The role of miRNAs in lung adenocarcinoma (LUAD) and microarray data analysis to identify miRNA biomarkers and explore the molecular mechanisms in LUAD carcinogenesis

Doctoral (Ph.D.) Thesis

Yongan Song

Head of the Doctoral School: Prof. Dr. Lajos Bogár

Head of the Doctoral Program: Prof. Dr. István Kiss

Supervisors: Prof. Dr. István Kiss

Department of Public Health Medicine

Doctoral School of Clinical Medical Sciences

Medical School, University of Pecs, Pecs, Hungary

Pécs

2025

Introduction

Lung cancer has been a major public health problem and is the leading cause of cancer deaths in the past few decades worldwide. Lung cancer presents a huge challenge to modern oncology. It has been accounted for a significant proportion of cancer-related fatalities globally and classified as one of the most prevalent form of cancer; approximately 2.2 million new cases are recorded annually with 1.8 million fatalities. Histologically, lung cancer can be divided into small cell lung cancer (SCLC) and non–small cell lung cancer (NSCLC). The latter comprises more than 80% of lung cancer cases and is subdivided into adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma. Lung adenocarcinoma (LUAD) is the most common form of lung cancer, accounting for approximately 40% of all cases. Unfortunately, the prognosis for most LUAD cases remains poor, with about two-thirds of LUAD diagnoses being at an advanced, inoperable stage due primarily to challenges in early detection and the absence of effective treatments. Therefore, it is imperative to identify biomarkers and therapeutic targets related to LUAD in order to improve early detection and facilitate gene-targeted therapy for lung cancer.

MicroRNAs (miRNAs) are small, single-stranded, non-coding RNA molecules with an approximate length of 20-25 nucleotides. They can interfere with mRNA translation by forming base pairings with the 3′ untranslated region (UTR) of the target mRNA, causing mRNA degradation or translational repression. Though miRNAs represent just 2% of the human genome, they play a crucial role in determining cell fate. They regulate the expression of nearly one-third of genes across almost all biological processes, including embryonic development, cell proliferation, apoptosis, immune responses and tumorigenesis. There is increasing evidence suggesting their multifaceted regulatory effects, their role in a variety of cellular regulatory processes, and certain miRNAs function as oncogenes or tumor suppressors. Several miRNAs have demonstrated considerable clinical potential in oncology, especially those that can bind and regulate multiple mRNAs, making them promising oncology-related biomarkers.

Long noncoding RNAs (lncRNAs) are noncoding RNAs with length of 200 to 10 000 base pairs (bp). LncRNAs are not translated into proteins, but rather regulate the expression of target genes at transcriptional and posttranscriptional level. However, the exact role of lncRNA in tumors still needs further research. According to the competitive endogenous RNA (ceRNA) hypothesis, lncRNAs can have a sponge effect on miRNAs and weaken the influence of miRNAs on mRNAs. In addition, studies have shown that the networks of lncRNAs, miRNAs, and mRNAs play an important role in the pathogenesis and progression of cancer. Nevertheless, the study of large-scale samples in LUAD is not common. Therefore, screening of miRNAs that are relevant in LUAD and building a ceRNA network with miRNAs as the core is very important for the early diagnosis and treatment of LUAD patients.

In this PhD thesis, we first review the comprehensive roles played by miRNAs in LUAD. MiRNAs play a crucial role in gene regulation, and their involvement in cancer has been extensively explored. While some reviews have been published on miRNAs and cancer, there remains a gap in the literature regarding miRNAs specifically in LUAD. Our review fills this gap by providing a comprehensive summary of the current understanding of miRNAs implicated in LUAD progression. We not only highlight the potential diagnostic, prognostic, and therapeutic implications of miRNAs in LUAD, but also present an inclusive overview of the extensive research conducted on miRNAs in this particular context. We believe that our review will facilitate researchers in comprehending the advancements made in the study of miRNAs in lung adenocarcinoma. Additionally, we have compiled an appendix encapsulating nearly all miRNAs pertinent to the development and treatment of lung adenocarcinoma, facilitating an easy reference and further investigation of the correlation between specific miRNAs and LUAD. Subsequently, in this PhD thesis, we have analyzed the expression of genes (lncRNAs, miRNAs, and mRNAs) in LUAD gene expression profiles taken from the GEO database. In addition, we have established a ceRNA network in LUAD through bioinformatics methods in order to find new potential targets for cancer therapy. The use of high-throughput microarrays for expression profiling has become a widely used technology; it can be used to measure the expression of

thousands of genes at once and to identify new cancer biomarkers. In this study, we reported a comprehensive analysis of miRNA, mRNA, and lncRNA expression by reanalyzing the public data sets from GSE135918, GSE136043 and GSE130779. Compared to control, differentially expressed miRNAs (DEMis), mRNAs (DEMs), and lncRNAs (DELs) were identified in LUAD samples. We tried to predict the interactions between DEMis and DEMs, and then performed functional enrichment analysis to construct miRNA-gene regulatory networks and ceRNA networks. Through comprehensive bioinformatics analysis, we expected to find new therapeutic targets and biomarkers for LUAD.

