Integrated bioinformatics analysis of microarray data from non-small cell lung cancer

Lulu Feng¹, Wenping Cai¹, Shan Jin¹, Caipu Chun², Chengyan Wang¹, Luping Ma¹, Hao Peng¹, Xingxing Dong¹, Jinfang Jiang¹, Xianling Lu¹*, Lijuan Pang¹,²*

¹Treatment of Central Asia High Incidence Diseases (First Affiliated Hospital, School of Medicine, Shihezi University)
²Department of Pathology, Central People’s Hospital of Zhanjiang and Zhanjiang Central Hospital, Guangdong Medical University, Zhanjiang, Guangdong, China

# Lulu Feng, Wenping Cai and Shan Jin have equally contributed to this work.

ARTICLE INFO

Original paper

Article history:
Received: January 15, 2023
Accepted: June 06, 2023
Published: July 31, 2023

Keywords:
Non-small cell lung cancer (NSCLC), diagnostic markers, miRNAs, cancer-related pathways, miRNA-hub gene networks

Non-small cell lung cancer (NSCLC), with its high mortality rate, lack of early diagnostic markers and prevention of distant metastases are the main challenges in treatment. To identify potential miRNAs and key genes in NSCLC to find new biomarkers and target gene therapies. The GSE102286, GSE56036, GSE25508, GSE53882, GSE29248 and GSE101929 datasets were obtained from the Gene Expression Omnibus (GEO) database and screened for differentially co-expressed miRNAs (DE-miRNAs) and lncRNAs (DElncRs) by GEO2R and R software package. Pathway enrichment analysis of DE-miRNAs-target genes was performed by String and Funrich database to construct protein-protein interaction (PPI) and competing endogenous RNA (ceRNA) network and visualized with Cytoscape software. Nineteen co-expressed DE-miRNAs were screened from five datasets. The 7683 predicted up- and down-regulated DE-miRNAs-target genes were significantly concentrated in cancer-related pathways. The top 10 hub nodes in the PPI were identified as hub genes, such as MYC, EGFR, HSP90AA1 and TP53, MYC, and ACTB. By constructing miRNA-hub gene networks, hsa-miR-21, hsa-miR-141, hsa-miR-200b and hsa-miR-30a, hsa-miR-30d, hsa-miR-145 may regulate most hub genes and hsa-miR-141, hsa-miR-200, hsa-miR-145 had higher levels in the miRNA and ceRNA regulatory networks, respectively. In conclusion, the identification of hsa-miR-21, hsa-miR-141, hsa-miR-200b, hsa-miR-30a, hsa-miR-30d and hsa-miR-145 provides a new theoretical basis for understanding the development of NSCLC.

Doi: http://dx.doi.org/10.14715/cmb/2023.69.7.35

Introduction

Lung cancer is the malignancy with the highest mortality rate, although immunotherapy has greatly improved the prognosis of lung cancer, the lack of early diagnosis and recurrence of distant metastases are the main challenges for treatment. The most common type of lung cancer is NSCLC (1-4), and since research on NSCLC is still in its early stages, it is urgent to explore its etiology and identify novel biomarkers to understand NSCLC (5). A number of molecular biomarkers have been applied to aid in the diagnosis of NSCLC. For example, miR-1 and miR-128-3p inhibit tumor growth and chemotherapy resistance (6,7). MiR-330-3p and miR-661 promote metastasis in NSCLC and are potential non-invasive NSCLC biomarkers (8,9). The findings suggest that miRNAs are closely associated with the development of NSCLC. The miRNAs-expression profiles of human paired tissues and their regulatory roles have not been systematically studied.

As the role of GEO in bioinformatics analysis has gradually increased, new functional miRNAs have been identified (10). DE-miRNAs were obtained by analyzing the GSE102286, GSE56036, GSE22508, GSE53882, GSE29248 and GSE101929 datasets. The up- and down-regulated DE-miRNAs-center genes were predicted separately, and their localization, function and enrichment pathways in NSCLC were determined. PPI, DE-miRNAs-hub genes and ceRNA regulatory networks were constructed. The aim is to identify dysregulated microRNAs and provide more information for exploring new therapeutic strategies for NSCLC.

