

Bioinformatics Analysis of Differential Gene and MicroRNA Expression in Lung Adenocarcinoma: Genetic Effects on Patient Prognosis, as Indicated by the TCGA Database

Cancer Informatics
Volume 21: 1–10
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769351221082020



Bingqing Sun  and Hongwen Zhao

Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of China Medical University, Shenyang, China.

ABSTRACT

OBJECTIVE: To investigate the differential expression of genes and microRNAs (miRNAs) in patients with lung adenocarcinoma and the relationship between such changes and patient prognosis.

METHODS: We analyzed the expression levels of genes and miRNAs in lung adenocarcinoma tissues and adjacent normal tissues using The Cancer Genome Atlas database (TCGA). We analyzed the function of the differentially expressed genes and miRNAs in a co-expression network. Finally, we performed survival analysis of differential genes and miRNAs in the co-expression network using clinical data from the TCGA database.

RESULTS: We successfully identified 6064 differentially expressed genes: 5324 upregulated genes and 740 downregulated genes. And we identified 161 differentially expressed miRNAs: 126 upregulated miRNAs and 35 downregulated miRNAs. We identified several genes that were related to each other in the co-expression network. Further analysis revealed that the high expression levels of *G6PC*, *APOB*, *F2*, *PAQR9*, and *PAQR9-AS1* genes were associated with poor prognosis. However, there was no significant correlation between the expression of *hsa-mir-122* with regards to patient prognosis.

CONCLUSIONS: Our data showed that *hsa-mir-122* and a number of related genes may affect the prognosis of patients with lung adenocarcinoma by regulating the cytoskeleton, thus promoting tumor angiogenesis and the metastasis of tumor cells. The high expression levels of some differentially expressed genes was associated with the low survival rate in patients with lung adenocarcinoma. However, the levels of *hsa-mir-122* were not correlated with patient prognosis.

KEYWORDS: Lung adenocarcinoma, differentially expressed genes, differentially expressed microRNA, Kaplan-Meier analysis, biomarkers

RECEIVED: November 6, 2021. **ACCEPTED:** January 26, 2022.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the National Natural Science Foundation of China (Grant no. 81472194) and by the Project of Liaoning Distinguished Professor [Grant no. (2013) 204] to Hongwen Zhao.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Hongwen Zhao, Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of China Medical University, Shenyang, 110001, China. Email: hwzhao2007@163.com

Introduction

Of all forms of malignancy, lung cancer is associated with the highest rates of morbidity and mortality.¹ Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer patients. The most common form of non-small cell lung cancer is lung adenocarcinoma (LUAD); this particular form accounts for approximately 40% of all lung cancers.^{2–5} With the use of advanced screening equipment, and the improvement of techniques that can be used to prevent and diagnose NSCLC at an early stage, the prognosis of patients with this particular form of cancer is gradually improving.^{6,7} However, most patients tend to be diagnosed when they are in an advanced stage of the disease. Without the possibility of surgery, the only options for such patients are radiotherapy and chemotherapy. Tumor development is a complex process that involves a multitude of genes and different stages of development. Consequently, it is very important that we investigate oncogenes and miRNAs as these play critical roles in the pathogenesis, targeted therapy,

and prognostic evaluation, of lung cancer. The Cancer Genome Atlas (TCGA) is a publicly funded project and aims to provide a comprehensive cancer genome map, to allow researchers to explore cancer-related knowledge, to help us to understand the mechanisms underlying cancer, and to facilitate diagnosis or therapeutic interventions. Based on this database, we investigated the expression levels of key genes and related miRNAs in patients with lung adenocarcinoma to ascertain whether these molecules exerted effect on prognosis. Our ultimate aim was to identify new biomarkers for lung adenocarcinoma that may facilitate clinical diagnosis and treatment.

Methods

The Acquisition of Gene Transcriptome Samples

We used the TCGA database (<https://cancergenome.nih.gov/>) to acquire transcriptomic data (Level 3) from patients with lung adenocarcinoma. A number of different factors were used



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

to screen the database, as follows: disease type (adenomas and adenocarcinomas; project ID (TCGA-LUAD), workflow types (HTSeq-Counts), data type (transcriptome profiling), and experimental strategy (RNA-Seq).

The Acquisition of MicroRNA Sequences

Sequence data (Level 3) relating to miRNAs were acquired for patients with lung adenocarcinoma from the TCGA (<https://cancergenome.nih.gov/>). The screening conditions were as follows: disease type (adenomas and adenocarcinomas); Project ID (TCGA-LUAD), data type (miRNA expression quantification), and experimental strategy (miRNA-Seq).

Statistical Methods

We chose to use the edgeR package (in R3.6.2 software, URL: <https://www.R-project.org/>) to standardize and analyze the data acquired from the TCGA database (<https://cancergenome.nih.gov/>). We used strict screening conditions ($|\log_{2}FC| \geq 2$; corrected P value $< .05$) to identify differentially expressed genes and miRNAs. We use the gplots software package (version 3.0.1.1; <https://CRAN.R-project.org/package=gplots>) to create volcanic and heat maps. Next, we calculated suitable soft thresholds in the WGCNA package (version: 1.69; <https://CRAN.R-project.org/package=WGCNA>) in order to create a co-expression matrix for differentially expressed genes and miRNAs. Next, we created a co-expression network for the differentially expressed genes and miRNAs using Cytoscape 3.7.2 (<https://cytoscape.org/>) software. Functional enrichment analysis, was carried out using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses and data were presented using the ggplot2 package (Version: 3.3.2; <https://CRAN.R-project.org/package=ggplot2>). Next, we used TargetScan (<http://www.targetscan.org>; http://www.targetscan.org/vert_72/), miRDB (<http://mirdb.org/>), PicTar (<https://pictar.mdc-berlin.de/>), and miRanda (<http://www.microrna.org/microrna/getGeneForm.do>) to predict the target genes for the miRNA we chose to analyze. Meanwhile, we used a Venn Diagram to visualize the predictive results arising from the target genes of differentially expressed miRNAs within the co-expression network. Finally, we performed survival analysis of differentially expressed genes and miRNAs within the co-expression network using log-rank test methods and clinical data from patients. $P < .05$ was considered to indicate statistical significance.