The role of miRNAs in lung adenocarcinoma

MiRNAs are small non-coding RNAs that were first identified in 1993 during studies of *Caenorhabditis elegans*. It was quickly recognized that these seemingly conserved miRNA sequences play a crucial role in regulatory pathways in eukaryotes. MiRNAs can interfere with mRNA translation through complementary base pairing with the 3' UTR of target mRNAs, leading to either mRNA degradation or translational repression. Around 70% of miRNAs are transcribed from specialized miRNA loci, while the remaining fraction of miRNAs is processed from introns of protein-coding genes. In most cases, miRNA genes are transcribed in the nucleus by RNA polymerase II (Pol II). This leads to the formation of primary miRNA (pri-miRNA) that undergoes capping, splicing, and polyadenylation. One pri-miRNA can generate either a single miRNA or a cluster containing two or more miRNAs. These long pri-miRNAs require cleavage by a microprocessor complex, primarily composed of the RNase III enzyme DROSHA and the double-stranded RNA (dsRNA) binding protein DiGeorge syndrome critical region 8 (DGCR8). The microprocessor cleaves one strand of the dsRNA at the base of the stem-loop secondary structure within the pri-miRNA, releasing a hairpin-shaped precursor miRNA (pre-miRNA) of approximately 60-70 nucleotides.

While the core components of the microprocessor, DROSHA and DGCR8, are essential for the biogenesis of almost all miRNAs in cells, there are several cofactors that also play a role in this process. The pre-miRNA is exported from the nucleus to the cytoplasm by the exporter protein 5 (XPO5) and subsequently processed by DICER1, a RNase III enzyme that cleaves the pre-miRNA at both its 5' and 3' ends. The pre-miRNA is cleaved into miRNA duplexes, and one of the strands is selected as the mature miRNA, which is loaded into the RNA-induced silencing complex (RISC) to act as a negative regulator of gene expression. The other strand is eventually degraded (Figure 1). The degradation or translational inhibition of target mRNAs by miRNAs largely depends on the complementarity between the miRNA's 5'-seed sequence and the mRNA's 3'-UTR element. Moreover, defects in the miRNA biogenesis machinery may contribute to tumorigenesis. Multiple lines of evidence have demonstrated that

miRNAs have diverse cellular regulatory roles, with some miRNAs being recognized as oncogenes or tumor suppressor genes. Numerous studies have shown that human cancers, including LUAD, exhibit a large number of dysregulated miRNAs. Therefore, these miRNAs may serve as potential diagnostic or prognostic markers and could even guide therapeutic interventions. MiRNAs can be utilized to subclassify tumors, as miRNA expression profiles serve as powerful indicators of pathological parameters and reliable biomarkers in LUAD.

Numerous public databases of miRNAs have been established, which accumulate data on various aspects of thousands of annotated human miRNAs, including an increasing number of miRNAs associated with LUAD. There is a growing interest in identifying and characterizing miRNAs from different types of body fluids due to the ease of access to these samples. MiRNAs can also be obtained from tissue samples. Particularly, miRNAs derived from formalin-fixed and paraffin-embedded (FFPE) samples exhibit greater resistance to degradation compared to mRNAs. Consequently, these samples stored in hospitals have a significant advantage for miRNA research. On the other side, while miRNAs are exciting new target for treatment of cancer, adaptation of miRNAs for therapy presents a challenge because of the lack of specificity. One miRNA typically targets a cluster of genes, so manipulating its expression can bringabout undesired consequences.

