Original Article
A panel of five-miRNA signature as a potential biomarker for predicting lung cancer

Fanlu Meng, Linlin Zhang

Department of Oncology, General Hospital of Tianjin Medical University, Tianjin, PR China
Received March 6, 2016; Accepted June 5, 2016; Epub September 15, 2016; Published September 30, 2016

Abstract: Lung cancer is one of the leading malignant tumors and the common cause of cancer-related deaths. In this study, in order to explore the pathogenesis of lung cancer, differentially expressed miRNAs (DEMs) were screened using high throughput data of lung cancer, combining with the function of the target genes and pathways. miRNA high-throughput data was downloaded from The Cancer Genome Atlas (TCGA) to screen miRNAs between lung cancer and normal tissues, the DEMs were subjected to perform principal component analysis and ROC curves of the important miRNAs were analyzed so as to judge the accuracy of cancer diagnosis by miRNA. Then target genes were predicted by Targetscan software, and the up and down-regulated target genes were performed function enrichment analysis by DAVID respectively. After that, pathways annotations were executed on the target genes of DEMs through KEGG database, and the genes involved in significantly related pathways were further researched. Finally, survival analysis was accomplished between DEMs and the patient’s survival time, and the miRNAs with prediction potentials were obtained. A total of 140 DEMs were screened, 58 DEMs were up-regulated including hsa-mir-519a-2, hsa-mir-522 and hsa-mir-520a; while 8 were down-regulated including hsa-mir-202, hsa-mir-675 and hsa-mir-323b. 5 miRNAs with diagnostic and preventive abilities were significantly associated with survival time. The selected important molecular targets may provide great clinical application value for the diagnostic and prognostic markers and targeted therapy, and it was expected to provide new ideas for the treatment of lung cancer.

Keywords: Lung cancer, differentially expressed genes, survival analysis

Introduction
Lung cancer is one of the malignant tumors and the leading cause of cancer-related death. It is reported that about 70% of patients with lung cancer show symptoms caused by locally advanced or metastatic disease [1, 2]. Although great advances have been made in lung cancer therapy, the average 5-year survival rates for all individuals with lung cancer still remains relatively poor (15% or so), possibly because lung cancer is often diagnosed at advanced stage and treatment options are limited [2, 3]. In order to improve the survival of patients with lung cancer, multiple studies on early diagnosis and treatments have been performed, however none have showed significant effects on the overall mortality [4, 5].

MicroRNAs (miRNAs) have gained increasing attention in recent years, as they have the potentials to regulate their target genes and control gene expression through translational repression and degradation. miRNAs are a class of small non-coding (18-25 nucleotides), naturally occurred endogenous RNA molecules that are thought to regulate other genes' expression [6, 7].

It has been suggested that the miRNAs are involved in various biologic processes, such as cell proliferation, differentiation, death and stress resistance. And mis-regulation of miRNAs can contribute to the development of human diseases including cancer [4, 7, 8]. Moreover, several studies have proved that the genome-wide expression profiling of miRNAs is significantly different between lung cancers and normal tissues [9, 10]. There were some specific miRNAs including hsa-miR-21, hsa-miR-146, hsa-miR-192 and so on were over-expressed in lung cancer samples, indicating these miRNAs may be considered as diagnostic or therapy markers [11]. Many miRNA expression studies in lung cancer cell lines are mainly focused on the let-7 family of miRNAs, which
miRNA biomarker for predicting lung cancer

acts as tumor suppressors by repressing cell proliferation [12]. Previous studies also identified various miRNA expression profiles associated with lung cancer survival [11, 13], however their underlying molecular mechanisms still remain poorly understand. It is of great importance to identify new and specific molecular markers that may contribute to the early diagnosis or prognosis of lung cancer.

In this study, we screened differentially expressed miRNAs (DEMs) between lung cancer and normal samples, and analyzed the target genes of these DEMs by DAVID. Through performing survival analysis between the DEMs and patients' survival time and constructing a mutual exclusivity signaling module, we aimed to find miRNAs that may contribute to the early diagnosis and therapy of lung cancer.