Results

Genes Showing Differential Expression When Compared Between Lung Adenocarcinoma and Adjacent Normal Tissues

We downloaded, collated, and then analyzed, transcriptome data from the TCGA database. A total of 551 transcriptome

samples were identified, including 497 adenocarcinoma tissue samples and 54 para-cancerous normal tissue samples. According to our selection conditions ($|\log_{2}FC| \geq 2$ and a corrected P value $< .05$), we were able to use the edgeR package in R3.6.2 software to identify differentially expressed genes between cancerous and adjacent normal tissues. We identified 6064 differentially expressed genes, including 5324 upregulated genes and 740 downregulated genes (Figure 1).

miRNAs Showing Differential Expression When Compared Between Lung Adenocarcinoma and Adjacent Normal Tissues

We downloaded, collated, and then analyzed, miRNA data from the TCGA database. We acquired a total of 528 miRNA sequences, including 483 adenocarcinoma tissue samples and 45 para-cancerous normal tissue samples. According to our selection conditions ($|\log_{2}FC| \geq 2$ and a corrected P value $< .05$), we were able to use the edgeR package in R3.6.2 software to identify differentially expressed miRNAs between cancer and adjacent normal tissues. Finally, we identified 161 differentially expressed miRNAs, including 126 upregulated miRNAs and 35 downregulated miRNAs (Figure 2).

Construction of a Co-expression Network Relating to Differentially Expressed Genes and miRNAs

According to the levels of differentially expressed genes and miRNAs, we were able to use the WGCNA package and topological network analysis to set a soft threshold for a co-expression matrix between differentially expressed genes and miRNAs. It was important that the link between genes and miRNAs in the network diagram followed the distribution of a non-scaled network as this would mean that the co-expression network would be more meaningful in many biological aspects. To do this, we set a soft threshold of 18 with which to create a co-expression matrix between differentially expressed genes and miRNAs. Based on this co-expression matrix, the co-expression network between differentially expressed genes and miRNAs was created using Cytoscape 3.7.2 software (Figure 3). Connections between nodes in the network represent the presence of an association related to expression.

Figure 3 shows a range of differentially expressed genes, including *VTN*, *ALB*, *PAQR9*, *AFP*, *CREB3L3*, *PRODH2*, *AC080128.1*, *HRG*, *APOB*, *PAQR9-AS1*, *LGALS14*, and *G6PC*. Only *hsa-mir-122*, when upregulated, showed a significant correlation with these differentially expressed genes (Figure 3A). Previous studies have shown that *hsa-mir-122* is associated with the development of several different types of cancer, including breast cancer,⁸ renal cancer,⁹ colorectal cancer,^{10,11} and gastric cancer.¹² For instance, the overexpression of *hsa-mir-122* in patients with gastric cancer has been shown to inhibit the proliferation, metastasis, and invasion of gastric cancer cells, while the downregulation of *hsa-mir-122*

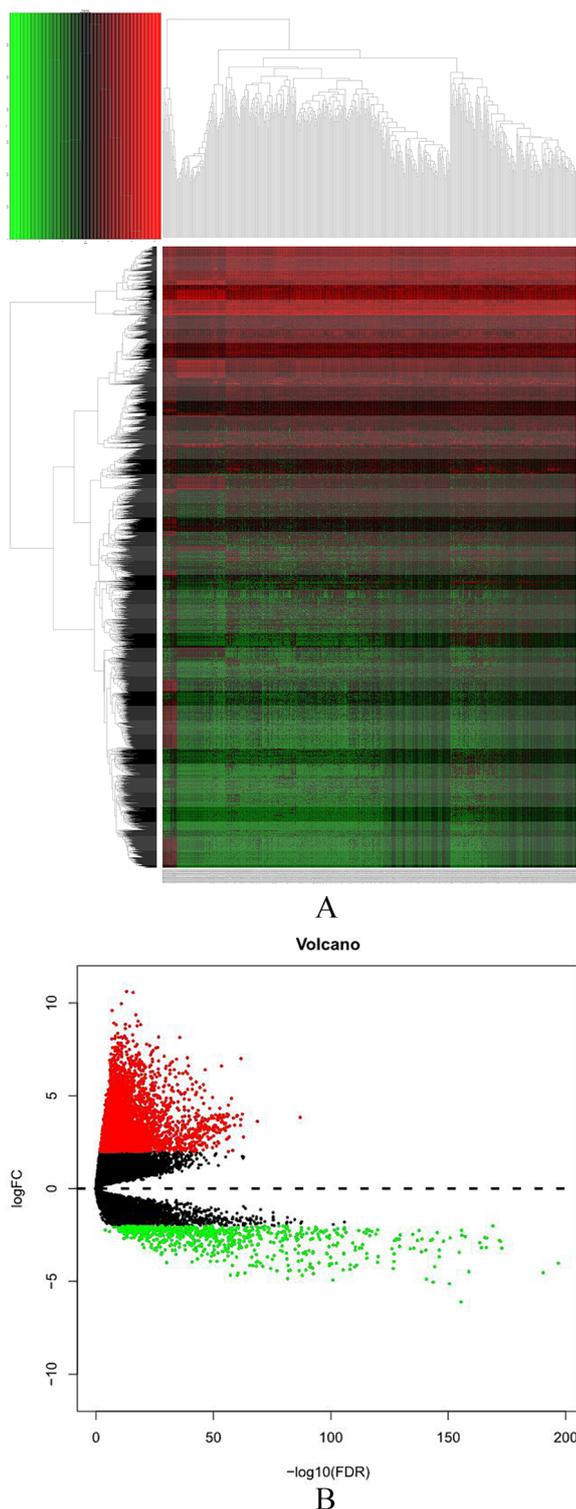


Figure 1. Gene expression in lung adenocarcinoma tissues and adjacent normal tissues: (A) heat map of differentially expressed genes between cancer tissues and adjacent normal tissues. The horizontal axis represents sample identifiers. While the vertical axis refers to gene names. Red represents high gene expression levels, while green indicates low expression levels and (B) volcanic map of gene expression levels in lung adenocarcinoma. The horizontal axis represents $-\log_{10}(\text{FDR})$, while the vertical coordinate represents $\log_{10}(\text{FC})$. The red points represent upregulated genes while the green points represent downregulated genes.