MiRNAs have emerged as a promising tool for cancer therapy due to their ability to regulate multiple biological pathways. Dysregulation of numerous miRNAs has been observed in various types of cancer, and even subtle changes in their levels can significantly impact disease outcomes. In the context of LUAD, miRNAs act as potent inhibitors of gene expression, effectively interfering with cancer cell growth and survival. Furthermore, miRNAs exhibit greater stability in serum, plasma, and FFPE-preserved samples compared to mRNAs, making them ideal noninvasive biomarkers for monitoring disease progression and classifying cancer subtypes. Increasing evidence suggests that miRNAs have the potential to combat chemotherapy-induced drug resistance. However, it is important to consider that miRNA-based therapies may entail unpredictable side effects, as each miRNA can target hundreds of mRNAs, even

beyond the intended specific target mRNA. In accordance with that only a handful of miRNAs have entered clinical trials and almost all have been abandoned before they reached phase 3. Also, in preclinical trials, some miRNAs have been found to be both, tumor suppressors and oncogenes, demonstrating the importance of expression landscape into which they are introduced. Identifying downstream elements to which miRNAs involved in cancer converge and targeting them with more specifically designed siRNAs might be a more promising strategy. Moreover, miRNAs derived from exosomes hold great promise in the diagnosis, prognosis, and treatment of LUAD. Further research on exosomal miRNAs in LUAD is expected to enhance our understanding of this previously overlooked miRNA fraction.

To effectively translate these foundational research findings into clinical practice, a comprehensive understanding of miRNA biology is crucial. Researchers have focused on identifying miRNA signatures that may offer new insights into longstanding questions. However, ensuring the safety and efficacy of miRNA-based therapies necessitates targeted delivery to tumor sites, efficient uptake by cancer cells, and minimizing off-target effects. Establishing standardized methods for miRNA detection, improving our understanding of the interactions between miRNAs and other genomic elements, and developing biocompatible delivery vehicles for miRNAs to target lung lesions are paramount. There are still obstacles to overcome on the path to translating miRNA research into clinical practice. However, with persistent efforts and emerging research findings, we can overcome these challenges and pave the way for a new era of comprehensive miRNA application in LUAD in the near future.

Aims of the study

More and more studies have shown that miRNAs are involved in gene regulation. Many miRNAs have been demonstrated to affect the occurrence and development of lung cancer. The purpose of this study is to use bioinformatics analysis to identify new miRNAs associated with LUAD and to explore certain molecular mechanisms regarding their effects.

The use of high-throughput microarrays for expression profiling has become a widely used technology; it can be used to measure the expression of thousands of genes at once and to identify new cancer biomarkers. In our study, we reported a comprehensive analysis of miRNA, mRNA, and lncRNA expression by reanalyzing the public data sets from GSE135918, GSE136043 and GSE130779. Compared to control, DEMis, DEMs, and DELs were identified in LUAD samples. We tried to predict the interactions between DEMis and DEMs, and then performed functional enrichment analysis to construct miRNA-gene regulatory networks. Through comprehensive bioinformatics analysis, we expected to find new therapeutic targets and biomarkers for LUAD.

Methods

1 Microarray Data

Gene expression data sets were obtained from the GEO repository (https://www.ncbi.nlm.nih.gov/geo/), which is a public database containing gene expression data from high-throughput experiments. The expression record sets of human miRNAs and mRNAs, containing expression profiles of lung cancer tissue and adjacent tissue of the same 5 lung adenocarcinoma patients, were obtained from NCBI GEO (GSE135918, GSE136043). The lncRNA microarray data of 8 lung adenocarcinoma patients were downloaded from NCBI GEO (GSE130779), which also includes record of lung cancer tissues and adjacent tissues. Data analysis from the GEO database does not require ethics committee approval.

2 Data preprocessing and screening of differential expression

The GEOquery package of the R platform was used to download the miRNA, mRNA, and lncRNA data, and the data was imported into the R statistical environment. The limma package was used (156) for data preprocessing, including extraction of expression matrix, clinical information, platform annotation files and deletion of missing data. For miRNA data, the results of nonhuman miRNA probes were removed, and the data were $\log_2$ converted. The lncRNA data was effectively processed using normalization and $\log_2$ conversion. The platform annotation information was obtained from the GEO database, and the chip probe IDs were converted into gene symbols. The expression matrix is divided into a cancer tissue group and an adjacent nontumor group. The limma package was used to calculate the p-value of the difference in gene expression between the cancer tissue group and the adjacent nontumor group, and the differentially expressed miRNAs (DEMis), mRNAs (DEMs), and lncRNAs (DELs) were selected in turn. According to p-value < 0.05 and $\log_2$ fold change (FC) > 1, DEMis,

DEMs, and DELs were filtered out. In order to visualize DEMis, DEMs, and DELs, the FactoMineR, factoextra, ggplot2, ggplotify, and pheatmap packages in the R platform were used to draw principal component analysis (PCA) maps, volcano maps, and heat maps.

