International Journal of General Medicine Dovepress

Open Access Full Text Article ORIGINAL RESEARCH

Comprehensive Analysis of circRNA-Related mRNAs as Prognostic Factors in Non-Smoking Women with Lung Adenocarcinoma

Hao Wang[1,2], Song Wei[1,2], Lijun Wang[3], Zhihong Zhang[1,2]

[1]Department of Respiratory Oncology, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230031, People's Republic of China; [2]Department of Respiratory Oncology, Anhui Provincial Cancer Hospital, Hefei, Anhui, 230031, People's Republic of China; [3]Department of Respiratory Disease, Building 8 of Tongling People's Hospital, Tongling, Anhui, 244000, People's Republic of China

Correspondence: Lijun Wang; Zhihong Zhang, Email junliy890@163.com; ahszlyyhxk@163.com

Background: Non-smoking women with lung adenocarcinoma(NSWLA) is a significant health problem globally; the carcinogenesis and prognostic signature remain poorly understood. Circular RNAs (circRNAs) are gradually implicated in cancer formation through sponging miRNAs to regulate mRNAs.

Methods: Tumor and paracancerous specimens from non-smoking women after lung adenocarcinoma surgery were collected for high-throughput sequencing of circRNA. miRNA and mRNA datasets were downloaded from TCGA. A circRNA-miRNA-mRNA network was built using differentially expressed circRNAs (DEcircRNAs), miRNAs (DEmiRNAs), and mRNAs (DEmRNAs). Following that, GSEA was applied to analyze the biological function of mRNAs in the ceRNA network. Utilizing the mRNAs associated with prognosis, we created an original prognostic risk score model. The expression of DEmRNA in the ceRNA network was verified by mRNA-seq and scRNA-seq data (GSE131907). The expression of BTBD3 and EIF4EBP2 was then verified by immunohistochemistry.

Results: 16 pairs of adenocarcinoma tissues and their corresponding para-tumor samples were collected from thoracic surgery. We created a circRNA related ceRNA in NSWLA. The hsa_circ_0002346 regulation of the cancer cell proliferation may through the hsa_circ_0002346/miR-96-5p/EIF4EBP2 axis. Hsa_circ_0072309 may affect proliferation of lung adenocarcinoma and activate Nature Killer cells by targeting miR-32-5p to regulate PCMTD1 expression. Based upon mRNA (BTBD3, CFL2, EIF4EBP2, EVI5, PCMTD1, SYDE2) related to overall survival, we also created a predictive signature. According to mRNA-seq, scRNA-seq, and immunohistochemical data, the expression of BTBD3 and EIF4EBP2 was lower in tumor samples than in normal tissues nearby.

Conclusion: The circRNA related mRNAs played an important role in predicting the overall survival of NSWLA. The circRNA in the ceRNA network might unravel the pathogenesis and be potential novel targets for NSWLA.

Keywords: women, lung adenocarcinoma, circRNA, miRNA, ceRNA, prognosis

Introduction

Lung cancer remains the leading cause of cancer death, with an estimated 1.8 million deaths in 2020 worldwide[1] and mainly due to tobacco exposure.[2] Lung adenocarcinoma is the main pathological type of lung cancer. However, lung adenocarcinoma incidence in never-smokers remains a significant health problem globally, especially in East Asian women.[3] Cooking factors, infectious elements, frequent alcohol drinking, and meat-based diet were verified are the main risk factors for lung cancer in Asians.[4–6] In non-smoking women, a number of significant genes, including GREM1, MMP11, SPP1, FOSB, and IL33, were highly linked to lung cancer.[6,7] Some metabolic pathways, including 1-carbon metabolism, nucleotide metabolism, oxidative stress, and inflammation were also associated with lung cancer in non-smoking women.[8] Understanding of the lung adenocarcinoma-related carcinogenesis and prognostic signature in non-smoking women is still lacking.[9]

Received: 13 September 2024
Accepted: 15 November 2024
Published: 3 December 2024

The ends of an individual RNA molecule are covalently linked to form circRNAs, a subclass of non-coding RNAs.[10] CircRNAs function as a sponge for microRNAs(miRNAs) and regulate gene expression by reducing the ability of miRNAs to target genes.[11] Numerous recent investigations have confirmed that circRNA is essential for the development of tumors.[12] In ovarian cancer,[13] circ-BNC2 could regulate the miRNA-223-3p/FBXW7 axis to suppress the cancer cell progression. In hepatocellular carcinoma,[14] circ_0008285 could promote the malignant biological behaviors of cancer cells through the miRNA-384/RRM2 axis. Via tripartite motif-containing protein 3, circVPRBP acts as a sponge for miRNA-106b-5p to control the growth and spread of cervical cancer cells.[15] Circ_0001058 represses the progression of lung adenocarcinoma by governing the miRNA-486-5p/TEK signaling axis.[16] These results demonstrate that circRNA interacts with miRNA and mRNA in the ceRNA network and is critical for the development and spread of cancer. The circRNA-associated mRNA in NSWLA, nevertheless, has not been systematically analyzed and needs more research.

For high-throughput sequencing, we collected tumor and paracancerous specimens from non-smoking women after lung adenocarcinoma surgery and obtained circRNA and mRNA expression profiling data. The mRNA and miRNA expression profiles were also collected from the TCGA database. Then, using bioinformatics evaluation, we built a ceRNA network associated with circRNA. In addition, we explored the protein-protein network and biological processes based on the mRNAs in the ceRNA network. Depending on the mRNAs associated with the prognosis, we created an innovative prognostic risk score model. Ultimately, separate data sets and immunohistochemistry were used to confirm the levels of circRNA and mRNA in the ceRNA network. This work offers fresh perspectives on the role of circRNAs in the development of adenocarcinoma of the lungs in smoking-free women. Figure 1 depicts the flowchart for the creation of the ceRNA network.

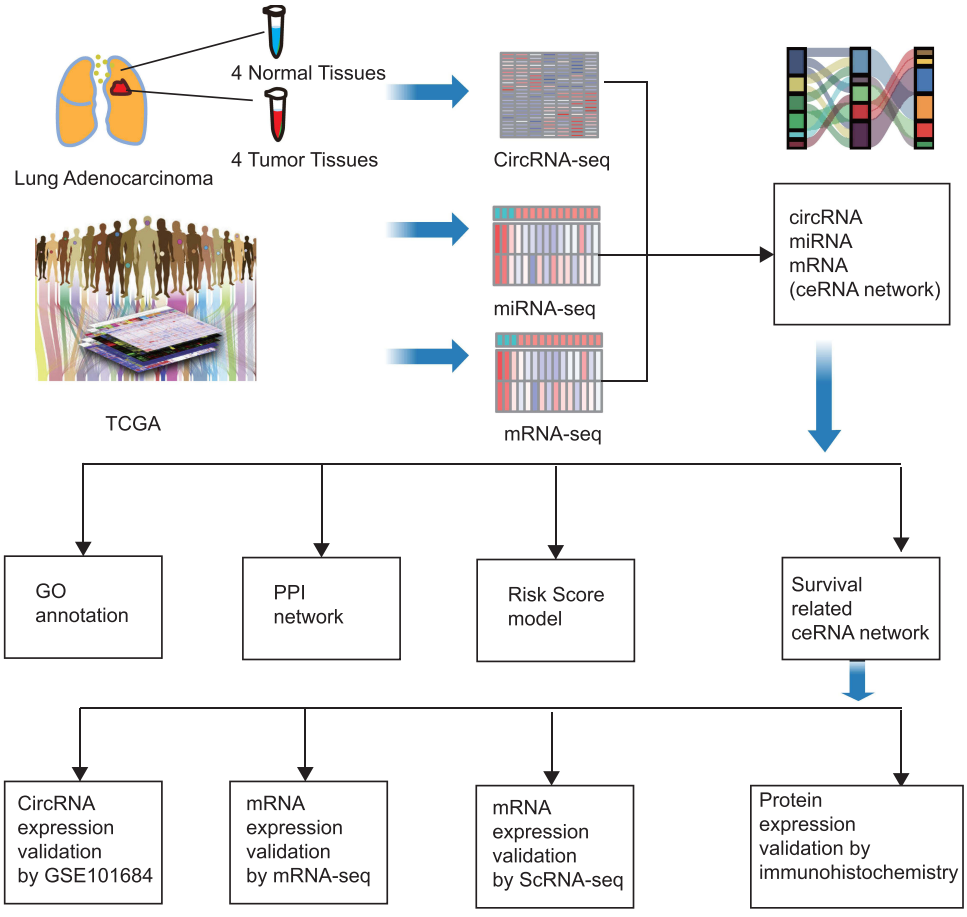

Figure 1 Schematic plot for the identification and analysis of prognostic circRNAs in the ceRNA network.
Abbreviations: NSWLA, Non-smoking women with lung adenocarcinoma; GO, Gene Ontology; PPI, protein protein interaction network; GSEA, Gene Set Enrichment Analysis; IHC, Immunohistochemistry.

