Comparative Gene Expression Profiling and Character Development in stable Angina Pectoris Disease

Xiaolong Zhu1, Yimin Wang2, Jingyi Nan2, Jian Guo2, Haixia Dang2, Yifei Qi4, Wangmin Qiao1, Fei Teng1* and Zhong Wang4*

1The Beijing Genomics Institute (BGI), Beishan Industrial Zone, Yantian District, Shenzhen, Guangdong, China
2Shaanxi Buchang Pharmaceutical Co, Ltd, China
3China Academy of Chinese Medical Sciences, Nanxiaojie, Dongzhimennei, Beijing, China
4Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Nanxiaojie, Dongcheng District, Beijing, China

Abstract

Objective: To investigate the characteristics and find more effective biomarkers at gene expression level of Stable Angina (SA) pectoris.

Methods: Here blood samples from 45 angina pectoris patients and 20 healthy controls were collected and produced by next generation sequencing and then compared gene expression levels between these two groups to find the significant regulated genes. Several methods or software were used to identify the features of angina pectoris relative to healthy people at transcriptase level.

Results: 874 differential expressed genes may likely to be the candidate biomarkers for angina pectoris patients, 459 genes can be searched from the database Online Mendelian Inheritance in Man (OMIM) which providing information about the human genes and genetic disorders especially pay attention to genetic variants and gene expression.

Discussion: Several different expressed genes are associated with angina pectoris related signal pathways. Such as PDGFA and KDR which play a role in atherosclerosis and already as a target for potential treatment. The function of rest 323 genes still needs to continue the function study in disease. These DEGs can act as candidate biomarkers for diagnose or treat, but still need further validate, refine and extend the process and function in coronary heart disease angina. The identification of biomarkers can be applied to the diagnosis, prognosis, and treatment of stable angina pectoris, that’s what we are counting on.

Keywords: Biomarkers; Gene expression; Stable angina

Introduction

Previous reports suggest that the most common type of heart disease is Coronary Artery Disease (CAD). Chronic stable angina is a main manifestation in approximately half of all patients who present with coronary heart disease, accounting for more than 80% of the cases worldwide [1]. Recently, some studies have demonstrated that miRNA can be involved in regulation of development of CAD patients by contrasting with controls [2,3]. And genomic mutations like SNP have been reported associated with myocardial infarction through Genome-wide association study [4].

This research aimed to investigate the distinctive mRNA features in the stable angina pectoris, and explore the novel biomarkers for disease diagnosis especially for Chinese population.

Methods

Sample collection

Serum samples from 45 patients with stable angina in 301 Hospital and Xuanwu Hospital China were randomly selected into the trials and twenty healthy volunteers joined as controls. The clinical characteristics of patients like sex/age/hyperlipidemia are summarized in table 1. And patients provided informed consent in accordance with local IRB requirements.
Library and Sequencing

Total RNA was isolated from the serum samples and was pooled together to construct cDNA Libraries. These libraries were subsequently sequenced throw Illumina HiSeq sequencing with pair-end reads of length 2*100bp according to manufacturer’s instructions as previously described.

Data analysis and Statistics

The widely used software Bowtie2 and RSEM were performed to do the alignments and calculate FPKM value of genes respectively. Normalized mRNA expression data were used to perform the clustering analysis, and only kept the gene which expressed in 60% or more samples which as input for R package ‘Consensus Cluster Plus’[5].

The normalized and log-transformed gene expression data then provided as input to the R package ‘iCluster’[6] and we reduce the dataset to top 6,000 most variable genes which measured by median absolute deviation.

The expressed genes were further analyzed between the groups which separated by clustering analysis, and represented the features of patients with stable angina and healthy respectively. Based on the gene expression level, differentially expressed genes were detected by software ‘DESeq2’ [7] with fold-change larger than 2 and p-value smaller than 0.05.

According to the DEGs, we next performed Gene Ontology (GO) classification and KEGG pathway classification and functional enrichment by using R package ‘phyper’. For each P-value, we correct for multiple comparisons by controlling the False Discovery Ratio. And the terms for which the FDR was not greater than 0.01 were defined as statistically significantly enriched. Meanwhile, online software ‘metascape’ [8] also used to perform the pathway and process enrichment analysis.

Results

Expression and clustering

Throw transcriptome sequencing we identified an averaged 15,747 genes for each sample. Based on the expressed RNAs hierarchical, clustering analysis were performed to show the relationship between angina pectoris and healthy. The figure1 showed great divergence between these two groups although some sample from patient has clustered into another group. The clustering analysis indicated that the expression of genes may be involved in the SA disease progression and required further research (Figure 1).

