Exploring Biomarkers of Coronary Heart Disease from Gene and MicroRNA Expression Profiles

Austin H Chen*, Shu-Yi Huang, and Pei-Shan Hong

Tzu-Chi University, achen@mail.tcu.edu.tw
Tzu-Chi University, zxc781215@yahoo.com.tw
Tzu-Chi University, sour1031@gmail.com

Abstract

According to the statistics from World Health Organization, one third global populations die from heart diseases; among these diseases, coronary heart disease (CAD) is one of the main causes. In recent years, biomarkers are gradually playing the key criteria in the clinical assessment of heart diseases. Recent studies reveal that microRNAs play important regulators of development and stress responsiveness of heart. Discovering microRNA biomarkers of heart disease is an increasingly aware issue. In this paper, we used statistic methods, functional similarity scores, and target prediction systems to discover the potential microRNA and gene CAD biomarkers. In summary, we success identified 8 microRNA biomarkers (has-mir-663, has-mir-122, hsa-mir-10a, has-mir-132, has-mir-20a, has-mir-584, has-mir-181d, and has-mir-21) and 5 gene biomarkers (ABCC6, ADD1, TNFAIP3, APOB48R, and LDLRAD1) that have biological evidence to heart disease.

Keywords: Biomarker, microarray expression profiles, microRNA, data mining, feature selection, heart disease

1. Introduction

MicroRNAs are a small and noncoding RNAs that control expression of complementary target mRNAs. In many current studies, microRNAs have been found to play important roles in biological processes, including cellular differentiation, apoptosis and proliferation [1], also in synaptic plasticity and cardiovascular development [2]. In addition, the microRNAs expression changes are related to disease pathogenicity including cancer and cardiovascular disease.

Recently, MicroRNAs were found to exist in peripheral blood flow [3]. Although many nucleic acid RNA enzyme activity with a high degree of sensitivity, micro-RNA in plasma or serum, however, was found to have strong stability [4][5]. Functional studies have shown a role for microRNAs in cardiac fibrosis, hypertrophy, angiogenesis, and heart failure.

Circular microRNAs can be easily detected in serum or plasma of patients. It suggests that microRNAs may fulfill biological functions outside the cell and serve as potential biomarkers for diseases. Table 1 summaries some current studies in the roles of microRNAs related to heart diseases, such as acute myocardial infarction, viral myocarditis, and heart failure [6].

MicroRNAs were considered to act as intracellular RNAs to control gene expression on a posttranscriptional level. Recent studies demonstrated that microRNAs can be detected in circulating blood and may be useful as biomarkers for disease. However, how to identify the microRNA biomarkers from gene expression profiles and microRNA expression profiles remain as a key challenge. In this paper, we propose three methods that can identify both gene biomarkers and microRNA biomarkers from gene expression and microRNA expression profiles. These three methods are functional similarity method, PITA network method, and miRTarBase network method.
Table 1. Roles of circulating microRNAs in various heart diseases

<table>
<thead>
<tr>
<th>Disease type</th>
<th>Up-regulated microRNAs</th>
<th>Down-regulated microRNAs</th>
<th>References Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute myocardial infarction (AMI)</td>
<td>miR-1, miR-133a, miR-133b, miR-499-5p</td>
<td>miR-122, miR-375</td>
<td>[7]</td>
</tr>
<tr>
<td>Acute myocardial infarction (AMI)</td>
<td>miR-208a, miR-1, miR-133a, miR-499</td>
<td></td>
<td>[8]</td>
</tr>
<tr>
<td>Acute myocardial infarction (AMI)</td>
<td>miR-499</td>
<td></td>
<td>[9]</td>
</tr>
<tr>
<td>Acute myocardial infarction (AMI)</td>
<td>miR-1</td>
<td></td>
<td>[10]</td>
</tr>
<tr>
<td>Acute myocardial infarction (AMI)</td>
<td>miR-208b, miR-499, miR-133a</td>
<td></td>
<td>[11]</td>
</tr>
<tr>
<td>Viral myocarditis (VM)</td>
<td>miR-208b, miR-499</td>
<td></td>
<td>[12]</td>
</tr>
<tr>
<td>Heart failure acute</td>
<td>miR-499, miR-122</td>
<td></td>
<td>[11]</td>
</tr>
<tr>
<td>Heart failure</td>
<td>miR-423-5p</td>
<td></td>
<td>[12]</td>
</tr>
</tbody>
</table>