Materials and Methods

Sample Collection and circRNA Sequencing

From June 2020 to June 2021, smoke-free females with LUAD who underwent thoracic surgery at the provincial cancer hospital in Anhui were the subjects of all tissue samples. Pathologists identify pathological types of lung adenocarcinoma. Tumor and paracancerous tissue were stored in a −80°C freezer for further sequencing. TRIzol was used to isolate the whole RNA from fresh tissues (Thermo Fisher Scientific, Inc.), in accordance with the instructions provided by the manufacturer. RNA was tested using the NanoDrop 2000 to determine its level of concentration and quality. The RNA preparation process started with 3 g of RNA for each sample. Following preparing the libraries and pooling the various samples, Illumina sequencing (Illumina Novaseq 6000) was performed on the samples. First, filter low-quality data to get high-quality (HQ) clean reads. And then alignment of HQ Clean Reads with reference genomes. Finally, the alignment results were submitted to the findcirc software (version 1.2.0) to identify circular RNAs. With the help of HISAT2 and the software Stringtie (version 1.3.4), researchers were able to rebuild transcripts and identify genes. mRNA sequencing for ceRNA network identification. The Human Investigation and Ethical Committees of the Anhui Provincial Cancer Hospital gave their approval for the current study (Ethics Number-41/2020). All participants provided written informed consent.

Data Collection and Processing

The Limma program and Student's *t*-test between women with LUAD who do not smoke and neighboring normal lung tissues (DFR0.01 and |Log2FC|>3) were used to identify the differentially expressed circRNAs (DEcircRNA). The expression of miRNA (73 non-smoking female LUAD and three normal tissues) and mRNA (72 non-smoking female LUAD and 11 normal tissues) were retrieved from the TCGA dataset. The Limma package and Student's *t*-test were used to identify differentially expressed miRNA (DFR<0.05 and |Log2FC|>2) and differentially expressed mRNA (DFR<0.05 and |Log2FC|>2). By removing unknown stages, age, and survival time, 68 non-smoking female LUAD patients were used to test the predictive value.

Constructing the circRNA-Associated ceRNA Network

We constructed the ceRNA network based on the DEcircRNA, DEmiRNA, and DEmRNA. To predict the DEcircRNA-associated miRNA, the cancer-specific circRNA database (CSCD, http://gb.whu.edu.cn/CSCD/) was employed. As a follow-up study, we selected a sequence that is 100% consistent with the position of DEcircRNA to create high-quality circRNAs that can serve as miRNA targets. And then, these predicted target miRNAs were further screened by DEmiRNA in the TCGA database. Depending on the data from TargetScan, Mortar, and miRDB databases, the remaining DEmiRNAs were utilized to estimate the interactions between DEmiRNA and mRNA. These predicted DEmRNA screened target mRNAs in the TCGA database. Depending on the DEcircRNA – DEmiRNA pairs and DEmiRNA – DEmRNA pairs, the novel circRNA – miRNA – mRNA network was structured.

PPI Network and GO Functional Enrichment

To explore the interaction of mRNAs and hub genes in the ceRNA network, we uploaded the mRNA in the ceRNA network to the STRING database (https://cn.string-db.org/). The minimum required interaction score in the PPI network was 0.4. We also hide disconnected nodes in the network. We uploaded the PPI network to probe hub genes in the PPI network and calculated key genes with Cytoscape plug-in cytoHubba. We used GO functional enrichment analysis relying on the DAVID database (https://david.ncifcrf.gov/) to inquire into the genetic basis of circRNA in the ceRNA network connected to smoke-free females with lung cancer. GO functional enrichment includes biological processes(BP), cellular components(CC), and molecular functions(MF). Each category demonstrates the biological functions of genes at a different level.

Survival Analysis

According to the mRNAs contained within the ceRNA network, we applied the univariate Cox regression analysis to test the overall survival-associated genes. The predictive hallmark was created utilizing the least absolute shrinkage and selection operator (LASSO) Cox regression analysis and those genes that had a strong association with OS. Then the candidate genes and weighted genes expression value were identified. Risk score = sum of coefficients* the expression of genes. A calculation of the risk score, smoke-free females with LUAD were divided into low- and high-risk categories. 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. To implement the Kaplan-Meier (K-M) assessment of survival, we used R's "survival package". The "suvivalROC" package was utilized to carry out the receiver operating characteristic(ROC) curve. The age, stage, primary tumor (T), and lymph node metastasis (N) were considered in the univariate and multivariate Cox regression analysis to remove the deviation brought on by clinical factors and ascertain whether OS could be predicted independently using the risk score model.

Gene Set Enrichment Analysis (GSEA)

The clusterProfiler (Version 4.10.0) was used to conduct the Gene Set Enrichment Analysis. The unigenes were utilized as the gene set and the gene expression matrix was constructed based on the KEGG database. The quantitative measures of enrichment magnitude and statistical significance were determined using the Normalized Enrichment Score (NES) and False Discovery Rate (FDR, p-adjust), respectively.

circRNA and mRNA Validation

To confirm that circRNA is expressed in the ceRNA network, the transcription of circRNA was retrieved from GSE 101684. The Limma package and Student's *t*-test were used to identify DEcircRNA and DEmRNA(DFR<0.05 and |Log2FC|>2). First of all, we validate the expression of mRNA by mRNA-seq. The Limma package and Student's *t*-test identified the differentially expressed mRNA between non-smoking females with LUAD and adjacent normal lung tissues (DFR<0.01 and |Log2FC|>3). And then, the expression of mRNA was validated at the single-cell level. The GEO database (GEO 131907) was additionally utilized to retrieve the scRNA-seq data of 19 smoke-free females with adenocarcinoma of the lungs (13 tumor tissues and 6 normal tissues). The harmony package batch corrected the data of the samples. The highest 2000 extremely diverse genes were found using the FindVariableFeatures program, and it was done using principal component analysis. To construct clusters, the top 30 fundamental elements and a resolution of 0.6 were employed in the discover clusters program. The characteristic genes in every cluster were identified using the FindAllMarkers tool.

Pathological Sample Collection and Immunohistochemistry

Thoracic surgery was used to collect an overall of 12 pairs of adenocarcinoma of the lungs tissues and their corresponding para-tumor samples at the Anhui Provincial Cancer Hospital. Prior to being inserted in paraffin, adenocarcinoma of the lungs and para-tumor tissues were fixed for 24 hours in 10% neutralized formaldehyde. The paraffin-embedded specimens were divided into slices that were 3 mm thick. The primary antibodies (BTBD3, BiossANTIBODIES, 1:200 and EIF4EBP4, BiossANTIBODIES, 1:100) were incubated at 4°C overnight. The slides were then counterstained with hematoxylin after being treated with the secondary antibody for an hour at room temperature. Immunolabeling staining evaluation; Staining intensity: negative= 0 points; mild positive= 1 point; moderate positive, 2 points; strong positive= 3 points. Staining ratio:<5%= 1 point; 6% −25%= 2 points; 26–75%= 3 points; >76%= 4 points. The total score equals Staining intensity multiplied by the Staining ratio: negative=0 points, weakly positive=1-4 points, positive=5-8 points, strongly positive=9-12 points.

Statistical Analysis

The statistical package R performed all statistical analysis (version 4.1.0, https://www.r-project.org/). The *t*-test was used to distinguish between malignancy and normal tissue RNA expression. To determine whether the fact was connected to prognosis, univariate and multivariate Cox were suggested. A P value of 0.05 or less was regarded as statistically meaningful.

