Original Article
Screening of diagnostic biomarkers for lung cancer by bioinformatics analysis

Bao Liu\textsuperscript{1}, An Yan\textsuperscript{1}, Leiguang Ye\textsuperscript{1}, Yina Gao\textsuperscript{1}, Fang Liu\textsuperscript{1}, Weibin Jing\textsuperscript{2}, Limin Zhang\textsuperscript{3}, Yan Yu\textsuperscript{1}, Li Zhong\textsuperscript{1}

\textsuperscript{1}Department of Oncology, Harbin Medical University Cancer Hospital, Haerbin 150081, China; \textsuperscript{2}Department of Burn, Hei Long Jiang Provincial Hospital, Haerbin 150036, China; \textsuperscript{3}Head and Neck Surgery, Harbin Medical University Cancer Hospital, Haerbin 150081, China

Received May 11, 2015; Accepted March 5, 2016; Epub February 15, 2017; Published February 28, 2017

Abstract: Lung cancer was a leading cause of cancer-related death, this study aimed to explore target genes and specific biomarkers associated with the lung cancer for early diagnosis and treatment. Firstly, the gene expression profile GSE32863 was downloaded from Gene Expression Omnibus database (GEO), which included 15 lung cancer tissue samples and 10 normal lung tissue samples. In addition, the miRNA microarray profile GSE17681 with 10 lung cancer and 10 normal tissue samples was also downloaded. Then the probe-level data were pro-processed by Significance Analysis of Microarray (SAM), and the differentially expressed genes (DEGs) and miRNAs were screened with multi-test package in R language. The selected DEGs were further subjected to functional enrichment analysis using DAVID online tool. Following that, the verified target genes based on screened miRNAs were selected from miRTarBase and miRecords databases. Then miRNA-target gene regulation network was constructed to identify symbolic miRNAs of lung cancer. miR-29a with target genes such as FGG and COL4A1, miR-7 with SLC7A5 and miR-222 with ICAM-1 were found differentially expressed in lung cancer compared with the normal samples. The discovery of meaningful miRNAs and their differentially expressed target genes has the potential to be used in clinic for diagnosis and treatment of lung cancer.

Keywords: Lung cancer, differentially expressed gene, miRNA

Introduction

Lung cancer was the most common cause of cancer-related death worldwide especially in industrialized countries, despite improvements in diagnosis and therapy, the overall 5-year survival was still 15% [1], possibly because lung cancer was often diagnosed at advanced stage and treatment options were limited. Smoking was the major risk factor for lung cancer [2], which can alter the activity of chemo-preventive drugs [3, 4], stimulate the clearance of selected targeted anticancer therapies [5], reduce the efficacy of cancer treatment, and increase the risk of second primary tumors. Although other factors, such as environmental exposure (e.g., chemicals, physical agents, and radiation), clinical history of lung diseases (e.g., chronic bronchitis, emphysema, pneumonia, and tuberculosis; ref. [6], familial tumor history [7], or diet [8, 9], may also be associated with the development of lung cancer [10].

As detection of early lung cancer based on symptoms was not very effective, so studying the pathogenesis of lung cancer and exploring biomarker and new therapeutic targets had important practical significance [11].

Several molecular markers associated with lung cancer progression have been identified, including TGF, MET, TP53, HIF1A, APC, KRAS, EGFR and so on [12].

In recent years, MicroRNAs (MiRNAs) have become a hot research topic. MiRNAs were post-transcriptional regulators of gene expression [13]. The primary miRNA (primiRNA) was transcribed from the genome by RNA polymerase II and processed by Drosha, yielding the precursor miRNA transcript [14]. It was reported expression of at least 20-30% of human protein-coding genes were modulated by miRNAs [15].
Studies found miR-486, miR-30d, miR-1 and miR-499 from the serum may serve as noninvasive predictors for the overall survival of NSCLC [16]. With the penetrating study of miRNA in complex diseases, it was found that miRNA plays a great role during disease development [17, 18]. A recent report indicated disease may be cured by importing exogenous synthetic miRNA, therefore it was of great significance to identify the disease-related miRNAs and take them as targets for future therapies [19].

