Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI: 10.26479/2021.0705.03**

IDENTIFICATION OF POTENTIAL MIRNA BIOMARKERS FOR NON-SMALL CELL LUNG CANCER DIAGNOSIS USING MICROARRAY DATASETS AND BIOINFORMATICS METHODS

Imteyaz Ahmad Khan[1], Raziuddin Khan[2], Imteyaz Ahmad[3], Mohammad Ahmad Ansari[4], Vinita Kumar Jaggi[5], Srikant Sharma[1]*

1. Department of Biotechnology, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India.
2. Department of Paramedical Sciences, Al-Falah University, Faridabad, Haryana, India
3. Department of Biochemistry, Kurukshetra University, Kurukshetra, Haryana, India.
4. University College of Medical Sciences (UCMS) & GTB Hospital, University of Delhi, India.
5. Department of Surgical Oncology, Delhi State Cancer Institute, Dilshad Garden, Delhi, India.

ABSTRACT: Non-small cell lung cancer (NSCLC) accounts for approximately 80% of lung cancers and is the leading cause of cancer deaths worldwide. Although recent advances in treatment have improved, overall 5-year survival rates have not improved significantly. Therefore, early diagnosis is still essential for patient survival. Circulating miRNAs might act as noninvasive blood-based biomarkers for NSCLC diagnosis and prognosis. Using Gene Expression Omnibus (GEO) data, we identified 12 miRNAs that are down-regulated in the tumor tissue and blood of NSCLC patients. This study identified three miRNAs that could serve as biomarkers in the diagnosis of NSCLC: miR-140-3p, miR-29c, and miR-199a. Functional enrichment analysis of the miRNA's target transcript identified several overrepresented pathways related to cancer progression. Seven target genes were identified as hub genes of the protein-protein interaction (PPI) network and hold significant prognostic power. A combination of three genes (IL6, SNAI1, and CDK6) has a hazard ratio greater than 1(hr=1.5) with a p-value less than 0.002. Since the expression levels of these three miRNAs were significantly decreased in both tissue and blood, detecting the expression level of miRNA in the blood provides information on its expression in tissue as well. Therefore, these miRNAs may be useful as diagnostic and prognostic biomarkers for NSCLC.

Keywords: Non-small cell lung cancer, miRNA, biomarkers, Gene Expression Omnibus, protein-protein interaction network.

Peerreviewunder responsibilityofLife Science Informatics Publications

Article History: Received: Sept 12, 2021; Revised: Sept 22, 2021; Accepted: Oct 16, 2021.

Corresponding Author: Dr. Srikant Sharma* Ph.D.
Department of Biotechnology, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India.
Email Address:shribioinfo@gmail.com

1. INTRODUCTION

Lung cancer is the leading cause of cancer deaths in the past few decades and has become one of the most serious malignant tumors in the world[1]. The incidence rate of lung cancer is higher in men than in women, and the death rate is nine times higher in men as compared to women[2,3]. Based on histology, lung cancers are divided into two major groups: non-small cell lung cancer (NSCLC), which accounts for at least 80% of lung malignancy cases and small cell lung cancer (SCLC)[4]. Furthermore, NSCLC is divided into three subtypes: adenocarcinoma (LAC), squamous cell carcinoma (LSCC), and large cell carcinoma (LCLC)[4]. Although the diagnostic practices and treatments for NSCLC have improved, the mortality rate is still high. The 5-year survival rate of patients with NSCLC is only 21%[5]. The lethality of NSCLC is often attributed to a lack of early diagnosis, metastasis, and the occurrence of drug resistance. Currently, NSCLC is generally diagnosed at an advanced stage when cancer has already metastasized. Therefore, it is essential to elucidate the molecular mechanisms of NSCLC pathogenesis and identify early diagnostic or predictive biomarkers. Currently, the identification of diagnostic and prognostic blood-based biomarkers for NSCLC is of great interest. Several biomarkers have been investigated in diverse mediums: cerebrospinal fluid (CSF), tissue, urine, and saliva. Blood samples (including plasma and serum), are attractive sources of cancer biomarkers, due to the non-invasive nature of blood sapling. Different types of biologically relevant biomarkers are available in the blood, such as microRNAs (miRNA), circulating tumor DNA (ctDNA), and metabolites[6]. MicroRNA s (miRNAs) represent a class of endogenous, highly conserved small non-coding RNAs, 20-24 nucleotides in length, which specifically bind to the 3′ UTR of mRNA targets to inhibit post-transcriptional gene regulation[7,8]. Mature miRNAs are loaded into RNA-induced silencing complex (miRISC), are complementary to the 3'untranslated region of the target gene to cleave the target mRNA and a single miRNA typically regulates hundreds of genes [9,10]. According to the study, over 50% of the known miRNAs are located on the genome at a tumor-associated fragile site and are associated with cancer cell progression, differentiation and apoptosis[11]. Therefore, miRNAs that are secreted by malignant cells can be used as non-invasive biomarkers for different

Peerreviewunder responsibilityofLife Science Informatics Publications

stages and different types of cancer. In a previous study, several differentially expressed miRNAs (DEmiRNAs) have been identified in the patients of NSCLC blood and tissue samples[12–14]. MicroRNAs (miRNAs) are stable in blood, and their unique expression profiles can serve as non-invasive biomarkers for the early detection of cancer[15,16]. Blood-based miRNAs can be profiled using different techniques, including next-generation sequencing (NGS) and microarray profiling. Previously, microarray-based miRNA expression detection has been used to select biomarkers for different types of cancer, such as pancreatic cancer, breast cancer, colon cancer, and lung cancer [16–18]. Many studies have identified specific differentially expressed miRNAs in NSCLC blood that could serve as biomarkers. A study conducted by Xue et al. found that serum miR-1228-3p and miR-181a-5p showed a promising result for the early detection of lung cancer, indicating that these miRNAs might be used as non-invasive biomarkers for lung cancer[19]. Studies have demonstrated significant interactions between gene alterations and tumorigenesis and cancer progression in many types of tumors[20]. Notch3 and CD44 expression were high in NSCLC patients and involved in the progression and migration of NSCLC[21,22]. One study showed the genes SPAG6 (Sperm Associated Antigen 6) and L1TD1 (LINE-1 Type Transposase Domain Containing 1) are tumor-specifically methylated in NSCLC [23]. A recent study conducted by Morris et al. showed that the expression levels of FPR1 gene in blood samples predict both NSCLC and small cell lung cancer[24]. In a recent study, miR-30d has been shown to be a tumor suppressor in the progression of NSCLC[25]. A study conducted by Yang et al. showed that miR-598 suppressed the proliferation, invasion and migration in NSCLC and functions as a tumor suppressor[26]. However, the precise roles of miRNAs and genes in NSCLCs are still not properly understood[27]. High-throughput, parallel gene expression analysis through microarray has become a widely used technology to obtain more global views on oncogenes and to identify novel diagnostic cancer biomarkers [28]. In this study, 3 blood miRNA profiling datasets (GSE137140, GSE93300, and GSE94536) and one tissue miRNA profiling dataset (GSE53882) were analyzed to identify differentially expressed miRNAs (DEmiRNAs) in both tissue and blood of NSCLC patients. Target prediction analysis was performed to identify the target genes of the identified deregulated miRNAs by integrating all three public online databases (TargetScan, PicTar, and miRanda) and the resulting genes were analyzed in Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes). Furthermore, the protein-protein interaction (PPI) network of DEGs was constructed and investigated to understand the molecular mechanisms of the targeted genes and role in the progression of NSCLC. Survival analysis was conducted to evaluate the potential effects of the hub genes on NSCLC prognosis. Finally, in order to further validate the potential of the identified miRNAs as candidate biomarkers of NSCLC, we performed survival analysis and receiver operating characteristic (ROC) analysis.

Peerreviewunder responsibilityofLife Science Informatics Publications

2. MATERIALS AND METHODS

Dataset Selection

The Gene Expression Omnibus (GEO) at http://www.ncbi.nlm.nih.gov/geo/ is a publicly accessible online database that contains high-throughput functional genomic data from various studies[29,30]. GEO datasets were searched for the keywords blood (serum or plasma), NSCLC, Human, and miRNA profiling via the NCBI website. Three datasets, GSE137140, GSE93300, and GSE94536 were selected for analysis (Table 1) [31–33]. The GSE137140 dataset contained serum miRNA profiling data consisting of 1566 NSCLC samples and 1774 control samples, while the other two datasets (GSE93300 and GSE94536) contained plasma miRNA profiling data. In addition, we selected tissue miRNA data from the GSE53882 dataset, consisting of 397 NSCLC tissue and 151 control samples[34]. All four studies used the microarrays gene expression technique to collect data. Microarrays technique that utilizes oligonucleotide probes to detect nucleic acids captured by array probes.