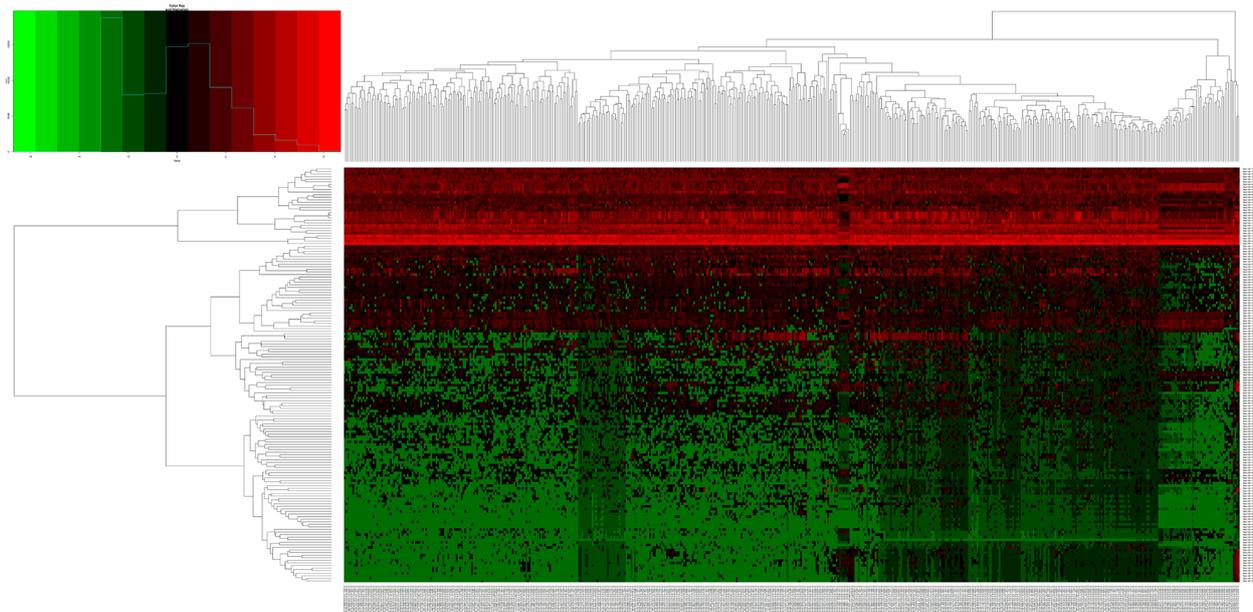
promotes the process of cell proliferation.¹² However, the over-expression of *hsa-mir-122* in patients with colon cancer has been shown to promote development, metastasis, and invasion of tumors.¹¹ In order to study the role of *hsa-mir-122* and its associated genes in the pathogenesis of lung adenocarcinoma further, we selected *hsa-mir-122* and several related genes (Figure 3A) in our co-expression network. This network was created using differentially expressed genes and miRNAs between lung adenocarcinoma and adjacent normal tissues.

Functional Enrichment Analysis of Differentially Expressed Genes Related to *hsa-mir-122*

We used the DAVID website (<https://david.ncifcrf.gov/>)^{13,14} to perform GO (gene ontology) enrichment analysis of genes in our co-expression network diagram of *hsa-mir-122* and its associated differentially expressed genes (Figure 3A). Of the differentially expressed genes shown in Figure 3A, *APOB*, *ALB*, and *VTN* were all significantly enriched in 1 particular biological process: receptor-mediated endocytosis (GO:0006898). We also found that *ALB*, *HRG*, and *VTN* were all significantly enriched in a certain cell component: blood cell microparticles (GO:0072562). *APOB*, *HRG*, and *VTN* genes were significantly enriched in 1 particular molecular function: binding heparin (GO:0008201) (Figure 4A). Next, we used the KOBAS website (kobas.cbi.pku.edu.cn/)¹⁵ to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.¹⁶ This showed that the pathway in which *CREB3L3*, *G6PC*, and *VTN* were most significantly enriched was the PI3K-Akt signaling pathway (Figure 4B).

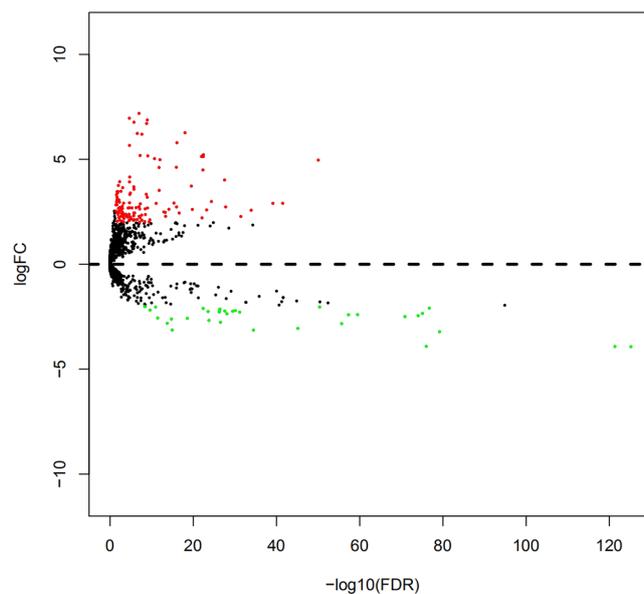
The Prediction and Functional Analysis of Target Genes for *hsa-mir-122*

We Used TargetScan,¹⁷ miRDB,^{18,19} PicTar,^{20,21} miRanda^{22,23} to predict the target genes for *hsa-mir-122*. The miRanda website predicted 1872 target genes, while the PicTar, TargetScan, and miRDB, websites predicted 115, 1129 and 516 target genes, respectively. Analyses showed that 27 target genes were consistently found on all 4 database: *ALDOA*, *CCNG1*, *GIT1*, *CS*, *DDR2*, *G3BP2*, *IQGAP1*, *STX16*, *DICER1*, *FOXP2*, *P4HA1*, *PAK3*, *MIPOL1*, *NPEPPS*, *OCLN*, *BACH2*, *DR1*, *SMYD4*, *EPO*, *MOSPD1*, *BRPF1*, *CPEB1*, *MASP1*, *MAF1*, *LRP10*, *LAMC1*, and *FUNDC2* genes. These 27 target genes were used for functional analysis to further explore the molecular function of *hsa-mir-122* in lung adenocarcinoma. Results showed that these target genes were significantly enriched in certain GO annotated entries. For biological process, there was significant enrichment for the negative regulation of transcription by RNA polymerase II (GO:0000122); this refers to any biological process that prevents or attenuates the frequency, rate, or extent of transcription mediated by RNA polymerase II. For cellular component, there was significant enrichment for the nucleus (GO:0005634). Chromosomes are



A

Volcano



B

Figure 2. miRNA expression in lung adenocarcinoma tissues and adjacent normal tissues: (A) heat map of the differentially expressed miRNAs between lung adenocarcinoma and adjacent normal tissues. The horizontal axis represents the sample identifiers, while the vertical axis represents the miRNAs. Red represents high miRNA expression levels while green represents low levels of expression and (B) volcanic map showing miRNA expression in lung adenocarcinoma. The horizontal axis represents $-\log_{10}(\text{FDR})$ while the vertical axis represents $\log_{2}(\text{FC})$. The red points represent upregulated miRNAs while the green points represent downregulated miRNAs.

preserved and copied in the membrane-bound organelles of eukaryotic cells. In most cells, the nucleus contains chromosomes other than organelle chromosomes, in which RNA is synthesized and processed. In some species or special types of cells, the process of RNA metabolism or DNA replication may not occur in the nucleus. For molecular function, there was significant enrichment for protein domain specific binding (GO:0019904). This refers to interactions with specific domains of proteins that occur in a selective and non-covalent

manner. Of these target genes, KEGG pathway analyses showed that the most significant enrichment for the *GIT1*, *IQGAP1*, and *PAK3* genes was the actin cytoskeleton pathway.

