Identification of novel genetic biomarkers and treatment targets for arteriosclerosis-related abdominal aortic aneurysm using bioinformatic tools

Fang Niu1,†, Zongwei Liu1,†, Peidong Liu2,†, Hongrui Pan1, Jiaxue Bi1, Peng Li1, Guangze Luo1, Yonghui Chen1, Xiaoxing Zhang1 and Xiangchen Dai1,*

1 Department of General Surgery, General Hospital of Tianjin Medical University, Tianjin Medical University, Tianjin, China
2 Department of Neurosurgery, First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

† The authors contributed equally to this work.

* Correspondence: Email: 13302165917@163.com; Tel: +8613302165917.

Abstract: A large number of epidemiological studies have confirmed that arteriosclerosis (AS) is a risk factor for abdominal aortic aneurysm (AAA). However, the relationship between AS and AAA remains controversial. The objective of this work is to better understand the association between the two diseases by identifying the co-differentially expressed genes under both pathological conditions, so as to identify potential genetic biomarkers and treatment targets for atherosclerosis-related aneurysms. Differentially-expressed genes (DEGs) shared by both AS and AAA patients were identified by bioinformatics analyses of Gene Expression Omnibus (GEO) datasets GSE100927 and GSE7084. These DEGs were then subjected to bioinformatic analyses of protein-protein interaction (PPI), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Finally, the identified hub genes were further validated by qRT-PCR in AS (n = 4), AAA (n = 4), and healthy (n = 4) individuals. Differential expression analysis revealed a total of 169 and 37 genes that had increased and decreased expression levels, respectively, in both AS and AAA patients compared with healthy controls. The construction of a PPI network and key modules resulted in the identification of five hub genes (SPI1, TYROBP, TLR2, FCER1G, and MMP9) as candidate diagnostic biomarkers and treatment targets for patients with AS-related AAA. AS and AAA are indeed correlated; SPI1, TYROBP, TLR2, FCER1G and MMP9 genes are potential new
genetic biomarkers for AS-related AAA.

**Keywords:** gene expression analysis; bioinformatics; biomarkers; therapeutic targets; arteriosclerosis-related abdominal aortic aneurysm

### 1. Introduction

Abdominal aortic aneurysm (AAA) has a high prevalence in western countries and leads to a large number of cardiovascular deaths [1]. Aneurysm is defined as an increase in artery diameter by no less than 1.5 times. Aneurysm can appear in any part of arteries in the human body, with AAA being the most common subtype [2]. AAA will cause artery rupture and continuous bleeding, eventually leading to deaths. The mortality rate of AAA after artery rupture is 85–90% [3]. If the diameter of artery is more than 5.5 cm in AAA, surgery is usually performed to prevent artery rupture. AAA is a disease with complex pathological mechanisms, therefore, it is challenging to develop efficient therapeutic strategies to prevent or treat AAA [4]. Although several lines of evidence have suggested that AAA is a multi-gene and multi-factor disease, its pathogenesis is not completely clear [5].

The very association of arteriosclerosis (AS) and AAA may be up for debate. Previous studies have also indicated that arteriosclerosis (AS) is likely negatively associated with the growth of AAA but positively correlated with its presence [6–8]. It is widely believed that AAA is merely a consequence/manifestation of AS, but this notion is yet to be fully substantiated. Although revascularization and risk factor control methods are extensively utilized for the treatment of AS-related AAA clinically, the disease is still the most common cause of death. Therefore, in the era of precision medicine, novel strategies are urgently needed to prevent and treat AS-related AAA. However, the biomarkers and therapeutic targets for AS and AAA are yet to be sufficiently explored. Besides, a comprehensive analysis of the associations between AS and AAA at the transcriptomic level is still lacking.

Herein, we utilized gene expression profiles retrieved from the Gene Expression Omnibus (GEO) database to identify differentially-expressed genes (DEGs) in both AS and AAA patients. The physiopathological functions of these DEGs were then characterized bioinformatically to discover novel diagnostic biomarkers and treatment targets for AS-related AAA.