Table 1: Summary of four datasets employed in this study.

Dataset ID	Platform	Number of NSCLC samples	Number of control samples	References
GSE137140	GPL21263	1566	1774	[31]
GSE93300	GPL21576	9	4	[32]
GSE94536	GPL21576	6	3	[33]
GSE53882	GPL18130	397	151	[34]
Total	NA	1978	1932	

Identification of Differentially Expressed miRNAs

All the datasets were screened for DEMiRNAs in R Studio using the programming language R (R version 3.0, www.r-project.org) and the Limma software package in the Bioconductor Package (http://www.bioconductor.org/)[20]. The statistical significance of the fold change was calculated for each miRNA by Student's t-test. The normal distribution of gene expression values was verified using the online tool GEO2R, provided by the Gene Expression Omnibus. On the basis of differential expression analysis, the DEMiRNAs were separated based on upregulation or downregulation. The cutoffs for differentially expressed miRNAs were set as absolute fold-change >2 and a P-value of less than 0.05. Venn diagrams of the differentially expressed miRNAs in each dataset were constructed via Jvenn software [21]. Finally, the Log fold change (FC) data from the four datasets was retrieved for each DEmiRNA using Python (http://www.python.org.org) and GraphPad Prism 4.0 (GraphPad Software, San Diego, CA) was used to construct heatmap.

Target Gene Prediction and Functional Enrichment Analysis

Candidate target genes of the DEmiRNAs were predicted by using three online databases: miRDB, mirTarBase and MiRWalk[35]. The target genes identified by these three databases were used for GO (Gene Ontology) analysis[36]. The list of genes was analyzed with PANTHER using default options and calculates which biological entities those genes are overrepresented in[37]. The target gene list was submitted to PANTHER for enrichment analysis of the significantly overrepresented GO biological processes and molecular function terms. Fisher's Exact test was used to determine statistical significance ($p<0.05$) and the statistical correction for false positives was completed using the false discovery rate (FDR) procedure. Significance level $p < 0.05$ was regarded as a statistical significance cut-off for overrepresentation.

Network Analysis

The candidate genes identified in this study was investigated in the search tool for the retrieval of interacting Proteins (STRING, https://string-db .org), to construct a network of the protein-protein interactions (PPI networks)[38]. STRING is a comprehensive database and user-friendly bioinformatics tool that scans several protein databases to create a network visualization of input genes and how their protein products interact. The confidence level of PPI was set to the highest (0.9). Then, Cytoscape (http://www.cytoscape.org/) was used to create and visualize the PPI network graphs. The Cytoscape was used to screen hub DEGs with the node degree and clustering coefficient. Genes with a degree of ≥ 20 were considered hub genes. In addition, the MCODE plug-in was used within Cytoscape to filter important modules in the PPI network with a degree cutoff $\geq$ 2, node score cutoff = 0:2, K − core $\geq$ 2, and max: depth = 100 as the cutoff criteria. For functional enrichment analysis on individual modules, the STRING plugin was used.

Analysis of Hub Genes

Kaplan Meier survival curves were constructed to visualize NSCLC survival curves for lung adenocarcinoma and squamous cell carcinoma patients[40]. Kaplan-Meier plotter and an in silico online tool was used to combines gene expression data from the well-established The Cancer Genome Atlas (TCGA), the Gene Expression Omnibus (GEO), and the European Genome-Phenome Archive. Differentially expressed hub genes were identified using Cytoscape and submitted to the Kaplan-Meier tool and the survival curves were used to analyze the survival of patients. Hazard ratio (HR) with 95% confidence intervals and log-rank P-value was calculated for each gene. GEPIA (Gene Expression Profiling Interactive Analysis) tool (http://gepia.cancer-pku.cn/) was used for the analysis of hub genes[41].

MiRNA Biomarker Selection

Survival estimates were generated for the differentially expressed miRNAs using Kaplan-Meier survival curves. Patient survival was determined using the median split of expression scores and

overall survival metrics. MiRNAs with statistically significant (log-rank p-value) prognostic potential were further analyzed using miR-TV, an interactive miRNA target viewer for miRNA and target gene expression interrogation for human cancer studies [42]. Plots of miRNA expression in lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) were generated after which the log2 transformation was applied. TCGA Wanderer database (http://maplab.imppc.org/wanderer/) was used to analyze the differential methylation of candidate miRNAs and to further analyzed the impact of methylation in NSCLC gene expression[43]. GraphPad Prism 6.0 (GraphPad Software, La Jolla, CA, USA) was used to create the ROC curve (receiver operating characteristic curve).

3. RESULTS AND DISCUSSION

Identification of differentially expressed miRNAs

GSE137140, GSE93300, GSE94536 and GSE53882 were selected for analysis (Table 2). Additional quantile normalization was performed for GSE93300 datasets. The GSE94536 dataset showed no statistically significant over-expressed miRNAs. The down-regulated differentially expressed miRNA (DEmiRNAs) from each of the four datasets were inputted into the web-based program VENNY 2.0 (http://bioinfogp.cnb.csic.es/tools/venny/) to generate a 4-way Venn diagram and determine overlaps (Figure 1). 13 DEmiRNAs were common to all four datasets and 54 DEmiRNAs were found in three out of four datasets. The 13 overlapping DEmiRNAs were selected for further analysis and their logFC values were retrieved from each dataset (Figure 2).

Table 2: Number of up-regulated and down-regulated DEmiRNAs in each dataset.

Dataset ID	Number of up-regulatedmiRNAs (logFC>0)	Number of down-regulated miRNAs (logFC<0)
GSE137140	280	2200
GSE93300	2161	58
GSE94536	0	242
GSE53882	550	564

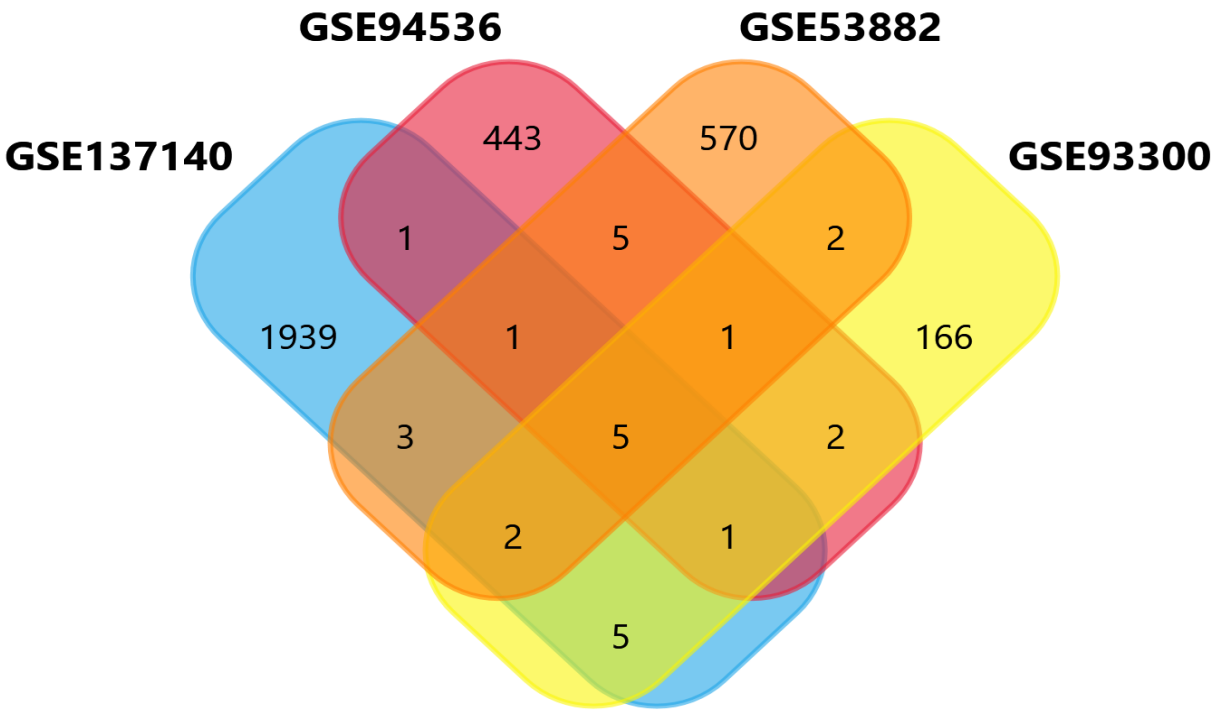

Figure 1: 4-way Venn diagram showing overlap DEmiRNAs.

