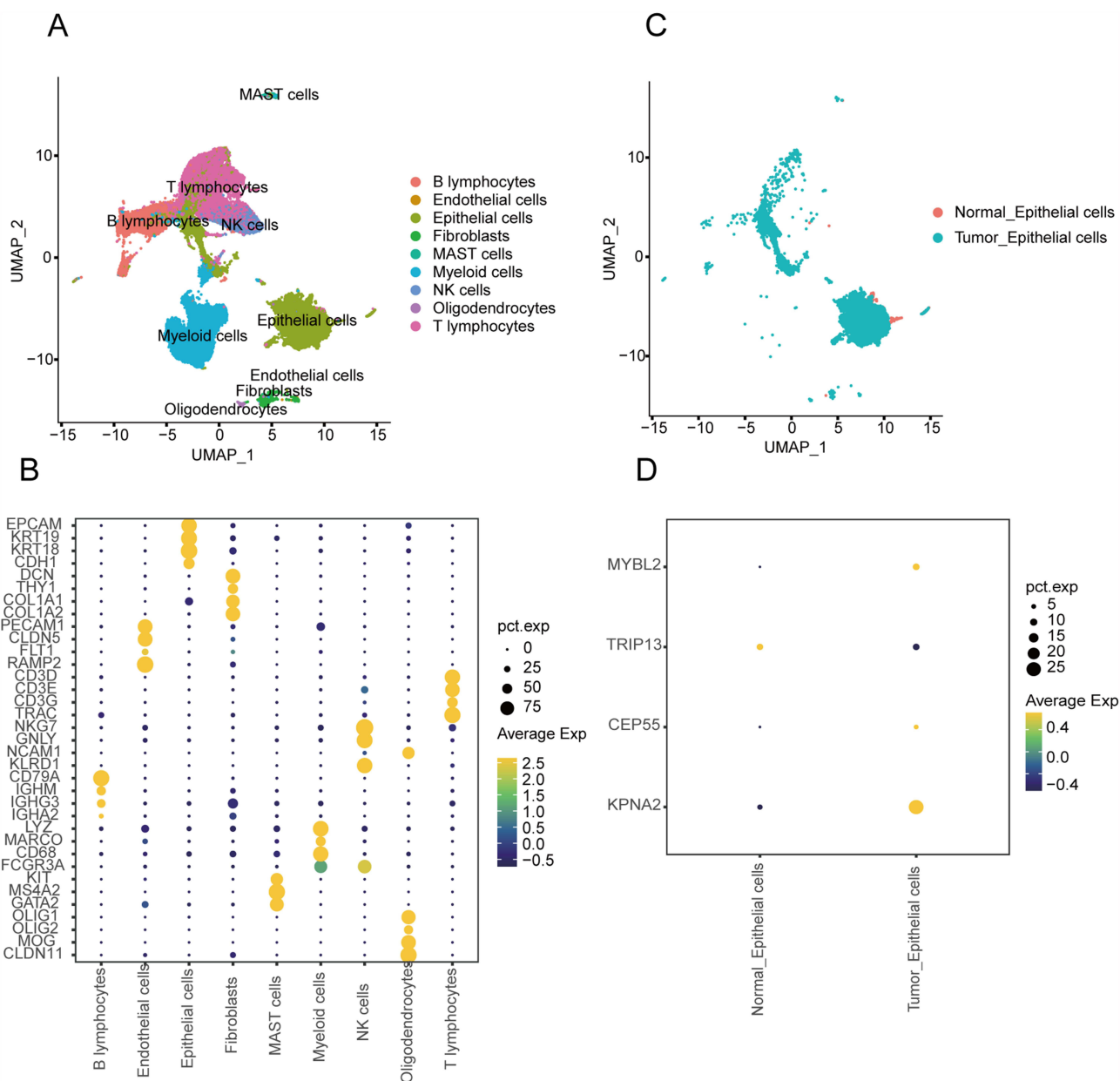
Figure 9 Validation of the expression level of prognosis-related DETGs. (**A**) UMAP plot of nine distinct cell lineages. (**B**) Dot plot showing the expression of representative genes in major cell types. (**C**) TSNE plot of tumor-epithelial cells and normal-epithelial cells in GSE 131907. (**D**) The expression level of prognosis-related DETGs in GSE 131907.

Abbreviation: DETGs, target genes of the common differentially expressed miRNA.