2. Methods

2.1. Flowchart and data sources

![Flowchart](image)

Figure 1. Flowchart in this paper

Figure 1 displays the flowchart we propose in this paper. First, we select one mRNA expression profile and one microRNA expression profile of heart disease from Gene Expression Omnibus (GEO). The top significant genes and microRNAs from these two datasets are selected using the t-Test method. The significant microRNAs are further screened by the q-value method based on minimum false discovery rate (FDR). By calculating the functional similarity, we can identify the potential microRNA biomarkers. We then identify the target genes related to these potential microRNA biomarkers. The common genes are retrieved from both these target genes and significant genes obtained from gene expression dataset. We then use the Fisher’s inverse chi-square method that can do the meta-analysis
of these common genes. Finally, by comparing with the biological evidence, we can identify both the potential gene biomarkers and microRNA biomarkers.

In this paper, two kinds of datasets were selected from Gene Expression Omnibus. One dataset is gene expression profile and the other dataset is microRNA expression profile. Table 2 displays the characteristics of these two datasets.

<table>
<thead>
<tr>
<th>GEO Number</th>
<th>Feature Type</th>
<th># of Patients (Heart disease/Health)</th>
<th># of Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE12288</td>
<td>mRNA</td>
<td>110/112</td>
<td>22283</td>
</tr>
<tr>
<td>GSE28858</td>
<td>microRNA</td>
<td>12/12</td>
<td>1146</td>
</tr>
</tbody>
</table>

2.3. Feature selection – the q-value method

In microarray data set, thousands of features are often tested against some null hypothesis, where a number of features are expected to be significant. A p-value threshold of 5% yields a false positive rate of 5% among all null features in the data set. Since p-value is based on the false positive rate, a p-value cutoff says little about the content of the features actually called significant. The q-value is similar to the well known p-value, except it is a measure of significance in terms of the false discovery rate rather than the false positive rate. The q-value approach avoids a flood of false positive results.

The q-value was first proposed by Storey and Tibshirani in 2002 [13]. The significant features are ranked by their q-values that are calculated based on the minimum false discovery rate (FDR).

\[
q\text{-value} = \min \text{FDR}(t) = \frac{\text{tm}p_i}{#(p_\leq t)}
\]

Where i, pi, t, and FDR represent feature i, p-value of feature i, threshold value, and false discovery rate individually. m is the number of features, \( \pi \) is true null distribution.

The q-value provides a measure of each feature’s significance, automatically taking into account the fact that thousands are simultaneously being tested. In contrast to p-value, a q-value threshold of 5% results in a FDR of 5% among the significant features, directly provide a meaningful measure among the features called significant. Because significant features will likely undergo some subsequent biological verification, a q-value threshold can be used as the proportion of significant features that turn out to be false leads.

2.4. Human microRNA functional similarity

Genes with similar functions are very often associated with similar diseases. This concept could also be applied in microRNAs. The MesH descriptors (http://www.nlm.nih.gov/) provide a comprehensive system for disease classification. We can then research into disease relationship via structure of directed acyclic graph (DAG) based on disease classification on MeSH. By measuring the semantic value of microRNAs associated diseases, we can infer the microRNAs functional similarity [14]. The stronger microRNAs functional similarity is, the higher potential microRNA biomarkers will be. A disease can be represented as a graph. For example, disease A can be shown as DAG_A = (A, T_A, E_A), where T_A and E_A denote all parents nodes of A and corresponding links of A respectively. The semantic value of disease A can be calculated via the equation:

\[
DV(A)=\Sigma_{t \in T_A} D_A(t).
\]