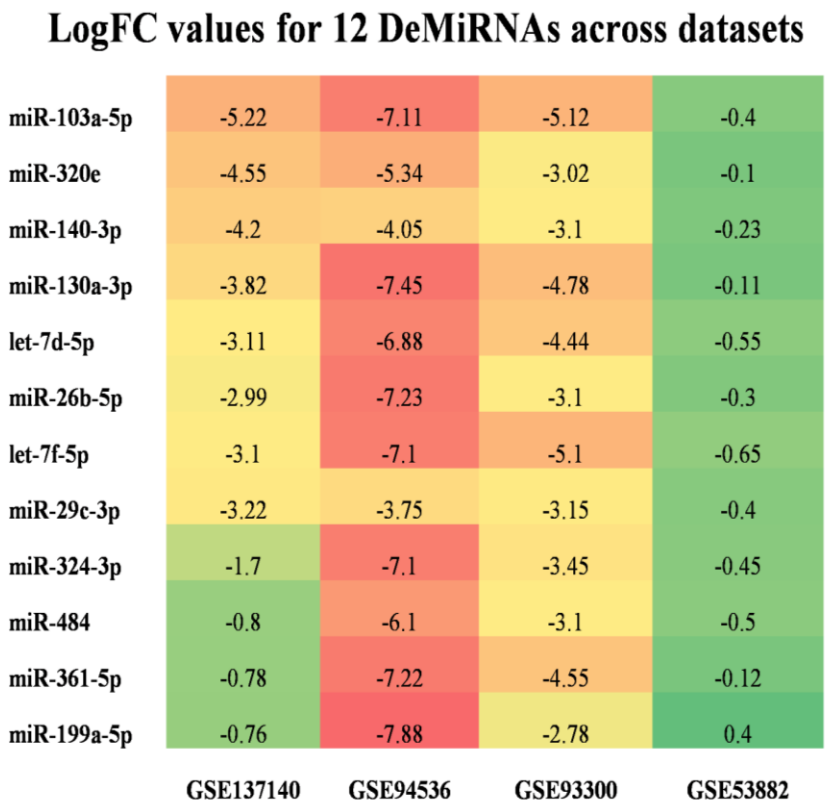

Figure 2: Heatmap of the log fold change of the DEmiRNAs. The green colour indicates a higher expression value. All logFC values shown on the heatmap are statistically significant ($p<0.05$).

Target Gene Prediction and Functional Enrichment Analysis

A total of 289 target genes were identified for twelve DEmiRNAs using TargetScan, mirTarBase, miRdb and miRwalk. The expression of these genes is found to be up-regulated in NSCLCs because the miRNAs that regulate them are downregulated. PANTHER analysis (http://www.pantherdb.org/) of the predicted target genes of these differentially expressed miRNAs revealed significantly enriched GO several molecular functions and biological processes (Table 3 and 4). The top significant enriched biological processes and molecular functions are generally involved in cancer progression, invasion, and metastasis. It was observed that most of the enriched biological processes were related to the epithelial to mesenchymal transition (EMT), phosphorylation of pathway-restricted SMAD proteins(17.11 times), heterochromatin assembly (12.21 times), cell growth and proliferation, negative regulation of gene silencing by miRNAs(11.45 times), and response to cholesterol. In addition, many other candidate genes related to cellular and molecular functions were also identified, which relate to transcription, protein kinase activity, transcription and chromatin binding. Additionally, PANTHER analysis applied to the GO cellular component terms showed that the RISC complex was enriched by 14.57 times.

Table 3: The top 21 GO terms based on biological processes and ranked by fold-enrichment.

GO biological process complete	Number of Genes in List	Expected Number of Genes	Fold Enrichment	P-Value	FDR
positive regulation of EMT involved in endocardial cushion formation (GO:1905007)	2	0.07	32.11	2.00E-03	1.90E-02
regulation of EMT involved in endocardial cushionformation (GO:1905005)	2	0.1	26.00	2.78E-03	2.56E-02
positive regulation of cardiac EMT (GO:0062043)	2	0.16	18.41	5.77E-02	3.69E-02
regulation of cardiac EMT (GO:0062042)	2	0.13	17.22	7.99E-04	3.36E-02
Pathway- restricted SMAD protein	3	0.24	16.13	1.11E-04	1.98E-02

Peerreviewunder responsibilityofLife Science Informatics Publications

GO biological process complete	Number of Genes in List	Expected Number of Genes	Fold Enrichment	P-Value	FDR
phosphorylation (GO:0060389)					
heterochromatin assembly (GO:0031507)	5	0.56	11.17	1.82E-04	1.67E-03
negative regulation of gene silencing by miRNA (GO:0060965)	3	0.37	10.77	5.16E-03	2.99E-02
response to cholesterol (GO:0070723)	5	0.51	10.65	2.44E-05	2.77E-03
negative regulation of posttranscriptionalgene silencing (GO:0060149)	4	0.44	10.12	5.10E-04	3.33E-02
negative regulation of gene silencing byRNA (GO:0060967)	3	0.24	12.82	5.10E-04	3.87E-02
miRNA metabolic process (GO:0010586)	2	0.36	9.89	7.55E-04	4.66E-02
heterochromatin organization (GO:0070828)	6	0.68	9.45	1.22E-05	1.30E-03
response to sterol (GO:0036314)	5	0.59	8.87	3.66E-05	4.78E-03
production of miRNAs involved in gene silencing by miRNA (GO:0035196)	6	0.44	8.21	2.50E-04	1.65E-02
ventricular septum morphogenesis (GO:0060412)	6	0.68	8.11	1.47E-05	1.78E-03

Peerreviewunder responsibilityofLife Science Informatics Publications

Table 4: The top 19 GO terms based on molecular functions and ranked by fold-enrichment.

GO molecular function complete	Number of Genes in List	Expected Number of Genes	Fold Enrichment	P-Value	FDR
5'-deoxyribose-5-phosphate lyase activity (GO:0051575)	2	0.13	26.44	2.97E-03	2.95E-02
TGF-beta-activated receptor activity(GO:0005024)	3	0.19	17.67	1.47E-03	1.66E-02
activin binding (GO:0048185)	3	0.24	14.21	1.98E-03	1.97E-02
1-phosphatidylinositol-3-kinase regulator activity (GO:0046935)	3	0.26	13.71	2.88E-03	2.45E-02
TGF- beta binding (GO:0050431)	4	0.37	11.33	7.22E-04	1.34E-01
transmembrane receptor protein serine/threonine kinase activity (GO:0004675)	3	0.31	11.21	4.22E-02	3.23E-02
SMAD binding (GO:0046332)	8	1.35	5.57	1.99E-04	2.21E-02
catalytic activity, acting on DNA (GO:0140097)	14	3.45	3.45	1.77E-04	5.44E-03
protein phosphatase binding (GO:0019903)	9	2.33	3.30	8.65E-04	1.66E-01
phosphatase binding (GO:0019902)	11	3.22	2.81	4.22E-04	5.68E-02
kinase regulator activity (GO:0019207)	11	3.64	3.91	1.67E-03	1.88E-01
protein serine/threonine kinase activity (GO:0004674)	19	7.52	1.98	3.47E-04	5.60E-03
DNA-binding transcription activatoractivity, RNA polymerase II-specific(GO:0001228)	19	7.13	1.92	3.00E-04	4.66E-02
Protein kinase activity (GO:0004672)	23	9.66	1.23	3.21E-04	3.96E-02

PPI Network Analysis

The Search Tool for Retrieval of Interacting Genes/Protein (STRING) database was used to collect protein interaction data to construct protein-protein interaction (PPI) network from 290 target

genes. The result identified 678 interactions between nodes (genes) with an average node degree of 3.89(Figure 3). The expected number of interactions was 498; the p-value for this high number of overlaps was < 1.2E-15. Hub gene network connections were exported to Cytoscape for visualization via the Network Analyzer and MCODE tools. MCODE was used to identify the functional modules of the highly interconnected clusters of genes in the Cytoscape network. By using MCODE algorithms, we identified 8 modules. Among them, the top 2 modules are displayed (Figure 4). Module 1 consists of ten genes, including MAPK1, PPARG, SNAI1, IL6, FGF2, HGF, ESR1, MET, HIF1A, and SMAD2 with 42 interactions between them. KEGG pathway enrichment analysis of the module revealed that 9 out of ten genes excluding SANI1 were significantly enriched in cancer-related pathways (Table 3). Furthermore, Gene Ontology (GO) analysis revealed that all these genes are part of biological processes, such as positive regulation of transcription, and signal transduction pathway. In addition, Module 2 consists of fifteen genes with 35 interactions between them. Among them, five out of 15 genes were involved in chromatid separation. Moreover, hub genes in the network were selected using Network Analyzer; we identify eight hub genes that pass the cutoff criteria: MAPK1, IL6, FGF2, SMAD2, SNAI1, DICER1, CDK6, HGF.