Prognostic Analysis of Survival in Patients With Lung Adenocarcinoma

We downloaded clinical data from patients with lung adenocarcinoma from the TCGA database; 486 samples were acquired.

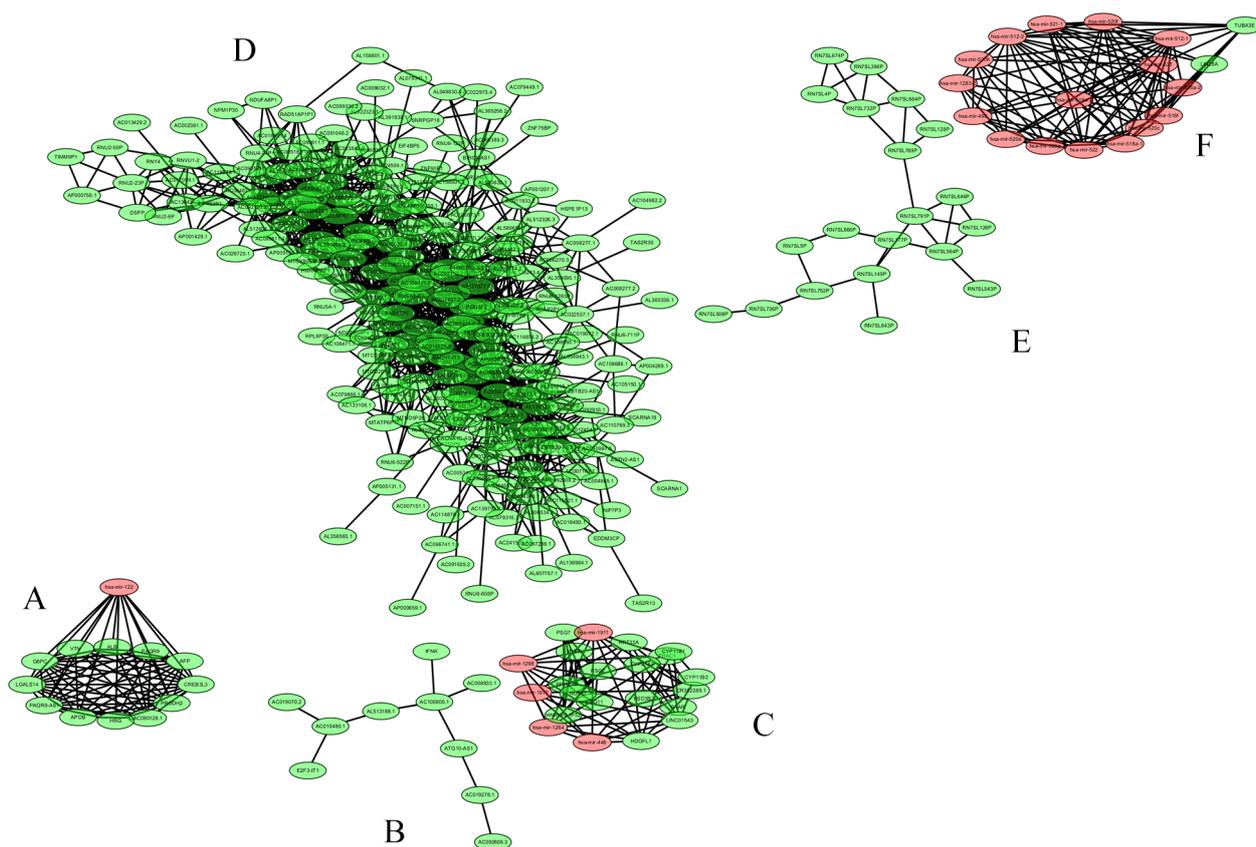


Figure 3. A co-expression network map of differentially expressed genes and miRNAs between lung adenocarcinoma and adjacent normal tissues. The green nodes represent differentially expressed genes while the red nodes represent miRNAs. The lines between the genes and miRNAs indicates a direct correlation. (A) A co-expression network diagram showing up-regulated *hsa-mir-122* and its associations with differentially expressed genes. The lines in the network indicate a quantitative correlation between the genes and miRNAs. All of the genes shown in Figure 3A were upregulated in lung adenocarcinoma tissue and correlated directly with high expression levels of *hsa-mir-122*. (B-F) The co-expression network diagrams showing differential microRNAs and genes between lung adenocarcinoma and adjacent normal tissues.

We carried out survival analysis for *hsa-mir-122*, and the differentially expressed genes that are associated with *hsa-mir-122* using the survival package in R3.6.2 software. We divided the patients into 2 different groups according to the median expression level. In other words, patients with expression levels that were higher than the median were placed into the high expression group, while those with expression levels that were lower than the median were placed into a low expression group. Comparative analysis of the survival status of the 2 groups showed that the expression levels of *G6PC*, *APOB*, *F2*, *PAQR9*, and *PAQR9-AS1* genes exerted significant effects on the survival prognosis of patients and that low expression levels are associated with a better prognosis ($P < .05$). However, there was no significant correlation between the expression of *hsa-mir-122* and the prognosis of patients (Figure 5).

Discussion

Lung cancer is considered to be the leading cause of cancer-related death.²⁴ Targeted therapy to driver gene mutations has improved the clinical outcome of some patients.²⁵ Over recent years, immunotherapy has become a new therapeutic method for patients without targeting mutated genes.²⁶ Despite the development of such treatments, the 5-year survival rate for

patients with lung adenocarcinoma remains unsatisfactory.²⁷ Most patients are in the median or advanced stage of disease when diagnosed and therefore the therapeutic effect is not as effective. Therefore, there is a significant need to diagnose patients during the early stages of disease as this would offer a better choice of treatment and is normally associated with a better outcome. In order to detect lung cancer earlier, it is necessary to investigate genes that are related to lung cancer and identify molecular biomarkers that can facilitate the routine diagnosis and treatment of lung cancer.