### 2. Materials and methods

#### 2.1. Artery specimen acquisition

This research was supervised and supported by the research ethics committee of the Tianjin Medical University General Hospital. Moreover, this study was conducted in accordance with the declaration of Helsinki. The participants were granted informed consent prior to their enrollments in our study. Artery samples were collected from popliteal arteries of 4 AS and abdominal aortas of 4 AAA patients treated with lower limb amputation and artificial vessel replacement, respectively. Normal artery specimens from visceral arteries of 4 healthy donors were also collected as controls. After arterial resection, total RNA was extracted from tissues rapidly frozen in liquid nitrogen.
2.2. Data sources

Two GEO (https://www.ncbi.nlm.nih.gov/geo/) datasets, namely GSE100927 (deposited by Marja) and GSE7084 (deposited by Gerard) were employed in this research. The GSE100927 dataset contains 104 samples, including 69 atherosclerotic peripheral artery samples and 35 control samples; while the GSE7084 dataset includes gene expression profiles of 15 abdominal aortas samples obtained from 7 AAA patients and 8 healthy controls.

2.3. Differential expression analysis

We utilized the “affy”, “affyPLM”, and “limma” packages in R (http://www.bioconductor.org/packages/release/bioc/html/affy.html) [9] to analyze the GSE100927 and GSE7084 datasets. The data were background-adjusted, quantile-normalized, subjected to probe summarization, and log2 transformed, to generate a robust multi-array average (RMA) and log-transformed perfect match (PM) and mismatch (MM) probes. The initial p-values were adjusted by employing the Benjamini-Hochberg method, and fold changes (FCs) were computed on the basis of the false discovery rate (FDR). A gene was considered as a DEG if it had a |log2 FC| > 1 and an adjusted p value < 0.05. Subsequently, DEGs shared by the GSE100927 and GSE7084 datasets were isolated, based on which a Venn diagram was established using Venny 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

2.4. Functional annotations of the DEGs

To annotate the functions of DEGs shared by the GSE100927 and GSE7084 datasets, we performed gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses on them using R. We also used the “clusterProfiler” software in R for gene set enrichment analysis (GSEA) [10]. All visualizations were processed by the “ggplot2” package in R.

2.5. Module screening and hub gene identification based on the protein-protein interaction (PPI) network

The comprehensive information of the hub proteins retrieved from STRING (https://string-db.org/), an online database for the retrieval of inter-gene and inter-protein interactions, was utilized to establish a PPI network, which was visualized using the Cytoscape software (Version 3.7.2) [11]. Subsequently, most important modules within the PPI network were identified by the “Molecular Complex Detection (MCODE)” function of Cytoscape with the following identification criteria: an MCODE score of more than 5, degree and node score cutoff values of 2 and 0.2, respectively, and a maximum depth of 100.

Based on the cutoff criteria of degree computed by the “cytoHubba” function of Cytoscape, genes that played important roles in the PPI network were the hub genes. The DEGs were scored by the 12 scoring methods of the “cytoHubba” function, and the top ten genes with the highest scores in each scoring method were registered. Based on the 12 ranking lists obtained, upset charts were constructed and the number of times that a gene was registered was counted.
2.6. Quantitative real-time polymerase chain reaction (qRT-PCR)

We extracted total RNA from the above-mentioned artery samples using the TRIzol reagent (TransGen Biotech, China). The total RNA was then reversely transcribed into cDNA using a real-time PCR mRNA detection kit (TIANGEN Biotech, China) before PCR amplification on a Roche Photorecycler 480 Real-time PCR system. The reaction consisted of the following steps: one step of 95°C for 30 sec and 40 rounds of 95°C for 5 sec and 60°C for 20 sec. GAPDH was included as the internal reference. Relative transcript levels of target genes were determined by the $2^{-\Delta\Delta Ct}$ method. The PCR primers were designed and synthesized by AoKe Dingsheng (Beijing, China) and their sequences are shown in Table 1.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence (5′→3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPI1</td>
<td>F: GTGCCCTATGACACGGATCTA</td>
</tr>
<tr>
<td></td>
<td>R: AGTCCCAGTAATGGTCGCTAT</td>
</tr>
<tr>
<td>TYROBP</td>
<td>F: ACTGAGACCGAGTCGCCTTAT</td>
</tr>
<tr>
<td></td>
<td>R: ATACGGCCTCTGTGTGTGGAG</td>
</tr>
<tr>
<td>TLR2</td>
<td>F: ATCCTCCAATCAGGCTTCTTCT</td>
</tr>
<tr>
<td></td>
<td>R: GGACAGGTCAAGGCTTTTTACA</td>
</tr>
<tr>
<td>FCER1G</td>
<td>F: AGCAGTGGTCTTGCTTTACT</td>
</tr>
<tr>
<td></td>
<td>R: TGCCTTTTCGCACTTGGATCTT</td>
</tr>
<tr>
<td>MMP9</td>
<td>F: TGTACCGCTATGGTTACACTCG</td>
</tr>
<tr>
<td></td>
<td>R: GGCAGGGACAGTTGCTTTCT</td>
</tr>
<tr>
<td>GAPDH</td>
<td>F: ATGGCTACAGCAACAGGGT</td>
</tr>
<tr>
<td></td>
<td>R: TTATGGGGTCTGCGGATGG</td>
</tr>
</tbody>
</table>