Result

Identification of DEcircRNAs, DEmiRNAs and DEmRNAs

Through high-throughput sequencing, we obtained the expression data of circRNA between 4 paired non-smoking females with LUAD and adjacent non-tumor tissues. Utilizing the threshold parameters of $p < 0.01$ and $|logFC| > 3$, 30 DEcircRNAs (14 upregulated circRNAs and 16 downregulated circRNAs) were found, as shown in Figure 2A and B. The TCGA dataset was used to obtain miRNA and mRNA transcript data. Using $|logFC| > 2$ and $p < 0.05$ as the cutoff

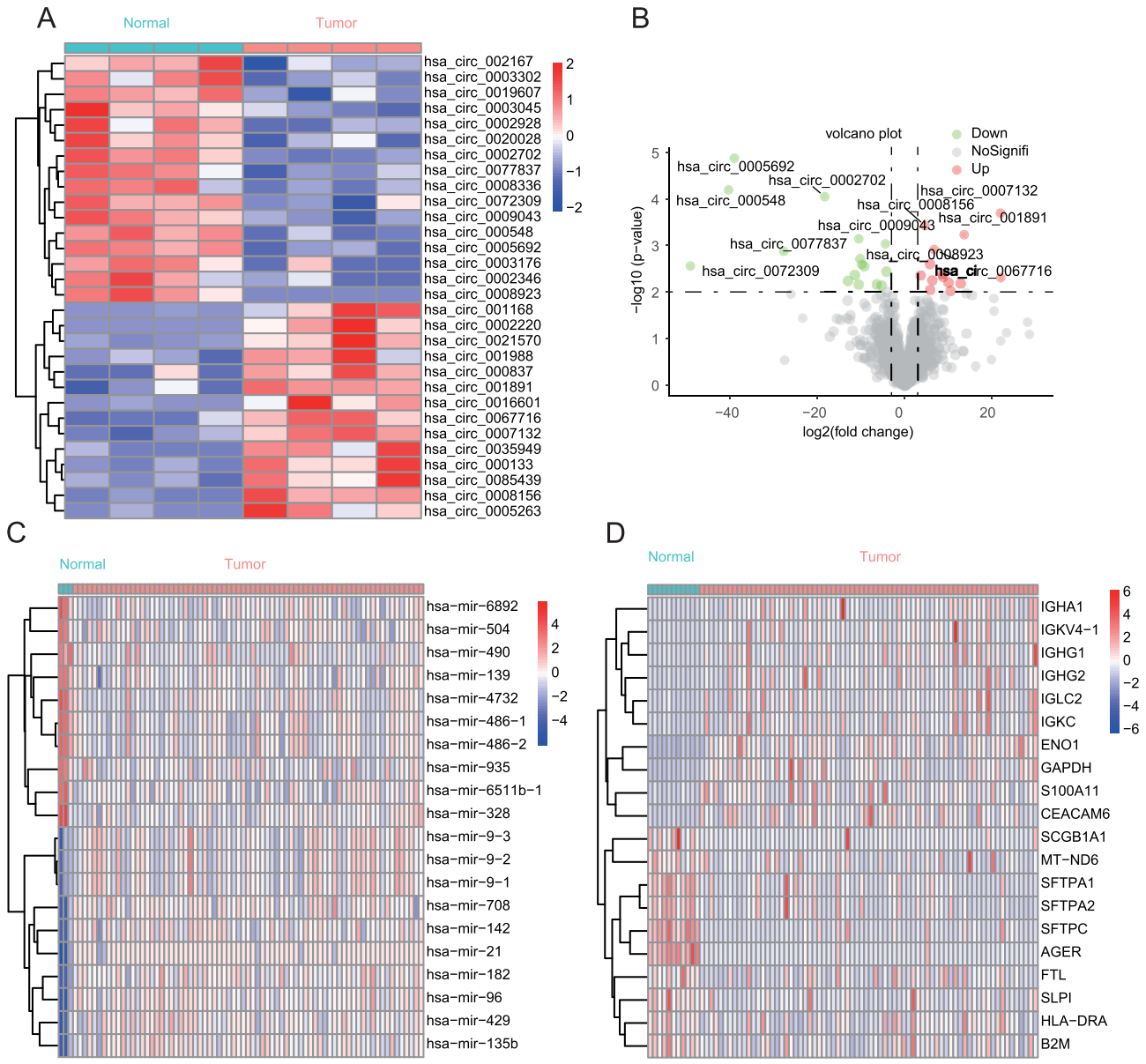

Figure 2 Differentially expressed circRNA, miRNA, and mRNA. (**A**) Heatmap of top 30 differentially expressed circRNA in non-smoking females with postoperative lung adenocarcinoma. (**B**) Volcano plots of differentially expressed circRNA. (**C**) Heatmap of the top 20 differentially expressed miRNA in non-smoking females with lung adenocarcinoma of TCGA dataset. (**D**) Heatmap of the top 20 differentially expressed mRNA in non-smoking females with lung adenocarcinoma of TCGA dataset.

criteria, 148 DEmiRNAs were identified (75 upregulated miRNAs and 73 downregulated miRNAs), and 4279 DEmRNAs were identified (2614 upregulated mRNAs and 1665 downregulated mRNAs) as shown in Figure 2C and D.

Construction of the circRNA-Associated ceRNA Network

To explore the effect of circRNAs on mRNAs in combination with miRNAs among nonsmokers with lung adenocarcinoma, we constructed the ceRNAs network. miRNAs that might bind to circRNAs were predicted by the circBase dataset. using 30 DEcircRNAs and 1180 predicted miRNAs, we discovered 1757 circRNA-miRNA pairs. After overlapping the predicting miRNAs and DEmiRNAs in the TCGA database, only 38 miRNAs remained. We further predicted the mRNAs of the 38 miRNA by TargetSacn. The overlapped predicting mRNAs and DEmRNAs in TCGA datasets were selected, and 386 mRNA

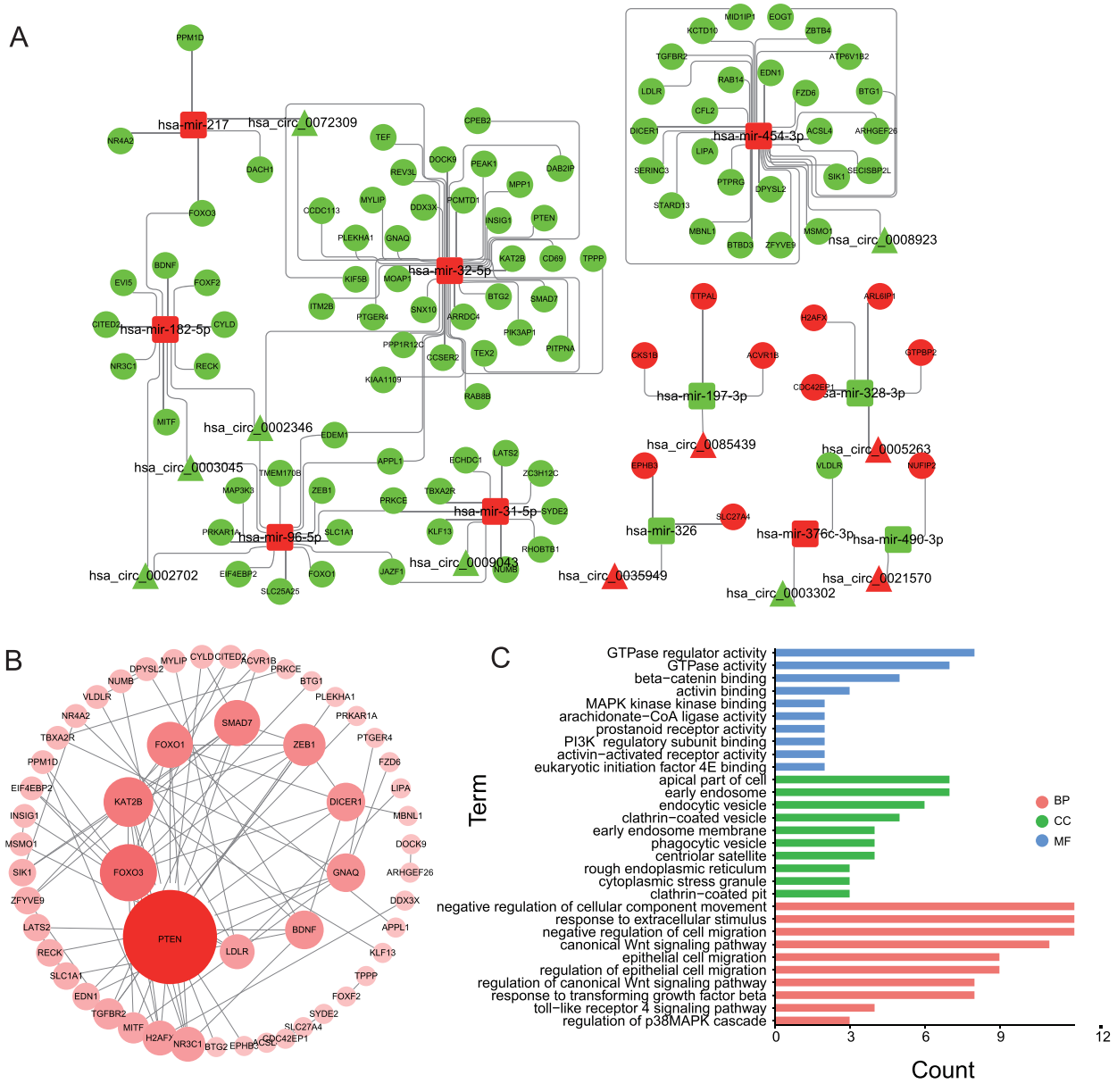

Figure 3 The ceRNA network in non-smoking females with lung adenocarcinoma. (**A**) The ceRNA network by DEcircRNA, DEmiRNA and DEmRNA. Circular represents mRNA. Square represents miRNA. Triangle represents miRNA. Red indicates upward adjustment. Green indicates a downward adjustment. (**B**) PPI network of the 102 DEmRNAs in the ceRNA network. Small circle represents hubgenes. The size and color depth of circles represent the weights of DEmRNA. (**C**) Gene Ontology(GO) of DEmRNAs in the ceRNA network.

