Identification of biomarkers for ischemic cardiomyopathy based on microarray data analysis

Yushuang Yang¹, Wei Yang², Wenxin Huo³, Pengfei Huo⁴, Hailing Yang⁵

¹Cardiovascular Department, China-Japan Union Hospital, Jilin University, Changchun 130031, China
²The Sixth Hospital of Changchun City, Changchun 130052, China
³The People’s Hospital of Jilin Province, Changchun 130000, China
⁴Intensive Care Unit, China-Japan Union Hospital, Jilin University, Changchun 130031, China
⁵Emergency Department, China-Japan Union Hospital, Jilin University, Changchun 130031, China

Abstract

Background: The aim of this study was to explore the biomarkers and potential mechanism underlying ischemic cardiomyopathy (ICM).

Methods: Using the GSE42955 Affymetrix microarray data accessible from the Gene Expression Omnibus database, the differentially expressed genes between 12 ICM tissue samples and 5 normal controls were identified. To investigate the function changes in the course of disease progression, Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed on the differentially expressed genes, followed by analysis of the protein–protein interaction (PPI) network and modules.

Results: A total of 50 up-regulated and 179 down-regulated genes were identified. The biological processes of immune response, response to virus, and cell adhesion molecules (CAMs) were significantly altered by the differentially expressed genes. The PPI network revealed certain hub nodes such as CXCL10, IRF1, STAT1, IFIT2, and IFIT3.

Conclusions: Candidate biomarker genes such as CXCL10, IRF1, STAT1, IFIT2, and IFIT3 may be suitable therapeutic targets for ICM. Further study of the CAMs pathway and immune response biological processes will be helpful in understanding the pathogenesis of ICM. (Cardiol J 2017; 24, 3: 305–313)

Key words: ischemic cardiomyopathy, differentially expressed genes, pathways, protein–protein interactions

Introduction

Ischemic cardiomyopathy (ICM) develops at the final stage of coronary artery disease, and is characterized by the narrowing of coronary arteries, myocyte death, reactive cellular hypertrophy, and ventricular scarring [1]. This form of cardiomyopathy represents a significant health concern owing to the high rate of sudden cardiac death in ICM patients worldwide [2]. Traditional treatment methods for ICM mainly include drug therapy, interventional angioplasty, and surgical vascular bypass [3]. However, the vascular lesions of some patients show small vessel diseases, which are not suitable for surgical operation or vascular obstructive intervention [4]. Thus, new therapies for ICM that satisfy current clinical needs are necessary.

Recent progress has been made in the discovery of novel targets for heart-related diseases. Transforming growth factor-beta signaling has been reported to be responsible for repressing the synthesis of inflammatory genes and preventing the
inflammatory infiltration in myocardial infarction [5]. The adaptive and proapoptotic pathways triggered by endoplasmic reticulum stress have been found to be involved in the development of cardiovascular disease; thus, endoplasmic reticulum stress has been proposed as a therapeutic target for cardiovascular disease [6]. The mitochondria have also been suggested as a drug target for ICM [7]. In addition, gene therapy targeting some important biomarkers has shown potential for altering the myocardial tissue microenvironment and improving mechanical cardiac function [8]. However, studies reporting gene targeted therapies of ICM are rare.

Gene expression profiling has been used to accurately predict the cardiomyopathy etiology [9]. In addition, some potential disease-related gene markers have been discovered based on the topological structure of biological networks [10, 11]. Therefore, adoption of a bioinformatics approach could help to reveal some novel biomarkers for cardiomyopathy therapy. Accordingly, in this study, a microarray data analysis was performed based on the publicly available gene expression profile (GSE42955), aiming to explore and identify novel biomarkers. In addition, it was hoped that further insight would be gained into the molecular mechanisms of ICM by screening the new biomarkers, which could facilitate selection of an appropriate treatment strategy and development of a new therapies for ICM.

Methods

Microarray data acquisition

The publicly available microarray dataset GSE42955 was downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), which was contributed by Molina-Navarro et al. [12]. The mRNA profiling was performed using 17 left ventricular transmural samples, including 12 obtained from ICM patients and 5 from normal donors, based on the platform of Affymetrix Human Gene 1.0 ST Array. All of the patients included in this study were functionally classified according to the New York Heart Association criteria and were receiving medical treatment following the guidelines of the European Society of Cardiology. The heart samples from patients were collected during cardiac transplantation. Non-diseased donor hearts that were not suitable for transplantation due to incompatibility of blood type or size were used as controls.

Data preprocessing

The raw CEL format files were preprocessed using the Affy package [13] in R, which included the processes of background correction and normalization. The statistical comparison of gene expression profiles between the ICM and control group was performed using the Limma package [14] in R. Genes with \( p < 0.01 \) and \( \log FC > 0.58 \) were considered significantly different. Hierarchical clustering [15, 16] was performed on the expression levels of differentially expressed genes based on the Euclidean distance using the heatmap program [17] in R.