3 Functional enrichment analysis

We used the clusterProfiler package in the R platform to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis on DEMs data. The results of the functional enrichment analysis were used to analyze the changes in molecular biological functions of lung cancer tissues.

4 Prediction of target lncRNAs and mRNAs of DEMis

The interaction between lncRNAs and miRNAs is predicted by LncBase Predicted v.2 of DIANA Tools. By setting the threshold of interaction score to 0.8, the predicted lncRNAs of lncRNA-miRNA pairs were further filtered (the score ranges from 0 to 1), and the information of DELs-DEMis pairs was obtained. Next, the targeted mRNAs of DEMis were retrieved from MiRTarBase and Targetscan. Both miRNA reference databases are reliable. Through the prediction results of these two databases, we then got the information of the DEMis-DEMs pairs.

5 DEMis Validation

The TCGA database provides a wealth of clinical information and miRNA expression quantification from a huge sample size. Therefore, the data set "TCGA-LUAD" was used to screen DEMis obtained from GSE135918. We obtained the file containing all TCGA-LUAD clinical information and miRNA expression quantification

from the database (https://portal.gdc.cancer.gov/). In the R platform, the DESeq2 package was used to calculate the p-value and $\log_2$ fold change value of the gene expression difference between the tumor group and the normal group to verify DEMis results obtained in the GEO database.

6 Construction of ceRNA network

The DELs-DEMis-DEMs network was reconstructed by integrating the prediction results of DEMis' target lncRNAs and mRNAs. The Cytoscape software was used to visualize the DELs-DEMis-DEMs network.

7 The survival analysis of DEMis

In order to explore whether the DEMs we selected are related to overall survival of the patients, the LUAD data in TCGA were used for analysis. Another two R packages, survival and survminer, were used to calculate the overall survival analysis for DEMs. We separated the patients into high and low groups and stratified the miRNA expression levels of cancer patients through the surv_categorize function in the R platform. In addition, the chart showed the hazard ratio with the 95% confidence interval (CI) and the p-value.

Results

1 Identification of DEMis

We extracted miRNA data from GSE135918 for analysis. The expression levels of miRNA in lung cancer tissues and their adjacent tissues of 5 lung adenocarcinoma patients were studied. The cutoff for log FC of miRNA was 1, and the cutoff for p-value was 0.05. A total of 272 upregulated miRNAs and 353 downregulated miRNAs were identified by removing the missing and duplicate data in the GSE135918. Supplemental Digital Content (http://links.lww.com/MD/H189) is a complete file containing all miRNAs information.

2 Identification of DEMs

The mRNA data were extracted from GSE136043, and the mRNA expression levels in lung cancer tissues and their adjacent tissues of 5 lung adenocarcinoma patients were analyzed. In order to explore the interaction between miRNA and mRNA more accurately, mRNA and miRNA data were from the same 5 patient samples. Like miRNA, the cutoff for log FC of mRNA was 1, and the cutoff for p-value was 0.05. We deleted the missing and duplicate values in the GSE135918 data. The analysis results showed that 1659 mRNAs were upregulated and 1476 mRNAs were downregulated. We show the complete DEMs file in Supplemental Digital Content (http://links.lww.com/MD/H190).

3 Identification of DELs

The lncRNA data came from GSE130779. This study investigated the expression levels of lncRNA in the lung cancer tissues and their adjacent tissues of 8 lung adenocarcinoma patients. According to the screening criteria of miRNA and mRNA,

the cutoff for log FC of lncRNA was 1, and the cutoff for p-value was 0.05. We also deleted missing and duplicate values in the data. The limma package in the R platform was used to normalize and analyze the expression levels of all lncRNAs. The results showed that 4054 lncRNAs were upregulated and 5543 lncRNAs were downregulated. We provided the complete DELs file in Supplemental Digital Content (http://links.lww.com/MD/H191).

4 Functional enrichment analysis of DEMs

The ClusterProfiler package in the R platform was used to perform KEGG analysis on DEMs, elucidating the possible mechanisms involved in the development of lung adenocarcinoma. The threshold of p-value was set to 0.05, yielding 70 KEGG signal pathways. A complete file is provided in Supplemental Digital Content (http://links.lww.com/MD/H192) and Supplemental Digital Content (http://links.lww.com/MD/H193), including their GeneRatio, p-value, geneID and count. Among them, Wnt, PI3K-Akt, and Notch signaling pathways were implicated in the development of lung cancer.