remained. Finally, a ceRNAs network was constructed, including 11circRNAs, 11miRNAs, and 102 mRNAs, and was visualized in Figure 3A.

PPI Network and Functional Enrichment Analysis of DEmRNAs in the ceRNA Network

To investigate the relationships between the ceRNA network's 102 DEmRNAs, we submitted the DEmRNAs to the string database and built the PPI network. We removed the disconnected nodes, and the PPI network included 54 nodes, and 85 edges, as shown in Figure 3B. The PPI network's top 10 essential genes were also identified by cytoHube, as depicted in Figure 3B. The 102 DEmRNAs in the ceRNA network's Gene Ontology (GO) were also built. Figure 3C displays the results of the functional enrichment. Many studies on the development and spread of cancer mentioned terminology like "regulation of epithelial cell migration", "canonical Wnt signaling pathway", "response to TGF-beta", and "regulation of p38-MAPK cascade". These results confirm that DEmRNAs within the network play an important role in nonsmoking females with lung adenocarcinoma. Furthermore, the lung adenocarcinoma tumorigenesis and development in smoke-free females may include the circRNA-associated ceRNA network.

Construction of Risk Scoring Model and Survival-Related ceRNA Network

We next explored the relationship between the 102 DEmRNAs and overall survival using univariate Cox regression analysis. The high expression of EIF4EBP2, RAB14, EDN1, PCMTD1, EVI5, SYDE2, and TPPP is related to a bad outlook. On the other hand, BTBD3 and CFL2 expression levels are high, indicating a favorable prognosis (Figure 4A). The nine genes linked to survival were then examined using LASSO Cox regression analysis (Figure 4B). Using the minimum requirement, six genes (EIF4EBP2, BTBD3, CFL2, PCMTD1, EVI5, and SYDE2) were selected to build the risk-scoring model. This is how the risk score was determined: risk score = (EIF4EBP2 expression * 0.5779) - (BTBD3 expression * 0.8044) - (CFL2 expression * 0.1348) + (PCMTD1 expression * 0.1468) + (EVI5 expression * 0.3891) + (SYDE2 expression *0.5690). Consistent with our study, the high expression of CFL2 predicted improved overall survival in leiomyosarcomas.[17] Additionally, those with carcinoma of the liver who had high EVI5 expression had shorter overall survival times.[18] We also constructed a survival-related ceRNAs network, including six circRNAs, five miRNAs, and six survival-related mRNAs, visualized in Figure 4C. In order to better understand the potential mechanism of EIF4EBP2 and PCMTD1 in NSWLA, we performed GSEA analysis in KEGG. The results revealed that EIF4EBP2 and PCMTD1 is negatively correlated with the cell cycle and DNA replication. It is worth noting that EIF4EBP2 was notably enriched in MAPK signal pathways and PCMTD1 was enriched in Nature Killer cell mediated cytotoxicity (Figure 4D and E).

Risk Scoring Model and Survival

After calculating each patient's risk score, the patients were divided into two categories: high-risk (n = 34) and low-risk (n = 34), based on the median risk score. OS in the high-risk group was shown to be lower than in the low-risk score group, according to Kaplan-Meier survival analysis (Figure 5A). We employed receiver operating characteristic (ROC) curve analysis to evaluate the model's precision. The 1-, 3-, and 5-year ROC curves' area under the curve (AUC) values were 0.705, 0.647, and 0.680, respectively, demonstrating the model's moderate accuracy and specificity. (Figure 5B). Both in train-dataset (Figure 5C) and test-dataset (Figure 5D), patients in the lower-risk group had significantly longer overall survival than that in the higher-risk group. In the univariate and multivariate Cox regression analyses, we employed the age, stage, primary tumor (T), and lymph node metastasis (N) to determine whether the risk score model was a separate predictor for OS. In the univariate Cox regression analysis, the hazard ratio was 1.402, $p < 0.001$ (Figure 5E). After the multivariate Cox regression analysis took the patients' clinical characteristics into account, the risk score model also demonstrated statistical significance (Figure 5F).

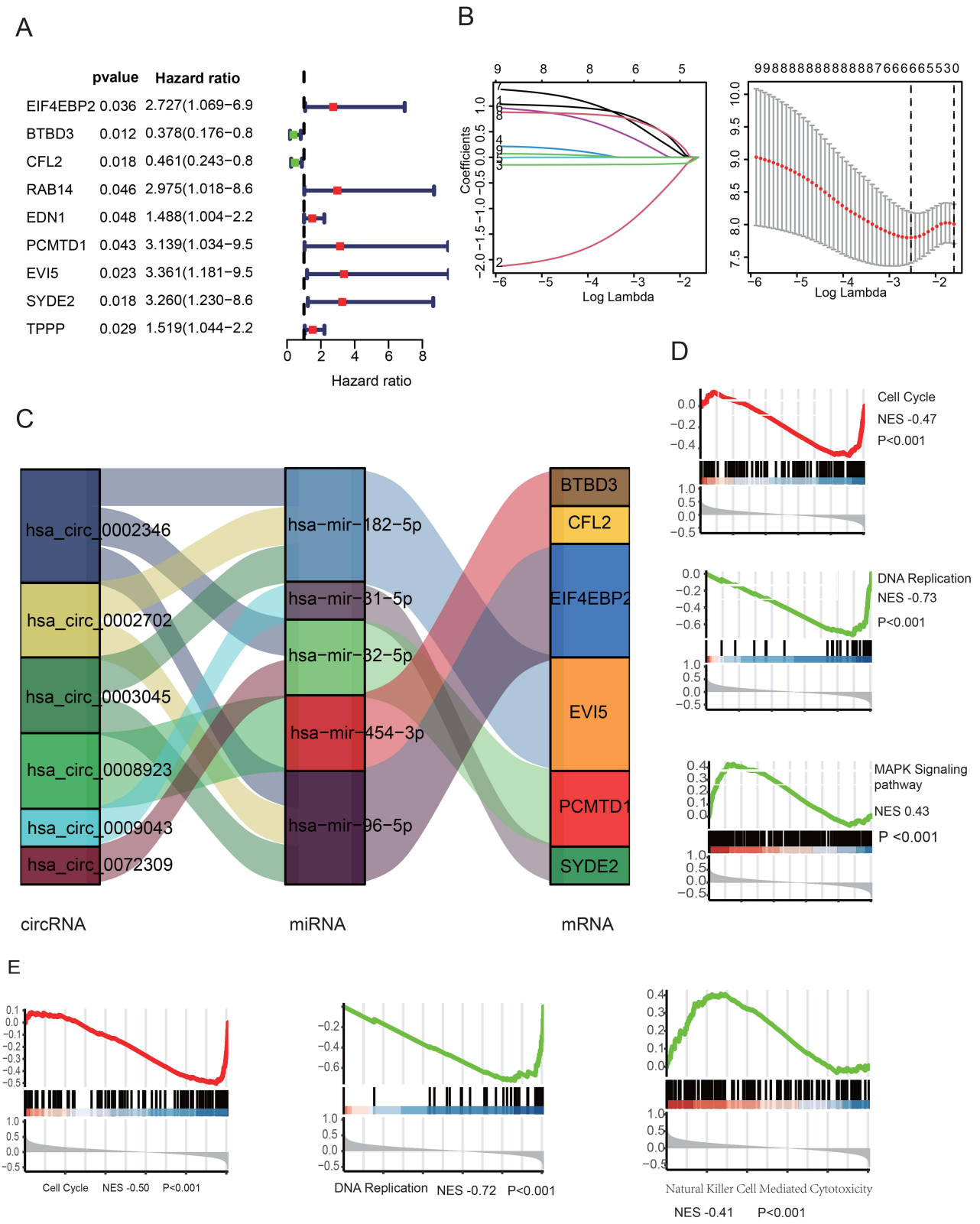

Figure 4 Survival-related ceRNA network in non-smoking females with lung adenocarcinoma. (**A**) Univariate Cox proportional-hazards regression analyses were used to evaluate the overall survival of DEmRNAs in the ceRNA network. (**B**) The minimum criteria calculated by LASSO Cox regression. (**C**) Survival-related ceRNA network. (**D**) GSEA analyses of pathways that significantly affected by EIF4EBP2. (**E**) GSEA analyses of pathways that significantly affected by PCMTD1.