OS, PFS and were also hub genes in non-smoking females with LUAD. Hsa-mir-200a, hsa-mir-21 and hsa-mir-584 were significantly associated with overall survival. We constructed a prognostic prediction model including the 3 DEmiRNA. Both univariate and multivariate Cox regression analysis showed the risk score of the prognostic prediction model were independent factors in non-smoking females with LUAD.

KPNA2 is a nuclear transport protein, could regulate tumor cell proliferation, differentiation, migration, immune response and DNA transcription.[13,14] The high expression of KPNA2 was significantly related to prognosis in lung adenocarcinoma, brain lower grade glioma, skin cutaneous melanoma and mesothelioma.[15,16] In this paper, we added that KPNA2 is highly expressed in non-smoking females with LUAD and was negatively correlated with prognosis. At the same time, the common DEmiRNA (hsa-mir-495) could regulate the expression of KPNA2. So we have reason to believe
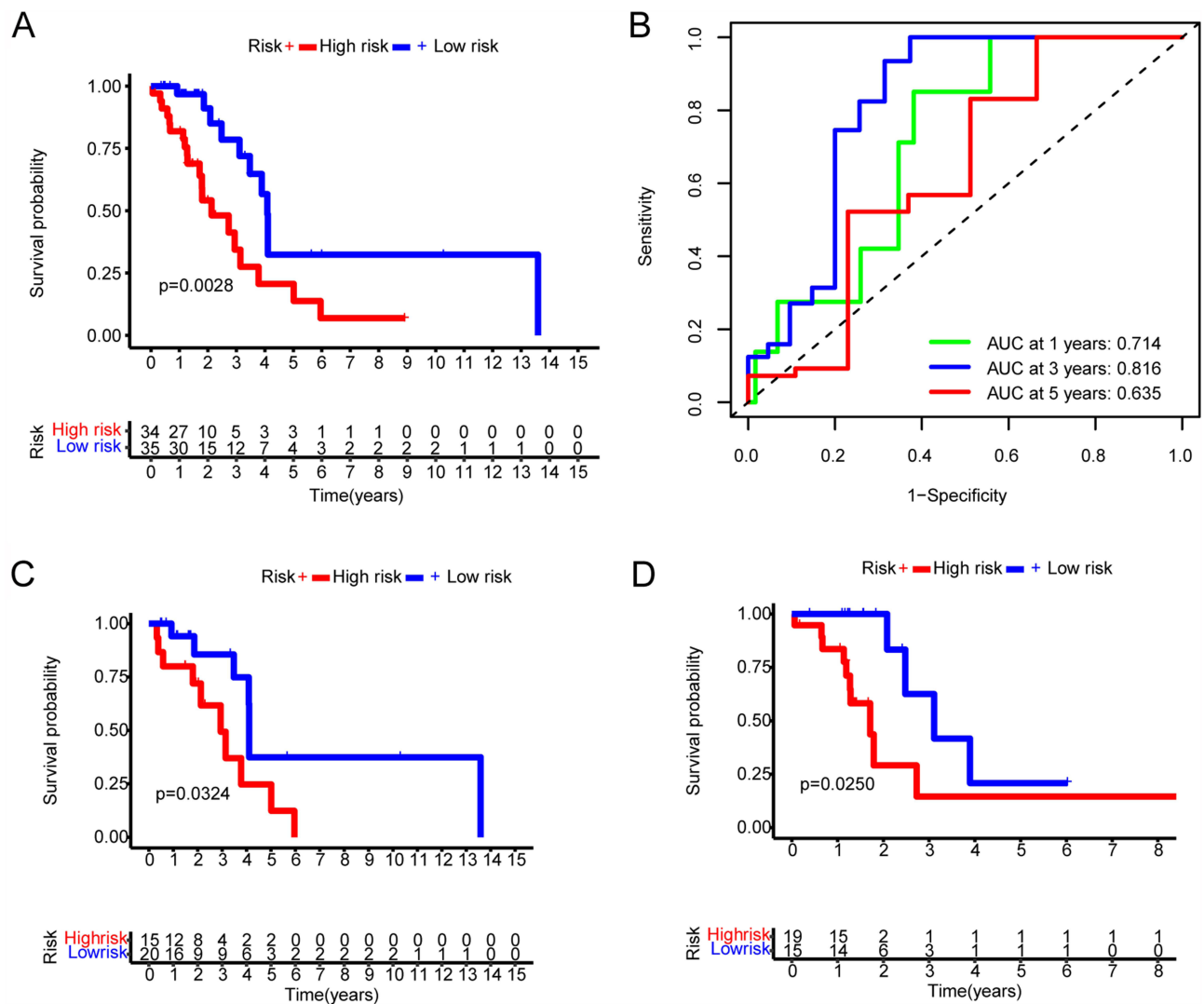
Figure 10 Prognostic risk model in non-smoking females with lung adenocarcinoma. (**A**) Kaplan–Meier survival curve display the overall survival of non-smoking females with lung adenocarcinoma according to the median cutoff value of risk score in all-dataset. (**B**) Time-dependent receiver operating characteristic curves (ROC) assess the prognosis of the prognostic model at 1, 3, and 5 years. (**C**) Kaplan–Meier survival curve display the overall survival of non-smoking females with lung adenocarcinoma according to the median cutoff value of risk score in train-dataset. (**D**) Kaplan–Meier survival curve display the overall survival of non-smoking females with lung adenocarcinoma according to the median cutoff value of risk score in test-dataset.