2.7. Statistical analysis

GraphPad Prism 8 was used to compare data from test and control samples. To analyze the clinical characteristics of AS patients, AAA patients and healthy controls, the quantitative data are expressed as means ± standard deviations (SDs). The Student’s t-test and Wilcoxon rank sum test were implemented for analysis of data as appropriate. P values less than 0.05 were indicative of statistical significance. All statistical analyses in this study were conducted with the SPSS 26.0 software (IBM, USA).

3. Results

3.1. DEGs identified in both GEO datasets

Microarray data from both GEO datasets were normalized and subjected to principal component analysis (PCA) for evaluating differences in biological characteristics between samples (Figure 1A,B), thereby ensuring data accuracy. Genes with a $|\log_{2}\text{FC}| > 1$ and an adjusted p value < 0.05 were identified using the “Limma” package in R. Compared with controls, 324 and 477 overexpressed and 113 and 370 underexpressed genes were identified in the peripheral artery samples of AS and AAA patients in the GSE100927 and GSE7084 datasets, respectively (Figure 2). The DEGs are also presented by volcano plots and heat maps shown in Figure 1C–F. The DEGs shared by the GSE100927 and GSE7084 datasets were exhibited by a Venn diagram (Figure 3A), including 169 upregulated and 37 downregulated DEGs.
**Figure 1.** Identification of differentially expressed mRNAs analysis. Principal component analysis (PCA) graph in (A) GSE7084 and (B) GSE100927. Volcano plot of differentially expressed mRNAs in (C) GSE7084 and (D) GSE100927. The red dot represents upregulated mRNAs and the blue dot represents downregulated mRNAs. A heatmap of 85 differentially expressed genes between (E) AAA patients and normal individuals, (F) AS patients and normal individuals. Red represents upregulated genes, and green represents downregulated genes.
Figure 2. Flow diagram of the study approach. mRNA microarray analyses were performed on aortic specimens obtained from GSE100927 (69 AS patients and 35 control group) and GSE7084 (7 AAA patients and 8 control group). From among differentially-expressed genes (DEGs), network analysis of gene expression both in AAA and AS identifies 206 co-DEGs. Afterwards, functional and pathway enrichment analysis were performed for the co-DEGs. The top 5 genes with the highest scores in cytoHubba module were taken as hub genes. At last, validation quantitative polymerase chain reaction (qRT-PCR) was performed for the 5 hub genes in samples obtained from our institute, including 4 organ donors, 4 AAA patients and 4 AS patients.
3.2. Functional annotation of the identified DEGs

The functions of the identified DEGs were explored via GO and KEGG enrichment analyses in R. The results suggested that the DEGs shared by the GSE100927 and GSE7084 datasets were
significantly enriched with biological processes (BP) terms of "neutrophil activation", "immune response", and "neutrophil degranulation", a molecular function (MF) term of "actin binding", as well as a cellular components (CC) term of "secretory granule membrane" in the GO analysis (Figure 3B). Our KEGG analysis suggested that these shared DEGs were associated with 10 pathways, such as "lipid and atherosclerosis", "phagosome", and "osteoclast differentiation" (Figure 3C).