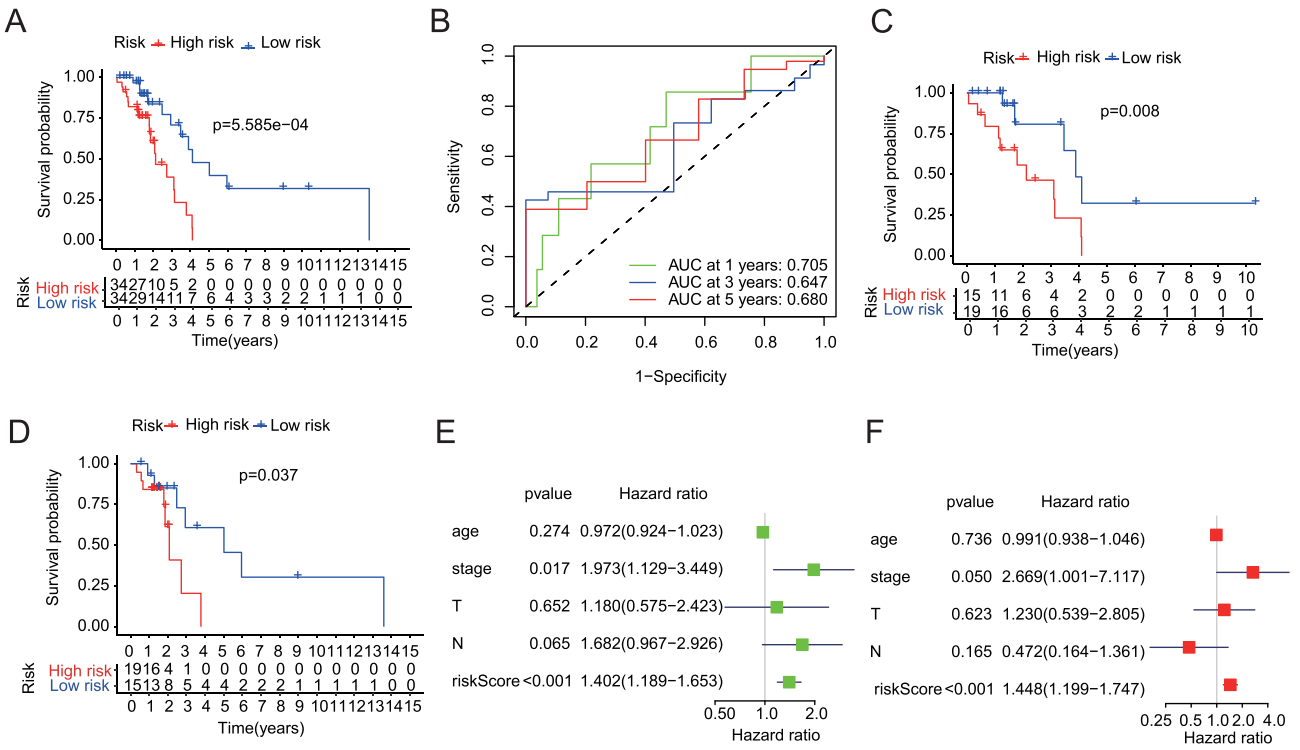

Figure 5 The relationship between risk score and survival. (**A**) Kaplan–Meier survival curve display the overall survival of non-smoking females with lung adenocarcinoma according to the median cutoff value of risk score in all-dataset (68 patients). (**B**) Time-dependent receiver operating characteristic 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. (**E**) Forest plots show the univariate Cox regression results between risk score and OS. (**F**) Forest plots show the multivariate Cox regression results between risk score and OS.

The Expression of circRNA and mRNA in the ceRNA Network

To validate the expression level in the circRNA, we analyzed the GSE 101684 datasets. According to our earlier investigation, the expression of hsa_circ_0009043 and hsa_circ_0072309 was significantly reduced in tumor tissues (Figure 6A and B). According to the mRNA-seq findings, tumor tissues expressed considerably less BTBD3 and EIF4EBP2 than normal tissues (Figure 6C and D). To compare the mRNA expression of healthy and malignant epithelial cells, we investigated the GSE 131907 datasets. As shown in Figure 7A, the cells were split into nine separate cell lineages, and their canonical marker gene expression was annotated. Epithelial cells in tumor tissues are called tumor-epithelial cells, whereas epithelial cells in normal tissues are called normal epithelial cells (Figure 7B). In line with our earlier research, tumor epithelial cells showed down-regulation of BTBD3 and EIF4EBP4 (Figure 7C).

Validation of the Expression of BTBD3 and EIF4EBP2 in Immunohistochemistry

The IHC staining was performed on 12 primary lung adenocarcinoma tissues coupled with para-tumor tissues. BTBD3 shows nuclear plasma positive expression in both lung adenocarcinoma and normal lung tissue. In adenocarcinoma tissue of the lung, the main manifestation is nuclear cytoplasmic positivity. Still, the degree of nuclear positivity differs from that of the normal alveolar epithelium, decreasing overall staining intensity. EIF4EBP2 showed membrane plasma positive expression in both normal lung tissue and lung adenocarcinoma tissue. The immunohistochemical staining results indicate that EIF4EBP2 expression is downregulated in lung adenocarcinoma tissue compared to the normal alveolar epithelium (Figure 8).