With the development of DNA microarray technology, it was now possible to screen many genes with expression alterations simultaneously. As a global approach, DNA microarray analysis was applied to investigate physiological mechanisms in health and disease [20]. Gene expression profiling based on microarrays was a robust and straightforward way to study the molecular features of cancer at a systemic level.

In this study, we aimed to identify target genes and specific biomarkers for identification and treatment of lung cancer with DNA microarray, and construct a miRNA co-expression network using miRNA micro-array data obtained by high-throughput means, so as to better understand the molecular mechanisms and explore new therapy strategies.

Materials and methods

Data resources and preprocessing

The gene expression profile GSE32863 [21] was downloaded from GEO (Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/)) database, which included 15 lung cancer tissue samples from previously untreated patients and 10 normal lung tissue samples, and miRNA microarray profile GSE17681 [22] with 10 lung cancer and 10 normal tissue samples was also downloaded. The test platform was GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array). The probe-level data in CEL files were firstly converted into expression measures by ReadAffy of Affy package of R software. Then, we imputed the missing data [23] and normalized the quartile data [24]. The probes without gene annotation or having more than one gene annotation were filtered out; the average value of multiple probes corresponding to the same gene was calculated as a unique value of the gene.

Screening of differentially expressed genes (DEGs)

The multistep package [25] in R language was used to identify DEGs between normal and lung cancer tissues. The Benjamin and Hochberg (BH) method [26] was used to adjust the raw P-values into false discovery rate (FDR) so as to avoid the multi-test problem which may cause too many false positive results. The FDR < 0.05 and |logFC| > 1 were used as the cut-off thresholds.

Disease-related miRNA screening

Pearson correlation coefficients (that was co-expression intensity values) for each two miRNAs could be calculated by formula, and then miRNA co-expression networks were filtered out through a given threshold 0.7. A specific miRNA co-expression network was constructed with expression profiles of miRNA.

Enrichment analysis for DEGs and disease-related miRNA

DAVID (the Database for Annotation, Visualization and Integrated Discovery) was a software with built-in rich graphical display, clustering the significant gene collections according to their functions, and it had abundant public database links [27]. The functional enrichment analysis of specific DEGs was performed by DAVID based on hypergeometric distribution algorithm (P < 0.05).

If the set of disease-related miRNA screened was M, a known disease-associated miRNA set was N, fisher test was used to test whether the screened disease-related miRNA could enrich the already known miRNA based on P-value, then DAVID was used to do GO [28] functions and KEGG pathway enrichment analysis, so as to validate the function of the screened miRNA.

Target gene selection of disease-related miRNAs

According to different algorithms, each miRNA had different corresponding target genes. In order to find target genes with high confidence level, the mutual genes from two miRNA data-
Figure 1. Box graph of the standardized expression profile data. A. The blue, pink, orange and green columns represent gene expression profiles of adenocarcinomas, small cell lung cancer, squamous cell carcinomas and normal tissue samples, respectively. B. The blue and pink columns represent miRNA expression profiles of normal and unclassified lung cancer tissue samples, respectively.
bases (miRecords and miRTarBase) were screened as the target genes of miRNAs.

MiRecords was a database collecting target interactions of animal miRNAs, including the target genes verified by artificial collection experiments [29]. MiRTarBase was also a database which comprehensively collects miRNA targets based on experimental verification [30].

**Correlation analysis between DEGs and disease-related miRNAs**

Target genes of disease-related miRNAs were firstly corresponded to the screened DEGs. According to the corresponding DEGs, unclassified miRNAs associated with lung cancer were categorized. Then miRNA-target gene regulation network was constructed to determine symbolic miRNAs of lung cancer. The reliability score of interaction relations more than 0.4 was regarded as an index to screen the reliable interaction relationship.

