Candidate genes related to atherosclerosis

Identificationof theCandidateGenesinthe Progression ofAtherosclerosisbyBioinformaticsAnalysis

Running title: Candidate genes related to atherosclerosis

Highlights:

1. A total of 670 DEGs in atherosclerotic samples were obtained.
2. KIAA0101 may be a crucial molecule in atherosclerosis by binding with PCNA.
3. TRAC may participate in atherosclerosis by binding with MHC antigen.
4. PLXNC1 and HLA-DQB1 may play a key role in atherosclerosis via immune system.

Abstract

Objective: The purpose of our study was to identify candidate genes in the progression ofatherosclerosis. In addition, we aimed to explore the molecular mechanism of the development and progression of atherosclerosis.

Materials and Methods: The gene expression profile of GSE28829 was downloaded from Gene Expression Omnibus (GEO) database. The differentially expressed genes (DEGs) in early and advance atherosclerotic samples were analyzed with limma package. Cluster analysis of the screened DEGs was performed through affinity propagation cluster method (APCluster). The DEGs enrichment was obtained via Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Finally, protein-protein interaction (PPI) network of the high relative genes was constructed using the Cytoscape software.

Results: Totally, 670 DEGs in atherosclerotic samples were obtained. After cluster analysis of DEGs, 28 genes were selected from the 69 clustering gene sets. The most significant gene wasmajor histocompatibility complex, class II, DQ beta 1 (HLA-DQB1) in the 14 biological pathways based on the pathway enrichment analysis of KEGG. Totally, 3 DEGs ofKIAA0101, plexin C1 (PLXNC1) andT cell receptor alpha constant (TRAC) were gained after the attributive analysis of protein-protein interaction (PPI) network.

Conclusion: KIAA0101, PLXNC1 and TRAC may be candidate genes for regulating the progression of atherosclerosis.

Keywords: atherosclerosis; bioinformatics analysis; differentially expressed genes

Introduction

Atherosclerosis is an ongoing process which already starts in childhood [1]. It is a progressive multifaceted inflammatory disease that affects large- and medium-sized arteries by thickening and hardening these arteries [2-4]. Atherosclerosis can mainly induce myocardial infarction or ischemic stroke, which is considered as the most common reason of death worldwide [5]. Besides, numerous lines of evidence suggest that many other diseases, such as coronary heart disease, diabetes mellitus and dyslipidemia, are associated with atherosclerosis [6-8]. Thus, atherosclerosis has been a worldwide threat to public health.

Atherosclerosis has been reported to be a disorder with multiple genetic and environmental contributions [9]. Genetic-epidemiologic studies have identified the activation of protein kinase C (PKC) was induced by elevated

levels of diacylglycerol, which resulted from a high concentration of glucose and nonesterified fatty acids. This event is considered to be a link between altered vascular cell signaling and abnormal metabolism, finally lead to atherosclerosis [10]. Furthermore, PKC isoforms can induce expression of pro-inflammatory cytokine and activation of nuclear factor κB and NADPH oxidase, as well as increase signaling proteins, eg. extracellular signal-regulated kinase (ERK) and p38 mitogen-activated protein kinase (MAPK), thus contribute to vascular inflammation and atherosclerosis [11]. Angiotensin II (ATII) has been demonstrated to have direct effects on development of atherosclerosis through stimulation of monocyte recruitment, activation of macrophages, and enhanced oxidative stress, all of which have been linked to increase of the atherogenic process [12-14]. On the other hand, angiotensin-(1-7) (AT-(1-7)) opposes AT II action to prevent atherosclerosis [15]. In spite of the expanded efforts to study the genetic bases of atherosclerosis, the molecular mechanisms of the development and progression still needed further study.

In the present study, we aimed to identify the differentially expressed genes (DEGs) from the early- and advanced-stage atherosclerotic samples and explore the molecular mechanisms in the onset and progression of atherosclerosis. Furthermore, the in-depth understanding of this disease can provide the basic for appropriate treatment.