Materials and methods

miRNA chip data

The miRNA expression data and the corresponding patients' medical information were obtained from The Cancer Genome Atlas (TCGA) database, consisting of 567 samples (46 of normal samples and 521 of lung cancer samples) and medical information for 521 patients. The standardized miRNA data were Level three and the data were sequenced by IlluminaHiSeq system.

The miRNA data with zero expression values were removed from the standardized miRNA data. miRNA expression data at level three were downloaded, including 1046 commented miRNA expression values. As level three miRNA data had already been standardized in samples, then the data between samples were standardized using the generalized linear model in R language of Limma package so as to eliminate the batch effects.

Screening of DEMs

The DEMs between normal samples and lung cancer samples were screened by R software in the SAMR [14] package. The degree of expression differences was showed by logFC and p values, Log2FC indicated the differential degree of miRNAs between differentially expressed tumor samples and normal samples. The down- and up-regulated miRNAs were showed as logFC < -1 and logFC > 1 respectively, with FDR < 0.05. In addition, the principal component analysis method was applied to observe whether the normal samples could be distinguished from the lung cancer samples.

Survival analysis

All medical information for patients was summarized and subjected to perform statistical analysis so as to determine the cutoff value of medical information. The distribution of survival time and the differences of survival ability under various diseased states were studied and defined by Kaplan-Meier and log-rank methods respectively. The relationship between DEMs and patients' survival time was studied using a single variable Cox regression mode.
miRNA biomarker for predicting lung cancer

The original data were arranged sequentially as survival days, survival state (status, death was 1 while survival was 2), cumulative survival, standard error of the survival rate, cumulative event and number remaining of each sample. Eventually, the survival rates of patients were obtained.

Screening and verification of molecular markers for miRNA

The data were divided into training and testing cohorts, and the expression profiling of miRNAs in training group was analyzed by Cox regression model in order to acquire survival-related miRNAs (P < 0.01) and 5 miRNAs significantly associated with survival (P < 0.005). Patients’ survival abilities were predicted by a multivariable Cox regression model constructed based on these 5 miRNAs. As each miRNA parameter possessed a Cox regression correlation coefficient, thus every patient had a risk factor. The patients with higher risk scores showed weaker viability comparing to those with lower risk scores. The median risk score was considered as the boundary, the risk score above the boundary was defined as high-risk, while it was low-risk below the boundary. The survival time distribution for each variable was observed by Kaplan-Meier model, and the significances among various classifications under the same variable were determined by log-rank detection method.

Analysis of the target genes of miRNAs

The target genes of DEMs in Mirtarbase [3] database were extracted firstly, then they were subjected to do GO functional annotations and KEGG pathway analysis by DAVID [4] online tool. P value of function enrichment and P value of multiple corrections after testing (Benjamini correction) corresponding to each GO value could be calculated by DAVID software.

Analysis of interaction network of miRNA and target genes

The topological properties of protein function Network were analyzed by Network Analyzer plugin in Cytoscape software [7], and the modularization of network function was fulfilled by Clusterone. The first 5 modules with p-values less than 1.0 e-05 were selected to perform function analysis.

Construction of mutual exclusivity signaling modules

Exome-seq (517 samples) and SNP (518 samples) level three data in TCGA were downloaded, MutSig software was utilized to identify mutation sites (Q-value < 0.1) while GISTIC software was applied to obtain fragments of DNA amplification and deletion. Then the mutual exclusivity signaling module was constructed using MEMo algorithm method (the detailed flowchart was shown in Figure 1).