5 MiRNA screening

The 50 upregulated and downregulated miRNAs with the most significant differential expression were selected and verified with the miRNA data in the TCGA database. All clinical information on miRNA expression quantification of lung adenocarcinoma (TCGA-LUAD) was downloaded from https://portal.gdc.cancer.gov/. The results showed that there were 567 LUAD samples, including 46 normal samples and 521 tumor samples. After comparison of TCGA data to our top 50 downregulated and top 50 upregulated miRNAs, we verified the downregulation of 8 DEMis (hsa-miR-101-3p, hsa-miR-195-5p, hsa-miR-30a-3p, hsa-miR-451a, hsa-miR-144-3p, hsa-miR-15b-5p, hsa-miR-193a-3p, hsa-miR-145-5p) and the upregulation of 6 DEMis

(hsa-miR-665, hsa-miR-369-3p, hsa-miR-224-3p, hsa-miR-381, hsa-miR-3944-3p, hsa-miR-3652). Because these 14 miRNAs were differentially expressed in both GEO and TCGA databases, they had a high degree of credibility. In order to analyze whether the 14 differentially expressed miRNAs were related to the occurrence of lung cancer, we selected mRNAs related to Wnt signaling pathway, PI3K-Akt signaling pathway, and Notch signaling pathway from the target mRNA predicted by miRNA to explore the interactions between miRNA and these three signaling pathways.

6 Construction of ceRNA network

Targetscan and MiRBase were used to predict the target genes of the 14 miRNAs we identified. We selected the key mRNAs in the aforementioned three signaling pathways from these target genes. These 14 miRNAs and these key mRNAs were used to construct a miRNA-mRNA regulatory network (Figure 1A, 1B ,1C). Then LncBase Predicted v.2 of DIANA Tools was used to predict the interaction between lncRNAs and miRNAs. The DELs interacting with these 14 miRNAs were selected, and the lncRNA-miRNA regulatory network was constructed (Figure 1D).

7 Survival analysis

We used the TCGA database and R package to evaluate the DEMis we selected. Our results showed that low expression level of hsa-miR-101, hsa-miR-195, hsa-miR-30a, hsa-miR-145 ($p < 0.05$) was associated with poor OS in LUAD. In contrast, high expression level of hsa-miR-381, hsa-miR-3944 ($p < 0.05$) was also associated with poor OS in LUAD. Then we used Kaplan-Meier Plotter (http://kmplot.com/) to perform an OS analysis for hsa-miR-101, hsa-miR-195, hsa-miR-30a, hsa-miR-145, hsa-miR-381, hsa-miR-3944. The same result is displayed in the Kaplan-Meier Plotter.

