Identification of key genes and pathways in calcific aortic valve disease by bioinformatics analysis

Yiran Zhang, Liang Ma

Department of Cardiovascular Surgery, the First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Contributions: (I) Conception and design: All authors; (II) Administrative support: L Ma; (III) Provision of study materials or patients: Y Zhang; (IV) Collection and assembly of data: Y Zhang; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Background: Calcific aortic valve disease (CAVD) is the most common type of valvular heart disease in the elderly. This study is aimed to explore molecular mechanism of CAVD via bioinformatics analysis.

Methods: The gene expression profiles of GSE51472 (including 5 normal aortic valve and 5 calcified aortic valve) and GSE83453 (including 8 normal aortic valve and 19 calcified aortic valve) were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were screened using the MetaDE package in R software. Functional and pathway enrichment analysis were performed based on Gene ontology (GO) and KEGG pathway database. Then, STRING database, Cytoscape and MCODE were applied to construct the protein-protein interaction (PPI) network and screen hub genes. Pathway enrichment analysis was further performed for hub genes and gene clusters identified via module analysis.

Results: A total of 107 DEGs were identified in CAVD (53 up-regulated genes, and 54 down-regulated genes), and they were mainly enriched in the terms of immune response, extracellular matrix organization, leukocyte transendothelial migration, cell adhesion molecules (CAMs), and fatty acid metabolism. Five hub genes including VCAM1, MMP9, ITGB2, RAC2, and vWF were identified via PPI network, which were mainly enriched in terms of leukocyte transendothelial migration and cell adhesion. An independently down-regulated protein cluster containing ALDH2, HIBCH, ACADVL, ECHDC2, VAT1L, and MAOA was also identified via PPI network.

Conclusions: The present study identified VCAM1, MMP9, ITGB2, RAC2, vWF and ALDH2 as key genes in the progression of CAVD. Immune cells infiltration might play a key role in the progression of CAVD, while ALDH2-mediated detoxification effect might play a protective role in CAVD. Further studies are needed to elucidate the pathogenesis of CAVD.

Keywords: Calcific aortic valve disease (CAVD); Gene Expression Omnibus (GEO) data; differentially expressed genes (DEGs); integrated bioinformatics analysis

doi: 10.21037/jtd.2019.11.57
View this article at: http://dx.doi.org/10.21037/jtd.2019.11.57

Introduction
Calcific aortic valve disease (CAVD) is the most common type of valvular heart disease in the elderly (1). Calcific aortic valve stenosis occurred in 2% of people above 65 years old; while the incidence of aortic valve sclerosis, the early stage of CAVD without hemodynamic consequence, was as high as 26% (2). Even without hemodynamically significant obstruction of left ventricular outflow, CAVD still can be an independent risk factor for cardiovascular mortality, all-cause mortality and myocardial infarction (1,3). Currently, the only effective therapeutic option to improve clinical outcomes in CAVD patients is early surgery (4).
Histological studies suggested that the early lesion of CAVD was an active inflammatory process shared some similarities with atherosclerosis, such as lipid deposition, macrophage and T-cell infiltration, and basement membrane disruption (5). While the end-stage of CAVD usually involved heterotopic ossification consisting of mature lamellar bone formation and active bone remodeling (6). Several serum biomarkers such as lipoprotein(a) (2,7), LDL-C (2), and serum phosphate (8) had been found to be associated with the prevalence of CAVD. Various molecular, cellular, and biomechanical mechanisms had been found involved in the progression of CAVD, such as bicuspid aortic valve, osteogenic mechanisms, endocrine mechanisms, dysregulated mineral metabolism, osteoclast deficiency, and developmental signaling (9). Growing evidences suggested that the progression of CAVD was a actively regulated process, which might be suspended or reversed via potential pharmacologic intervention (10). However, thus far, there have been no identified pharmacologic interventions or presumably underlying pathways that affect the CAVD progression process. Besides, the key genes and signaling pathways underlying the progression of CAVD remain poorly understood.