Research has shown that the dysregulation of genes and non-coding RNA molecules is associated with the development of many types of cancer, of which circRNAs and miRNAs are promising tumor biomarkers. miRNAs are non-coding RNAs of small molecules and can regulate the expression of genes and participate in some biological processes, such as transcription and endonuclease processes in the nucleus or cytoplasm.²⁸ Studies have shown that the abnormal expression of miRNAs is closely related to various types of cancers and plays a key role in cell proliferation, apoptosis, and metastasis. The regulation of gene and miRNA expression can explain many cancer-related molecular mechanisms, such as those involved in colorectal cancer, lymphatic cancer, and

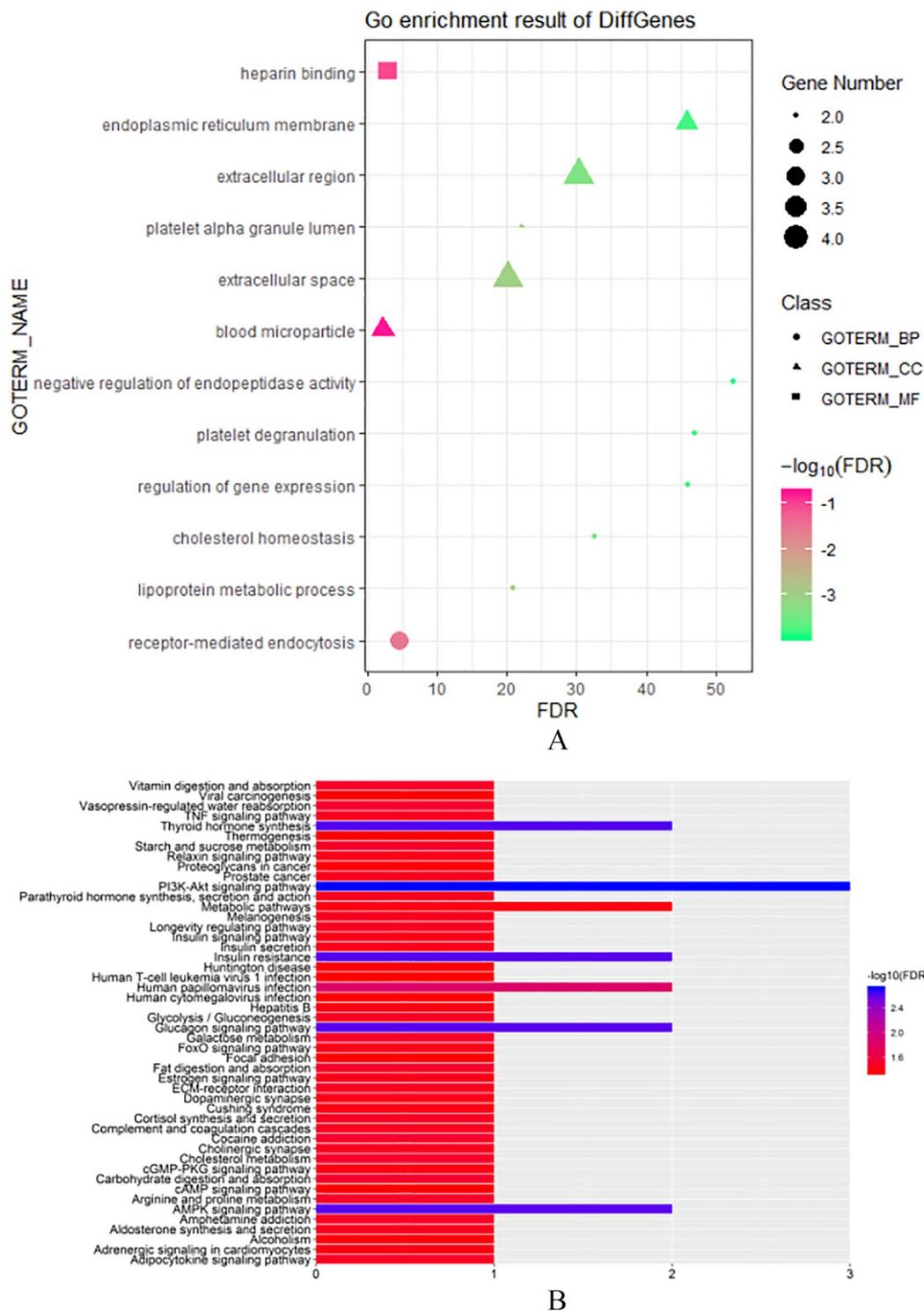


Figure 4. Enrichment analysis of differentially expressed genes related to *hsa-mir-122*: (A) data arising from gene ontology functional enrichment analysis of differentially expressed genes related to *hsa-mir-122*. The horizontal axis refers to FDR while the vertical axis refers to GOTERM_NAME. Different shapes represent biological processes, cellular components, and molecular functions, respectively. The circles refer to biological processes, the triangles refer to cellular components, while the squares refer to molecular functions. The size of the shapes represents the number of genes enriched on the gene ontology term. Coloration reflects $-\log_{10}(\text{FDR})$; the FDR value of red is less than green and (B) results arising from KEGG pathway analysis of differentially expressed genes related to *hsa-mir122*. The bar chart shows the results of KEGG pathway analysis. The vertical axis refers to the names of pathways that were enriched by differentially expressed genes associated with *hsa-mir-122*. The length of the bar represents the number of genes enriched in the pathway. Coloration reflects $-\log_{10}(\text{FDR})$; the FDR value of blue is less than red.

lung cancer.^{29,30} Some studies have predicted the survival prognosis of patients by constructing the expression profiles of miRNAs.³¹

In order to identify miRNAs, and their associated genes, in lung adenocarcinoma, and to investigate their roles in the pathogenesis of lung adenocarcinoma, we analyzed differences