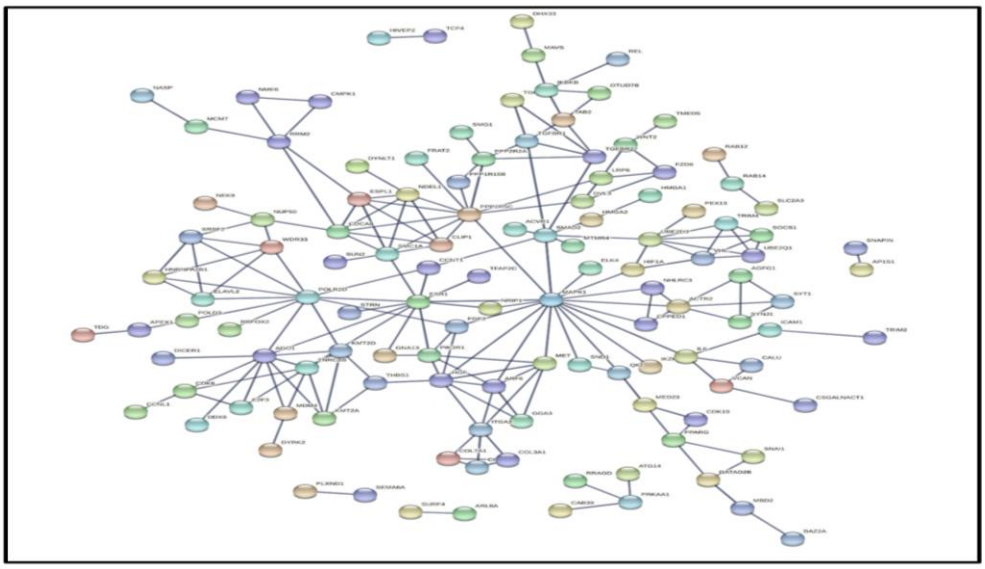

Figure 3: PPI interaction graph created by STRING. Disconnected nodes are not taken and minimum interaction confidence is set to 0.6 for clarity. Each circle represents one gene and its product (protein), and edges represent interaction.

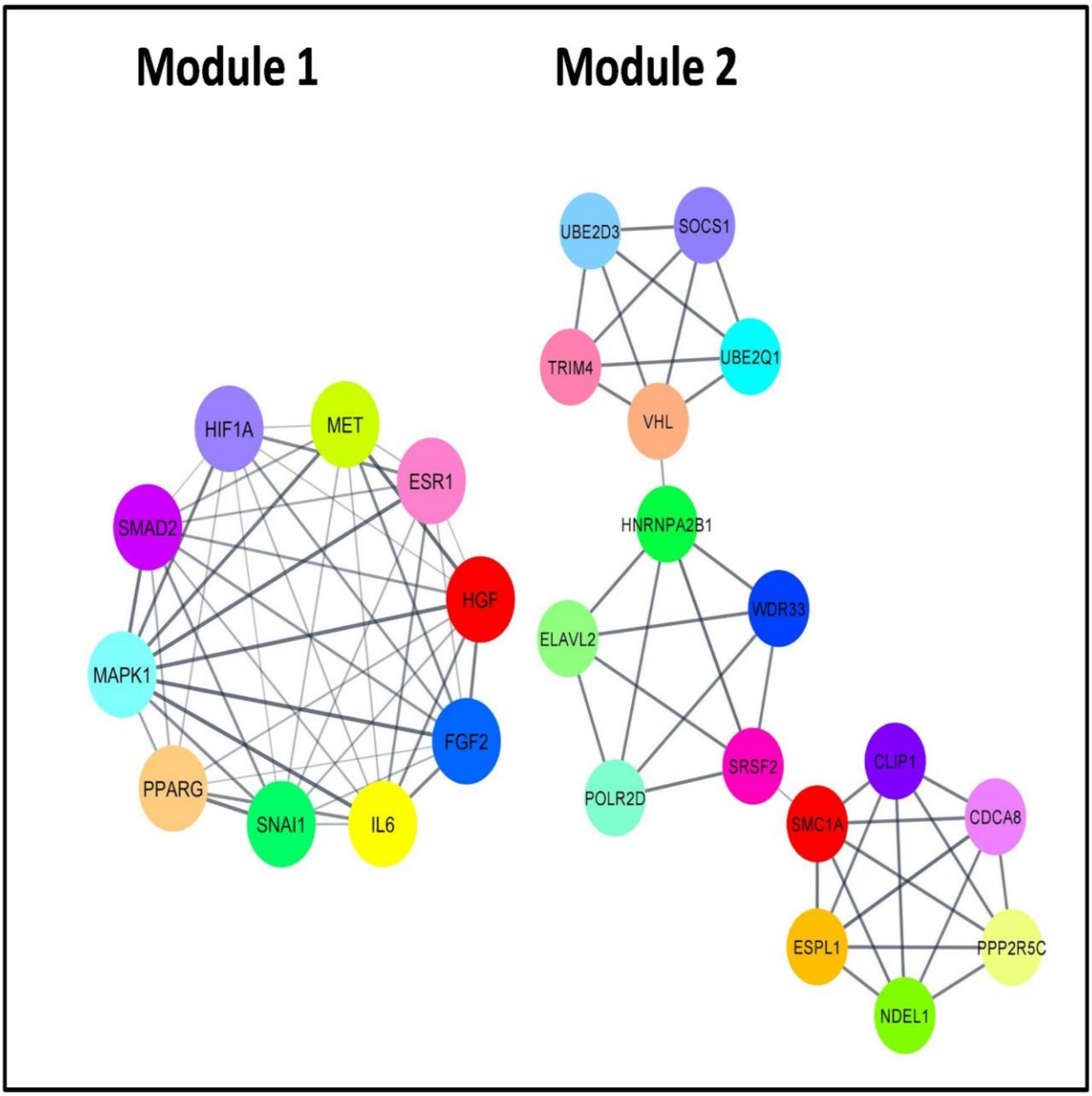

Figure 4: The highly interconnected genes in module 1 and module 2 are displayed. Colours are arbitrary and each line indicates relationships between groups of genes. The first module consisting of ten genes with 42 interactions, and the second module consist of fifteen genes with 35 interactions.

Hub Gene Survival Analysis

For each hub gene, Kaplan-Meier survival curves for patients with lung adenocarcinoma and squamous cell carcinoma were plotted, and the log-rank test was used to assess statistical significance. Patients were split by the median expression of each gene into two groups, namely patients with high or low expression. The survival analysis identified 7 of the 10 genes which were statistically significantly different between low and high expression for patients with NSCLC. Consistent with these results, we also found that higher expression of IL6, SNAI1, and CDK6 were

associated with poorer prognosis and lower expression of ESR1, FGF2, SMAD2, DICER1, and HGF were associated with poorer prognosis (Figure 5). Among the entire hub gene, SMAD2 showed the lowest hazard ratio (0.67) with a p-value of 2.9E-06. This result indicates that differential expression of SMAD2 has a particularly large effect on patient survival. Taken together this study suggests that these 8 genes could hold promise as prognostic biomarkers of NSCLC. Among the 8 most significant gene networks, IL6 showed the highest hub gene node degree from the PPI network. Survival analysis was performed on NSCLC patients using a panel of three genes (IL6, SNAI1, and CDK6). The difference in overall survival rate between high and low expressions was statistically significant ($p = 0.005$). The hazard ratio for the combined 3 gene survival curve was higher as compared to each of the 3 genes alone, indicated that their combination lends prognostic power (Figure 6).