Gene expression microarrays have been widely used to study gene expression profiles in many human diseases, which provide an effective way for studying disease-associated genes and discovering novel therapeutic targets and biomarkers (11). Recently, several studies had compared the mRNA expression profile of CAVD (12,13). However, the results were inconsistent due to small sample size, different technological platforms, different data processing and sample heterogeneity. The disadvantage of these single studies can be overcome by integrated bioinformatics analysis, which combines different microarray dataset to obtain a more steady result (14). MetaDE package developed by Wang et al. contains more than 12 popular genomic meta-analysis methods, which is an effective integrative analysis solution (15).

In the present study, we applied integrated bioinformatics methods to identify key genes and pathways in CAVD. Two microarray profile datasets [GSE51472 (12) and GSE83453 (13)] were downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were screened using the MetaDE package in R software. Gene ontology (GO) and pathway enrichment analysis of DEGs were then performed. A protein-protein interaction (PPI) network was established, and key genes and pathways associated with CAVD were analyzed.

**Methods**

**Microarray data**

The present study was a bioinformatics study based on publicly accessible microarray datasets. Thus, an ethics approval was not required. The gene expression profiles of GSE51472 (12) and GSE83453 (13) were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo). The platform for GSE51472 is GPL570, Affymetrix Human Genome U133 Plus 2.0 Array, which includes 5 normal aortic valve samples, 5 fibrotic aortic valve samples and 5 calcified aortic valve samples. The control group contained aortic valves from patients underwent surgery due to ascending aortic aneurysm or dissection, while the valves were macroscopically smooth, pliable and opalescent, and without visible calcifications. The fibrotic group contained valves that were macroscopically thickened and stiff, while the patients had no significant transvalvular pressure gradient. The calcified group contained aortic valves from patients with severe aortic valve stenosis, and the valves were presented with varying degrees of calcification. Only the normal and calcified aortic valve samples were used for our bioinformatics analysis. The dataset information is shown in Table 1. The data were normalized and log2 transformed.

**Data preprocessing**

The probe IDs was converted into international standard names for genes (gene symbol). For multiple probes that mapped to a single gene, interquartile range (IQR) of each probe was calculated with MetaDE.match package in R software, and the probe with largest IQR was selected to represent the expression level of the gene. Next, MetaDE.merge package was used to merge the gene expression profiles to obtain the commonly profiles genes between the two datasets. Finally, either 30% un-expressed genes (small mean intensity) or 30% un-informative genes (small
standard deviation) were filtered out by MetaDE.filtering package, and the remaining 7,946 genes were remained for further analysis.

**Integration of microarray data**

MetaDE package was used to perform the integrate analysis (15). Moderate t test was used to determine the gene expression difference in each dataset, and the maximum P value method (maxP) was used to combine the P values from moderate t test. An adjusted P value of <0.0001, based on the false discovery rate (FDR) using Benjamini-Hochberg method was used to select DEGs.

**GO and pathway enrichment analysis of DEGs**

GO annotations were performed using DAVID database (https://david.ncifcrf.gov/) (16). KEGG (17) pathway enrichment analysis was performed using the KOBAS online analysis database (http://kobas.cbi.pku.edu.cn/).

Up- and down-regulated DEGs were subjected to GO and pathway enrichment analysis separately. P value <0.05 was used as threshold value.

**PPI network analysis**

The STRING database (http://string-db.org/) (18) was used to identify the interactions between proteins encoded by DEGs based on experimental data, databases, text mining, and predictive bioinformatics data. A combined score of >0.4 was set as threshold value. PPI networks were constructed with Cytoscape software (19). The genes with higher degrees of interaction were considered as hub genes, and pathway enrichment analysis was performed for these hub genes. The plug-in Molecular Complex Detection (MCODE) was used to screen the modules of PPI network in Cytoscape software. MCODE score >3 and number of nodes >4 were considered as threshold value. Pathway enrichment analysis were performed for genes in the modules.