that has-mir-495 could negatively regulate the expression of KPNA2, and then, regulate the biological function in non-smoking females with LUAD. CEP55, a key regulator of cell division, has been found to play important roles in various types of cancer cells.[17] The high expression of CEP55 predicted favorable prognosis in colorectal cancer,[18] while in non-small cell lung cancer and renal cell carcinoma, the low expression of CEP55 was related to better overall survival.[19,20] Our research showed that the high expression of CEP55 was predicted poor prognosis, consistent with the result of non-small cell lung cancer and renal cell carcinoma.

Both of thyroid hormone receptor interacting protein 13 (TRIP13) and MYB protooncogene-like 2 (MYBL2) are proto-oncogenes.[21,22] TRIP13 could sponge miRNA615 and miRNA495 to promote the progression of prostate cancer.[23,24] MYBL2 could sponge miRNA30a to promote the prostate cancer cell growth.[25] In lung adenocarcinoma, miRNA30c targeted the MYBL2 and regulated the cell cycle signal pathway.[26] Our study indicated that both TRIP13 and MYBL2 were negatively related to prognosis of non-smoking females with lung adenocarcinoma. The KEGG of DETGs enrichment in cell cycles, showing that the common DEmiRNAs could regulate female lung adenocarcinoma cells by monitoring cell signal pathways. However, more basic experiments are needed to be confirmed.
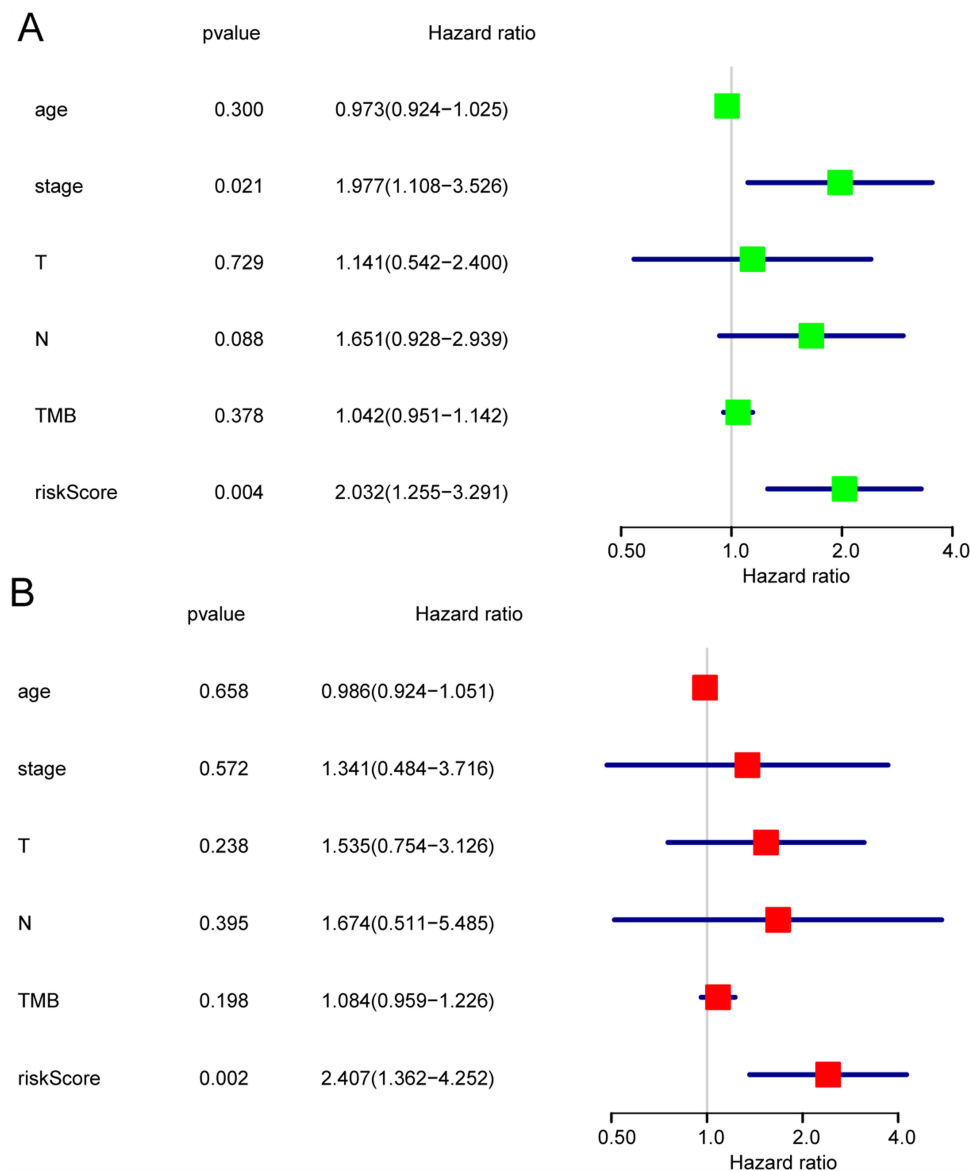
Figure 11 Independent prognostic of risk score. The univariate Cox regression (**A**) and multivariate Cox regression (**B**) analysis of the associations between the risk scores and clinical parameters and the overall survival (OS) of non-smoking females with lung adenocarcinoma.