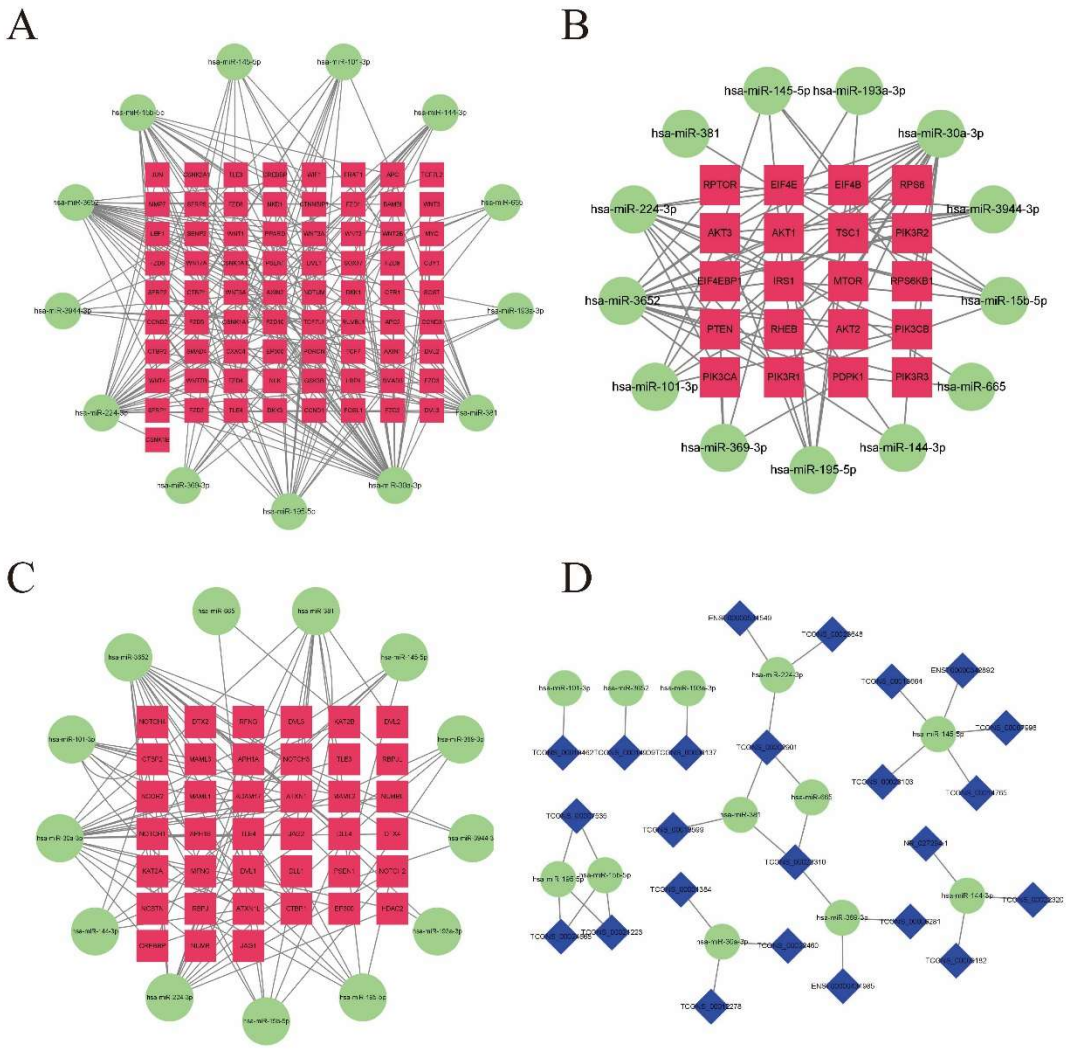

Figure 1. **CeRNA regulatory network in LUAD.**

Discussion

We screened the miRNA and mRNA microarray data of 5 lung adenocarcinoma patients' tumor tissues and adjacent normal tissues from the GEO database. The miRNA and mRNA data were screened from the same patients to better analyze and predict the relationship between miRNA and lung cancer. We evaluated the changes in miRNA and mRNA expression in LUAD by integrating miRNA and mRNA expression profiles. A total of 625 DEMis (272 upregulated miRNAs and 353 downregulated miRNAs) and 3135 DEMs (1659 upregulated mRNAs and 1476 downregulated mRNAs) were identified. These mRNAs and miRNAs significantly participate in 70 signaling pathways. Among them, Wnt signaling pathway, PI3K-Akt signaling pathway, and Notch signaling pathway are implicated in the occurrence of lung cancer. We wanted to evaluate these pathways in the selected cohort of 5 lung cancer patients. Then we listed the top 50 upregulated and top 50 downregulated miRNAs and validated them with the TCGA-LUAD database. A total of 14 miRNAs were found to also be differentially expressed in the TCGA-LUAD database, including 8 downregulated miRNAs and 6 upregulated miRNAs. These 14 miRNAs may play a role in the pathogenesis of LUAD and have the potential to become biomarkers of LUAD.

Next, we used Targetscan and MiRBase to predict the target genes of these 14 miRNAs. We chose three relevant signaling pathways in cancer development (Wnt, PI3K-Akt, and Notch), and scanned for proteins involved in them that are potential targets of the aforementioned miRNAs.

Among the 8 downregulated miRNAs, the target genes of hsa-miR-30a-3p are abundant in the Wnt signaling pathway, including *WNT*, *FZD*, *DVL*, *LEF*, *CCND*, *PLC* and other family-related mRNAs, whose high expression can activate the Wnt signaling pathway to promote the occurrence of lung adenocarcinoma. The low expression of hsa-miR-30a-3p may lead to increased expression of these mRNAs, thereby promoting the activation of the Wnt signaling pathway. In addition, *FZD4* and *DVL1* are the target genes of hsa-miR-144-3p. *WNT2B*, *WNT3A*, *WNT4*, *WNT7A*, *FZD4*, *FZD6*, *CCND1*, *CCND2* and *CCND3* are the target genes of hsa-miR-195-5p and hsa-miR-15b-5p.