**Results**

**Identification of DEGs**

Three method were used to detect common DEGs between the two microarray datasets: the maximum P value method (maxP), the Fisher’s method (Fisher), and the adaptive weight method (AW). The detection competency curves (Figure 1) suggested all the three methods were useful to detect common DEGs between the two datasets. The maxP method was chosen in this study. Under the threshold of adjusted P value of <0.0001, 107 DEGs were identified in CAVD. Among the 107 DEGs, 53 were up-regulated genes, and 54 were down-regulated genes (Table 2). Heatmap showed the expression pattern of these DEGs (Figure 2), suggesting the consistency of the 107 DEGs in each dataset.

**GO analysis**

GO terms were divided into three functional groups:
Table 2 Screening DEGs in CAVD by integrated analysis of microarray

<table>
<thead>
<tr>
<th>DEGs</th>
<th>Gene names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>TAGLN, SMOC2, A2M, VCAM1, MFAP5, TNFRSF11B, SH3KBP1, MMP9, IBSP, STMN2, IL7R, HTRA1, THBS2, THY1, CTHRC1, SCG2, CCL19, PLOD2, SLCO20A1, ARL4C, VWF, IGJ, SYTL2, PTGER4, FCGR1B, SERPINA1, ITGB2, SLAMF8, CA12, CD53, PLAUR, FCGR2A, CD93, ENPP2, HLA-DPA1, HLA-DRA, HLA-DMB, CD14, HLA-DMA, RGS19, PLAU, SPP1, RAC2, CD74, CYBA, WAS, ARHGAP30, CXCL16, C5AR1, LAPT5M, COR01A, TYROBP, NCF4</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>F5, AFF1, KLF4, TMEM100, RHOU, HIBCH, CDH19, FOXO1, LIMCH1, MEIS3P1, SNX1, MEIS2, TMOD1, LPAR1, IL17D, SPON1, IGSF10, ECHDC2, MXI1, TSPAN8, SCARAS, EHB1, PRFDC1, LOCT729680, CLDN11, MASP1, MAOA, ALDH2, TCEAL2, PLCE1, PTGDS, SHC4, VAT1L, GPR83, DCLK1, ITM2A, PGMS, ABCA8, PPA2B, ABLIM1, NFI, ANGPTL7, COL6A6, COL4A3, RHOB, PMP22, NTRK2, GPM6B, RBMS3, SEXTD1, GSN, SPTBN1, FEZ1, ACADVL</td>
</tr>
</tbody>
</table>

DEGs, differentially expressed genes; CAVD, calcific aortic valve disease.

Figure 2 The heatmap of the actual expression profiles for the 53 up- and 54 down-regulated DEGs. The heatmap was generated using MetaDE package in R. The expression profiles greater than the mean are colored in red and those below the mean are colored in green. 0, normal aortic valve; 1, calcified aortic valve; DEG, differentially expressed gene.

biological processes (BP), cellular component (CC), and molecular function (MF). The top 5 enriched GO terms in each category were listed in Table 3. For BP, the up-regulated DEGs were mainly enriched in GO terms of immune response, extracellular matrix organization, antigen processing and presentation via MHC class II, positive regulation of T cell proliferation, and chemotaxis; while the down-regulated DEGs were mainly enriched in GO terms of fatty acid beta-oxidation, cell adhesion, cytoskeleton organization, cellular response to hydrogen peroxide.
CC, the up-regulated DEGs were mainly enriched in GO terms of MHC class II protein complex, plasma membrane, cell surface, extracellular exosome, and extracellular space; while no GO-CC term was enriched from down-regulated DEGs. For MF, the up-regulated DEGs were mainly enriched in GO terms of MHC class II binding and receptor activity.

Pathway enrichment analysis

The most significantly enriched pathways of DEGs analyzed by KEGG analysis were listed in Table 4. The up-regulated DEGs were enriched in phagosome, complement and coagulation cascades, leukocyte transendothelial migration, cell adhesion molecules (CAMs), and antigen
Table 4 KEGG pathway enrichment analysis of DEGs