Materials and methods

Data source

Theatherosclerotic gene expression profiles ofGSE28829 [16] was downloaded from Gene Expression Omnibus (GEO) database (http://www.

ncbi. nlm. nih. gov/geo/) based on Affymetrix Human Genome U133 Plus 2. 0 Array (GPL570[HG-U133_Plus_2]). A total of 29 samples, consisting of 13 early atherosclerotic samples and 16 advanced atherosclerotic samples, were available in the present analysis.

Data pre-processing

The original probe-level data in CEL files were converted into expression measures using the affy package [17] in R language. Quality analysis was constructed to confirm the available chips. Background correction and quartile data normalization were performed by gcrma package [18] to obtain the expression profile data.

Identification of differentially expressed genes (DEGs)

The Affymetrix Microarray Suite 5 (mas5) calls was applied to get the expressive gene in at least one chip by gene screening. After that, differentially expressed genes (DEGs) were identified through the contrastive analysis of early- and advanced -stage samples using the limma package [19]. Fold change value (| $\log_2$ FC|) of DEGs larger than 2. 0 and P -value less than 0. 05 were used as the cut-off criterion.

Cluster analysis of DEGs

The genes in relation to atherosclerosis were downloaded from Online Mendelian Inheritance in Man (OMIM). Then, cluster analysis of the screened DEGs was performed through affinity propagation cluster method (APCluster) [20], which works based on consideration of all the data points as potential cluster centers [21].

Functional enrichment analysis

For the screened DEGs, functional analysis of GO (gene ontology) term, using biological process (BP) GO term, was firstly performed using the online DAVID (database for annotation, visualization, and integrated discovery) [22], GOEAST (Gene Ontology Enrichment Analysis Software Toolkit) [23] and Toppgene [24]. Reliable results were obtained via the comprehensive analysis of the three results. Then, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways of the DEGs enrichment were also obtained using the online DAVID. P-value less than 0. 05 and false discovery rate (FDR) less than 0. 05 were considered as the cut-off criterion.

Protein-protein interaction (*PPI) network construction*

The screened DEGs were further selected combining with the known atherosclerosis genes. PPI data were downloaded from human PPI datasets, such as HPRD (Human Protein Reference Database) [25], BioGRID (Biological General Repository for Interaction Datasets) [26], MINT (Molecular INTeraction database) [27], InAct [28] through which the highest reliability PPI data was obtained. The PPI networks were constructed using the Cytoscape software [29] based on the PPI relationship.

Results

Identification of DEGs

We used Limma package in R to construct the contrastive analysis between the early- and late-stage samples. According to the cut-off criteria of $|\log_2 FC| > 2.0$ and P-value < 0. 05, we finally gained 670 DEGs.

Cluster analysis of DEGs

There were 4 atherosclerotic genes from OMIM (table 1), and only arachidonate 5-lipoxygenase (ALOX5) was one of the screenedDEGs. After cluster analysis of DEGs, a total of 69clustering gene sets were gained, andthe clustering heat map shown in Fig 1. Finally, 28 genes were selected from the 69 clustering gene sets based on whether ALOX5 was in the clustering gene sets (table 2).

Functional enrichment analysis

A total of 187 DEGs were obtained based on a $P < 0.05$ after the functional enrichment analysis of BP GO terms. Following the pathway enrichment analysis of KEGG, 14 biological pathways (shown in table 3) were chosen according to FDR < 0. 05. The most significant enrichment pathway in atherosclerosis was systemic lupus erythematosus, and the most significant gene was major histocompatibility complex, class II, DQ beta 1 (HLA-DQB1).

PPI network

The resources of PPI networks were showed in table 4. The PPI network was obtained with 4, 888 nodes and 88, 423 edges. Then, the node degree, betweenness centrality and closeness centrality in the PPI network were calculated to select high relative genes. The top 10 key genes in each group were showed in table 5. Totally, there were 17 genes including 3 DEGs of KIAA0101, plexin C1 (PLXNC1) and T cell receptor alpha constant (TRAC) and 1 known relative gene of estrogen receptor 1 (ESR1). At last, akey sub-network, consisting of 3, 859 nodes and 75, 916 interaction edge, was constructed using the 17 genes and their neighbor nodes.