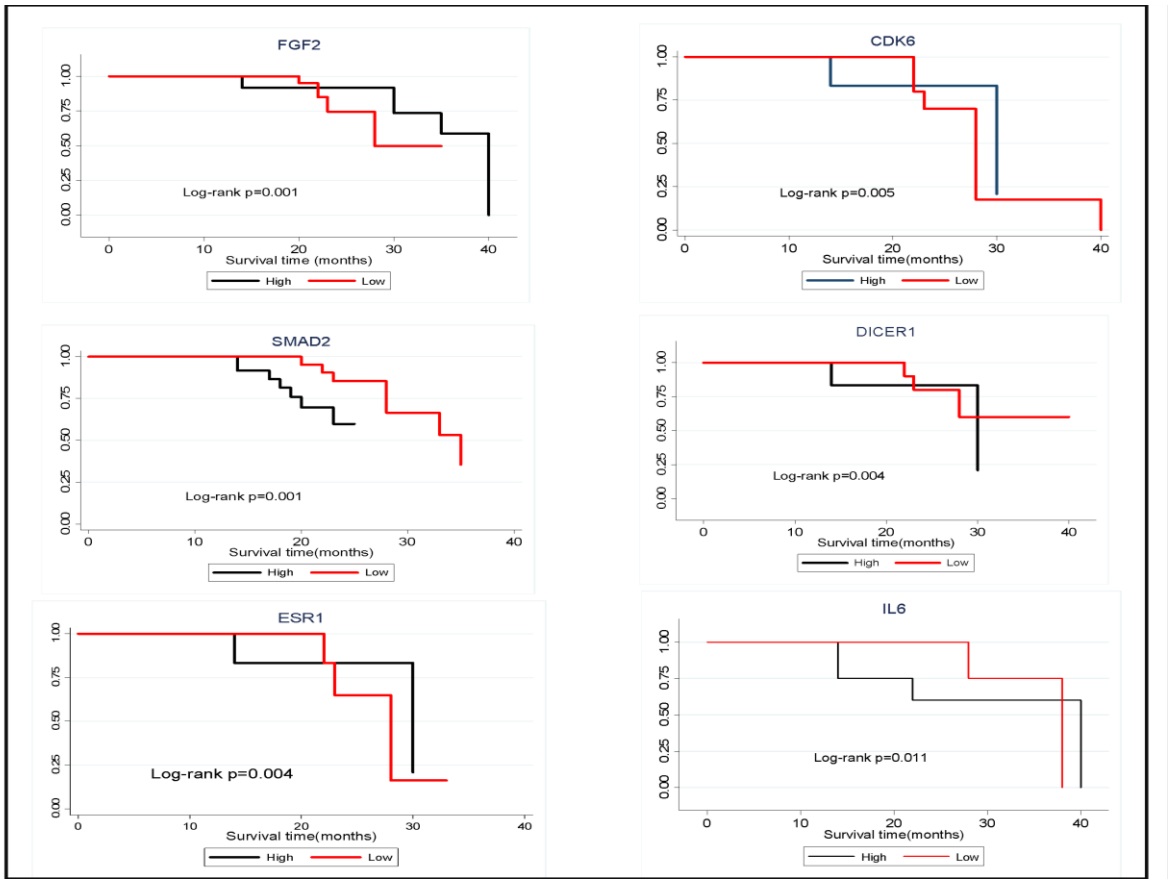

Figure 5: Kaplan Meier survival plots for six genes. The black curve represents the survival of patients with high expression and the red curve represents the survival of those with low expression.

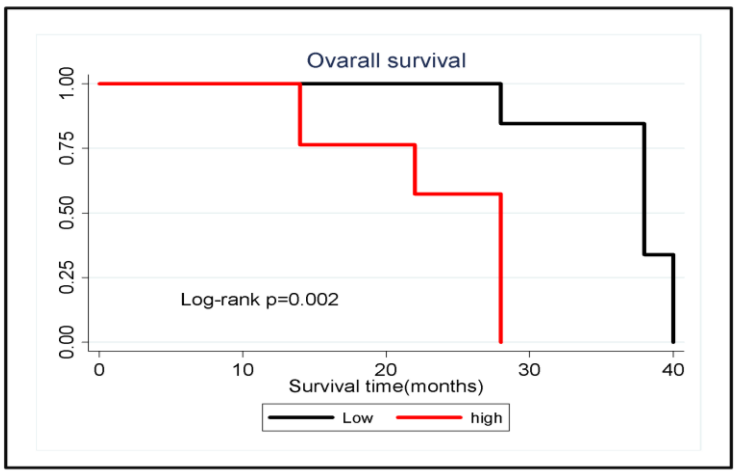

Figure 6: A combined survival curve analysis from the panel of three genes (IL6, SNAI1, and CDK6). Blue colour indicating low and red indicates high expression. Higher expression of these three genes was associated with a poorer prognosis.

We performed a correlation analysis of transcript-level expression for IL6 and SNAI1 in LUAD patients (Figure 7). The results showed a significant correlation ($p<0.02$). The expression levels of these two genes were positively correlated with a Spearman coefficient of 0.58. Further GEPIA analysis showed that IL6 and SNAI1 are down-regulated in lung cancer patients based on The Cancer Genome Atlas (TCGA) database.

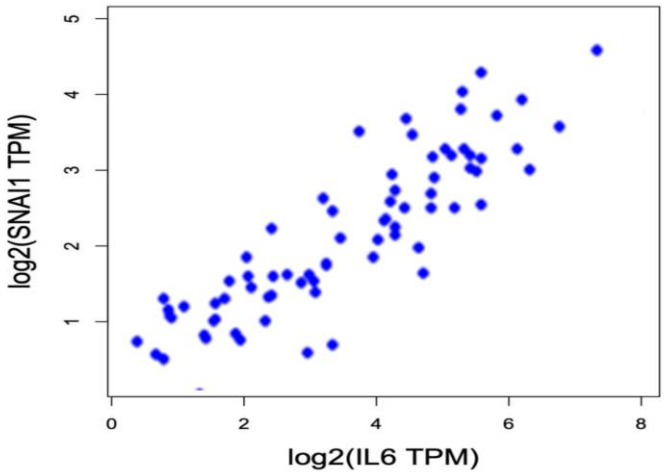

Figure 7: Scatter plots of correlations between IL6 and SNAI1 expressions in LUAD patients with a log base 2 transformations applied, a positive correlation was observed. Spearman's rank correlation coefficients and p-value are displayed.

MiRNA Candidate Biomarker Selection

Kaplan-Meier survival analysis was used for the 12 DEmiRNAs on 871 patients (LUAD or LUSC). The results identified miRNAs (miR-140, miR-29c, and miR-199a) with p-values lower than 0.05 and hazard ratio 0.69 (Figure 8). However, the other nine miRNAs did not have significant hazard

ratios. Moreover, we found that high expression of all three of these miRNAs was associated with better NSCLC outcomes. The expression of miR-140-3p was significantly down-regulated in both LUAD and LUSC tumor tissue according to TCGA data analysis. The other miRNA that was found to be significantly down-regulated in LUSC patients was miR-29c-3p, while it was not significantly different for LUAD (Figure 9). Expression of miR-199a-5p was significantly increased in both LUAD and LUSC patients. The areas under the receiver operating characteristic curves (AUC) for each of the three miRNAs were generated. AUC values of 0.81, 0.69, and 0.75 were observed for miR-140-3p, miR-199a-5p and miR-29c-3p, respectively (Figure 10).

