Identification of key genes and crucial modules associated with coronary artery disease by bioinformatics analysis

XUEMEI ZHANG\(^1\), XIAOSHU CHENG\(^2\), HUIFENG LIU\(^3\), CHUNHUA ZHENG\(^1\), KUNRUI RAO\(^1\), YI FANG\(^1\), HAIRONG ZHOU\(^4\) and SHENGHE XIONG\(^1\)

\(^1\)Department of Cardiology, Third Affiliated Hospital of Nanchang University; 
\(^2\)Department of Cardiology, Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi 330008; 
\(^3\)Department of Cardiology, Xiaolan People's Hospital, Zhongshan, Guangdong 528415; 
\(^4\)Department of Cardiology, Second People's Hospital, Mudanjiang, Heilongjiang 157000, P.R. China

Received December 19, 2013; Accepted May 28, 2014
DOI: 10.3892/ijmm.2014.1817

Abstract. The aim of this study was to identify key genes associated with coronary artery disease (CAD) and to explore the related signaling pathways. Gene expression profiles of 110 CAD and 112 non-CAD, healthy patients [CAD index (CADi)>23 and =0, respectively] were downloaded from the Gene Expression Omnibus (GEO) database (accession: GSE12288). The differentially expressed genes (DEGs) in CAD were identified using t-tests, and protein-protein interaction (PPI) networks for these DEGs were constructed using the Search Tool for the Retrieval of InteractiNg Genes (STRING) database. The Database for Annotation, Visualization and Integrated Discovery (DAVID) tool was used to identify potentially enriched biological processes (BP) among the DEGs using Gene Ontology (GO) terms, and to identify the related pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. In addition, expression-activated subnetworks (crucial modules) of the constructed PPI networks were identified using the jActiveModule plugin, and their topological properties were analyzed using NetworkAnalyzer, both available from Cytoscape. The patient specimens were classified as grade I, II and III based on CADi values. There were 151 DEGs in grade I, 362 in grade II and 425 in grade III. In the PPI network, the gene GRB2, encoding the growth factor receptor-bound protein 2, was the only common DEG among the three grades. In addition, 10 crucial modules were identified in the PPIs, 4 of which showed significant enrichment for GO BP terms. In the 12 nodes with the highest betweenness centrality, we found two genes, encoding GRB2 and the heat shock 70 kDa protein 8 (HSPA8). Moreover, the chemokine and focal adhesion signaling pathways were selected based on their relative abundance in CAD. The GRB2 and HSPA8 proteins, as well as the chemokine and focal adhesion signaling pathways, might therefore be critical for the development of CAD.

Introduction

Coronary artery disease (CAD), also called coronary arteriosclerosis, is the most common type of heart disease (1). It is the leading cause of death in the United States in both genders (2,3). CAD occurs when the coronary arteries which supply blood to the heart muscle become partially blocked or clogged, which then leads to inducative and stenotic coronary arteries (4). As with the development of atherosclerosis, reduced amounts of blood and oxygen are transferred to the heart muscles, resulting in angina or a heart attack (5). Most heart attacks occur when a blood clot suddenly cuts off the blood supply to the heart, causing permanent heart damage. Over time, CAD can also weaken the function of heart muscles and contribute to heart failure and arrhythmias (6).

Survival following coronary artery bypass graft surgery (CABG) and medical therapy in patients with CAD has been studied in both randomized trials and observational treatment comparisons (7-11). A number of randomized trials have shown that treatment with percutaneous coronary intervention (PCI) or mineral technologies (MT) is associated with higher rates of angina and subsequent revascularization compared to CABG treatment, but no significant difference has been found in mortality or rates of myocardial infarction (MI) between these types of treatment (12). Despite the lack of concluding evidence from randomized trials and prospective observational comparisons, the application of coronary angioplasty (PTCA) has increased dramatically over the past decade (13). Current pharmacologic therapy for CAD patients presents multiple limitations, such as compliance issues and medication side-effects. For revascularization procedures, there is often a need to repeat the procedures, which is accompanied by organ damage (14). Therefore, the development of new methods for the treatment of CAD is necessary.
Gene therapy offers such an attractive solution. It has the potential to locally and continuously provide therapeutic proteins at the disease site, and can potentially lead to reversal of the pathophysiology associated with the disease. Therefore, in order to identify key genes and related signaling pathways associated with CAD, which could potentially serve in gene therapy, we performed a meta-analysis on the GSE12288 CAD gene expression profiling dataset, downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) and related protein-protein interaction (PPI) networks were identified, along with expression-activated subnetworks (crucial modules) of the PPI networks, enriched biological functions and the related signaling pathways.