Results

Expression difference of miRNA among various samples

There were 1046 miRNA expression values for 187 lung cancer and 13 normal samples. Through principal component analysis and cluster analysis, we could observe the difference between normal and cancer samples, and they were separated available (Figure 2).
miRNA biomarker for predicting lung cancer

Table 1. Six significantly differentially expressed miRNAs

<table>
<thead>
<tr>
<th>Type</th>
<th>miRNA ID</th>
<th>logFC</th>
<th>P_value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>up regulated miRNA</td>
<td>hsa-mir-519a-2</td>
<td>3.87</td>
<td>1.88E-33</td>
<td>1.97E-30</td>
</tr>
<tr>
<td></td>
<td>hsa-mir-522</td>
<td>3.66</td>
<td>3.34E-32</td>
<td>1.75E-29</td>
</tr>
<tr>
<td></td>
<td>hsa-mir-520a</td>
<td>3.45</td>
<td>1.02E-31</td>
<td>3.54E-29</td>
</tr>
<tr>
<td></td>
<td>hsa-mir-202</td>
<td>-1.7</td>
<td>3.04E-13</td>
<td>1.18E-11</td>
</tr>
<tr>
<td>Down regulated miRNA</td>
<td>hsa-mir-675</td>
<td>-1.47</td>
<td>1.84E-12</td>
<td>6.63E-11</td>
</tr>
<tr>
<td></td>
<td>hsa-mir-323b</td>
<td>-1.39</td>
<td>1.28E-10</td>
<td>3.82E-09</td>
</tr>
</tbody>
</table>

Survival analysis

As the censoring rate (61.54%) was relatively high (Table 2), indicating that there were still many patients alive at the end of the study or they directly went away. A survival curve was drawn according to the patients’ survival time (Figure 3). From which, we know that although there was a high cutoff value, the medical information (Table 2) was enough and suitable to predict the biological targets of miRNAs in the following study.

Kaplan and log-rank methods were taken to verify the relationship between the survival time and various variables (including age, sex, T stage, R stage, M stage, N stage, and tumor stage), the results were demonstrated in Figure 4. Significant differences were found between different tumor status and the overall survival time. The results in Figure 4 demonstrated that univariate cox regression model could be applied to analyze the differences between miRNA and survival time of patients under various states. The correlation coefficients were demonstrated in heat map Figure 5. The molecular markers of miRNAs were selected with P < 0.1 and there were significant results at least in two separate categories.

Through univariate survival analysis, the miRNAs associated with patients’ overall survival under different tumor status were filtered, as shown in Figure 6. 5 miRNAs including hsa-mir-516-a-1, hsa-mir-519-a-1, hsa-mir-675, hsa-mir-323-b and hsa-mir-935 that may be considered as biomarkers were identified.

Construction of predictive disease model

Cox multivariate regression model was utilized to construct a mathematical model for miRNA and survival time using data in training group, the prognosis formula was Prognostic score = (-11.4 × expression level of hsa-mir-516a-1)+(-7.993 × expression level of hsa-mir-519a-1)+(+13.176 × expression level of hsa-mir-675)+(11.192 × expression level of hsa-mir-323b)+(-8.077 × expression level of hsa-mir-935). Among which, 4 miRNAs belonged to risky type, 1 belonged to protective type.

Log-rank test was taken to analyze the differences between patients in high-risk and low-risk groups, the results were shown in Figure 7. Significant differences between high-risk and low-risk patients were found in training group, also the same with test group, which indicated this model was feasible in actual prediction process. The prediction results of miRNA in patients were displayed in Figure 7.

Analysis of the target genes for the 5 five miRNAs

2147 target genes were predicted by Targetscan prediction software. These target genes were subjected to conduct GO annotation and KEGG pathway analysis by David (Figure 8). And GO annotation results demonstrated the main functions of these target genes were enriched in ion binding and cation binding; while KEGG analysis showed they were mainly associated with Graft-versus-host disease and Folate biosynthesis.

The calculation of Network topology was shown in the following figures, the node degree distribution (Figure 9A), the shortest path length distribution (Figure 9B), close to the center (Figure 9C), topological degree (Figure 9D). Through these parameters, we found the network node distribution of protein of the differential genes conformed to power-law distribution form, and
miRNA biomarker for predicting lung cancer