Materials and Methods

miRNA and mRNA microarray

Five miRNAs and one mRNA microarray dataset were selected from GEO by keywords, "NSCLC miRNA" or "NSCLC mRNA". The miRNA dataset GSE102286 was derived from the GPL23871 platform. It includes 91 NSCLC samples and 88 normal control samples (11). GSE56036 by using the GPL15446 platform contained 29 samples of NSCLC tissues and 29 samples of normal tissues (18 adenocarcinomas and 11 squamous cell carcinomas) (12), and GSE25508 based on the platform of...
GPL7731 consisted of 32 tissue samples of NSCLC and 34 matched normal tissue samples (13). In the GPL18130 platform, the GSE53882 dataset contains 397 tissue samples of NSCLC and 151 normal tissue samples (14). GSE29248, based on the GPL8179 platform, contains 12 tissue samples of NSCLC and 12 normal tissue samples, including 6 cases of adenocarcinoma and 6 cases of squamous cell carcinoma (15). The gene expression profile GSE101929 included 32 NSCLC and 34 normal control samples with GPL570 as the platform (11).

DElnRs-DEmiRs analysis

DEmiRs were screened by GEO2R based on R analysis with the critical values adjusted p<0.05 and |logFC|>0. In addition, the Series matrix file and platform file of the mRNA were downloaded. Based on the platform information, the probes are transferred to gene symbols, and DElnRs were screened. P-values for expression differences were calculated using Limma. P<0.05 and |logFC|>2 was selected as the thresholds for screening DElnRs.

Survival data of DEmiRs

Survival analysis of DEmiR was performed by the R package, containing 1782 LUAD patients from TCGA. Survival curves were generated based on the high and low expressions of DEmiR. p < 0.05 was considered significant.

Prediction of DEmiRs’ targets

The DEmiRs were intersected by Venn Diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) and then DEmiRs-target genes were predicted by miRTarBase (16,17).

Functional analysis

The functional and pathway enrichment of DEmiRs-target genes was analyzed using Funrich software, and P<0.05 indicated statistical significance (18).

PPI construction

The PPI network was constructed by the Sting database (19). The visualization was then completed using Cytoscape (version 3.7.2). We used the CytoHubba plugin to screen the center genes. The top 10 genes were used as hub genes by the degree algorithm.

Expression of DEmiRs-Hub genes

The expression information of each target gene was analyzed using the UALCAN website that performs gene expression and survival analysis of tumor subgroups (20). The comparison between the tumor sample and the normal sample was statistically analyzed, and the logarithmic rank P values were observed in the database (21).

Construction of ceRNA network

The lncRNAs corresponding to DEmiRs were obtained by using the miRcode website which can predict miRNA targets. Based on the predicted DEmiR-DElnRs relationship and DEmiR-hub genes regulatory relationship, the ceRNA network was constructed by Cytoscape (22).

Results

Identification of DEmiR-target genes

In the GSE102286, GSE56036, GSE25508, GSE53882 and GSE29248 datasets, the up-regulated DEmiRs in NSCLC tissues were miR-21, miR-210, miR-200, miR-148a, miR-93, miR-141, miR-425, miR-193b, miR-301a, miR-720, miR-339-5p, miR-588 (Table 1), and the down-regulated DEmiRs were miR-30a, miR-145, miR-30d, miR-342-3p, miR-486-5p (Table 2). A total of 20 DEmiRs (lung adenocarcinoma tissue vs. normal tissue), 10 up-regulated and 10 down-regulated, were screened according to the criteria of |logFC| > 2. The up-regulated DEmiRs contained miR-21, miR-588, and the down-regulated DEmiRs contained miR-30a and miR-486-5p. In lung squamous
cell carcinoma, the up-regulated DE-miRs were miR-93 and the down-regulated DE-miR was miR-342-3p. 5558 and 2125 DE-miRs-target genes (14 up-regulated and 5 down-regulated DE-miRNAs) were predicted using the miRTarBase website, respectively. The volcano plot of DE-miRNAs in the five datasets has been shown in Figure.1 and Table 3.

**Functional enrichment analysis**

The functions of up-regulated DE-miRs target genes mainly include the regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism in Biological Process (BP); these target genes were expressed in the nucleus, cytoplasm, lysosome and exosomes in Cellular Component (CC); and transcription regulator activity, DNA binding and ubiquitin-specific protease activity in Molecular Function (MF) (Figure. 2A-2C). As shown in Figure.3A-3C, the functions of down-regulated DE-miRs target genes include regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism and regulation of cell cycle in the BP category; the CC showed that they were particularly expressed in the nucleus, cytoplasm, lysosome, exosomes and mitochondrion; and transcription factor activity, RNA binding, receptor signaling complex scaffold activity and Protein binding in MF category.