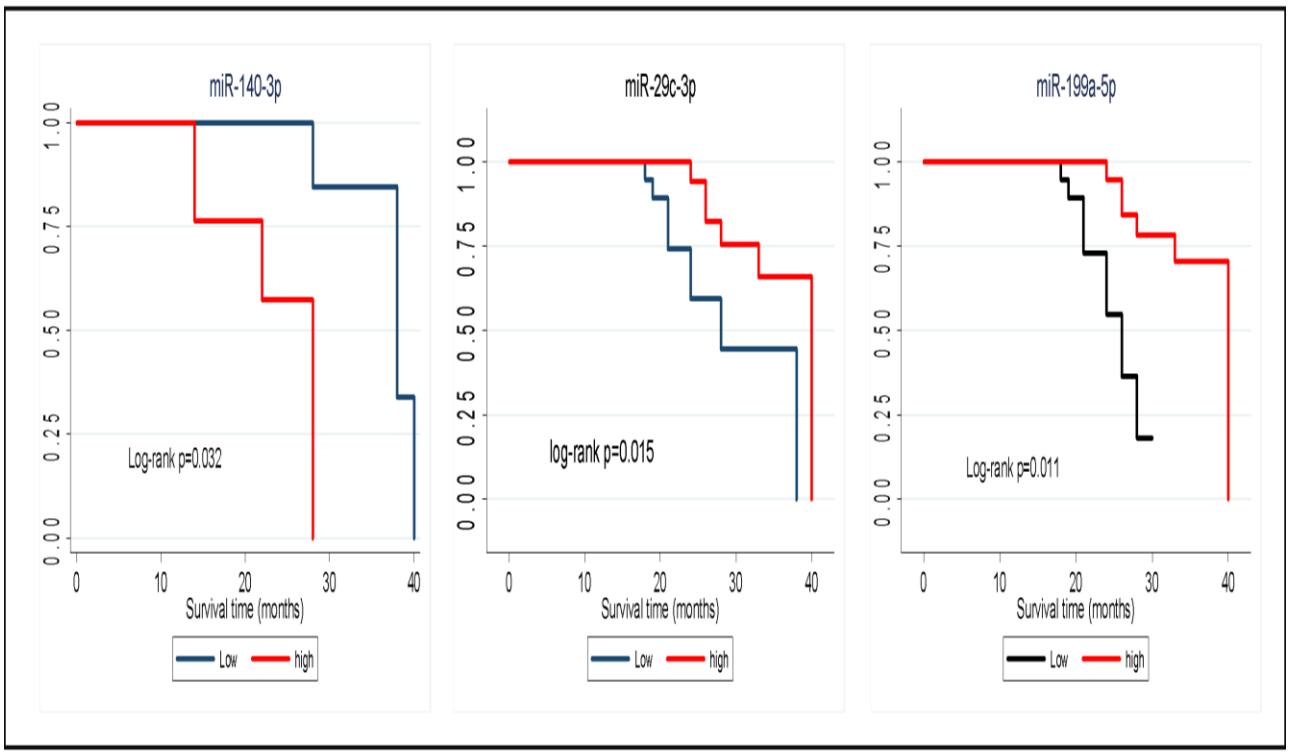

Figure 8: Kaplan Meier survival plots for miR-140, miR-29c, and miR-199a. The black curve shows patients with low miRNA expression in tissue while the red curve represents high expression.

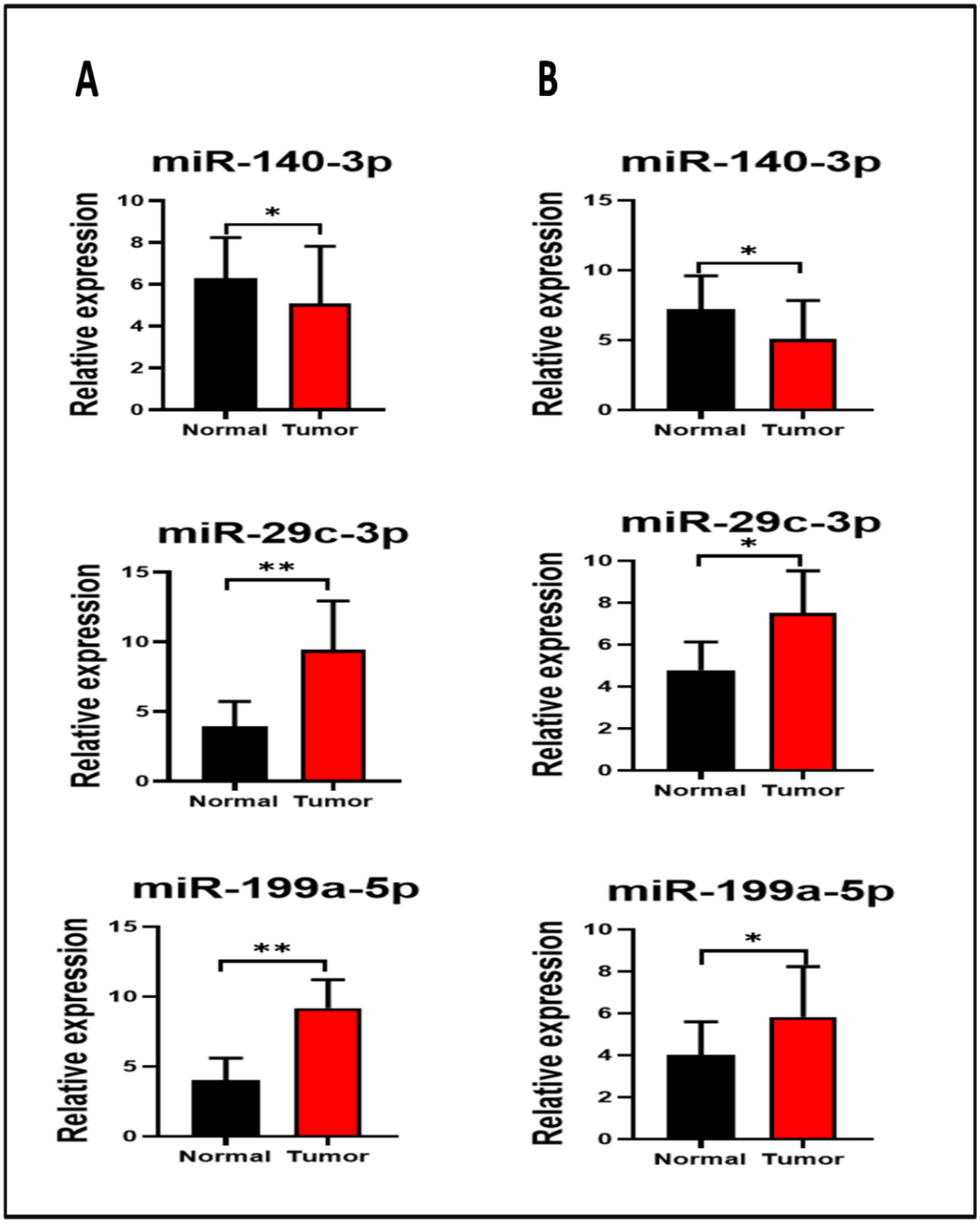

Figure 9: Relative expression of miRNAs. LUAD vs control (A), LUSC vs controls (B).

Peerreviewunder responsibilityofLife Science Informatics Publications
2021 Sept – Oct RJLBPCS 7(5) Page No.54

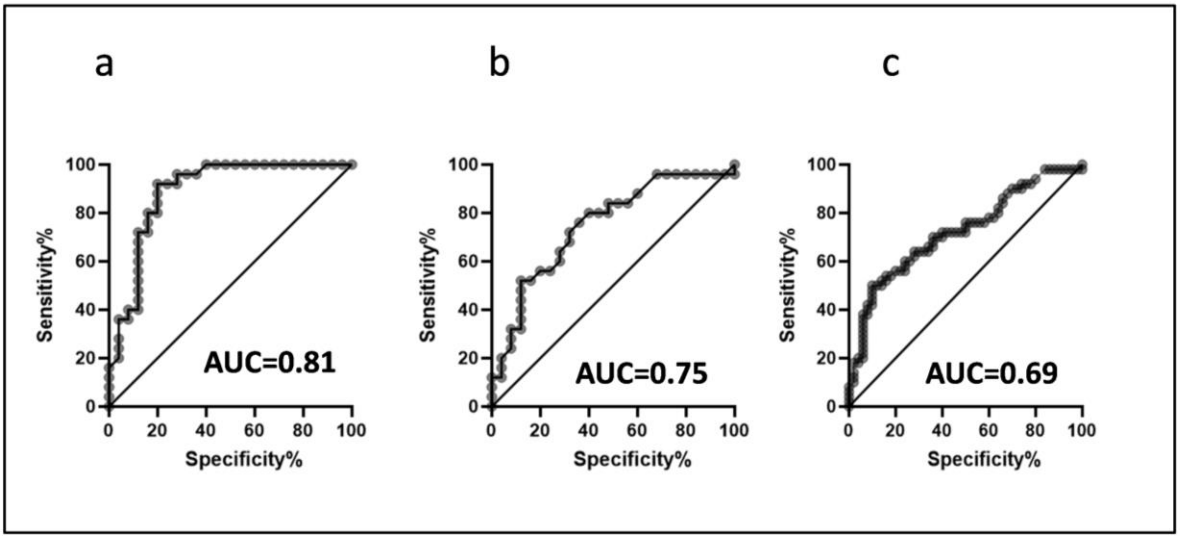

Figure 10: ROC for (a) miR-140-3p, (b) miR-199a-5p, and (c) miR-29c-3p based on expression data from controls and NSCLC patients.