Table 2. Summary of the medical information for lung cancer patients

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Category</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, no (%)</td>
<td>&lt; 60</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>≥ 60</td>
<td>361</td>
</tr>
<tr>
<td>Gender, no (%)</td>
<td>Male</td>
<td>242</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>279</td>
</tr>
<tr>
<td>Vital status</td>
<td>Alive</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Dead</td>
<td>65</td>
</tr>
<tr>
<td>Tobacco_smoking_history_indicator</td>
<td>Current reformed smoker for &lt; or = 15 years</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>Current reformed smoker for &gt; 15 years</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>Current Reformed Smoker, Duration Not Specified</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Current smoker</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>Lifelong Non-smoker</td>
<td>75</td>
</tr>
<tr>
<td>Lymph node involvement, no (%)</td>
<td>N0</td>
<td>334</td>
</tr>
<tr>
<td></td>
<td>N1</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>N3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>NX</td>
<td>11</td>
</tr>
<tr>
<td>M stage, no (%)</td>
<td>M0</td>
<td>352</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>M1a</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>M1b</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>MX</td>
<td>140</td>
</tr>
<tr>
<td>T stage, no (%)</td>
<td>T0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>T1</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>281</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>19</td>
</tr>
<tr>
<td>R stage, no (%)</td>
<td>R0</td>
<td>346</td>
</tr>
<tr>
<td></td>
<td>R1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>RX</td>
<td>26</td>
</tr>
</tbody>
</table>

Figure 3. The overall survival status changes of patients. The survival curves were drawn according to the patients’ survival time. The vertical axis showed survival time and the horizontal axis showed survival rates, the red crosses on curve were cut off values. See from the figure, patients’ viability decreased gradually over time.

had the characteristics of small world structure (short average shortest path and the large average aggregation coefficient).

The miRNAs (P < 0.05) with top 3 node degrees were selected by software to perform function analysis. In the down-regulated miRNAs, the top 3
miRNAs were hsa-miR-483-5p (173 target genes), hsa-miR-675 (157 target genes) and hsa-miR-139-3p (140 target genes); in the upregulated miRNAs, the top 3 miRNAs were hsa-miR-598 (104 target genes), hsa-miR-625 (97 target genes) and hsa-miR-187 (94 target genes).

Mutual exclusivity signaling module construction

A total of 1315 mutagenic genes were identified through MutSig software. The mutation situation of Driver gene in different samples
miRNA biomarker for predicting lung cancer

Figure 10. After analysis of GISTIC2 software, 74 sites with significant differences of copy number were achieved, including 41 deletion changes, 33 amplification changes. Mutations presenting mutual exclusivity (Figure 11) indicated that a gene with mutations in the signaling pathway could induce tumor genesis, and mutations of PIK3CA, PIK3R1, PTEN, EGFR or FRBB2 could cause cancer.

Discussion

In the study, a total of 66 differentially expressed miRNAs between normal and lung cancer samples were screened. And according to the medical information of patients, 5 miRNAs including hsa-mir-516-a-1, hsa-mir-519-a-1, hsa-mir-675, hsa-mir-323-b and hsa-mir-935 that could be considered as molecular targets were predicted. Finally, through construction of
miRNA biomarker for predicting lung cancer

**Figure 7.** Analysis of the risk coefficients predicted by the 5 miRNAs. A: Risk coefficients analysis of training set data; B: Heat map for expression levels of 5 miRNAs in training set; C, D: The relationship between the patients' survival time and risk coefficients; E, F: Distribution of risk coefficients.

**Figure 8.** GO and KEGG analysis of the target genes of differentially expressed miRNAs. A: GO analysis of the target gene. B: KEGG analysis of the target gene. The abscissa represented significant degree, while the vertical axis represented functional annotation, the greater the significant degree, the greater correlation between the target gene and annotation.
miRNA biomarker for predicting lung cancer

Figure 9. Structure analysis of the Network topology. A: The node degree distribution; B: The shortest path length distribution; C: Closeness to the center; D: Topological degree.
A mutual exclusivity signaling module, we identified that PIK3CA, PIK3R1, PTEN, EGFR or FRBB2 genes may be closely associated with the pathogenesis of lung cancer and they could be served as molecular targets for the prognosis and treatments of lung cancer.