miRNAs are a group of endogenous conservative small non-coding RNAs (ncRNAs), and they could negatively regulate the expression of genes.[27] In the research of Huang,[28] miR-200a was low expressed in NSCLC, and over-expression of miR-200a could inhibit the cell proliferation and promoted apoptosis of NSCLC. However, in the study of Wang et al,[29] they found that lncRNA-ATB promotes the apoptosis of NSCLC cells by inhibiting the expression of miR-200a. Therefore, more experiments need to verify the role of miR-200a in non-small cell lung cancer. Our result showed that miR-200a is a protector effector in females with lung adenocarcinoma, consistent with Huang et al. In thyroid carcinoma, glioma and gastric cancer, miR-584 acted as a suppress factor.[30,31] However, the function of miR-584 in females with lung adenocarcinoma still unclear. In lung adenocarcinoma, Siriwardhana et al[32] have proven that miR-584 showed a worsening of survival, consisting with our paper.

Many studies have constructed miRNA base risk score model as a prognosis signature in lung cancer[12] and in NSCLC.[33] However, using miRNA as a risk model is still blank in non-smoking females with lung adenocarcinoma. We constructed the risk model by three miRNAs with good predictive power, and it can be used as an independent prognostic factor for non-smoking females with lung adenocarcinoma.

Limitations

There are several limitations in this study. Only eight samples were collected for high-throughput sequencing, and a large number of samples are needed for validating our findings. Healthy females and smoker females with LUAD were not included in the control group. The risk model based on miRNA is needed for further validation. The molecular mechanism of these miRNAs was needed further exploration.

Conclusions

In conclusion, we identified common DEmiRNA both in our RNA-seq data and TCGA dataset and explored the prognostic value and function enrichment of target genes of the common DEmiRNA. Moreover, we constructed three miRNA (hsa-mir-200a, hsa-mir-21 and hsa-mir-584) based risk model that could effectively predict overall survival and it can be used as an independent prognostic factor for non-smoking females with lung adenocarcinoma.

Abbreviations

LUAD, lung adenocarcinoma; DEmiRNA, differentially expressed miRNA; DETGs, target genes of the Common DEmiRNAs; OS, overall survival; NSCLC, non-small cell lung cancer; TCGA, The Cancer Genome Atlas; PPI, protein–protein interaction; ROC: receiver operating characteristic curves.