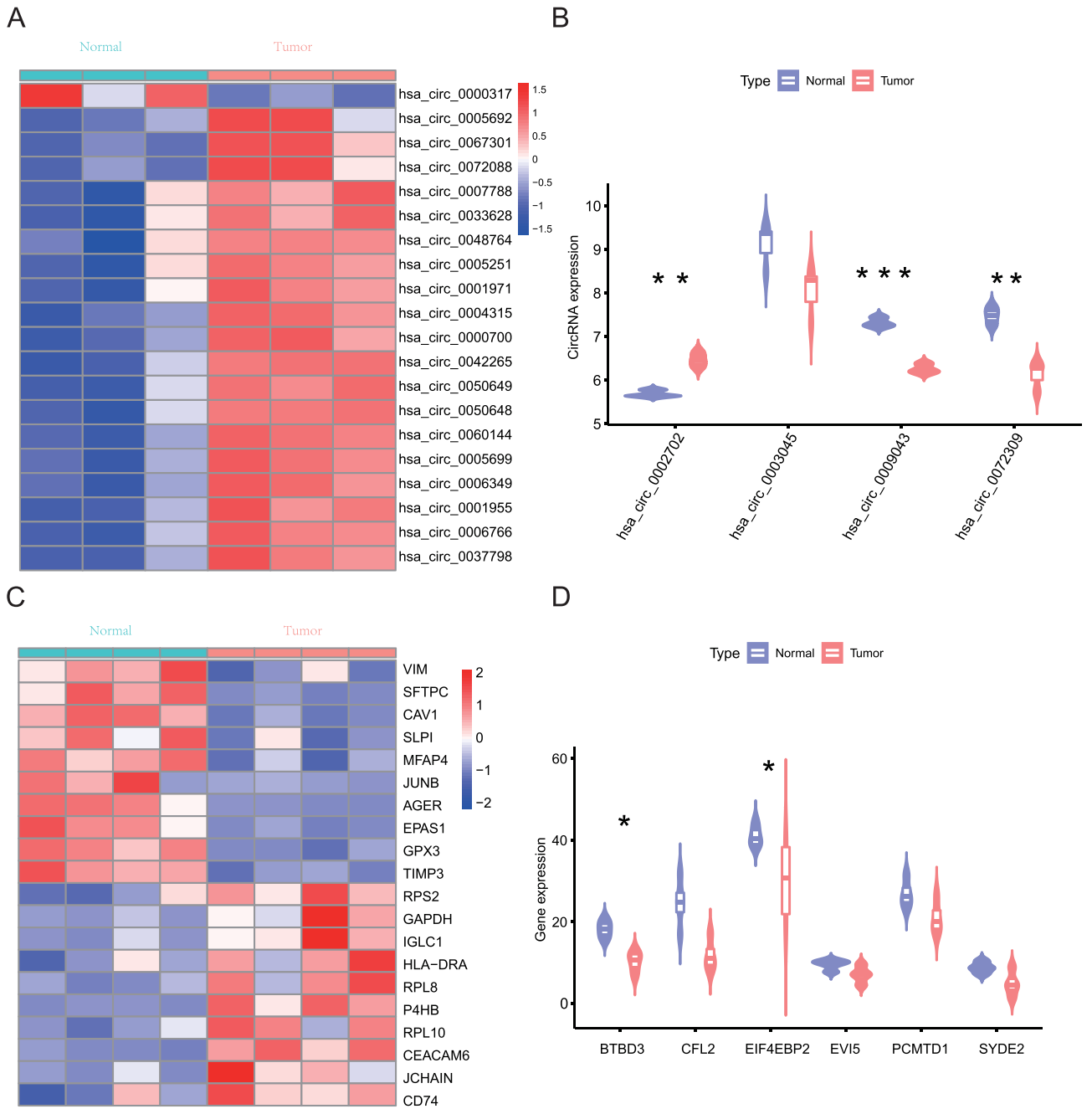

Figure 6 Validation of the expression level of the circRNA and mRNA in the survival-related ceRNA network. (**A**) Heatmap of top 20 differentially expressed circRNA in non-smoking females with lung adenocarcinoma in GSE 101684 dataset. (**B**) Violin plot of circRNA in the survival-related ceRNA network of GSE 101684 dataset. (Wilcox. test, *p-value < 0.05; **p-value <0.01; ***p-value <0.001). (**C**) Heatmap of top 20 differentially expressed mRNA in non-smoking females with lung adenocarcinoma in mRNA-seq data. (**D**) Violin plot of mRNA in the survival-related ceRNA network of mRNA-seq data. (Wilcox.test, *p-value < 0.05; **p-value <0.01; ***p-value <0.001).
Abbreviations: Normal, normal tissue; Tumor, Lung adenocarcinoma tissue.

Discussion

Adenocarcinoma of the lungs in non-smoking women(NSWLA) is becoming more common. Further exploring the causes would help to improve the prevention and protect non-smoking women with lung adenocarcinoma. Our previous studies have shown that hsa-mir-200a, hsa-mir-21, and hsa-mir-584 can serve as potential prognostic predictors and based on the three DEmiRNAs a novel prognostic model was also constructed which showed good performance in NSWLA.[19] CircRNAs could also be employed as biomarkers for cancer diagnosis and prognosis, according to an increasing number of studies,[20,21] however, circRNAs have been less studied in NSWLA. As ceRNAs, the circRNAs

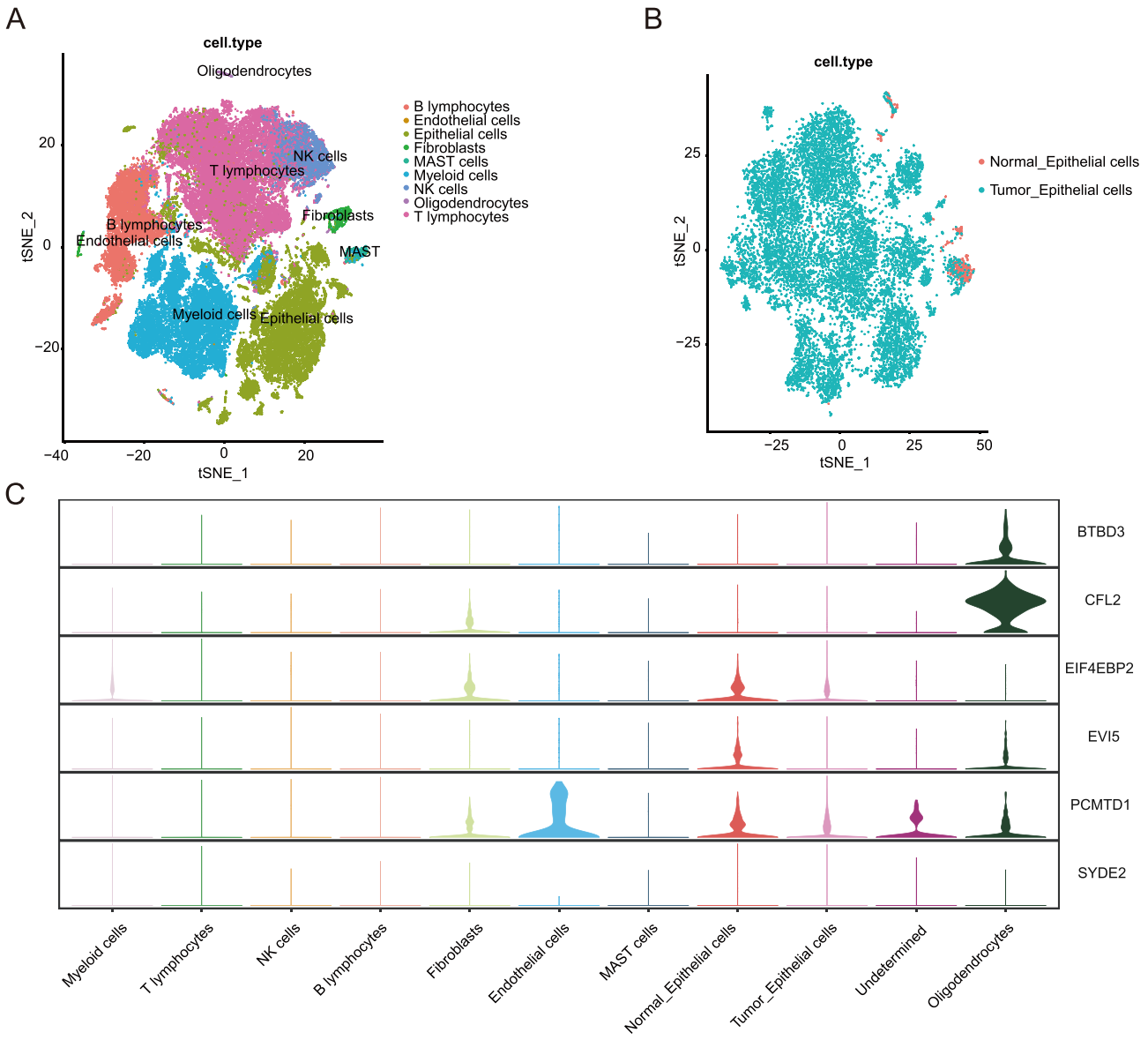

Figure 7 Validation of the expression level of mRNA in the survival-related ceRNA network at the single-cell level. (**A**) TSNE plot of nine distinct cell lineages annotated with canonical marker gene expression in GSE 131907. (**B**) TSNE plot of tumor-epithelial cells and normal-epithelial cells in GSE 131907. (**C**) mRNA expression of the survival-related ceRNA network in GSE 131907.

play crucial roles in the emergence and progression of lung cancer.[22–24] Many articles proved that circRNAs could regulate various cell biological processes by absorbing miRNAs.[25,26] Has_circ_002346 derived from exons 2, 3, and 4 of the CRIM1 gene, also named circCRIM1. Circ_0002346 upregulation in NSCLC reduced the cancer cells' ability to spread through interacting with miRNA-582-3p and inducing the production of STXBP6.[27] In lung adenocarcinoma, interaction between hsa_circ_0002346 and miRNAs 93 and 182 to increase leukemia inhibitory factor receptor expression, a well-known tumor suppressor, which could inhibit the invasion and metastasis of lung adenocarcinoma cancer cells.[28] It was discovered that Hsa_circ_0002346 decreased in non-smoking women with adenocarcinoma of the lungs, which is in line with past studies. Additionally, we discovered that the DEmRNA in the hsa_circ_0002346 related ceRNA network participated in cell migration and MAPK signal pathway, leading us to hypothesize that hsa_circ_0002346 may act through MAPK signal pathway to influence tumor cell migration in lung adenocarcinoma. However, in the GSE101684 dataset, We did not find the expression of hsa_circ_0002346, and more experiments are needed to verify the function of hsa_circ_0002346.