In recent years, more and more miRNAs have been demonstrated to be diagnostic, prognostic markers and therapeutic target in several types of cancers such as lung cancer and so on [11]. For example, Masahiro et.al reported that up-regulation of miR-21 was found to be associated with EGFR (epidermal growth factor receptor) mutations, suggesting it may be a potential therapeutic strategy in combination with EGFR-TKI (tyrosine kinase inhibitors) treatment. They also improved that over-expression of miR-21 was an early event in lung cancer development. Their findings showed miR-21 was one of major miRNAs contributed to lung carcinogenesis [15].

In our study, we identified 5 miRNAs that may be considered as diagnostic and prognostic markers. Among which, expression of miR-675 has been shown in recent studies to be up-regulated in several cancers like gastric cancer [16, 17], hepatocellular cancer [18] and colorectal cancer [19]. While some other studies found miR-675 was down-regulated in metastatic prostate cancer cells. He et. al also found in lung cancer samples the expression level of miR-675 was significantly reduced comparing with that of normal samples [20]. Furthermore, down-regulation of miR-675 could promote cell growth, proliferation and migration; whereas up-regulation of miR-675 exhibited the contrary effects. All these findings suggested the potential for miR-675 as a therapeutic target in lung cancer.

In order to further analyze the selected miRNAs and explore the possible pathogenesis of lung cancer, the target genes of these miRNAs were also predicted. We achieved a total of 2147 target genes, and the main functions of these target genes were associated with ion binding and cation binding through GO annotation; while the major pathways of these genes were enriched in Graft-versus-host disease and Folate biosynthesis by KEGG analysis. Through construction of a mutual exclusivity signaling module, we understood that mutations of any PIK3CA, PIK3R1, PTEN, EGFR or FRBB2 gene may cause lung carcinogenesis. High frequencies of mutations of PIK3CA (phosphatidylinositol 3-kinase catalytic subunit) gene have been found in many human cancer patients including brain [21], breast [22] and lung cancer patients [23, 24]. In addition, it was reported approximately 4% of lung cancer samples harbored mutations in the exon 9 of the PIK3CA gene, indicating that PIK3CA may play a pivotal role during the oncogenesis of lung cancer [24].

PTEN: MMAC1 (Phosphatase and Tensin Ho-mologue deleted on chromosome 10: Mutated in Multiple Advanced Cancers 1) gene was also a candidate tumor suppressor gene in various human cancers, such as gliomas and renal carcinomas [25, 26]. Several studies have reported that PTEN:MMAC1 gene was changed in lung cancer [27], deletion and point mutations of PTEN:MMAC1 was identified in nearly 8% and 9% of small cell lung cancer tissues respectively. Taken all these into consideration, we may predict that alteration of the PTEN:MMAC1 gene may be involved in the development, progression and metastasis of lung cancer.
Another gene, EGFR, was one of the first molecular targets used for the therapy of lung cancer [27]. EGFR was often found to be up-regulated in non-small-cell lung cancer and other tumors [28], however mutations of EGFR occurred almost exclusively in lung adenocarcinomas [29, 30]. EGFR mutation rendered cancer cells dependent on the EGFR pathway, therefore making them very sensitive to EGFR inhibitors like anti-EGFR antibodies. For example, cetuximab, an anti-EGFR antibody, has been evaluated in phase 3 clinical trials for treatments of patients with lung cancer [27, 31].

To conclude, we screened differentially expressed miRNAs between normal and lung cancer samples in this study and selected 5 miRNAs including hsa-mir-516-a-1, hsa-mir-519-a-1, hsa-mir-675, hsa-mir-323-b and hsa-mir-935 with predictive and diagnostic potentials in lung cancer. Moreover, we analyzed the target genes of these 5 miRNAs, which may provide new insights into cancer development and find new candidate biomarkers and therapeutic targets. Although further studies are still needed, the results in this study could help better understand the pathogenesis of lung cancer.

Disclosure of conflict of interest

None.

Address correspondence to: Fanlu Meng, Department of Oncology, General Hospital of Tianjin Medical University, No. 154 Anshan Road, Heping District, Tianjin, PR China. Tel: 8613652076473; E-mail: fanlumengfjm@163.com

References

miRNA biomarker for predicting lung cancer