Discussion

Generally, atherosclerosis is a systematic and widespread disease [30]. It is frequently associated with chronic heart disease of the arteries with high death rates all over the world [31]. Nowadays, the diagnosis of atherosclerosis still remains difficult for its asymptomatic early-stage [32]. Therefore, genetic therapies or new drug targets in the early detection of atherosclerosis is urgently needed based on the molecular pathogenesis. In the current study, we analyzed 670 DEGs upon gene expression profile of early and advance stage atherosclerosis by bioinformatics analysis. Finally, 3 DEGs including KIAA0101, PLXNC1 and TRAC was significantly identified. By analysis of the 14 biological pathways enriched by DEGs, we found that HLA-DQB1 might be also an important gene associated with atherosclerotic development.

The gene of KIAA0101 located in nucleus and mitochondrion [33, 34]. It encodes a 15 kDa protein, which is related to the proliferative activity of human atherosclerotic lesions [35, 36]. The KIAA0101 protein can bind with conserved proliferating cell nuclear antigen (PCNA) competitively with p21WAF, a cell cycle regulator [37]. PCNA has been reported to play a key role in DNA replication and damage repair as an indispensable factor for DNA polymerase, that is, the DNA repair, apoptosis and cell cycle progression is partly attributed to PCNA [33, 38, 39]. In addition, PCNA, as an essential component of the DNA synthesis machinery, promotes the proliferation human vascular smooth muscle cells [36]. To the best of our knowledge, atherosclerotic lesions is fundamentally characterized by the accumulation of

cells within the intima [40]. Therefore, KIAA0101 may be a crucial molecule in the progression of atherosclerosis by binding with PCNA.

PLXNC1 encodes plexin C1 (a member of the plexin family), which is one of the semaphorins (SEMA) receptors, and the semaphorins is known as a large family including transmembrane and secreted signalling proteins [41]. Particularly, plexins are receptors for transmembrane semaphorins, which are frequently involved in immune response, from initiation to terminal inflammatory processes [42]. Immune responses consisting of adaptive and innate immunity have been evidenced to tightly participate in regulation during the progression of atherosclerosis [43], as immune activation is part of the disease process [44]. Besides, plexin can interact with some semaphorins on monocytes [45, 46]. As a result, monocytes are exposed to soluble SEMA4D/CD100 (a critical semaphorin in the immunoregulation), representing a significant down-modulation in pro-inflammatory cytokine production and leading to immune response [47]. These lead us to hypothesis that PLXNC1 may play an important role in the development of atherosclerosis through modulation in immune response.

TRAC, another significant DEG in our study, is a protein-coding gene. The TRAC protein can bind major histocompatibility complex (MHC) antigen. Aberrant MHC antigen expression in smooth muscle and endothelial cells may active T lymphocytes. Meanwhile, the activated T lymphocytes may modulate the functions of other cells in atherosclerotic plaque and the significant amounts of T lymphocytes are also an important cause of atherosclerosis [48]. In addition, overexpression of MHC antigen may also participate in the perpetuation of the atherogenetic autoimmune reaction

[49]. Therefore, TRAC may participate in the development of atherosclerosis via binding MHC antigen.

Besides, in the present study, HLA-DQB1 was shown as the most significant gene involved in several biological pathways. HLA-DQB1 belongs to the histocompatibility leukocyte antigen (HLA) class II beta chain paralogs. The protein encoded by HLA-DQB1 is essential for constructing the DQ heterodimer, which is a cell surface receptor playing a central role in the immune system [50, 51]. Moreover, HLA-DQ may contribute to the presentation of antigen to suppressor systems. The antigen-presenting function is possibly related to inflammatory mechanisms in atherosclerosis [50]. Thus, HLA-DQB1 may be a significative gene in the development of atherosclerosis involving in the immune system.

As a result of this preliminary study, the 3 DEGs of KIAA0101, PLXNC1 and TRAC may be candidate genes that tightly associated with the development and progression of atherosclerosis. In addition, HLA-DQB1 involved in biological pathways may be also an important gene that plays a pivotal role in atherosclerotic development. These findings may provide possible molecular mechanism for well treatment of atherosclerosis. However, further study is warranted to verify our conclusions with more genetic experiments of DEGs as no experiments is performed in the present study.