Materials and methods

Data source. The gene expression dataset GSE12288 (15) was downloaded from the GEO database (16). The 222 patients assessed in this dataset were divided into two groups based on their respective CAD index (CADi): the control group, with 112 healthy individuals (CADi =0), and the case group, with 110 patients (CADi >23). Samples from all individuals had been hybridized on the Affymetrix Human Genome U133A (HG-U133A) array on a GPL96 platform (Affymetrix, Santa Clara, CA, USA).

DEG identification. The GSE12288 dataset had been normalized using MA55 (Affymetrix) (15). Log (base 2) transformation of the downloaded raw data was performed with the GEOquery module, available in R (17). CAD groups comprising case samples with the same CADi were defined. Subsequently, a between-subjects t-test was performed to identify DEGs of each CAD group using the T-test module of MultiExperiment Viewer (18) and a significance threshold \( P<0.01 \) compared to the control group. For the pairwise alignment of genes differentially expressed between two CAD groups (dataset A and B), the Jaccard similarity coefficient or index was computed according to the following formula:

\[
J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]

Based on the proximity of DEGs as defined by the Jaccard index value, the case samples were reclassified in three CAD grades (Fig. 1). Subsequently, re-comparison between CAD grade samples and those of the control group was performed, and the DEGs were identified with a significance threshold \( P<0.01 \).

Construction of PPI. It was shown that functionally associated genes are co-expressed, with their gene expression levels varying depending on the cell type and state (19). Thus, the Search Tool for the Retrieval of Interacting Genes (STRING) (http://string-db.org/) database (20) was used to build PPI networks for the identified DEGs, so as to predict their interactions. A combination score of >0.3 was used as the threshold. The topological properties of the resulting PPI network were visualized and analyzed with Cytoscape (21). We used the jActiveModule plug-in of Cytoscape to further identify the significant modules in the network as previously described (22,23).

Enrichment analysis of significant modules. The Database for Annotation, Visualization and Integrated Discovery (DAVID) provides a comprehensive set of functional annotation tools for biological interpretation of large gene lists. DAVID was used here to group the functions of DEGs in modules, identify enriched biological processes, and identify the pathways associated with the DEGs in the most significant modules. Function and pathway terms were retrieved from the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, respectively (24,25).

Topological analysis of PPI networks. The topological properties of the PPI networks were analyzed using NetworkAnalyzer available in Cytoscape (26). NetworkAnalyzer allows to compute and display a comprehensive set of topological parameters, such as the number of nodes and edges, the betweenness and the stress centrality. Betweenness centrality, a measure of the centrality of a node in a network, is the fraction of shortest paths between node pairs that pass through the node of interest (27). The stress centrality is the number of shortest paths a node lies on (28).

Results

Identification of DEGs. All control samples were collected as group A. Based on their CADi values (23, 32, 37, 42, 48, 56, 63, 67, 74, 82 and 100), the case samples were grouped in CAD groups, each comprising samples of similar CADi. The CADi = 56 group comprised only one sample, while all other CAD groups comprised >1 samples. Following identification of genes differentially expressed in each CAD group compared to the control group, pairwise alignment of the genes differentially expressed between two CAD groups was conducted. Then, the case groups were reclassified into three grades based on their similarity degree: grade I comprised samples with CADi = 23-42; grade II with CADi = 48-67; and grade III, with CADi = 74-100. Finally, we identified the differentially expressed genes, defined as DEGs of each grade, through a new comparison between each grade and the control group. There were 151 DEGs in grade I, 362 in grade II and 425 in grade III.