3.3. PPI network analysis

To characterize the interactions between proteins encoded by these shared DEGs, a PPI network was built by STRING (Version 11.0) and visualized by Cytoscape (Figure 4A). The network consisted of 190 nodes (representing 190 proteins) and 1715 edges (representing 1715 interactions between the proteins). In addition, 3 essential modules were identified from the network by MCODE in Cytoscape (Table 2). According to the upset chart, five genes (namely SPI1, TYROBP, TLR2, FCER1G and MMP9) were registered more than 5 times and were regarded as hub genes (Figure 4B). Interestingly, these genes were also included in the modules with the highest MCODE scores (Figure 4C).

**Table 2. Module analysis of DEGs using Cytoscape.**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Score</th>
<th>Nodes</th>
<th>Edges</th>
<th>Node IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.6</td>
<td>21</td>
<td>196</td>
<td>LAPTM5, FCER1G, LY86, CCR1, CD53, FCGR2A, FCGR3A, CSF1R, TYROBP, C1QA, SPI1, IL10RA, LILRB2, CYBB, CTSS, LCP2, MNDA, CD74, MPEG1, AIF1, ITGAM</td>
</tr>
<tr>
<td>2</td>
<td>8.5</td>
<td>21</td>
<td>85</td>
<td>CCL5, ITGAL, CCL3, PARVG, CCR5, CCR7, IGSF6, BTK, CD83, TLR2, IL1B, CSF3R, HLA-DRA, VAV1, CD68, HK3, FGFR, CD14, MYO1F, CCL4, NCF2</td>
</tr>
<tr>
<td>3</td>
<td>6.207</td>
<td>30</td>
<td>90</td>
<td>EVI2B, MS4A6A, CCR2L2, IL7R, VAMP8, SYK, APOE, LILRB3, STK10, ADAM8, HCST, LYN, HMOX1, INPP5D, SLA, ALOX5, CXCR4, CORO1A, MMP9, CXCL16, NCF4, PTPN6, TREM1, IL1RN, HCLS1, SLC2A5, AQP9, PLAU, RAC2, VAV3</td>
</tr>
</tbody>
</table>
Figure 4. PPI network of co-DEGs constructed by the STRING database. (A) The PPI network of co-DEGs was constructed using Cytoscape. (B) Among the 12 scoring methods of the cytoHubba module, the intersection of the top ten genes in each group. (C) Identification of a sub-network using MCODE in Cytoscape software. Diamond nodes, hub genes; Red nodes, up-regulated genes; Blue nodes, down-regulated genes.

3.4. Verification of the identified DEGs

Differential expression of the DEGs, as revealed by bioinformatic methods, was further confirmed by qRT-PCR using cDNA obtained from artery samples from AS patients (n = 4), AAA patients (n = 4), and healthy controls (n = 4). As shown in Figure 5, SPI1, TYROBP, TLR2, FCER1G and MMP9 were significantly upregulated both in AS and AAA samples compared with normal controls.
Figure 5. Validation of the gene expression levels of hub genes between patients and healthy control group. (A–E) The expression of SPI1, TYROBP, TLR2, FCER1G and MMP9 in ASO and AAA arteries compared with normal arteries. The data are presented as the mean ± SEM, **P < .01 vs control group; ***P < .001 vs control group.

4. Discussion

A large number of clinical investigations have comprehensively exhibited an enhanced risk of AAA in patients with AS. However, the mechanism underlying the association between AS and AAA remains elusive, although it has been speculated to be associated with chronic inflammation, degradation of extracellular matrix, apoptosis of vascular smooth muscle cells, and thrombosis [12], physiological characteristics that also contribute to the formation of atherosclerotic plaques, suggesting that AS, which frequently affects the aneurysm wall, is a crucial risk factor for AAA. We herein discovered 5 key genes implicated in AS-related AAA, namely SPI1, TYROBP, TLR2, FCER1G and MMP9. These genes have the potential to become novel biomarkers and treatment targets for AS-related AAA. Our findings not only provide further insights into the pathogenic mechanisms of AS-related AAA, but also have clinical significance—they suggest several potential early diagnostic markers to prevent the progression of AAA.