Data Sharing Statement

The miRNA-seq data and clinicopathological data used in this study can be acquired free of charge from the corresponding authors by email, and the rest of the data can be retrieved from TCGA (https://portal.gdc.cancer.gov/), and GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131907).

Acknowledgments

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors declare there are no competing interests.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249. doi:10.3322/caac.21660
2. Behrend SJ, Giotopoulou GA, Spella M, Stathopoulos GT. A role for club cells in smoking-associated lung adenocarcinoma. *Eur Respir Rev*. 2021;30(162):210122. doi:10.1183/16000617.0122-2021
3. Liang SK, Chien LH, Chang GC, et al. Programmed death ligand 2 gene polymorphisms are associated with lung adenocarcinoma risk in female never-smokers. *Front Oncol*. 2021;11:753788. doi:10.3389/fonc.2021.753788
4. Shi K, Li N, Yang M, Li W. Identification of key genes and pathways in female lung cancer patients who never smoked by a bioinformatics analysis. *J Cancer*. 2019;10(1):51–60. doi:10.7150/jca.26908
5. Wang H, Zhang Z, Xu K, Wei S, Li L, Wang L. Exploration of estrogen receptor-associated hub genes and potential molecular mechanisms in non-smoking females with lung adenocarcinoma using integrated bioinformatics analysis. *Oncol Lett*. 2019;18(5):4605–4612. doi:10.3892/ol.2019.10845
6. Ziebarth NR. Lung cancer risk perception biases. *Prev Med*. 2018;110:16–23. doi:10.1016/j.ypmed.2018.01.020
7. Mishra S, Yadav T, Rani V. Exploring miRNA based approaches in cancer diagnostics and therapeutics. *Crit Rev Oncol Hematol*. 2016;98:12–23. doi:10.1016/j.critrevonc.2015.10.003
8. Ganju A, Khan S, Hafeez BB, et al. miRNA nanotherapeutics for cancer. *Drug Discov Today*. 2017;22(2):424–432. doi:10.1016/j.drudis.2016.10.014

9. Nie H, Xie X, Zhang D, et al. Use of lung-specific exosomes for miRNA-126 delivery in non-small cell lung cancer. *Nanoscale*. 2020;12 (2):877–887. doi:10.1039/c9nr09011h

10. Hong W, Xue M, Jiang J, Zhang Y, Gao X. Circular RNA circ-CPA4/ let-7 miRNA/PD-L1 axis regulates cell growth, stemness, drug resistance and immune evasion in non-small cell lung cancer (NSCLC). *J Exp Clin Cancer Res*. 2020;39(1):149. doi:10.1186/s13046-020-01648-1

11. Jiang C, Hu X, Alattar M, Zhao H. miRNA expression profiles associated with diagnosis and prognosis in lung cancer. *Expert Rev Anticancer Ther*. 2014;14(4):453–461. doi:10.1586/14737140.2013.870037

12. Cheng Y, Yang S, Shen B, et al. Molecular characterization of lung cancer: a two-miRNA prognostic signature based on cancer stem-like cells related genes. *J Cell Biochem*. 2020;121(4):2889–2900. doi:10.1002/jcb.29525

13. Asakawa Y, Okabe A, Fukuyo M, et al. Epstein-Barr virus-positive gastric cancer involves enhancer activation through activating transcription factor 3. *Cancer Sci*. 2020;111(5):1818–1828. doi:10.1111/cas.14370

14. Zeng F, Luo L, Li D, Guo J, Guo M. KPNA2 interaction with CBX8 contributes to the development and progression of bladder cancer by mediating the PRDM1/c-FOS pathway. *J Transl Med*. 2021;19(1):112. doi:10.1186/s12967-021-02709-5

15. Sun Y, Li W, Li X, Zheng H, Qiu Y, Yang H. Oncogenic role of karyopherin α2 (KPNA2) in human tumors: a pan-cancer analysis. *Comput Biol Med*. 2021;139:104955. doi:10.1016/j.compbiomed.2021.104955

16. Li H, Yuan H. MiR-1297 negatively regulates metabolic reprogramming in glioblastoma via repressing KPNA2. *Hum Cell*. 2020;33(3):619–629. doi:10.1007/s13577-019-00316-7

17. Putra V, Hulme AJ, Tee AE, et al. The RNA-helicase DDX21 upregulates CEP55 expression and promotes neuroblastoma. *Mol Oncol*. 2021;15 (4):1162–1179. doi:10.1002/1878-0261.12906