PPI networks and functional enrichment analysis. PPIs of DEGs in grade I, II and III were constructed by calculating confidence values using the STRING tool (29). In the resulting networks (Fig. 2), genes that simultaneously belonged to more than two grades were few, and only the gene encoding growth factor receptor-bound protein 2 (GRB2) was common to grades I, II and III. The expression values of the DEGs and the p-values from t-test comparisons were used to identify the expression activated subnetworks (crucial modules) of the PPI network using the jActiveModule in Cytoscape. Table I displays the 10 modules identified with this procedure, 4 of which were significantly enriched for biological processes (colored in Fig. 3). Module 4 genes belonged to grade I samples, module 2 and 3 belonged to grade II and module 1 belonged to grade III.
To gain further insights into the function of DEGs belonging to the crucial modules, we used the online biological classification tool DAVID to retrieve functional annotations from GO (biological process terms) and KEGG for the DEGs and perform an enrichment analysis for the GO terms. Table II lists the enriched biological processes in which the DEGs belonging to the crucial modules are involved. The crucial modules 1-4 were related to cell apoptosis (P=2.70E-03), protein catabolic process (P=1.40E-04), cell signaling (P=2.30E-03) and telomere maintenance (1.70E-06), respectively. In addition, according to the percentage of identified DEGs in each pathway relative to the total number of DEGs, which is defined as their relative abundance (RA), we ranked the KEGG pathways in the selected modules. Figure 4 shows the top 10 pathways, in which DEGs from the 4 modules are involved. These include the chemokine signaling pathway (RA=0.176), focal adhesion (RA=0.143), and regulation of actin (RA=0.099). The chemokine pathway and related genes as depicted in the KEGG Pathway database are shown in Fig. 5.

**Discussion**

In this study, in order to identify the key genes associated with CAD and the related signaling pathways, so as to ultimately gain knowledge on the disease that would be applicable to early diagnosis and therapy, we identified 151 DEGs in samples from grade I patients, 362 genes from grade II and 425 genes from grade III patients. The number of selected DEGs increased along with the rise in CADi values, and
there was only one overlapping gene (GRB2) among the three grades. To explore the potential molecular interactions among the DEGs, PPI networks were constructed. Analysis of their topological properties indicated that GRB2 and HSPA8 genes, showing the highest betweenness centrality, belong to crucial modules of the PPI network. The protein product of GRB2, Grb2, is widely distributed in various tissues and cells and participates in numerous biological functions (30,31). The docking proteins of the Grb2-associated binder family (Gab1, Gab2 and Gab3), serving as important signaling components, may amplify and integrate signal transduction pathways in response to various stimuli, such as cytokines (32). In addition, as a molecular switch, Grb2 is a ubiquitously expressed linker protein that couples growth factor receptor activation to down-

Figure 4. The top 10 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in the modules, selected by jActiveModule.

Figure 5. Chemokine signaling pathway. Red boxes denote differentially expressed genes (DEGs), and green boxes the non-DEGs.
Table I. The 10 crucial modules and the differentially expressed genes (DEGs).

<table>
<thead>
<tr>
<th>Module</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.0996</td>
</tr>
<tr>
<td>2</td>
<td>5.8049</td>
</tr>
<tr>
<td>3</td>
<td>3.7185</td>
</tr>
<tr>
<td>4</td>
<td>2.1734</td>
</tr>
<tr>
<td>5</td>
<td>2.0949</td>
</tr>
<tr>
<td>6</td>
<td>1.7816</td>
</tr>
<tr>
<td>7</td>
<td>1.7670</td>
</tr>
<tr>
<td>8</td>
<td>1.7518</td>
</tr>
<tr>
<td>9</td>
<td>1.6902</td>
</tr>
<tr>
<td>10</td>
<td>1.6280</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene list</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALAD, MAP2K7, PPARD, PTPRT, EEA1, HSPA8, ARRB2, CD68, HSPE1, TMEM11, GRM3, PPM1A, CCL15, RPS25, PYCRL, HSPD1, CBLB, ATAD3A, IL22, GLUL, FLT4, CHD1, TTC3, BCHE, MGMT, CTNND1, CAPN5, PPP2R4, HIST1H2A1, PDE6A, CCDC72, PXN, EIF3M, CABIN1, SUV420H1, SS8, PTK2B, GRK1, DNAJA3, REV3L, PKM2, GRB2, PITPNM1, TGM4, CETP, STK11, HDAC5, CALM3, IGTV1-69, MRPL19, ALB, HIST2H4A, HIST2H4B, IGHA1, VAV1, ACAN, DAP, LEPR, CS, TRPC, CCL14, C22orf28, IGHDIA, RPS27A, HIST1H2BH, HSPA5, S1PR4, CCR1, PTK2, CD6, SDC2, IGHD1, CCR7, SMARC8, CDKN1C, IARS, FANCA, STAT1, NCAPH, CTNNB1, HSPA8, COPS6, NIN, GATA1, PSMF1, FANCC, FANCF, USP14, DARS2, CTNND1, TES, SKP2, PTK2, CDH11, FBXO46</td>
</tr>
</tbody>
</table>

Table II. Modules with enriched biological processes as defined by Gene Ontology (GO) terms.