WNT2B, *FZD4* and *CCND2* are the target genes of hsa-miR-145-5p. These genes can positively regulate the Wnt signaling pathway. Next, we analyzed the upregulated miRNAs. *SOST*, *SOX17*, and *NLK* are the target genes of hsa-miR-665. *NKD1*, *GSK3B*, and *TLE4* are the target genes of hsa-mir-369. *DKK1*, *SFRP1*, *SFRP2*, *CXXC4*, *GSK3B*, *CTBP1*, *CTBP2* are the target genes of hsa-mir-224. *SOST*, *DKK3*, *SFRP2*, *CXXC4*, *CTBP2*, and *NLK* are the target genes of hsa-mir-381. *SFRP5*, *NOTUM*, *NKD1*, *AXIN1*, *APC2*, *CTNNBIP1*, *SOX17*, *CTBP1* are the target genes of hsa-mir-3652. These mRNAs can negatively regulate the Wnt signaling pathway. The high expression of these miRNAs may lead to a decrease in the expression of these target genes and activate the Wnt signaling pathway.

We also analyzed the relationship between miRNAs (8 downregulated miRNAs and 6 upregulated miRNAs) and the PI3K-Akt signaling pathway. Hsa-miR-30a-3p can target *PI3K*, *AKT*, *MTOR*, and other important genes in the PI3K-Akt signaling pathway. Among the 8 downregulated miRNAs, hsa-miR-30a-3p has the most target genes in the PI3K-Akt signaling pathway. *PIK3CB*, *AKT3*, and *MTOR* are the target genes of hsa-miR-101-3p. *IRS1* and *PIK3CB* are the target genes of hsa-miR-144-3p. *IRS1*, *PIK3R1*, and *AKT3* are the target genes of hsa-miR-15b-5p and hsa-miR-195-5p. *IRS1* and *AKT3* are the target genes of hsa-miR-145-5p. These genes can activate the PI3K-Akt signaling pathway, and the low expression of these miRNAs may lead to the upregulation of these genes. Among the 6 upregulated miRNAs, we analyzed several key tumor suppressor genes. *PTEN* is the target gene of hsa-miR-224-3p and hsa-miR-369-3p.*TP53*, *MAGI*, and *PTEN* are the target genes of hsa-miR-3652. *MAGI* and *TSC1* are the target genes of hsa-miR-3944.The upregulation of these miRNAs may result in the loss of *TP53*, *MAGI*, *TSC1*, and *PTEN* expression.

Finally, we analyzed the relationship between miRNA and the Notch signaling pathway. *DLL1*, *DLL4*, *NOTCH2*, and *RBPJ* are the target genes of hsa-miR-15b-5p and hsa-miR-195-5p. The low expression of these two miRNAs may lead to the activation of the Notch signaling pathway.

These 14 miRNAs may play a role in the pathogenesis of LUAD and have the potential to serve as biomarkers for this malignancy. Among the 8 downregulated

miRNAs, all have been reported to exhibit reduced expression in cancer or function as tumor suppressors. Specifically, hsa-miR-101-3p, hsa-miR-195-5p, hsa-miR-30a-3p, hsa-miR-144-3p, hsa-miR-193a-3p, and hsa-miR-145-5p have been identified as tumor suppressors in lung cancer. Our study further corroborated the role of these miRNAs in lung adenocarcinoma and proposed new and reliable molecular mechanisms for their action. Hsa-miR-30a was particularly noteworthy, as its low expression level was highly associated with poor overall survival (OS) in LUAD ($p < 0.001$). Hsa-miR-30a-3p may inhibit the occurrence and progression of lung cancer through the Wnt and Akt signaling pathways, making it a promising biomarker for LUAD. Although hsa-miR-451a and hsa-miR-15b-5p have been reported to play inhibitory roles in tumors, there is limited research on these two miRNAs in the context of lung adenocarcinoma. They have the potential to become novel tumor markers for this disease. Among the 6 upregulated miRNAs, hsa-miR-665, hsa-miR-369-3p, hsa-miR-224-3p, and hsa-miR-381 have been reported to exhibit elevated expression in cancer. Hsa-miR-665, hsa-miR-224-3p, and hsa-miR-381 have been found to act as oncogenes in lung cancer. Our study further confirmed the role of these miRNAs in lung adenocarcinoma and explored additional potential molecular mechanisms. Although hsa-miR-369-3p has been identified as an oncogene, there is a lack of research on its role in lung adenocarcinoma. Its close association with the Wnt signaling pathway suggests it could become a key gene of interest in this field. Furthermore, our results are the first to report high expression levels of hsa-miR-3944 and hsa-miR-3652 in LUAD, and we predicted their target genes. Notably, the high expression level of hsa-miR-3944 is associated with poor OS. Currently, there is no research on hsa-miR-3944 in the cancer field, indicating its potential as a novel biomarker for lung adenocarcinoma. This research provides a valuable reference for screening miRNA-related biomarkers for the diagnosis and treatment of lung cancer.