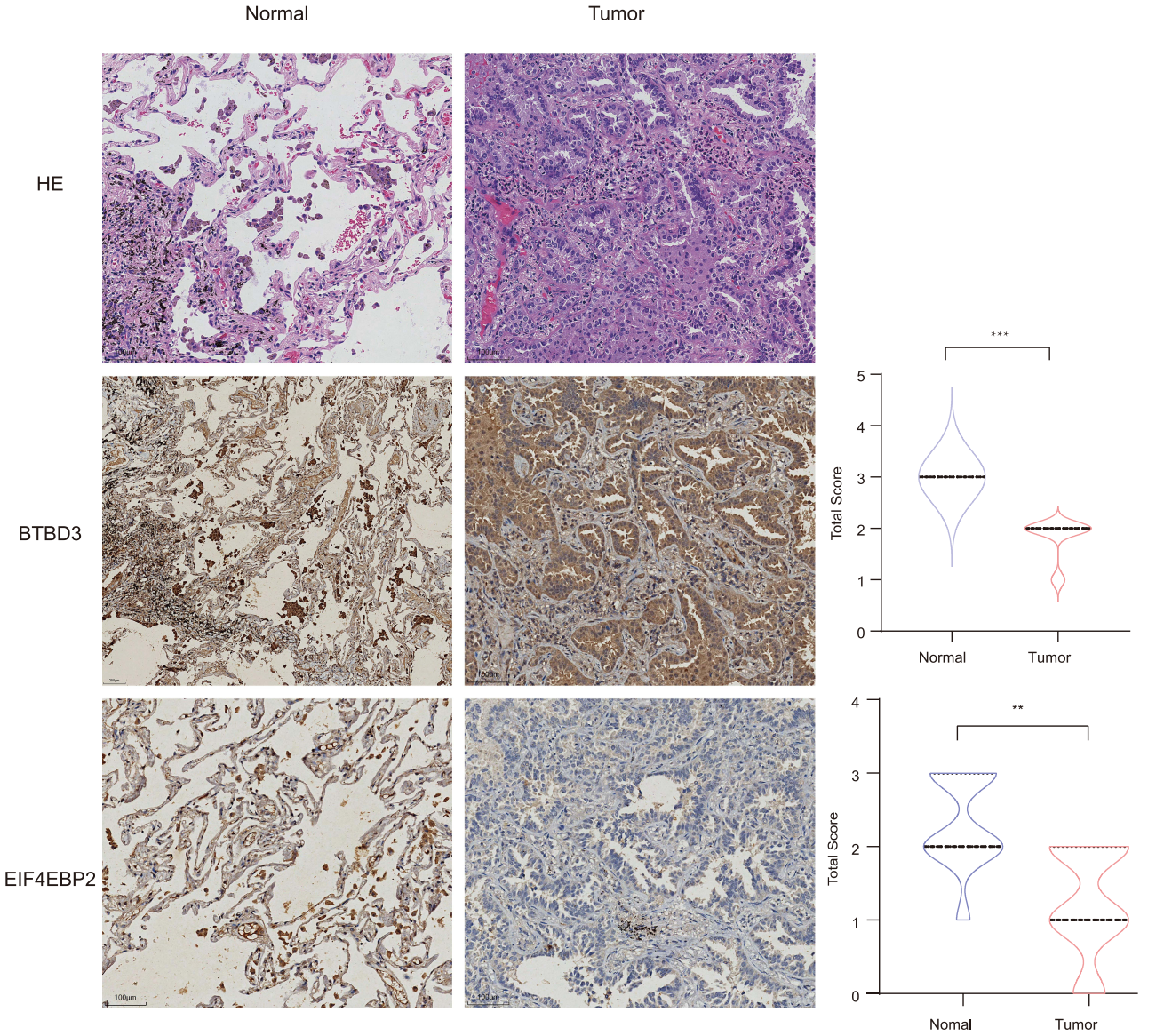

Figure 8 Validation of the expression of BTBD3 and EIF4EBP2 by immunohistochemistry Representative microscopic images (scale bar 250 um) demonstrate high expression of BTBD3 and EIF4EBP2 in normal tissues confirmed by IHC; the result of quantification of BTBD3 and EIF4EBP2 expression in IHC images is shown in the right part (Paired sample non parametric testing). (**p-value <0.01; ***p-value <0.001).
Abbreviations: Normal, normal tissue; Tumor, Lung adenocarcinoma tissue; Total score, Staining intensity multiplied by the Staining ratio.

Moreover, eIF4E-mediated translational activation by reducing EIF4E-BP2 via mRNA degradation leads to enhanced cell proliferation.[29] EIF4E-BP2 is downregulated in non-smoking women with adenocarcinoma of the lungs, which is consistent with our findings. In addition, hsa_circ_0002346 could interact with miRNA 96–5p to influence the expression of EIF4E-BP2 in the ceRNA network. Therefore, we speculate that has_circ_0002346 interaction with miRNA-96-5p reduces the expression of EIF4E-BP2, which could promote the proliferation of lung adenocarcinoma cancer cells.

Hsa_circ_0072309 is derived from exons 2–5 of the LIFR gene, also named circLIFR. Experimental evidence has shown that Hsa_circ_0072309 reduces in breast cancer,[30] renal carcinoma,[31] bladder cancer,[32] and lung adenocarcinoma,[33] indicating its significant tumor-suppressive functions in the pathophysiology of cancer. CircLIFR was demonstrated to suppress tumor metastasis in vivo and cell migration in vitro in models of liver.[34] Consistent with previous studies, our result found that Hsa_circ_0072309 was downregulated in non-smoking women with lung adenocarcinoma. By the way, we also got a comparable conclusion with the GSE101684 database. In breast cancer

and renal carcinoma, Hsa_circ_0072309 was found to function mechanistically as a sponge for miRNA.[33] To suppress tumor progression. In agreement with the previous paper, we found that Hsa_circ_0072309 participates in the migration of lung cancer cells through the Wnt signal pathway, and Hsa_circ_0072309 may act as a miRNA-32-5p sponge to inhibit PCMTD1 expression. Hsa_circ_0072309 in non-smoking women with adenocarcinoma of the lungs has a specific underlying mechanism, however it is still unclear and requires more research.

We uncovered survival-related mRNAs in the ceRNA network in order to investigate survival-related circRNAs. With the help of LASSO Cox regression analysis, we finally identified six (BTBD3, CFL2, EIF4EBP2, EVI5, PCMTD1, SYDE2) survival-related genes. A survival-related ceRNA network was built according to the six genes. Therefore, we speculate that the circRNAs (hsa_circ_0008923, hsa_circ_0002346, hsa_circ_0002702, hsa_circ_0003045, hsa_circ_0072309, hsa_circ_0009043) in the ceRNA network had survival prediction value. We also created a six-gene predictive signature for non-smoking women with lung adenocarcinoma patients using the six genes. The OS is often shorter for the high-risk scores. After age, stage, T, and N correction, the risk scores remained an independent risk predictor for OS.

There were some limitations in this study. The survival-related ceRNA network was initially built using bioinformatics and required experimental validation. Second, the survival-related circRNA needs to be verified by large-scale survival analysis. Finally, the circRNA-related signal pathway and GO analysis need more data validation.

Conclusion

In conclusion, we identified differentially expressed circRNA, miRNA, and mRNA based on high-throughput sequencing and public data. We created a brand-new predictive risk score model by circRNA related mRNAs. This research may shed fresh light on the etiology of lung cancer in non-smoking women and identify possible treatment options.

Abbreviations

circRNAs, circular RNAs; LUAD, lung adenocarcinoma; GO, Gene Ontology; BP, biological processes; CC, cellular components; MF, molecular functions; ROC, receiver operating characteristic curve analysis; AUC, the area under the curve; OS, overall survival; IHC, immunohistochemistry.

Data Sharing Statement

The miRNA-seq data, mRNA-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).

Ethics Statement

The study was approved by The Human Investigation and Ethical Committees of Anhui provincial cancer hospital (Ethics Number-41/2020) and was conducted in accordance with the principles outlined in the Declaration of Helsinki. All participants provided written informed consent.

Acknowledgments

Thanks to Lijun Wang from Tongling People's Hospital for kindly guidance.

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.

Funding

This study was supported by Bengbu Medical College Science and Technology Project Fund (2023byzd004), Tongling Science and Technology Plan Project (20230203070).