<table>
<thead>
<tr>
<th>Module</th>
<th>Enriched biological process</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regulation of caspase activity</td>
<td>5.90E-04</td>
</tr>
<tr>
<td></td>
<td>Programmed cell death</td>
<td>2.70E-03</td>
</tr>
<tr>
<td>2</td>
<td>Protein catabolic process</td>
<td>1.40E-04</td>
</tr>
<tr>
<td></td>
<td>Cellular response to hormone stimulus</td>
<td>8.90E-04</td>
</tr>
<tr>
<td>3</td>
<td>Integrin-mediated signaling pathway</td>
<td>3.40E-04</td>
</tr>
<tr>
<td>4</td>
<td>Telomere maintenance</td>
<td>2.30E-03</td>
</tr>
<tr>
<td></td>
<td>Double-strand break repair</td>
<td>1.70E-06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.00E-05</td>
</tr>
</tbody>
</table>

The analysis was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool.

Figure 6. Nodes with the highest betweenness centrality. The size of each dot indicates the degree of betweenness centrality.

Figure 7. Stress centrality.
stream mitogen-activated protein kinases (MAPK), including the c-Jun N-terminal kinase (JNK2), the extracellular-signal-regulated kinase (ERK) and p38 MAPK (33,34), which may play a role in cardiac hypertrophy, atherosclerosis and vascular restenosis, suggesting that Grb2 might be a pharmacological therapeutic target for CAD (35). The HSPA8 protein (also known as HSP73 or HSC70) displays 86% amino acid similarity with the heat-inducible protein of the same family HSPA1A (36). Constitutively expressed HSPA8 is involved in protecting cardiac muscle cells from injuries such as oxidative challenge (37,38). Moreover, heritable variations in the HSPA8 gene might increase the risk of coronary heart disease (39). These studies suggest that Grb2 and HSPA8 proteins might play a major role in CAD. Namely, Grb2 contributes to atherosclerotic lesion formation, while basal expression of HSPA8 contributes in reducing the risk of CAD.

Functional enrichment analysis of DEGs belonging to crucial modules showed that DEGs in CAD are mainly involved in protein synthesis, cell apoptosis and signal transduction. Analysis of the related KEGG pathways identified a number of signaling pathways, such as the chemokine signaling pathway. Notably, chemokines and chemokine receptors were reported as important mediators in atherosclerosis (40), not only by directing leukocytes into the vessel wall, which is a crucial feature of all stages of atherosclerosis, but also by activating these cells within the atherosclerotic lesion (41). Consequently, the chemokine pathway may represent an important therapeutic target in CAD. In addition, two additional processes with a high RA in our study, focal adhesion (42) and actin regulation (43), were previously reported as being involved in coronary heart disease and the development of atheroma. Briefly, integrin activation can lead to the formation of multi-protein focal adhesion signaling complexes and activate the non-receptor tyrosine kinases focal adhesion kinase (FAK) and c-Src (42); regulation of the actin cytoskeletal organization may also contribute to the differential roles of the Rho kinase (ROCK) isoforms in contractility. The RhoA/ROCK signaling pathway is now well recognized as a mediator of vascular smooth muscle contraction in response to vasoconstrictors, by inhibiting myosin phosphatase (MLCP) activity and increasing the phosphorylation of the myosin light chain (43).

UBC was the gene with the highest betweenness centrality in the topological analysis of PPIs. It connected two modules of grade I and III, which were associated with cell activity and signal transduction respectively, but it did not belong to any of the crucial modules. UBC is one of two stress-inducible polypeptides genes in mammals, which contributes, along with constitutive Uba genes, in maintaining ubiquitin (Ub) levels during cell proliferation and stress (44). We therefore hypothesize that this gene might act as a connective component for different factors involved in CAD.

In summary, genes differentially expressed in CAD, their potential functions and relevant pathways were identified in this study. The genes UBC, GRB2 and HSPA8 can be considered as critical genes for CAD, with potential to be used in gene diagnosis and therapy. Moreover, our study suggested that the chemokine pathway might play a vital role in the development of CAD, although further experimental validation is needed to confirm this hypothesis.

References