Among the identified hub genes, SPI1 encodes PU.1, an ETS-domain transcription factor that is required for the development of myeloid cells and can regulate intercellular communications to facilitate immune responses [13]. The role of the SPI1 in AAA progression is unclear. According to previous findings, the expression of SPI1 gene is upregulated during the differentiation of myeloid cells and remains persistently high in B cells, monocytes, mast cells, circulating neutrophils, and other granulocytes in human [14]. Pan et al. found that SPI1 expression was significantly increased in Tibetan minipigs subjected to a high-fat/cholesterol diet [15], further implying that the activation of myeloid cells is critical for AS development. TYROBP, which has a synonym of DNAX-activating protein of 12 kDa (DAP12), encodes a transmembrane receptor ubiquitously expressed in neutrophils, natural killer cells, and monocytes/macrophages that transduces immune signals via a tyrosine-based
activation motif [16]. DAP12 has been extensively implicated in the survival, reproduction, polarization, and differentiation of multiple types of immune cells, particularly the monocytes/macrophages [16]. Wang et al. observed that DAP12 was expressed at a high level in APOE mice, which facilitated the development of AS plaques via the TREM-1/DAP12 pathway [17]. Hinterseher et al. observed that TYROBP was highly expressed in human AAA and was involved in the pathogenesis of AAA through the NK cell-mediated cytotoxicity pathway [18]. TLR2 encodes a cell membrane protein that identifies microorganism invasions and regulates innate immune reactions [19]. Jabłońska et al. found significantly higher levels of TNF-α, IL-4 and TLR2 in AAA patients compared with healthy controls, and suggested that TLR2 may be involved in the aortic and systemic inflammatory responses in AAA patients [20]. Lee et al. showed that TLR2 could activate p38 and ERK1/2 signaling pathways, selectively regulate IL-6–mediated RANKL upregulation and OPG downregulation, and induce the differentiation of VSMCs into chondrocytes, thereby resulting in vascular calcification during AS development [21]. FCER1G (CD23) encodes a high-affinity cell membrane receptor for IgE, which is upregulated during aging [22]. FCER1G was demonstrated to be able to regulate hypertension and platelet hyperaggregability during thrombus formation [23–25]. Given that AAA is tightly associated with these diseases, it is reasonable that FCER1G may be implicated in the development of AAA. MMP9 encodes a matrix metalloproteinase that reorganizes and degrades extracellular matrix during collagen catabolic processes, leading to the instability and rupture of atherosclerotic plaques [26,27]. This molecule activates the immune responses that lead to atherosclerotic plaque destabilization by facilitating the degranulation, transendothelial migration, and differentiation of leukocytes, as well as activating interleukin signaling pathways. However, its role in AAA progression remains controversial. Hurks et al. found that the MMP9 levels remained stable in a study of inflammation in AAA artery wall [28]. Another study found that MMP9 level in AAA was correlated with intraluminal thrombus amount, implying that thrombus activated MMP9 [29].

In this study, we identified DEGs shared by the AS and AAA cohorts and discovered 5 hub genes (namely SPI1, TYROBP, TLR2, FCER1G and MMP9) by PPI network construction and identification of key modules. The relationship between these genes, AS and AAA needs to be further investigated in future studies.

5. Conclusions

By analyzing GEO datasets using bioinformatic methods and verifying the results by in vitro experiments, we not only identified 206 DEGs shared by both AS and AAA cohorts, but also detected 5 hub genes (SPI1, TYROBP, TLR2, FCER1G and MMP9) linking AS and AAA, which may be used as potential biomarkers for the diseases. More importantly, these hub genes may become potential treatment targets for AS-related AAA, which will facilitate the establishment of new therapeutic strategies, including early prevention and precision treatment.

Limitations

Our study has some limitations. Firstly, our research was based on microarray data retrieved from a public database. Secondly, since transcript abundances of genes may not be able to accurately reflect the corresponding protein levels, for example, we do not know which type of cells are mainly involved
in the expression of these hub genes, and whether the complex in vivo environment will mask some of more meaningful genes, additional in vitro and in vivo studies are required to characterize the functions of the key genes. Therefore, large-scale prospective clinical investigations can, to some extent, better support our findings.

Acknowledgements

We thank Yudi Liu for technical assistance.

Funding

This study was supported by the National Natural Science Foundation of China [Grant Numbers 82070489].

Ethics approval and consent to participate

This study was approved on June, 2021 by the Institutional Review Board of the Tianjin Medical University General Hospital with an approval number of IRB2021-WZ-080.

Conflict of interest

The authors declare that they have no competing interests.

References


©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)