Gene Ontology and pathway enrichment analysis

Genes showing differential expression were functionally classified based on the GO database using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool [18]. The significantly altered GO functions with a false discovery rate (FDR) < 0.05 (count = 2, EASE = 0.1) were identified.

To identify the pathways significantly associated with the genes of interest, the potentially altered pathways with an FDR < 0.05 were predicted using Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology-Based Annotation System (KOBAS) [19] based on the cumulative hypergeometric distribution algorithm.

Protein–protein interaction (PPI) network construction and hub protein screening

The Search Tool for the Retrieval of Interacting Genes (STRING) [20] database collects information of both experimental and predicted protein interactions. Differentially significant protein pairs of the expressed genes were predicted using String software [21]. Protein pairs with a combined score > 0.8 were assembled for PPI network construction with the application of Cytoscape software [22]. Subsequently node degrees (i.e., the number of neighbors) were calculated and hub nodes were screened out.

Modules analysis based on the PPI network

Based on the PPI network described above, the modules comprised of genes with similar biological functions were analyzed using Mcode [23] package in Cytoscape. A degree cutoff \( \geq 2 \) and K-core \( \geq 2 \) were set as the cutoff values for module analysis. The functionally associated genes in modules were classified using Bingo software [24]. The GO functions with an adjusted p-value < 0.05 were considered significant.
Results

Identification of differentially expressed genes in ICM samples compared with normal controls

A total of 229 genes with differentially significant expression were screened out between the ICM and control groups, of which 50 genes were up-regulated and 179 were down-regulated. The heat map for the hierarchical clustering of gene expression is shown in Figure 1, which illustrated distinct gene expression profiles between ICM sample and normal controls.

Gene Ontology function and KEGG enrichment analysis

The GO functions closely related with the genes of interest are shown in Table 1. A total of 8 biological functions were significantly enriched, such as immune response (GO: 0006955) and defense response (GO: 0006952). The majority of the annotated genes were related to immune response (Fig. 2).

KEGG pathway analysis showed that the cell adhesion molecules (CAMs; ID: hsa04514, p = 1.51E-07, FDR = 1.64E-04) and cytokine-cytokine receptor interaction (ID: hsa04060, p = 3.24E-07, FDR = 4.69E-07).
FDR = 3.51E-04) were significantly enriched by the significantly differentially expressed genes in ICM.

PPI network and hub proteins
A total of 170 protein pairs were in accordance with a combined score > 0.8. The PPI network was constructed, which comprised 170 edges and 88 nodes (14 up-regulated genes and 74 down-regulated genes) (Fig. 3). The hub nodes were assessed based on the fitting function \( y = 37.56 \times (-1.365) \), as shown in Figure 4; the X-axis indicates the number of nodes that were directly connected to one node, while the Y-axis indicates the number of genes in each degree. The top 10 hub nodes were selected and are shown in Table 2, including C-X-C motif chemokine ligand 10 (CXCL10; degree = 15), signal transducer and activator of transcription 1 (STAT1; degree = 14), interferon induced protein with tetratricopeptide repeats 2 (IFIT2; degree = 13), interferon induced protein with tetratricopeptide repeats 3 (IFIT3; degree = 13), and interferon regulatory factor 1 (IRF1; degree = 12).

Significant modules in the PPI network
The significant modules were further explored using the Mcode package in Cytoscape software. A total of 2 modules (degree cutoff ≥ 2 and K-core ≥ 2) were explored, as shown in Figure 5. The results revealed that 5 nodes in the modules had high degrees of connectivity: CXCL10, IRF1, STAT1, IFIT2, and IFIT3. The functional annotation for the modular genes indicated that module 1 associated with CXCL10 and IRF1 was related to the immune system process, while modules associ-
ated with STAT1, IFIT2, and IFIT3 were related to a response to virus (Table 3).

**Discussion**

Ischemic cardiomyopathy is the leading cause of sudden cardiac death and poses a significant public health threat worldwide. Unfortunately, the detailed pathophysiology of ICM remains to be clearly elucidated, and novel effective therapies are urgently needed. Gene therapies that alter the myocardial tissue microenvironment have shown positive effects for improving mechanical cardiac function by targeting certain important biomarkers [8]. Thus, the investigation of novel biomarkers for the development of new therapeutic strategies is needed.