**Analysis of pathway enrichment**

For up-regulated DE-miRs (including lung adenocarcinoma) (Figure.3a, Figure.S1), enriched pathways include cancer focal adhesion pathway-LKB1 signaling events, DE-miRs-hub genes were also enriched in thrombin/protease-activated receptor (PAR) and death receptor pathway-TRAIL signaling pathway in lung squamous cell carcinoma (Figure.S2); and the concentrated pathways for down-regulated DE-miRs (including lung adenocarcinoma) (Figure.3b, Figure.S3) included plasma membrane estrogen receptor signaling and nectin adhesion Pathway, adenocarcinoma was enriched in the Alpha9 beta1 inte-
gulated DEmiRs-center genes were MYC, EGFR, HSP90AA1, PTEN, JUN, EP300, STAT3, HSPA4, KRAS and CCND1. The down-regulated DE-miRNAs-center genes were TP53, ACTB, MYC, CTNNB1, EGFR, PTEN, KRAS, JUN, EP300 and ESR1. MYC and TP53 had the highest degrees, which were 282 and 267, respectively. It is suggested that MYC and TP53 may be key target.

Through the constructed DEmiRs-hub genes network, we identified five hub genes (MYC, EGFR, PTEN, STAT3 and CCND1) that could be regulated by hsa-miR-21. Hsa-miR-30a could potentially target six (TP53, MYC, CTNNB1, EGFR, KRAS and EP300) in 10 hub genes (Figure 5). Therefore, hsa-miR-21 and hsa-miR-30a may be two important regulatory factors in NSCLC. The expression of target genes of has-miR-21 and has-miR-30a in NSCLC were further evaluate by the UALCAN website in Figure 6. The results showed that PTEN gene expression was particularly down-regulated in NSCLC. PTEN and KRAS may be the most potential target genes for has-miR-21 and has-miR-30a, respectively.

cRNA network construction

The cRNA network is shown in Figure 7. The DElncRs of has-miR-21 and has-miR-30a were predicted, and 27 DE-miRNAs-DElncRs relationships were obtained, including 7 DE-miRNAs and 10 DElncRs. We compared the regulatory relationship of these two miRNAs in DEmiR-target genes and the integration of their interactions with DE-miRNAs-DElncRs. A total of 40 relationships were identified, including 5 up-regulated DE-miRNAs (Has-miR-21, Has-miR-210, Has-miR-31, Has-miR-7 and Has-miR-96), 2 down-regulated DE-miRNAs (Has-miR-30a and Has-miR-451), 20 target genes and 10 DElncRs. Hsa-miR-145 can be regulated by multiple DElncRs including C10orf25, LINC00115, MIAT and LINC00472.

Prognostic analysis of DEmiRs

To determine the clinical relevance of DEmiRs, the overall survival (OS) of patients was assessed by the TCGA database. As shown in Figure 8, high expression of miR-21 was associated with poorer OS, and miR-30d and miR-145 had a good prognostic effect on patients. Therefore, the 3 DEmiRs (miR-21, miR-30d and miR-145) should be further investigated.

Discussion

NSCLC is the most prevalent histological type of lung cancer (2,3). However, the etiology of NSCLC has not been fully elucidated. Therefore, it is urgent to identify dysregulated miRNAs in NSCLC to improve lung cancer survival and provide more effective therapeutic targets for the clinic. MiRNAs have been shown to be oncogenic or suppressive factors in tumors (23,24), and the etiology of NSCLC is closely related to miRNA dysregulation. However, a systematic analysis of the role of miRNAs in NSCLC should be performed. According to bioinformatic analysis, 19 DEmiRs (14 up-regulated and 5 down-regulated DEmiRs) were consistent in five datasets. For the up-regulated DEmiRs-target genes were clearly concentrated in cancer focal adhesion pathway, especially LKB1 signaling events. Down-regulated DEmiRs-target genes were particularly concentrated in plasma membrane estrogen receptor signaling and nectin adhesion Pathway.
Hsa-miR-21, hsa-miR-141, hsa-miR-200b and hsa-miR-30a, hsa-miR-30d, hsa-miR-145 were expressed at high levels in DE-miRNAs-hub genes and ceRNA regulatory network, respectively, and most of the hub genes were probably regulated by hsa-miR-21, hsa-miR-141, hsa-miR-200b, hsa-miR-30a, hsa-miR-30d and hsa-miR-145.