IClusterPlus is developed for integrative clustering analysis of multi-type genomic data, such as DNA copy number alterations and mRNA expressions. In this research, mRNA expression was used as input performing clustering algorithm to discover the subtype of all individuals. Table 2 a indicates the classification results by integrative clustering and figure 2c/d showed the comparing subtypes via PCA. Most samples from healthy are clustered to subgroup 3 and 66.7% of patient samples were divided to subgroup 1. Meanwhile table 2b and figure 2b represented the grouping situations of Consensus Plus were quite similar to result of iClusterPlus.

<table>
<thead>
<tr>
<th>ICluster subtypes</th>
<th>Histology class</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>0 (0%)</td>
<td>3 (15%)</td>
<td>17 (85%)</td>
<td></td>
</tr>
<tr>
<td>angina pectoris</td>
<td>30 (66.7%)</td>
<td>3 (6.6%)</td>
<td>12 (26.7%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consensus Cluster</th>
<th>Histology class</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>19 (95%)</td>
<td>1 (5%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>angina pectoris</td>
<td>15 (33.3%)</td>
<td>22 (48.8%)</td>
<td>8 (17.7%)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: a) Integrated subtype of iCluster. b) Subgroups of Consensus Cluster Plus.

Figure 2: A) The Consensus Matrix plot at k= 3 show a clean consensus distribution of items that have high intra-cluster consensus and low inter-cluster consensus. B) Comparing subtypes via PCA (Consensus Cluster) C) D) Comparing subtypes via PCA (iClusterPlus).

Consensus Cluster Plus is a tool for unsupervised class discovery and Gene expression features were used to detect the possible groups of items. Sample distribution of histology class and consensus cluster is shown for k=3 in table 2b. Percentages are the proportion of the pathology class in a consensus cluster. And found that the angina pectoris and healthy are mainly clustered into separated groups, especially 95% of healthy samples clustered into one group. Both methods showed the clearly divided for healthy and stable angina and can provide guidance for subsequent grouping analysis. The members of each group for the two methods can be found in supplementary material.
table 1. And supplementary figure1 explained why these samples should divide into three groups.

https://www.heraldopenaccess.us/fulltext/Alternative-Complementary-&-Integrative-Medicine/Supplements.xlsx

**Supplementary Tables 1&2:** clustering discovery result of iClusterPlus and consensusClusterPlus.

**Supplement Figure 1:** Graphical plots of Consensus ClusterPlus (A) item-consensus plot. The Item-consensus plot showed that k=3 has items with high consensus to their assigned cluster and low consensus to other clusters denoted by bars of mainly a single color. (B) The Delta Area Plot represents the change in the area under the CDF curves and shows that k=3 is the largest k with an appreciable increase in consensus. (C) cluster-consensus plot. Cluster-consensus (CLC) is the average pair wise IC of items in a consensus cluster. The CLC plot displays these values as a bar plot that are grouped at each k. According to the sum of this evidence, a cluster count of 3 is suitable for this project.

**Differential expressed genes**

Differential expression analysis between compared groups was performed to identify the candidate novel biomarkers of angina pectoris. Based on the gene expression level, R packages DESeq2 were used to detect the differentially expression genes.

Although it’s hard to find the significant biomarkers while comparing all the healthy and patients since some patient samples mixed in the healthy samples through the clustering. We performed the differential analysis for features based on the clustering results while samples from iCluster subtype 1 (subtype 2 and subtype 3 of consensus) as treatment and all healthy samples as control.

In total 874 genes were detected as significant differentially expressed in SA and among them 569 were up-regulated and 305 were down-regulated. It is obvious that a large proportion of genes were over expressed through figure 3 A/B. And based on the gene expression level distribution which showed in figure 3C and listed in table 3, the FPKM value of mainly over expressed genes are larger than 1 in SA group which indicate highly expressed genes may play the leading role in disease. As can be seen from the supplementary figure 2, most differential expressed genes were located in chromosomes 1 and 2 and 19 and 12, the function or other meaning was unclear and called for further research.