DISCUSSION

Despite many advances in non-small cell lung cancer (NSCLC) treatments over the last several decades, the prognosis for patients with NSCLC remains poor. This is mainly due to the lack of early and accurate diagnostic tests. We demonstrate that the blood and tissue miRNAs and mRNAs and their profiles can be developed as biomarkers in the diagnosis and prognosis of NSCLC. In this study, we examined 4 miRNA expression datasets and observed a substantial overlap of twelve differentially expressed miRNAs and 330 target genes were predicted for 12 miRNAs. Network pathway analysis and survival analysis were performed on the genes and miRNAs differentially expressed between NSCLC and healthy individuals. Functional enrichment analysis of the candidate miRNAs target genes disclosed several important biological processes and molecular functions. Dysregulated SMAD signalling has been linked to cancer[44–46]. Network analysis identified SMAD3, a member of the SMAD family as hub gene. The process of negative regulation of gene silencing by miRNA was significantly enriched. The main terms of molecular functions were related to transcription factor binding, transcription regulatory activity, protein kinase activity, and chromatin binding. In molecular function, 5'-deoxyribose-5-phosphate lyase activity was the most highly enriched. 5'-deoxyribose-5-phosphate lyase is an enzyme involved in base excision repair, a key pathway in the repair of DNA single-strand breaks that provide an important line of defence against cancer progression[47,48]. We observe a prominent enrichment for the RISC complex. MiRNAs regulate gene expression by recruiting the RNA-induced silencing complex (RISC) to a target mRNA with which it shares partial complementarily, causing a decrease in the stability of the mRNA. This indicated the possible role of epigenetic control of gene expression in

NSCLC. Eight genes including MPAL1, IL6, FGF2, SMAD2, DICER1, CDK6, HGF and FGF2 were identified as hub genes. Among these genes, MAPK1 and IL6 are already known to be implicated in cancer. These eight hub genes are proposed as blood-born mRNA biomarkers for NSCLC and should be experimentally validated in future work. This study identified three miRNAs that could serve as biomarkers in the diagnosis of NSCLC: miR-140, miR-29c, and miR-199a. MiR-140-3p has been found to play a role in cancer as tumor suppressor. This is in line with the results of the present study where decreasing levels of miR-140-3p were found in four separate NSCLC datasets. Recently, Huang et al. showed that the expression of miR-140-5p was down-regulated in the tissue of SCLC patients and was significantly correlated with survival and tumor stage[49]. Another study showed that miR-140-3p targets ATP6AP2 and inhibits proliferation, migration, and invasion of lung cancer cells. In their study, they found that miR-140-3p was down-regulated in lung cancer tissue compared to adjacent normal lung tissue [50]. In addition, miR-29c-3p also showed the potential of using it as a non-invasive blood-based biomarker for the detection of NSCLC. Studies found that miR-29c-3p inhibits colon cancer cell invasion, and suppresses hepatocellular carcinoma tumor progression[51,52]. In laryngeal squamous cell carcinoma, low expression of miR-29c-3p is associated with a poor prognosis [53]. Furthermore, low levels of miR-29c-3p in endometrial cancer cells led to reduced growth[54]. MiR-199a-5p has also been shown to be associated with cancer. Chen et al. identified the miR-199a-5p as a tumor suppressor in triple-negative breast cancer[55]. In addition, one study showed that miR-199a-5p targets MAP3K11 and suppress NSCLC progression[56]. Zhu et al. demonstrated that miR-199a-5p inhibit the growth of colorectal cancer cells[57]. Ma et al. showed that miR-199a-5p target SNAI1 and inhibit papillary thyroid carcinoma progression[58]. The main limitation of this study was that no normalization was performed between datasets. In addition, this analysis did not differentiate between males and females. Another important limitation is that the research only compared NSCLC patients with healthy controls, but not with other types of malignancy or lung diseases. This study identified mRNA and miRNA as biomarkers for NSCLC, however, not be able to discriminate between NSCLC and other diseases. Future large scale follow-up study comparing miRNA expression between NSCLC and other cancers and lung diseases is warranted. Future research should be designed to include more blood samples to evaluate the diagnostic and prognostic potential of the proposed miRNAs and mRNAs.

4. CONCLUSION

The present study identified twelve miRNAs that are down-regulated in the blood and tissue of NSCLC patients. Based on bioinformatics analysis, miR-140-3p, miR-29c, and miR-199a are candidate biomarkers. Functional enrichment analysis of the candidate miRNA target genes identified several overrepresented pathways associated with cancer. Eight target genes were hub

genes in the PPI network and possessed significant prognostic value. Further clinical validation of these identified potential candidate miRNAs and mRNAs is warranted.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The author confirms that the data supporting the findings of this research are available within the article.

FUNDING

None.

ACKNOWLEDGEMENT

Not applicable.

CONFLICT OF INTEREST

Authors have no conflict of interest.

REFERENCES

1. Xu F, Xu P, Cui W, Gong W, Wei Y, Liu B, et al. Neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios may aid in identifying patients with non-small cell lung cancer and predicting Tumor-Node-Metastasis stages. Oncol Lett. 2018;16: 483–490.
2. Hu W, Bi Z-Y, Chen Z-L, Liu C, Li L-L, Zhang F, et al. Emerging landscape of circular RNAs in lung cancer. Cancer Lett. 2018;427: 18–27.
3. Depierre A, Lagrange JL, Theobald S, Astoul P, Baldeyrou P, Bardet E, et al. Summary report of the Standards, Options and Recommendations for the management of patients with non-small-cell lung carcinoma (2000). Br J Cancer. 2003;89: S35–S49.
4. Sánchez-Jiménez C, Ludeña MD, Izquierdo JM. T-cell intracellular antigens function as tumor suppressor genes. Cell Death Dis. 2015;6: e1669.
5. Browse the Tables and Figures - SEER Cancer Statistics Review (CSR) 1975-2012. In: SEER [Internet]. [cited 3 Sep 2021]. Available: https://seer.cancer.gov/archive/csr/1975_2012/browse_csr.php?sectionSEL=15&pageSEL=sect_15_table.14
6. Khetrapal P, Lee MWL, Tan WS, Dong L, de Winter P, Feber A, et al. The role of circulating tumour cells and nucleic acids in blood for the detection of bladder cancer: A systematic review. Cancer Treat Rev. 2018;66: 56–63.

7. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961;3: 318–356.
8. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75: 843–854.
9. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev. 2006;20: 515–524.
10. Catto JWF, Alcaraz A, Bjartell AS, De Vere White R, Evans CP, Fussel S, et al. MicroRNA in prostate, bladder, and kidney cancer: a systematic review. Eur Urol. 2011;59: 671–681.
11. Iqbal MA, Arora S, Prakasam G, Calin GA, Syed MA. MicroRNA in lung cancer: role, mechanisms, pathways and therapeutic relevance. Mol Aspects Med. 2019;70: 3–20.
12. Ma J, Lin Y, Zhan M, Mann DL, Stass SA, Jiang F. Differential miRNA expressions in peripheral blood mononuclear cells for diagnosis of lung cancer. Lab Invest. 2015;95: 1197–1206.
13. Hennessey PT, Sanford T, Choudhary A, Mydlarz WW, Brown D, Adai AT, et al. Serum microRNA biomarkers for detection of non-small cell lung cancer. PLoS One. 2012;7: e32307.
14. Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, et al. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. Lab Invest. 2011;91: 579–587.
15. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. Clin Chem. 2010;56: 1733–1741.
16. Zhao H, Shen J, Medico L, Wang D, Ambrosone CB, Liu S. A Pilot Study of Circulating miRNAs as Potential Biomarkers of Early Stage Breast Cancer. PLOS ONE. 2010;5: e13735.
17. Szafranska AE, Davison TS, John J, Cannon T, Sipos B, Maghnouj A, et al. MicroRNA expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma. Oncogene. 2007;26: 4442–4452.
18. Galka-Marciniak P, Urbanek-Trzeciak MO, Nawrocka PM, Dutkiewicz A, Giefing M, Lewandowska MA, et al. Somatic Mutations in miRNA Genes in Lung Cancer-Potential Functional Consequences of Non-Coding Sequence Variants. Cancers (Basel). 2019;11.
19. Serum miR-1228-3p and miR-181a-5p as Noninvasive Biomarkers for Non-Small Cell Lung Cancer Diagnosis and Prognosis. [cited 6 Sep 2021]. Available: https://www.hindawi.com/journals/bmri/2020/9601876/
20. Riveiro ME, Astorgues-Xerri L, Vazquez R, Frapolli R, Kwee I, Rinaldi A, et al. OTX015 (MK-8628), a novel BET inhibitor, exhibits antitumor activity in non-small cell and small cell lung cancer models harboring different oncogenic mutations. Oncotarget. 2016;7: 84675–84687.
21. Henderson L-J, Coe BP, Lee EHL, Girard L, Gazdar AF, Minna JD, et al. Genomic and gene

expression profiling of minute alterations of chromosome arm 1p in small-cell lung carcinoma cells. Br J Cancer. 2005;92: 1553–1560.