**Enrichment analysis of specific DEGs**

DAVID was utilized to do enrichment analysis of specific DEGs, finally 9 GO functional nodes were obtained, such as regulation of cell morphogenesis, regulation of neuron differentiation and negative regulation of cellular component organization and so on, the results were listed in Table 2. The Gene Ontology enrichment map is shown in Figure 2.

**Target gene screening of miRNAs and correlation analysis with DEGs**

The miRNAs information and the corresponding target genes were downloaded from miRecords and miRTarBase database. As shown in Table 3, the 38 screened target genes of differentially expressed miRNAs and the common part between target genes and DEGs were identified successfully. Then the regulation network for miRNA-target gene was constructed, the results were presented in Figure 3.
We could see that hsa-miR-29a was differentially expressed with target genes such as FGG and COL4A1, has-miR-7 with its target gene SLC7A5 and hsa-miR-222 with target gene ICAM-1 were identified. In addition, enrichment analysis of target gene function found that the disease-related miRNAs were mainly associated with cell adhesion and metabolic, regulation of cell morphogenesis, cell cycle process, regulation of cell proliferation and so on, which indicated that there was potential link between lung cancer and these function changes of miRNAs.

Discussion

Lung cancer was the leading cause of cancer-related mortality worldwide and has become the largest threat to human health [31, 32]. Therefore, there was an urgent need to explore the mechanism of lung cancer and explore novel potential diagnostic and therapeutic targets. In this study, we linked DEGs with the disease-related miRNAs and concluded that has-miR-7 with SLC7A5, hsa-miR-222 with ICAM-1 and hsa-miR-29a with target genes such as FGG and COL4A1, etc. were differentially expressed in lung cancer, which indicated that these target genes may be considered as potential biomarkers for the early diagnosis and future therapy of lung cancer caused by smoking.

MiRNA was a class of non-coding small RNA molecules, with length of about 22 nucleotides. It combined with 3 untranslated region of its target gene through the silence complex RNA-induced, leading to degradation of target mRNA or preventing translation of target miRNA. Since the first discovery of miRNA in C. elegans (Caenorhabditis elegans) in 1993, a large number of studies have shown that miRNA can part in cell growth, differentiation, proliferation, apoptosis, stress response and other biological processes by fine regulation of gene expression. With the deep study of complex diseases, it was found miRNA plays an important role in disease [19, 33].

It was reported Hsa-miR-7 played an important role in apoptosis (ACIN1, BAD, CASP8, CRYAA, GLO1, HSPA5, INHA, PCSK6, RELA, UBE4B),
cell cycle (NUSAP1, STK11) and cell proliferation and differentiation (ALOX15B, TBX2) with its target genes [34]. In addition, its target gene SLC7A5 was part of a two-protein complex with SLC3A2, the heavy chain of a neutral amino-acid transporter implicated in nutrient transport at the blood-brain barrier and has been noted to be differentially expressed between adenocarcinomas and squamous cell lung carcinomas [35, 36]. SLC7A5 protein expression was positive in SCLC (small cell lung cancer) tumors [37], and suppression of SLC7A5 resulted in an increasing percentage of transfected cells in the G1 phase over time and a concomitant decrease in the percentage of cells in the G2/M phase. Therefore we concluded that hsa-miR-7 cooperated with SLC7A5 to regulate in lung cancer.

Hsa-miR-222 also found to play a critical role in cell cycle and multiple biological processes [38]. It has been reported to be involved in lung cancer metastases in vitro [39] and significantly down-regulated in the diseased serum samples [40]. ICAM-1 (Intercellular adhesion molecule-1), one target gene of hsa-miR-222 identified in this study, was a cell adhesion molecule with a key role in inflammation and immunosurveillance. It has been implicated in carcinogenesis by facilitating instability of the tumor environment [41] and increases significantly in NSCLC (non-small cell lung cancer) patients compared to healthy individuals [42]. The specific miRNAs and differentially expressed target genes found in lung cancer may be unique biomarkers.