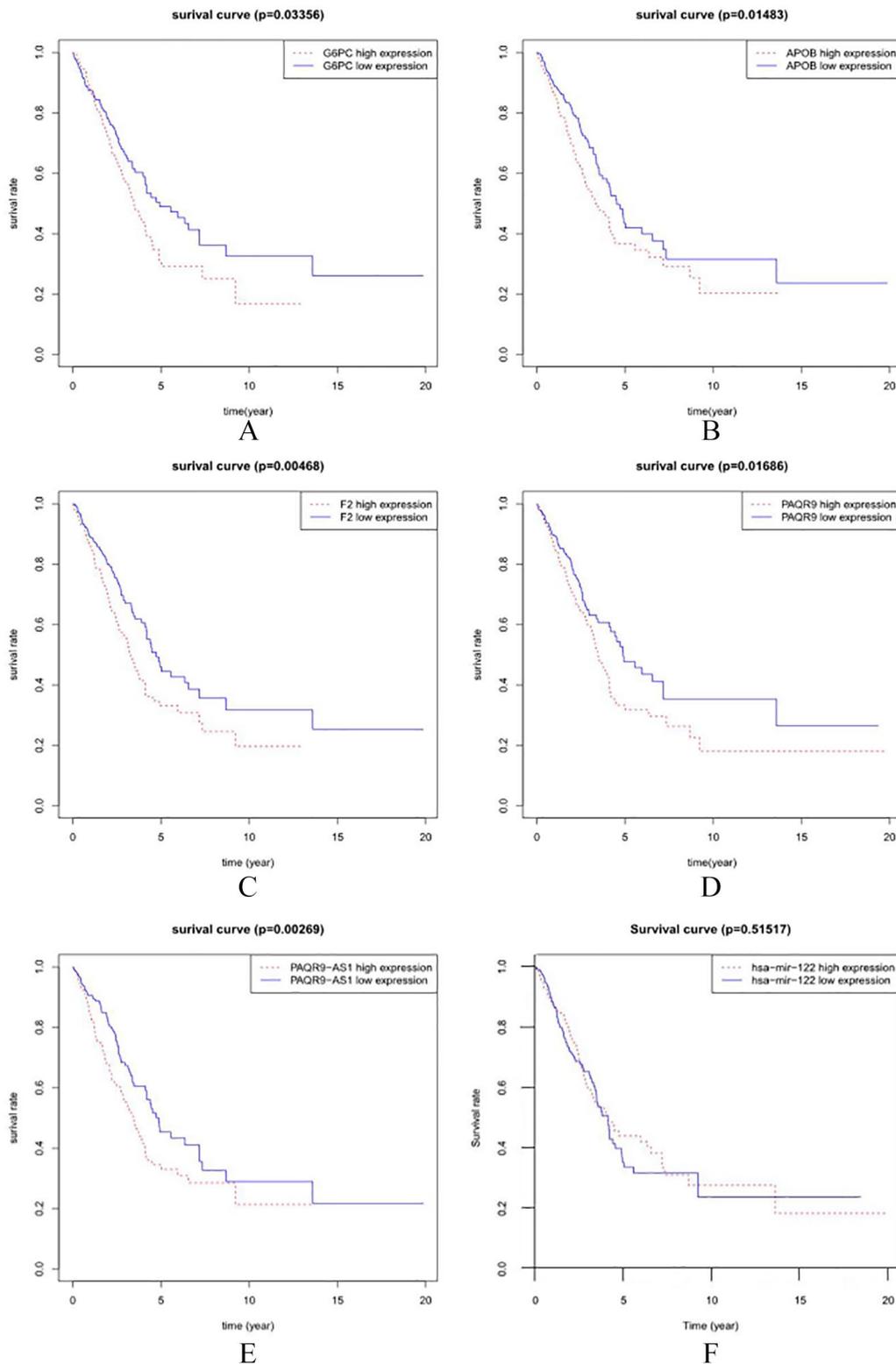


Figure 5. Kaplan-Meier survival curve for patients with lung adenocarcinoma. The horizontal axis represents the survival time (years) while the vertical axis represents survival rate (also known as cumulative survival probability or survival function).

in expression between lung cancer tissues and adjacent normal tissues by analyzing data from the TCGA database. We identified 6064 genes that were differentially expressed, including 5324 upregulated genes and 740 downregulated genes. We also identified 161 differentially expressed miRNAs, including 126

upregulated miRNAs and 35 downregulated miRNAs. Next, we mapped the co-expression network of differentially expressed genes and differential miRNAs with the WGCNA package and related software (Figure 3). We identified a number of differentially expressed genes that interacted with each other in

the network, including *VTN*, *ALB*, *PAQR9*, *AFP*, *CREB3L3*, *PRODH2*, *AC080128.1*, *HRG*, *APOB*, *PAQR9-AS1*, *LGALS14*, and *G6PC*. Only up-regulated levels of *hsa-mir-122* were correlated with these particular genes (Figure 3A). Previous studies have shown that *hsa-mir-122* is associated with the development of different types of cancer, including breast cancer,⁸ renal cancer,⁹ colorectal cancer,^{10,11} gastric cancer.¹² The overexpression of *hsa-mir-122* in gastric cancer patients has also been shown to inhibit the proliferation, metastasis, and invasion, of gastric cancer cells; its down-regulation promoted the proliferation of tumor cells.¹² However, the overexpression of *hsa-mir-122* in patients with colon cancer has been shown to promote the development, tumor metastasis and invasion.¹¹

In order to study the role of *hsa-mir-122* and its associated genes in the pathogenesis of lung adenocarcinoma further, we chose *hsa-mir-122* and genes that were associated with this miRNA and constructed a co-expression network using differentially expressed genes and miRNAs between lung adenocarcinoma and adjacent normal tissues (Figure 3A). We used the DAVID website^{13,14} to perform GO (gene ontology) enrichment analyses of genes within the co-expression network for *hsa-mir-122* and its associated differentially expressed genes (Figure 3A). Of the differentially expressed genes shown in Figure 3A, *APOB*, *ALB*, and *VTN* genes were found to be significantly enriched in terms of a specific biological process: receptor-mediated endocytosis (GO:0006898). This is a specific type of molecular transport that is mediated by cell surface receptors. Specific receptors on the cell surface can bind closely to extra-cellular macromolecules, such as ligands. This binding region contains a receptor-ligand complex for endocytosis; this creates a vesicle containing a receptor-ligand complex. This process usually occurs via clathrin pits and vesicles. *ALB*, *HRG*, and *VTN* genes were also significantly enriched in one particular cell component: blood cell microparticles (GO:0072562). This refers to particles in the blood that are derived from other types of cells, such as platelets, endothelial cells, or other types of cells. These particles express membrane receptors and other specific proteins from parental cells. These particles are also uneven in size and are devoid of nucleic acids. Studies have proposed that blood cell particles play a significant role in the pathophysiological mechanisms underlying inflammation and tumor metastasis.³² *APOB*, *HRG*, and *VTN* genes were significantly enriched in 1 specific molecular functional: binding heparin (GO:0008201). This refers to the selective and non-covalent binding to heparin. Heparin is a glycosaminoglycan and a key component of mast cells. Glycosaminoglycan is associated with malignant transformation in cells and tumor metastasis³³ (Figure 4A). The results of GO functional analysis for differentially expressed genes related to *hsa-mir-122* show that they are related to the development and metastasis of tumors.