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*. 2021;71(3):209–249. doi:10.3322/caac.21660
2. Bach PB. Smoking as a factor in causing lung cancer. *JAMA*. 2009;301(5):539–541. doi:10.1001/jama.2009.57
3. 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 Letter*. 2019;18(5):4605–4612. doi:10.3892/ol.2019.10845
4. Huang J, Yue N, Shi N, et al. Influencing factors of lung cancer in nonsmoking women: systematic review and meta-analysis. *J Public Health*. 2022;44(2):259–268. doi:10.1093/pubmed/fdaa254
5. Ko YH, Kim SJ, Kim WS, et al. Risk factors for primary lung cancer among never-smoking women in South Korea: a retrospective nationwide population-based cohort study. *Korean J Intern Med*. 2020;35(3):692–702. doi:10.3904/kjim.2019.283
6. Tsai LL, Chu NQ, Blessing WA, Moonsamy P, Colson YL. Lung cancer in women. *Ann Thorac Surg*. 2022;114(5):1965–1973. doi:10.1016/j.athoracsur.2021.09.060
7. Zhao Y, Dong Y, Zhao R, et al. Expression profiling of driver genes in female never-smokers with non-adenocarcinoma non-small-cell lung cancer in China. *Clin Lung Cancer*. 2020;21(5):e355–e362. doi:10.1016/j.cllc.2020.02.005
8. Seow WJ, Shu XO, Nicholson JK, et al. Association of untargeted urinary metabolomics and lung cancer risk among never-smoking women in China. *JAMA Netw Open*. 2019;2(9):e1911970. doi:10.1001/jamanetworkopen.2019.11970
9. Seow WJ, Matsuo K, Hsiung CA, et al. Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from Western populations. *Human Mol Genetics*. 2017;26(2):454–465. doi:10.1093/hmg/ddw414
10. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):7441):-333–338. doi:10.1038/nature11928
11. Qadir J, Li F, Yang BB. Circular RNAs modulate Hippo-YAP signaling: functional mechanisms in cancer. *Theranostics*. 2022;12(9):4269–4287. doi:10.7150/thno.71708
12. Wang L, Long H, Zheng Q, Bo X, Xiao X, Li B. Circular RNA circRHOT1 promotes hepatocellular carcinoma progression by initiation of NR2F6 expression. *Mol Cancer*. 2019;18(1):119. doi:10.1186/s12943-019-1046-7
13. Liu T, Yuan L, Zou X. Circular RNA circ-BNC2 (hsa_circ_0008732) inhibits the progression of ovarian cancer through microRNA-223-3p/FBXW7 axis. *Jovarian Res*. 2022;15(1):95. doi:10.1186/s13048-022-01025-w
14. Peng S, Yi L, Liao L, Bin Y, Qu W, Hu H. Circ_0008285 knockdown represses tumor development by miR-384/RRM2axis in hepatocellular carcinoma: running title: circ_0008285 knockdown inhibits hepatocellular carcinoma tumorigenesis. *Ann Hepatol*. 2022;27:100743. doi:10.1016/j.aohep.2022.100743
15. Jiang Y, Xu W, Tao J. Circular RNA circVPRBP serves as a microRNA-106b-5p sponge to regulate proliferation and metastasis of cervical cancer cells via tripartite motif-containing protein 3. *Anti-Cancer Drugs*. 2022;33:850–860. doi:10.1097/cad.0000000000001335
16. Li W, Wang H, Zheng Y. Circ_0001058 represses the progression of lung adenocarcinoma through governing of the miR-486-5p/TEK signaling axis. *Anti-Cancer Drugs*. 2022;33(8):710–719. doi:10.1097/cad.0000000000001337
17. Demicco EG, Boland GM, Brewer Savannah KJ, et al. Progressive loss of myogenic differentiation in leiomyosarcoma has prognostic value. *Histopathology*. 2015;66(5):627–638. doi:10.1111/his.12466
18. Tang J, Ou J, Xu C, et al. EVI5 is a novel independent prognostic predictor in hepatocellular carcinoma after radical hepatectomy. *Oncol Rep*. 2017;38(4):2251–2258. doi:10.3892/or.2017.5862
19. Wang H, Wang L, Sun G. MiRNA and potential prognostic value in non-smoking females with lung adenocarcinoma by high-throughput sequencing. *Int J Gene Med*. 2023;16:683–696. doi:10.2147/ijgm.S401544
20. Yu Z, Rong Z, Sheng J, et al. Aberrant non-coding RNA expressed in gastric cancer and its diagnostic value. *Front Oncol*. 2021;11:606764. doi:10.3389/fonc.2021.606764
21. Zhao Y, Jia Y, Wang J, et al. circNOX4 activates an inflammatory fibroblast niche to promote tumor growth and metastasis in NSCLC via FAP/IL-6 axis. *Mol Cancer*. 2024;23(1):47. doi:10.1186/s12943-024-01957-5
22. Wang XJ, Gao J, Yu Q, Zhang M, Hu WD. Multi-omics integration-based prioritisation of competing endogenous RNA regulation networks in small cell lung cancer: molecular characteristics and drug candidates. *Front Oncol*. 2022;12:904865. doi:10.3389/fonc.2022.904865
23. Tong Z, Wang Z, Jiang J, Tong C, Wu L. A novel molecular mechanism mediated by circ CCDC134 regulates non-small cell lung cancer progression. *Thoracic Cancer*. 2023;14:1958–1968. doi:10.1111/1759-7714.14942
24. Meng Q, Li Y, Sun Z, Yang X. CircRNA hsa_circ_0070659 predicts poor prognosis and promotes non-small cell lung cancer (NSCLC) progression via microRNA-377 (miR-377) / Ras-Associated Binding Protein 3C (RAB3C) pathway. *Bioengineered*. 2022;13(6):14578–14594. doi:10.1080/21655979.2022.2091572
25. Misir S, Hepokur C, Aliyazicioglu Y, Enguita FJ. Circular RNAs serve as miRNA sponges in breast cancer. *Breast Cancer*. 2020;27(6):1048–1057. doi:10.1007/s12282-020-01140-w

26. Meng S, Zhou H, Feng Z, et al. CircRNA: functions and properties of a novel potential biomarker for cancer. *Mol Cancer*. 2017;16(1):94. doi:10.1186/s12943-017-0663-2
27. Wang W, Lin Y, Zhang G, et al. circ_0002346 suppresses non-small-cell lung cancer progression depending on the regulation of the miR-582-3p/STXBP6 axis. *Int J Genomics*. 2021;2021:1565660. doi:10.1155/2021/1565660
28. Wang L, Liang Y, Mao Q, et al. Circular RNA circCRIM1 inhibits invasion and metastasis in lung adenocarcinoma through the microRNA (miR)-182/miR-93-leukemia inhibitory factor receptor pathway. *Cancer Sci*. 2019;110(9):2960–2972. doi:10.1111/cas.14131
29. Mizutani R, Imamachi N, Suzuki Y, et al. Oncofetal protein IGF2BP3 facilitates the activity of proto-oncogene protein eIF4E through the destabilization of EIF4E-BP2 mRNA. *Oncogene*. 2016;35(27):3495–3502. doi:10.1038/onc.2015.410
30. Yan L, Zheng M, Wang H. Circular RNA hsa_circ_0072309 inhibits proliferation and invasion of breast cancer cells via targeting miR-492. *Cancer Manage Res*. 2019;11:1033–1041. doi:10.2147/cmar.S186857
31. Chen T, Shao S, Li W, Liu Y, Cao Y. The circular RNA hsa-circ-0072309 plays anti-tumour roles by sponging miR-100 through the deactivation of PI3K/AKT and mTOR pathways in the renal carcinoma cell lines. *Artif Cells Nanomed Biotechnol*. 2019;47(1):3638–3648. doi:10.1080/21691401.2019.1657873
32. Zhang H, Xiao X, Wei W, et al. CircLIFR synergizes with MSH2 to attenuate chemoresistance via MutSα/ATM-p73 axis in bladder cancer. *Mol Cancer*. 2021;20(1):70. doi:10.1186/s12943-021-01360-4
33. Wang C, Liu WR, Tan S, et al. Characterization of distinct circular RNA signatures in solid tumors. *Mol Cancer*. 2022;21(1):63. doi:10.1186/s12943-022-01546-4
34. Cao C, Wang Y, Wu X, Li Z, Guo J, Sun W. The roles and mechanisms of circular RNAs related to mTOR in cancers. *J Clin Lab Analysis*. 2022;36(12):e24783. doi:10.1002/jcla.24783