18. Huang RH, Yang WK, Wu CM, Yeh CM, Sung WW. Over-expression of CEP55 predicts favorable prognosis in colorectal cancer patients with lymph node involvement. *Anticancer Res*. 2021;41(1):543–547. doi:10.21873/anticanres.14806

19. Feng J, Guo Y, Li Y, et al. Tumor promoting effects of circRNA_001287 on renal cell carcinoma through miR-144-targeted CEP55. *J Exp Clin Cancer Res*. 2020;39(1):269. doi:10.1186/s13046-020-01744-2

20. Li M, Liu Y, Jiang X, et al. Inhibition of miR-144-3p exacerbates non-small cell lung cancer progression by targeting CEP55. *Acta Biochim Biophys Sin*. 2021;53(10):1398–1407. doi:10.1093/abbs/gmab118

21. Banerjee R, Liu M, Bellile E, et al. Phosphorylation of TRIP13 at Y56 induces radiation resistance but sensitizes head and neck cancer to cetuximab. *Mol Ther*. 2021. doi:10.1016/j.ymthe.2021.06.009

22. Chen X, Lu Y, Yu H, et al. Pan-cancer analysis indicates that MYBL2 is associated with the prognosis and immunotherapy of multiple cancers as an oncogene. *Cell Cycle*. 2021;1–18. doi:10.1080/15384101.2021.1982494

23. Zeng L, Liu YM, Yang N, Zhang T, Xie H. Hsa_circRNA_100146 promotes prostate cancer progression by upregulating TRIP13 via Sponging miR-615-5p. *Front Mol Biosci*. 2021;8:693477. doi:10.3389/fmolb.2021.693477

24. Chen F, Liu L, Wang S. Long non-coding RNA NORAD exhaustion represses prostate cancer progression through inhibiting TRIP13 expression via competitively binding to miR-495-3p. *Cancer Cell Int*. 2020;20:323. doi:10.1186/s12935-020-01371-z

25. Li X, Jiao M, Hu J, et al. miR-30a inhibits androgen-independent growth of prostate cancer via targeting MYBL2, FOXD1, and SOX4. *Prostate*. 2020;80(9):674–686. doi:10.1002/pros.23979

26. Xu Q, Xu Z, Zhu K, Lin J, Ye B. LINC00346 Sponges miR-30c-2-3p to promote the development of lung adenocarcinoma by targeting MYBL2 and regulating CELL CYCLE signaling pathway. *Front Oncol*. 2021;11:687208. doi:10.3389/fonc.2021.687208

27. Su X, Lang C, Luan A, Zhao P. MiR-200a promotes proliferation of cervical cancer cells by regulating HIF-1α/VEGF signaling pathway. *J BUON*. 2020;25(4):1935–1940.

28. Huang Y, Bao T, Li Z, Ji G, Zhang L. Function of miR-200a in proliferation and apoptosis of non-small cell lung cancer cells. *Oncol Lett*. 2020;20 (2):1256–1262. doi:10.3892/ol.2020.11649

29. Wang T, Tang X, Liu Y. LncRNA-ATB promotes apoptosis of non-small cell lung cancer cells through MiR-200a/β-Catenin. *J BUON*. 2019;24 (6):2280–2286.

30. Xiang J, Wu Y, Li DS, et al. miR-584 suppresses invasion and cell migration of thyroid carcinoma by regulating the target oncogene ROCK1. *Oncol Res*. 2015;38(9):436–440. doi:10.1159/000438967

31. Ebrahimi Ghahnavieh L, Tabatabaeian H, Ebrahimi Ghahnavieh Z, et al. Fluctuating expression of miR-584 in primary and high-grade gastric cancer. *BMC Cancer*. 2020;20(1):621. doi:10.1186/s12885-020-07116-5

32. Siriwardhana C, Khadka VS, Chen JJ, Deng Y. Development of a miRNA-seq based prognostic signature in lung adenocarcinoma. *BMC Cancer*. 2019;19(1):34. doi:10.1186/s12885-018-5206-8

33. Zhang Y, Roth JA, Yu H, et al. A 5-microRNA signature identified from serum microRNA profiling predicts survival in patients with advanced stage non-small cell lung cancer. *Carcinogenesis*. 2019;40(5):643–650. doi:10.1093/carcin/bgy132