Moreover, from GO enrichment analysis of DEGs we found interleukin 6 (IL-6) and interleukin 8 (IL-8), which were expressed in pre-malignant epithelial cells, and their expression were associated with a poor prognosis in lung cancer patients [43, 44]. Evidence showed that inflammatory mediators contributed to the pathogenesis of many human cancers, including lung cancer, as under inflammatory stress, IL-6 and IL-8 participated in tumorigenesis by acting directly on lung epithelial cells via signaling through the nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) pathway [45]. In addition, IL-6 and IL-8 were expressed by lung cancer cells and acted in an autocrine and/or paracrine fashion to stimulate cancer cell proliferation [46, 47], migration, and invasion [48].

To conclude, in the present study, we explored the DEGs and differentially expressed miRNAs of lung cancer caused by smoking, and compared the DEGs with the target genes of the miRNAs, the results showed that IL6, IL8, COL4A1, ICAM-1 and so on may have potential to be used as biomarkers for the early diagnosis and future target treatment for lung cancer.

Although great efforts have been made in the research of lung cancer, yet it is not enough to fully understand the pathogenesis of lung cancer and effective diagnostic and therapeutic strategies are needed to be developed.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-199b-3p</td>
<td>KRT7</td>
</tr>
<tr>
<td>hsa-miR-222</td>
<td>ICAM1</td>
</tr>
<tr>
<td>hsa-miR-29a</td>
<td>FGG</td>
</tr>
<tr>
<td>hsa-miR-29b</td>
<td>FGG</td>
</tr>
<tr>
<td>hsa-miR-29c</td>
<td>FGG</td>
</tr>
<tr>
<td>hsa-miR-7</td>
<td>SNCA</td>
</tr>
<tr>
<td>hsa-miR-29a</td>
<td>BCL2</td>
</tr>
<tr>
<td>hsa-miR-29b</td>
<td>BCL2</td>
</tr>
<tr>
<td>hsa-miR-29c</td>
<td>BCL2</td>
</tr>
<tr>
<td>hsa-miR-210</td>
<td>CASP8B2</td>
</tr>
<tr>
<td>hsa-miR-210</td>
<td>CBX1</td>
</tr>
<tr>
<td>hsa-miR-126</td>
<td>CCNE2</td>
</tr>
<tr>
<td>hsa-miR-29b</td>
<td>COL1A1</td>
</tr>
<tr>
<td>hsa-miR-29c</td>
<td>COL1A1</td>
</tr>
<tr>
<td>hsa-miR-29a</td>
<td>COL4A1</td>
</tr>
<tr>
<td>hsa-miR-29b</td>
<td>COL4A1</td>
</tr>
<tr>
<td>hsa-miR-29c</td>
<td>COL4A1</td>
</tr>
<tr>
<td>hsa-miR-29a</td>
<td>DNMT1</td>
</tr>
<tr>
<td>hsa-miR-126-5p</td>
<td>SOX4</td>
</tr>
<tr>
<td>hsa-miR-29a</td>
<td>SPARC</td>
</tr>
<tr>
<td>hsa-miR-29c</td>
<td>SPARC</td>
</tr>
<tr>
<td>hsa-miR-29c</td>
<td>SRSF10</td>
</tr>
<tr>
<td>hsa-miR-29c</td>
<td>TDG</td>
</tr>
<tr>
<td>hsa-miR-19a</td>
<td>TGFB1</td>
</tr>
<tr>
<td>hsa-miR-30e</td>
<td>UBE2I</td>
</tr>
<tr>
<td>hsa-miR-222</td>
<td>KIT</td>
</tr>
<tr>
<td>hsa-miR-29a</td>
<td>GLUL</td>
</tr>
</tbody>
</table>
Biomarkers for lung cancer

Disclosure of conflict of interest

None.

Address correspondence to: Yan Yu, Department of Oncology, Harbin Medical University Cancer Hospital, 150 Haping Road, Nangang District, Harbin 150081, Heilongjiang Province, China. Tel: +86-45186298285; Fax: +8645186298285; E-mail: 153020426@163.com

References

Biomarkers for lung cancer


Biomarkers for lung cancer