While measuring semantic similarity of two diseases (A and B) would be

\[
S(A,B)=\Sigma_{t \in T_A \cup T_B} (D_A(t)+D_B(t)) / (DV(A)+DV(B)).
\]

In order to measure functional similarity between two microRNAs accurately, it also need to further consider the contributions from similar disease (disease groups) that associated with these two microRNAs. We can measure the microRNAs functional similarity from the following equation:
\[
\text{MiS}(\text{Mi1, Mi2}) = \Sigma S(dt_{1i}, DT_2) + \Sigma S(dt_{2j}, DT_1) / m+n \\
1 \leq i \leq m \quad 1 \leq j \leq n
\]

After ranking microRNAs functional similarity, we can then retrieve the highly potential microRNA biomarkers. Finally, microRNA biomarkers can be confirmed with biological evidence.

### 2.5. Target prediction network methods

In this paper, we use two target prediction network methods to predict the target genes of our identified potential microRNA biomarkers. The first method is Probability of Interaction by Target accessibility (PITA). PITA is a target prediction algorithm developed in 2007 by the Segal lab members. It computes the difference between the free energy gained from the formation of the microRNA-mRNA duplex and the energetic cost of unpairing the mRNA to make it accessible to the microRNA. Free energy refers to the minimum free energy and shows how strong the binding of a microRNA with its target is. Normally free energy is a negative real value. The lower the free energy, the firmer the binding structure is and the more likely it suggests the true binding. We chose PITA for microRNA target prediction because it had been demonstrated to reach high accuracy, and it takes advantage of the target accessibility but not conservation information to reduce false positive.

The second method is miRTarBase, developed by National Chiao Tung University and National Yang Ming University on 2010, is an integrated database that curates experimentally validated microRNA-target interactions. A collection of microRNA–target interactions (MTIs) with experimental support is essential to thoroughly elucidating microRNA functions under different conditions and in different species. miRTarBase contains the largest amount of validated MTIs by comparing with other similar, previously developed databases.

### 3. Results and Discussions

#### 3.1. Potential microRNA biomarkers

We identify 46 potential microRNA biomarkers using microRNA functional similarity method, 23 potential microRNA biomarkers using PITA network method, and 5 potential microRNA biomarkers using miRTarBase network method. Among the 46 potential microRNA biomarkers using microRNA functional similarity method, 7 microRNA biomarkers have biological evidence. Among 23 potential microRNA biomarkers using PITA network method, 3 microRNA biomarkers have biological evidence. Finally, 3 microRNA biomarkers have biological evidence among the 5 potential microRNA biomarkers using miRTarBase network method. In total, 8 microRNA biomarkers are identified using our three methods. Table 4 summaries the biological evidence of these 8 microRNA biomarkers.

<table>
<thead>
<tr>
<th>microRNA Biomarker</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-mir-663</td>
<td>Treatment of HUVECs with the miR-663 antagonist (miR-663-LNA) blocks OS-induced monocyte adhesion, but not apoptosis. In contrast, overexpression of miR-663 increases monocyte adhesion in LS-exposed cells</td>
</tr>
<tr>
<td>hsa-mir-122</td>
<td>MiR-122a antagonirs may stimulate bile acid synthesis to reduce serum cholesterol and triglycerides</td>
</tr>
<tr>
<td>hsa-mir-10a</td>
<td>Human atherosclerotic lesions show low abundance of miR-10a. Knockdown of miR-10a in human aortic endothelial cells increases the abundance of inflammation- or migration-promoting factors, such as nuclear-localized p65 (a nuclear factor–κB subunit), cytokines, chemokines, and adhesion molecules. In contrast, miR-10a overexpression reduces the basal abundance of vascular cell adhesion molecule–1 (VCAM-1) and E-selectin, molecules that mediate inflammatory cell adhesion and initiate atherogenesis</td>
</tr>
</tbody>
</table>