In the present study, we mined some new biomarkers for ICM based on gene expression profile data. The results indicated that ICM induced the differential expression of a total of 299 genes (50 up-regulated genes and 179 down-regulated genes), which were screened out and compared with healthy controls. Use of a functional classification tool allowed the classification of functionally associ-

**Table 2.** The top 10 selected hub proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Node</th>
<th>Protein</th>
<th>Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXCL10</td>
<td>15</td>
<td>DDX58</td>
<td>11</td>
</tr>
<tr>
<td>STAT1</td>
<td>14</td>
<td>GBP2</td>
<td>10</td>
</tr>
<tr>
<td>IFIT2</td>
<td>13</td>
<td>CCR1</td>
<td>9</td>
</tr>
<tr>
<td>IFIT3</td>
<td>13</td>
<td>C3</td>
<td>8</td>
</tr>
<tr>
<td>IRF1</td>
<td>12</td>
<td>IFITM1</td>
<td>7</td>
</tr>
</tbody>
</table>

**Figure 4.** Investigation of hub proteins. The X-axis indicates the number of nodes that are directly connected to one node, and the Y-axis indicates the number of genes in each node.

**Figure 5.** Significant modules in the protein–protein interaction network. The polygon represents differentially expressed genes with a high node degree. The lines represent the connections between polygons.
ated differentially expressed genes in 8 biological processes-related GO terms and 2 pathways. The differentially expressed genes were then mapped to a PPI network, and significant nodes such as CXCL10, IRF1, STAT1, IFIT2, and IFIT3 were screened out, followed by 2 significant modules.

The CAM pathway was the most significantly enriched pathway by the genes showing differential expression. Basic and clinical research indicates that CAMs play an important role in mediation of the immune and inflammatory processes of cardiovascular diseases [25]. Genome-wide association studies also suggest that human diseases such as atherosclerosis and Alzheimer’s disease are related to the processes of CAMs [26, 27]. Intercellular adhesion molecule-1 (ICAM-1), as a member of the immunoglobulin superfamily, plays an important role in mediating leukocyte adhesion and the development of atherogenesis [28]. Moreover, a high level of ICAM-1 was found to increase the risk of coronary heart disease [29]. Consequently, although there is no direct evidence for the critical role of CAMs in ICM, the results of this study suggests that they may play a key role in the pathological mechanism of ICM. However, further investigation is needed to confirm this finding.

Function analysis showed that immune response was the most significant GO term enriched by the differentially expressed genes in ICM. The current understanding of ICM pathogenesis involves concurrent changes in the cellular and humoral chains of immunity throughout the multi-staged course of the disease [30, 31]. Upon an immune inflammatory reaction of the myocardium, cellular and humoral immune responses are involved in the pathological remodeling of the heart, which further induce extracellular matrix degradation, collagen deposition, and myocardial cell apoptosis [32]. Collectively, these findings suggest that the immune response is a critical process in the pathogenesis of ICM.

Our data also showed that the genes in module 1 were significantly associated with immune response-related biological processes, such as CXCL10 and IRF1. CXCL10 is suggested to be the master regulator of myocardial inflammatory cell migration, and may contribute to the clinical progression to life-threatening cardiomyopathy [33]. Previous studies have demonstrated that some factors such as CXL10 are associated with immunity in the process of disease [34, 35]. The marked up-regulation of CXC chemokines (such as CXCL10) is the early event in ICM development [36]. CXCL10 with antifibrotic properties can prevent fibrosis development from inflammation in ICM [37].

IRF1 is a critical member of interferon signaling pathway and is proposed to be a genetic risk factor for several chronic inflammatory and autoimmune diseases [38, 39], especially blood-related