**Functional of DEGs**

Based on the significant differentially expressed genes and KEGG pathway annotation, Figure 4a below showed the pathway classification and functional information which may disturb by expression changes? Meanwhile, the number of up or down regulated genes in corresponding enrichment pathways are listed in Figure 4b. For each given gene list, pathway and process enrichment analysis has been carried out with the following ontology sources: KEGG Pathway, GO Biological Processes (Supplementary Table 2).
Discussion

In present, atherosclerotic disease is a leading cause of human death in the world, coronary heart disease like stable angina is one type of atherosclerotic disease. Here we found four significant higher expressed genes in SA samples which belong to Fluid shear stress and atherosclerosis pathway. The over expressed gene MAPK10/PDGFA/KDR and SUMO4 which showed in figure 5 can promote the occurrence and development of atherosclerosis and coronary heart diseases in signaling pathways.

A member of protein family involved both Platelet-Derived Growth Factors (PDGF) and Vascular Endothelial Growth Factors (VEGF) are encoded by gene PDGFA. Coding preproprotein is proteolytically processing generates subunit A of platelet-derived growth factor, can homodimerize or polymerization of platelet-derived growth factor subunit B. And the proteins and its receptor have been indicated likely to play important roles in multiple developmental processes. Previous scientific work reported that aberrant expressions of PDGF and its receptor (PDGFR) are often be considered as a potential drug target for development of diagnostic or therapeutic agents for multiple diseases including atherosclerosis [9]. Inhibition of stromal PDGF receptors can reduced angiogenesis throw depress the expression of PDGFA can be considered a target for potential treatment. And suppressed KDR expression was already as a biomarker of Nilotinib which may have effects on vascular endothelial functions [10].

The main pathogenic factor of atherosclerosis is lipid abnormalities. Hypercholesterolemia can promote atherosclerosis of blood vessels, the most dangerous of which is coronary atherosclerosis, which is the pathogenesis of angina pectoris. The DEGs in Fat digestion and absorption and Cholesterol metabolism should play great roles in the progress of SA disease. Statins are drugs always used to lower cholesterol in treatment of patients with stable angina. Previous research shows active steroid genesis holds an important status in the process of cholesterol movement from outer to the inner mitochondrial via the action of STAR [11]. The observation of disease-related abnormal gene expression also proved the accuracy of this research.

Some other different expressed genes have enriched in the heart-related pathways like adrenergic signaling in cardiomyocytes also have been reported to be related to illness. Previous study has suggested that CALN1 play essential roles in the physiology of neurons and

Table 3: Statistic of Differently Expressed Genes (DEseq2).

<table>
<thead>
<tr>
<th>Pathway Entry</th>
<th>Name</th>
<th>Class</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>map04270</td>
<td>Vascular smooth muscle contraction</td>
<td>Organismal Systems; Circulatory system</td>
<td>72(ACTG2);2983(GUCY1B1);8398(-CALN1);26279(PLAG2G10);83698(-CALD1);10398(MYL9);553(AVR1B)</td>
</tr>
<tr>
<td>map05418</td>
<td>Fluid shear stress and atherosclerosis</td>
<td>Human Diseases; Cardiovascular disease</td>
<td>5791(KDR);5602(MAPK10);5154(PDGFA);387082(SUMO4)</td>
</tr>
<tr>
<td>map04975</td>
<td>Fatty acid metabolism</td>
<td>Organismal Systems; Digestive system</td>
<td>8399(PLA2G10);26279(PLAG2G2)</td>
</tr>
<tr>
<td>map04979</td>
<td>Cholesterol metabolism</td>
<td>Organismal Systems; Digestive system</td>
<td>6770(STAR);11247(NXPH4)</td>
</tr>
<tr>
<td>map04261</td>
<td>Adrenergic signaling in cardiomyocytes</td>
<td>Organismal Systems; Circulatory system</td>
<td>83698(CALN1);23439(ATP1B4);90993(CREB3L1);85442(KND1);786(CACNG1)</td>
</tr>
</tbody>
</table>

Figure 4: a) Pathway classification of DEGs. X axis represents the number of regulated genes and the left Y axis represents the functional classification. b) The enrichment pathways of regulation genes. The X axis is the terms of pathway and the Y axis is the number of increased or decreased genes.

Table 4: Angina pectoris related pathways.
may have a great relationship with learning ability and memory [12] and ATP1B4 may associated with endometriosis pathogenesis. It is also reported that expression of CREB3L1 has correlation with survival time of the glioma patients [13] and the expression of CREB3L1 has been considered as a biomarker for triple negative breast cancer patients who may sensitive to drugs like doxorubicin-based chemotherapy. BMC Cancer 18: 1-7. 

Conflict of Interest Statement

Competing Interests: The authors have declared that no competing interests exist.