We also used the KOBAS website (kobas.cbi.pku.edu.cn/)¹⁵ to perform Kyoto Encyclopedia of Genes and Genomes

pathway enrichment analysis.¹⁶ This technique showed that the pathway in which *CREB3L3*, *G6PC* and *VTN* genes are most significantly enriched was the PI3K-Akt signaling pathway (Figure 4B). The PI3K-Akt signaling pathway is one of the complex regulatory pathways found in human malignant tumors; this pathway is involved in the regulation of tumor development, including cell proliferation, genomic instability and metabolism.³⁴ This pathway can affect the migration of endothelial cells by regulating recombination of the actin cytoskeleton.³⁵ Because of the importance of this mechanism in tumors, this particular pathway has become a promising target for the therapy of cancer. A number of drugs that target this pathway are now involved in clinical trials.³⁶ Many factors affect the production of vascular endothelial growth factor (VEGF) through the PI3K-Akt signaling pathway to promote angiogenesis. This pathway plays an important role in tumor development.³⁷⁻⁴⁰ Therefore, *CREB3L3*, *G6PC*, and *VTN* genes play an important role in tumor angiogenesis and metastasis via the PI3K-Akt signaling pathway.

In order to investigate the role of *hsa-mir-122* in the pathogenesis of lung adenocarcinoma further, we used TargetScan,¹⁷ miRDB,^{18,19} PicTar^{20,21} and miRanda^{22,23} to predict the target genes of *hsa-mir-122*. These analyses identified 27 target genes that were associated with *hsa-mir-122*, including *ALDOA*, *CCNG1*, *GIT1*, *CS*, *DDR2*, *G3BP2*, *IQGAP1*, *STX16*, *DICER1*, *FOXP2*, *P4HA1*, *PAK3*, *MIPOL1*, *NPEPPS*, *OCLN*, *BACH2*, *DR1*, *SMYD4*, *EPO*, *MOSPD1*, *BRPF1*, *CPEB1*, *MASP1*, *MAF1*, *LRP10*, *LAMC1*, and *FUNDC2* genes. Functional analysis of these predicted target genes showed that they were significantly enriched in the following GO annotated entries: biological process (negative regulation of transcription by RNA polymerase II; GO:0000122; cellular component (nucleus; GO:0005634); molecular function (protein domain specific binding; GO:0019904). Of the predicted target genes, *GIT1*, *IQGAP1*, and *PAK3*, were significantly enriched in the actin cytoskeleton pathway. Cytoskeleton proteins are the driving force of the movement of tumor cells and also the core component of cellular pseudopodia. Their structural characteristics are the key in determining the migration ability of tumor cells.^{41,42} Different types of pseudopodia play different roles in cellular metastasis and invasion. Of these, the lamellar pseudopodia play an important role in cellular migration.⁴³ Invasive pseudopodia can help tumor cells degrade the extracellular matrix in order to access blood vessels and complete cellular invasion.^{44,45} The production and increased number of pseudopodia in tumor cells are closely related to the metastasis and invasion of tumor cells. We hypothesize that *hsa-mir-122* may interact with *GIT1*, *IQGAP1*, and *PAK3* in the nucleus and bind to specific domains of RNA polymerase II in a selective and non-covalent manner to inhibit RNA polymerase II-mediated transcription and to regulate cytoskeleton-related pathways. Therefore, *hsa-mir-122* plays an important role in the development and metastasis of lung

adenocarcinoma. Consequently, *has-mir-122*, and its related genes, should be investigated with regards to its role in regulating the cytoskeleton.

In order to investigate the effects of *has-mir-122* and its related genes on the prognosis of patients with lung adenocarcinoma, we acquired 486 samples from patients with lung adenocarcinoma patients from the TCGA database, including patient survival time and status. By combining the expression of *has-mir-122* and its related genes with patient survival data, we were able to analyze the prognosis of patients with lung adenocarcinoma using the survival package in R3.6.2 software. We found that the expression levels of *G6PC*, *APOB*, *F2*, *PAQR9*, and *PAQR9-AS1* were upregulated in lung adenocarcinoma and also related to the prognosis of patients. The survival rate of patients with lung adenocarcinoma who had lower expression levels of these was higher (Figure 5). *has-mir-122* was also highly expressed in lung adenocarcinoma. However, there was no significant correlation between the expression of this miRNA and prognosis. These findings suggest that *has-mir-122* and its associated genes may affect the prognosis of patients by regulating the cytoskeleton, angiogenesis, and the metastasis of tumors.

To conclude, this study identified a specific miRNA molecule (*has-mir-122*) that is associated with lung adenocarcinoma and several genes (*VTN*, *ALB*, *PAQR9*, *AFP*, *CREB3L3*, *PRODH2*, *AC080128.1*, *HRG*, *APOB*, *PAQR9-AS1*, *LGALS14*, *G6PC* genes). Of these, *CREB3L3*, *G6PC*, and *VTN* are related to the PI3K-Akt signaling pathway. We also found that *G6PC*, *APOB*, *F2*, *PAQR9*, and *PAQR9-AS1* were related to patient prognosis. Furthermore, we found that *GIT1*, *IQGAP1*, and *PAK3* were associated with the actin cytoskeleton pathway. This suggests that *has-mir-122* and its related genes may affect the prognosis of patients with lung adenocarcinoma by regulating the cytoskeleton, tumor angiogenesis, and tumor cell metastasis. High expression levels of *G6PC*, *APOB*, *F2*, *PAQR9*, and *PAQR9-AS1*, are associated with low survival rates in patients with lung adenocarcinoma although there was no correlation between the expression of *has-mir-122* and prognosis. This paper provides useful information relating to *has-mir-122* and its related genes in the occurrence and development of lung adenocarcinoma and proposes new ideas for the targeted therapy of lung cancer, and potentially other forms of tumors.

Acknowledgements

The authors would like to express their gratitude to EditSprings (<https://www.editsprings.com/>) for the expert linguistic services provided.

Author Contributions

Conceptualization: Bingqing Sun. Data curation: Bingqing Sun. Formal analysis: Bingqing Sun. Funding acquisition: Hongwen Zhao. Methodology: Bingqing Sun. Project administration: Hongwen Zhao. Supervision: Hongwen Zhao.

Writing – original draft: Bingqing Sun. Writing – review & editing: Bingqing Sun, Hongwen Zhao.