22. Overexpression of CD44 is associated with the occurrence and migration of non-small cell lung cancer. [cited 6 Sep 2021]. Available: https://www.spandidos-publications.com/10.3892/mmr.2016.5636
23. Altenberger C, Heller G, Ziegler B, Tomasich E, Marhold M, Topakian T, et al. SPAG6 and L1TD1 are transcriptionally regulated by DNA methylation in non-small cell lung cancers. Molecular Cancer. 2017;16: 1.
24. Morris S, Vachani A, Pass HI, Rom WN, Ryden K, Weiss GJ, et al. Whole blood FPR1 mRNA expression predicts both non-small cell and small cell lung cancer. Int J Cancer. 2018;142: 2355–2362.
25. Hosseini SM, Soltani BM, Tavallaei M, Mowla SJ, Tafsiri E, Bagheri A, et al. Clinically Significant Dysregulation of hsa-miR-30d-5p and hsa-let-7b Expression in Patients with Surgically Resected Non-Small Cell Lung Cancer. Avicenna J Med Biotechnol. 2018;10: 98–104.
26. MiR-598 Suppresses Invasion and Migration by Negative Regulation of Derlin-1 and Epithelial-Mesenchymal Transition in Non-Small Cell Lung Cancer - FullText - Cellular Physiology and Biochemistry 2018, Vol. 47, No. 1 - Karger Publishers. [cited 6 Sep 2021]. Available: https://www.karger.com/Article/FullText/489803
27. Olivieri F, Capri M, Bonafè M, Morsiani C, Jung HJ, Spazzafumo L, et al. Circulating miRNAs and miRNA shuttles as biomarkers: Perspective trajectories of healthy and unhealthy aging. Mech Ageing Dev. 2017;165: 162–170.
28. Identification of potential diagnostic and prognostic biomarkers in non-small cell lung cancer based on microarray data - PubMed. [cited 6 Sep 2021]. Available: https://pubmed.ncbi.nlm.nih.gov/29731852/
29. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30: 207–210. 30. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41: D991-5.
31. Asakura K, Kadota T, Matsuzaki J, Yoshida Y, Yamamoto Y, Nakagawa K, et al. A miRNA-based diagnostic model predicts resectable lung cancer in humans with high accuracy. Commun Biol. 2020;3: 134.
32. Qu L, Li L, Zheng X, Fu H, Tang C, Qin H, et al. Circulating plasma microRNAs as potential markers to identify EGFR mutation status and to monitor epidermal growth factor receptor-tyrosine kinase inhibitor treatment in patients with advanced non-small cell lung cancer.

Oncotarget. 2017;8: 45807–45824.

33. Li L-L, Qu L-L, Fu H-J, Zheng X-F, Tang C-H, Li X-Y, et al. Circulating microRNAs as novel biomarkers of ALK-positive nonsmall cell lung cancer and predictors of response to crizotinib therapy. Oncotarget. 2017;8: 45399–45414.
34. Pu H-Y, Xu R, Zhang M-Y, Yuan L-J, Hu J-Y, Huang G-L, et al. Identification of microRNA-615-3p as a novel tumor suppressor in non-small cell lung cancer. Oncol Lett. 2017;13: 2403–2410.
35. Dweep H, Sticht C, Pandey P, Gretz N. miRWalk – Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes. Journal of Biomedical Informatics. 2011;44: 839–847.
36. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research. 2019;47: D330–D338.
37. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 2017;45: D183–D189.
38. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003;31: 258–261.
39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003;13: 2498–2504.
40. Nagy Á, Lánczky A, Menyhárt O, Győrffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep. 2018;8: 9227.
41. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017;45: W98–W102.
42. Pan C-Y, Lin W-C. miR-TV: an interactive microRNA Target Viewer for microRNA and target gene expression interrogation for human cancer studies. Database (Oxford). 2020;2020: baz148.
43. Díez-Villanueva A, Mallona I, Peinado MA. Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. Epigenetics & Chromatin. 2015;8: 22.
44. Bello-DeOcampo D, Tindall DJ. TGF-betal/Smad signaling in prostate cancer. Curr Drug Targets. 2003;4: 197–207.
45. Xie W, Rimm DL, Lin Y, Shih WJ, Reiss M. Loss of Smad signaling in human colorectal cancer is associated with advanced disease and poor prognosis. Cancer J. 2003;9: 302–312.
46. Cui Y, Song Y, Yan S, Cao M, Huang J, Jia D, et al. CUEDC1 inhibits epithelial-mesenchymal

transition via the TβRI/Smad signaling pathway and suppresses tumor progression in non-small cell lung cancer. Aging (Albany NY). 2020;12: 20047–20068.

47. García-Díaz M, Bebenek K, Kunkel TA, Blanco L. Identification of an intrinsic 5'-deoxyribose-5-phosphate lyase activity in human DNA polymerase lambda: a possible role in base excision repair. J Biol Chem. 2001;276: 34659–34663.
48. Longley MJ, Prasad R, Srivastava DK, Wilson SH, Copeland WC. Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair in vitro. Proc Natl Acad Sci U S A. 1998;95: 12244–12248.
49. Huang H, Wang Y, Li Q, Fei X, Ma H, Hu R. miR-140-3p functions as a tumor suppressor in squamous cell lung cancer by regulating BRD9. Cancer Lett. 2019;446: 81–89.
50. Kong X-M, Zhang G-H, Huo Y-K, Zhao X-H, Cao D-W, Guo S-F, et al. MicroRNA-140-3p inhibits proliferation, migration and invasion of lung cancer cells by targeting ATP6AP2. Int J Clin Exp Pathol. 2015
51. Chen G, Zhou T, Li Y, Yu Z, Sun L. p53 target miR-29c-3p suppresses colon cancer cell invasion and migration through inhibition of PHLDB2. Biochem Biophys Res Commun. 2017;487: 90–95.
52. miR-29c-3p regulates DNMT3B and LATS1 methylation to inhibit tumor progression in hepatocellular carcinoma | Cell Death & Disease. [cited 30 Sep 2021]. Available: https://www.nature.com/articles/s41419-018-1281-7
53. Fang R, Huang Y, Xie J, Zhang J, Ji X. Downregulation of miR-29c-3p is associated with a poor prognosis in patients with laryngeal squamous cell carcinoma. Diagnostic Pathology. 2019;14: 109.
54. Van Sinderen M, Griffiths M, Menkhorst E, Niven K, Dimitriadis E. Restoration of microRNA-29c in type I endometrioid cancer reduced endometrial cancer cell growth. Oncology Letters. 2019;18: 2684–2693.
55. miR-199a-5p confers tumor-suppressive role in triple-negative breast cancer | BMC Cancer | Full Text. [cited 30 Sep 2021]. Available: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-016-2916-7
56. MiR-199a-5p suppresses non-small cell lung cancer via targeting MAP3K11. [cited 30 Sep 2021]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6584351/
57. Zhu QD, Zhou QQ, Dong L, Huang Z, Wu F, Deng X. MiR-199a-5p Inhibits the Growth and Metastasis of Colorectal Cancer Cells by Targeting ROCK1. Technol Cancer Res Treat. 2018;17: 1533034618775509.
58. Ma S, Jia W, Ni S. miR-199a-5p inhibits the progression of papillary thyroid carcinoma by targeting SNAI1. Biochem Biophys Res Commun. 2018; 497: 181–186.