New findings of my PhD thesis

1 Our review addresses a significant gap in the literature concerning the role of miRNAs in lung adenocarcinoma, offering a comprehensive synthesis of current research on miRNAs implicated in LUAD progression.

2 We have identified 14 miRNAs associated with lung adenocarcinoma and have delved into their molecular mechanisms underlying LUAD carcinogenesis. We also found that 6 of these miRNAs were associated with poor OS in LUAD.

3 Among the 14 miRNAs we identified, 9 have been experimentally validated to play a role in lung cancer pathogenesis. Our study not only corroborates these findings but also elucidates novel molecular mechanisms underlying their involvement in lung cancer.

4 Three of the 14 miRNAs we identified have previously been reported to be associated with cancer. However, they have been rarely studied in lung adenocarcinoma. We analyzed these three miRNAs as potential biomarkers for LUAD and investigated their possible molecular mechanisms.

5 Our results reported for the first time that hsa-miR-3944 and hsa-miR-3652 were highly expressed in LUAD. And the high expression level of hsa-miR-3944 was associated with poor OS. Hsa-miR-3944 and hsa-miR-3652 may serve as new biomarkers in LUAD.

6 The miRNAs uncovered in our study hold transformative potential for the landscape of lung cancer treatment, offering insights that could catalyze the development of novel therapies and early-stage disease biomarkers.

List of publications

My doctoral thesis is based on the following publications:

1 **Song Y,** Kelava L, Kiss I. MiRNAs in Lung Adenocarcinoma: Role, Diagnosis, Prognosis, and Therapy. Int J Mol Sci. 2023 Aug 27;24(17):13302. doi: 10.3390/ijms241713302. PMID: 37686110. (IF=5.6, Q1)

2 **Song Y,** Kelava L, Zhang L, Kiss I. Microarray data analysis to identify miRNA biomarkers and construct the lncRNA-miRNA-mRNA network in lung adenocarcinoma. Medicine (Baltimore). 2022 Sep 9;101(36):e30393. doi: 10.1097/MD.0000000000030393. PMID: 36086747. (IF=1.817, Q3)

Other Publications:

Zhang L, Pozsgai É, **Song Y**, Macharia J, Alfatafta H, Zheng J, Li Z, Liu H, Kiss I. The relationship between single nucleotide polymorphisms and skin cancer susceptibility: A systematic review and network meta-analysis. Front Oncol. 2023 Feb 15;13:1094309. doi: 10.3389/fonc.2023.1094309. PMID: 36874118. (IF=4.7, Q2)

Acknowledgment

As time swiftly passes, the five-year journey of pursuing a PhD degree is on the brink of conclusion, appearing as a mere fleeting moment. Throughout this transformative period, I have encountered numerous experiences and garnered invaluable knowledge. I wish to extend my heartfelt gratitude to my esteemed PhD mentor, Professor István Kiss. It was five years ago when Professor István Kiss graciously offered me the esteemed opportunity to embark on the pursuit of a PhD degree, thereby guiding me through the profound journey of doctoral training. Professor István Kiss epitomizes both warmth and reverence as a mentor. Moreover, he is recognized as a fervent pursuer of scientific inquiry, particularly in the realms of fostering, challenging, and refining one's intellectual faculties. Over the course of my PhD studies, Professor István Kiss's erudition and unwavering commitment to scientific rigor have enriched my learning experience immensely. Throughout my doctoral training, he has consistently championed my ideas and provided invaluable guidance. Beyond academia, Professor István Kiss has imparted invaluable lessons on navigating life's intricacies and charting a course for personal growth. As I reflect on this juncture in my life, I consider myself profoundly fortunate to have been mentored by Professor István Kiss.

I would like to express my gratitude to my colleagues in our department: Professor Gyöngyi Zoltán, Professor Varga Csaba, Professor Kiss (Orsós) Zsuzsanna, Ms. Csörge Timea, Dr. Zhang Lu and other colleagues. I am thankful for their assistance and the pleasant memories we shared during laboratory work. Their patient support throughout my five years in Hungary has been invaluable. Whether in studying, working, or navigating life, their enthusiasm has consistently instilled in me confidence and a love for life in a foreign country.

Finally, I would like to express my gratitude to my family: to my parents, who instilled in me values of humanity, honor, hard work, and humility; to my dear sister, brothers, and my closest friends for accepting me just as I am.