ORCID iD

Bingqing Sun  <https://orcid.org/0000-0002-2148-1080>

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394-424.
- El-Zein RA, Abdel-Rahman S, Santee KJ, Yu R, Shete S. Identification of small and non-small cell lung cancer markers in peripheral blood using cytokinesis-blocked micronucleus and spectral karyotyping assays. *Cytogenet Genome Res.* 2017;152:122-131.
- Colling R, Bancroft H, Langman G, Soilleux E. Fully automated real-time PCR for EGFR testing in non-small cell lung carcinoma. *Virchows Arch.* 2019;474:187-192.
- Maemura K, Watanabe K, Ando T, et al. Altered editing level of microRNAs is a potential biomarker in lung adenocarcinoma. *Cancer Sci.* 2018;109:3326-3335.
- Kleczyk EK, Kwak JW, Schenk EL, Nemenoff RA. Targeting the complement pathway as a therapeutic strategy in lung cancer. *Front Immunol.* 2019;10:954.
- Hasan N, Kumar R, Kavuru MS. Lung cancer screening beyond low-dose computed tomography: the role of novel biomarkers. *Lung.* 2014;192:639-648.
- Kassem K, Shapiro M, Gorenstein L, Patel K, Laird C. Evaluation of high-risk pulmonary nodules and pathologic correlation in patients enrolled in a low-dose computed tomography (LDCT) program. *J Thorac Dis.* 2019;11:1165-1169.
- Fong MY, Zhou W, Liu L, et al. Breast-cancer-secreted miR-122 reprograms glucose metabolism in premetastatic niche to promote metastasis. *Nat Cell Biol.* 2015;17:183-194.
- Nie W, Ni D, Ma X, et al. miR-122 promotes proliferation and invasion of clear cell renal cell carcinoma by suppressing Forkhead box O3. *Internet J Oncol.* 2019;54:559-571.
- Iino I, Kikuchi H, Miyazaki S, et al. Effect of miR-122 and its target gene cationic amino acid transporter 1 on colorectal liver metastasis. *Cancer Sci.* 2013;104:624-630.
- Li H, Zhang X, Jin Z, et al. MiR-122 promotes the development of colon cancer by targeting ALDOA in vitro. *Technol Cancer Res Treat.* 2019;18:15330338 19871300.
- Meng L, Chen Z, Jiang Z, et al. MiR-122-5p suppresses the proliferation, migration, and invasion of gastric cancer cells by targeting LYN. *Acta Biochim Biophys Sin.* 2020;52:49-57.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat Protoc.* 2009;4:44-57.
- Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37:1-13.
- Xie C, Mao X, Huang J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39:W316-W322.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27-30.
- Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife.* 2015;4:e05005.
- Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 2020;48:D127-D131.
- Liu W, Wang X. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol.* 2019;20:18.
- Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet.* 2005;37:495-500.
- Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet.* 2006;38:1452-1456.
- Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 2008;36:D149-D153.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. *PLoS Biol.* 2004;2:e363.
- Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Abate D. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study. *JAMA Oncol.* 2019;4:1749-1768.

25. Ettinger DS, Wood DE, Aisner DL, et al. Non-small cell lung cancer, version 5.2017: clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2017;15:504-535.
26. Bodor JN, Kasireddy V, Borghaei H. First-line therapies for metastatic lung adenocarcinoma without a driver mutation. *J Oncol Pract*. 2018;14:529-535.
27. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016;66:7-30.
28. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol*. 2014;15:509-524.
29. Metzler M, Wilda M, Busch K, Viehmann S, Borkhardt A. High expression of precursor microRNA-155/BIC RNA in children with Burkitt lymphoma. *Genes Chromosomes Cancer*. 2004;39:167-169.
30. He L, Thomson JM, Hemann MT, et al. A microRNA polycistron as a potential human oncogene. *Nature*. 2005;435:828-833.
31. Yerukala Sathipati S, Ho SY. Identifying the miRNA signature associated with survival time in patients with lung adenocarcinoma using miRNA expression profiles. *Sci Rep*. 2017;7:7507.
32. Simak J, Gelderman MP. Cell membrane microparticles in blood and blood products: potentially pathogenic agents and diagnostic markers. *Transfus Med Rev*. 2006;20:1-26.
33. Basappa Rangappa KS, Sugahara K. Roles of glycosaminoglycans and glycanmimetics in tumor progression and metastasis. *Glycoconj J*. 2014;31:461-467.
34. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-674.
35. Zhang LJ, Tao BB, Wang MJ, Jin HM, Zhu YC. PI3K p110 α isoform-dependent Rho GTPase Rac1 activation mediates H2S-promoted endothelial cell migration via actin cytoskeleton reorganization. *PLoS One*. 2012;7:e44590.
36. Fruman DA, Rommel C. PI3K and cancer: lessons, challenges and opportunities. *Nat Rev Drug Discov*. 2014;13:140-156.
37. Tsai JL, Lee YM, Pan CY, Lee AY. The novel VEGF121-VEGF165 fusion attenuates angiogenesis and drug resistance via targeting VEGFR2-HIF-1 α -VEGF165/Lon signaling through PI3K-AKT-mTOR pathway. *Curr Cancer Drug Targets*. 2016;16:275-286.
38. Takahashi T, Yamaguchi S, Chida K, Shibuya M. A single autophosphorylation site on KDR/Flk-1 is essential for VEGF-A-dependent activation of PLC-gamma and DNA synthesis in vascular endothelial cells. *EMBO J*. 2001;20:2768-2778.
39. Narunsky L, Oren R, Bochner F, Neeman M. Imaging aspects of the tumor stroma with therapeutic implications. *Pharmacol Ther*. 2014;141:192-208.
40. Sharma S, Guru SK, Manda S, et al. A marine sponge alkaloid derivative 4-chloro faspalyisin inhibits tumor growth and VEGF mediated angiogenesis by disrupting PI3K/Akt/mTOR signaling cascade. *Chem Biol Interact*. 2017;275:47-60.
41. Svitkina T. The actin cytoskeleton and actin-based motility. *Cold Spring Harb Perspect Biol*. 2018;10:a018267.
42. Rottner K, Faix J, Bogdan S, Linder S, Kerkhoff E. Actin assembly mechanisms at a glance. *J Cell Sci*. 2017;130:3427-3435.
43. Innocenti M. New insights into the formation and the function of lamellipodia and ruffles in mesenchymal cell migration. *Cell Adh Migr*. 2018;12:401-416.
44. Murphy DA, Courtneidge SA. The 'ins' and 'outs' of podosomes and invadopodia: characteristics, formation and function. *Nat Rev Mol Cell Biol*. 2011;12:413-426.
45. Eddy RJ, Weidmann MD, Sharma VP, Condeelis JS. Tumor cell invadopodia: invasive protrusions that orchestrate metastasis. *Trends Cell Biol*. 2017;27:595-607.