Table 4. Biological evidence of 8 microRNA biomarkers
hsa-mir-132 | MiR-132 is shown to be involved in the inflammatory response of endothelial cells
---|---
hsa-mir-20a | AntagomiR directed against miR-20a restores functional BMPR2 signalling and prevents vascular remodelling in hypoxia-induced pulmonary hypertension
hsa-mir-584 | The expression of miR-584 is significantly lower in the blood of patients with CAD than healthy subjects
hsa-mir-181d | The expression of miR-181d is significantly lower in the blood of patients with CAD than healthy subjects
hsa-mir-21 | MiR-21 overexpresses at a significant level in neointimal lesions. It also involves in the promotion of proliferation and anti-apoptosis in vascular smooth muscle cell

Among the three methods used here, miRTarBase network method produce the highest percent of biological evidenced microRNA biomarkers (60% at 3 out of 5 potential biomarkers). Three common microRNA biomarkers appeared in each of these three methods. They are hsa-mir-663, hsa-mir-122, and hsa-mir-21. These biomarkers may deserve a further study.

3.2. Potential gene biomarkers

In this study, we identify 111 potential gene biomarkers using PITA network method, and 7 potential gene biomarkers using miRTarBase network method. Among these potential gene biomarkers, 5 gene biomarkers have biological evidence. Table 5 summaries the biological evidence of these 5 gene biomarkers.

<table>
<thead>
<tr>
<th>Gene Biomarker</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCC6</td>
<td>Frequent mutation in the ABCC6 gene is associated with a strong increase in the prevalence of coronary artery disease</td>
</tr>
<tr>
<td>ADD1</td>
<td>The ADD1 Gly460Trp polymorphism is significantly associated with an increased risk of coronary artery disease as well as blood pressure, indicating that ADD1 plays a role in the pathogenesis of coronary artery disease as well as hypertension</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>Variability in the TNFAIP3 gene works as a modulator of CAD risk in type 2 diabetes</td>
</tr>
<tr>
<td>APOB48R</td>
<td>ApoB-48R pathway may contribute to the spontaneous development of atherosclerotic lesions rich in macrophage-derived foam cells observed in apoE-deficient mice, a murine model of human atherosclerosis</td>
</tr>
<tr>
<td>LDLRAP1</td>
<td>Mutations in this gene lead to LDL receptor malfunction and cause the disorder autosomal recessive hypercholesterolaemia</td>
</tr>
</tbody>
</table>

We also identify some potential gene biomarkers, although not currently supported by medical evidence, have a potential to be a future true biomarkers due to the analysis of similar gene functionality. For examples, the function of PNPLA2 gene directly related to the activity of triglyceride lipase. The latter is the important dangerous factor of obesity, insulin resistance, and metabolic syndrome. FKBP8 gene can inhibit cell apoptosis by activating BCL2 gene. Oxidized LDL, a dangerous factor of CAD, promotes cell apoptosis by activating BCL2 gene. We may, therefore, deduct that FKBP8 gene could be a gene biomarker of CAD.

4. Conclusion

In clinical assessment of heart disease, biomarkers become key criteria. Recent studies reveal that microRNAs play important regulators of development and stress responsiveness of heart. How to identify microRNA biomarkers and gene biomarkers of heart disease is an important challenge. In this study, statistic methods, functional similarity scores, and target prediction systems are used to discover the potential microRNA biomarkers and gene biomarkers of coronary heart disease. In this study, we
identify 8 microRNA biomarkers: has-mir-663, has-mir-122, hsa-mir-10a, has-mir-132, has-mir-20a, has-mir-584, has-mir-181d, and has-mir-21 using microRNA functional similarity method, PITA network method, and miRTarBase network method. We also identify 5 gene biomarkers: ABCC6, ADD1, TNFAIP3, APOB48R, and LDLRAD1 using PITA network method, and miRTarBase network method.

5. Acknowledgments

The authors thank the National Science Council for its financial support for project NSC 101-2221-E-320-001 and Tzu Chi Foundation for its financial support for project TCRPP101002.

6. References