Hsa-miR-21 is upregulated in expression in many cancers, including breast, liver and colon cancers. It targets a variety of oncogenes such as PTEN and PDCD4, thereby affecting invasion, migration and radiotherapy sensitivity in cancer (25). Overexpression of miR-21 may promote the proliferation and migratory effects of cervical cancer and nephroblastoma cells through the upregulation of the STAT3 signaling pathway (26). Elevated miR-21 levels in glioma tissues were associated with EGFR and VEGF expression (27,28). Our results suggest that miR-21 is an oncogene in NSCLC and NSCLC patients with high levels of miR-21 have a poor prognosis. PTEN, STAT3 and EGFR may be potential target genes of miR-21. miR-21 may be regulated by WDFY3-AS2 and promote NSCLC development by targeting STAT3 and EGFR. EMT-associated miR-200b acts in lung adenocarcinoma by targeting the ZEB1 gene (29). Only in mice with alcoholic fatty liver was miR-200b found to promote its transcription by targeting the JUN gene and binding to the SREBP1 promoter region, causing increased lipid accumulation (30). Our results suggest that miR-200b may promote the development of NSCLC through WDFY3-AS2 / miR-200 / JUN. miR-141 (miR-200 family) is mostly overexpressed in plasma exosomes of lung patients such as bladder and esophageal squamous carcinomas, affecting patient prognosis and TNM staging maligning, and correlates with tumor drug resistance (31). miR-141 has a pro-oncogene role in lung adenocarcinoma and promotes the proliferation and invasion of lung adenocarcinoma cells (32,33). In this study, miR-141 may promote the growth, migration and cell cycle of NSCLC cells by targeting the PTEN/PI3K/AKT signaling pathway.

Evidence suggests that hsa-miR-30 (hsa-miR-30a, hsa-miR-30d) is a key regulator of NSCLC development. It is involved in the regulation of EMT in breast cancer by targeting ZEB2 in response to TP53 stimulation (34). We investigated that hsa-miR-30a may be regulated by LINC00155 to promote NSCLC progression by targeting TP53, KRAS and CTNNB1, and the expression level of hsa-miR-30d was positively correlated with survival time in NSCLC patients. miR-145 inhibits the proliferation and invasion of lung and gastric adenocarcinoma by regulating the target genes EGFR and c-Myc (35). Although the expression level of miR-145-5p in cancer tissues and serum of patients with non-small cell lung cancer is controversial (36). This genetic phenomenon may also occur in other cancers (37-40). This study showed that the expression level of miR-145-5p was significantly reduced in NSCLC, which may regulate the development of NSCLC by targeting oncogenes such as EGFR and MYC. GO annotation and pathway enrichment analysis showed that these target genes were enriched in tumor pathways, especially the cancer adhesion pathway. Therefore, these target genes may be among the key genes in mirna regulation of NSCLC disease progression.

In conclusion, we identified nineteen DE-miRNAs (miR-21, miR-210, miR-200abc, miR-148a, miR-93, miR-141, miR-425, miR-193b, miR-301a, miR-720, miR-339-5p, miR-588, miR-30a, miR-145, miR-30d, miR-342-3p, miR-486-5p) as potential regulators of NSCLC progression. In addition, further studies on the corresponding regulatory networks will contribute to a deeper understanding of NSCLC and have important clinical implications for the early diagnosis, treatment and prevention of NSCLC.

Acknowledgements
The authors would like to express their gratitude to all participants in this study.

Authors’ contributions
LF was responsible for the reliability of the submitted data and drafted the article. SJ, CC, and CW performed the statistical analysis and interpretation of the data. SL, JH, LW and HZ were responsible for the evaluation and guidance of the full text, and LP provided final approval of the submitted version. All authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (82060054) and the Science and Technology Cooperation Project of Corps (2020BC003).

Availability of data and materials
The Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) gene expression profile GSE102286, GSE56036, GSE25508 and GSE101929 were used to for differential expression analysis and gene interactions by applying systems biology molecular analysis methods.

Competing interests
The authors declare that they have no conflict of interest.
Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

References