<table>
<thead>
<tr>
<th>Module 1</th>
<th>GO ID</th>
<th>Description</th>
<th>P</th>
<th>Adjusted p</th>
<th>n</th>
<th>Genes in test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1</td>
<td>2376</td>
<td>Immune system process</td>
<td>9.11E-11</td>
<td>3.22E-08</td>
<td>10</td>
<td>C3AR1, C3, CCR1, CXCL9, IRF1, CXCL11, GBP2, IFI35, IFI6, CXCL10</td>
</tr>
<tr>
<td>Module 1</td>
<td>50896</td>
<td>Response to stimulus</td>
<td>4.32E-05</td>
<td>1.24E-03</td>
<td>10</td>
<td>C3AR1, IFITM1, C3, CCR1, CXCL9, CXCL11, GBP2, IFI35, IFI6, CXCL10</td>
</tr>
<tr>
<td>Module 1</td>
<td>50794</td>
<td>Regulation of cellular process</td>
<td>5.86E-03</td>
<td>2.65E-02</td>
<td>10</td>
<td>C3AR1, IFITM1, C3, RGS4, CCR1, CXCL9, IRF1, CXCL11, IFI6, CXCL10</td>
</tr>
<tr>
<td>Module 1</td>
<td>50789</td>
<td>Regulation of biological process</td>
<td>9.17E-03</td>
<td>3.61E-02</td>
<td>10</td>
<td>C3AR1, IFITM1, C3, RGS4, CCR1, CXCL9, IRF1, CXCL11, IFI6, CXCL10</td>
</tr>
<tr>
<td>Module 1</td>
<td>65007</td>
<td>Biological regulation</td>
<td>1.50E-02</td>
<td>4.94E-02</td>
<td>10</td>
<td>C3AR1, IFITM1, C3, RGS4, CCR1, CXCL9, IRF1, CXCL11, IFI6, CXCL10</td>
</tr>
<tr>
<td>Module 1</td>
<td>6955</td>
<td>Immune response</td>
<td>4.99E-09</td>
<td>8.80E-07</td>
<td>8</td>
<td>C3, CCR1, CXCL9, CXCL11, GBP2, IFI35, IFI6, CXCL10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Module 2</th>
<th>GO ID</th>
<th>Description</th>
<th>P</th>
<th>Adjusted p</th>
<th>n</th>
<th>Genes in test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 2</td>
<td>9615</td>
<td>Response to virus</td>
<td>9.44E-06</td>
<td>1.62E-03</td>
<td>3</td>
<td>DDX58, RSAD2, STAT1</td>
</tr>
<tr>
<td>Module 2</td>
<td>51707</td>
<td>Response to other organism</td>
<td>1.48E-04</td>
<td>1.28E-02</td>
<td>3</td>
<td>DDX58, RSAD2, STAT1</td>
</tr>
<tr>
<td>Module 2</td>
<td>9607</td>
<td>Response to biotic stimulus</td>
<td>2.89E-04</td>
<td>1.66E-02</td>
<td>3</td>
<td>DDX58, RSAD2, STAT1</td>
</tr>
</tbody>
</table>

Note: The adjusted p-value represents the p-value corrected for multiple testing; n represents the number of genes.
diseases [40]. Recently, Chapin et al. [41] indicated that the expression level of IRF1 was changed in the peripheral blood, which can be used as a marker to reflect the glucocorticoid response. IRF1 plays a key role in response to pressure or volume overload, which is involved in the pathogenesis of ischemic heart diseases [42]. In the present study, based on the PPI network, CXCL10 and IRF1 were identified as hub nodes with high degrees of connectivity, suggesting that these genes might play important roles in the immune response of ICM.

Moreover, functional module 2 was closely related with the biological process of response to virus. The functionally associated genes such as STAT1, IFIT2 and IFIT3 were classified in module 2. The protein encoded by the STAT1 gene is activated in the process of cell growth inhibition, and has multiple interactions with cytokines and growth factors [43]. Disruption of the STAT1 gene in mice has been found to affect innate immunity to a virus [44]. Additionally, STAT1 is activated after ischemia-reperfusion, which induces the expression of proapoptotic cytokines leading to cardiac cell death [45]. IFIT2 is believed to function as a suppressor in tumorigenesis [46]. In addition, a previous study indicated that IFIT2 does not function as a monomer, but rather forms complexes (such as ISG54/IFIT2) with itself or related ISG60/IFIT3 proteins to elicit complex cellular responses [47]. The apoptotic response to IFIT2 may contribute to other functions, including translational regulation, inhibition of tumor colonization, and protection against a lethal viral infection [47]. IFIT2 and IFIT3 are the interferon-related genes and are found to be differentially expressed in heart failure mice [48]. In this study, STAT1, IFIT2, and IFIT3 were identified as hub nodes and showed multiple interactions with other genes/proteins. The changes in the expression of these genes may disturb the biological processes related to a response to a virus and contribute to the occurrence of ICM. The functions of these significant nodes warrant further investigation.

Limitations of the study

In spite of these findings, some important limitations remain in this work. The microarray data from the GEO database were analyzed in this study and the data was not generated by this study. The diagnostic and inclusion criteria for the ICM patients and controls included in the database were not available. Another limitation was the lack of experimental validation in other samples or datasets, which might preclude any conclusive suggestions. The present study was also limited by a relatively low sample size.

Future directions for investigation

Further experimental validations of the presented findings in a large number of samples are needed in the near future. The expression levels of the biomarkers identified in this work should be assessed in patients at high risk of ICM and those that are already diagnosed with ICM to discover novel gene therapies for the prevention and treatment of ICM.

Conclusions

In conclusion, the potential ICM-related genes, including CXCL10, IRF1, STAT1, IFIT2, and IFIT3, were found to have multiple interactions with other genes. The differential expression of these target genes may increase the possibilities for detecting or predicting the occurrence of ICM. Furthermore, the CAMs-related pathway, immune response, and response to virus are proposed to be involved in the process of ICM. These candidate biomarker genes and pathways could be useful in further studies aiming to understand the genetic mechanisms of ICM and are expected to significantly assist in the development of therapeutic strategies.

Conflict of interests: None declared

References