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>No. of genes</th>
<th>P value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Up-regulated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa04145</td>
<td>Phagosome</td>
<td>11</td>
<td>$2.47 \times 10^{-16}$</td>
<td>FCGR2A, THBS2, CORO1A, HLA-DMA, HLA-DRA, HLA-DPA1, ITGB2, NCF4, CD14, CYBA, HLA-DMA</td>
</tr>
<tr>
<td>hsa04610</td>
<td>Complement and coagulation cascades</td>
<td>7</td>
<td>$2.11 \times 10^{-11}$</td>
<td>SERPINA1, VWF, ITGB2, A2M, PLAU, C5AR1, PLAU</td>
</tr>
<tr>
<td>hsa04670</td>
<td>Leukocyte transendothelial migration</td>
<td>7</td>
<td>$3.02 \times 10^{-10}$</td>
<td>ITGB2, MMP9, THY1, NCF4, VCAM1, CYBA, RAC2</td>
</tr>
<tr>
<td>hsa04514</td>
<td>CAMs</td>
<td>6</td>
<td>$4.94 \times 10^{-8}$</td>
<td>ITGB2, HLA-DMA, HLA-DMA, HLA-DRA, HLA-DPA1, VCAM1</td>
</tr>
<tr>
<td>hsa04612</td>
<td>Antigen processing and presentation</td>
<td>5</td>
<td>$8.39 \times 10^{-8}$</td>
<td>HLA-DRA, HLA-DMA, HLA-DPA1, CD74, HLA-DMA</td>
</tr>
<tr>
<td></td>
<td>Down-regulated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa04933</td>
<td>AGE-RAGE signaling pathway in diabetic complications</td>
<td>3</td>
<td>0.003</td>
<td>COL4A3, FOXO1, PLCE1</td>
</tr>
<tr>
<td>hsa00340</td>
<td>Histidine metabolism</td>
<td>2</td>
<td>0.0005</td>
<td>ALDH2, MAOA</td>
</tr>
<tr>
<td>hsa00410</td>
<td>beta-Alanine metabolism</td>
<td>2</td>
<td>0.0009</td>
<td>ALDH2, HIBCH</td>
</tr>
<tr>
<td>hsa00380</td>
<td>Tryptophan metabolism</td>
<td>2</td>
<td>0.0014</td>
<td>ALDH2, MAOA</td>
</tr>
<tr>
<td>hsa00071</td>
<td>Fatty acid degradation</td>
<td>2</td>
<td>0.0017</td>
<td>ALDH2, ACADVL</td>
</tr>
</tbody>
</table>

DEGs, differentially expressed genes; CAMs, cell adhesion molecules.

processing and presentation. It was noteworthy that four HLA molecules (HLA-DRA, HLA-DMA, HLA-DPA1, HLA-DMA) presented in three pathways associated with immune response, while ITGB2 and VCAM1 presented in two pathways associated with cell adhesion and migration. The down-regulated DEGs were enriched in AGE-RAGE signaling pathway, metabolism of amino acid (histidine, beta-Alanine, tryptophan) and fatty acid. ALDH2 presented in metabolism of amino acids and fatty acid.

**PPI network analysis**

A total of 61 nodes and 143 protein pairs were obtained with a combined score of >0.4 based on the STRING database (Figure 3A). The top 5 hub genes a degree of interaction $\geq$10 were VCAM1 (degree =16), MMP9 (degree =14), ITGB2 (degree =14), RAC2 (degree =13), and vWF (degree =10). Pathway enrichment analysis suggested that these hub genes mainly involved in leukocyte transendothelial migration and cell adhesion (Figure 3B). MCODE detected a significant protein cluster containing HLA-DPA1, HLA-DRA, HLA-DMA, HLA-DMB, and CD74 (Figure 4A). Pathway enrichment analysis showed that these genes mainly involved in antigen processing and presentation and CAMs (Figure 4B). We also conducted pathway enrichment analysis to the independently down-regulated protein cluster containing ALDH2, HIBCH, ACADVL, ECHDC2, VAT1L, and MAOA (Figure 4C), which suggested that these genes mainly involved in metabolism of fatty acid and various amino acids (Figure 4D).

**Discussion**

In the present study, a total of 107 DEGs (53 up- and 54 down-regulated genes) were identified in CAVD by integrated analysis of microarray datasets (GSE51472 and GSE83453). Further bioinformatics analysis suggested that VCAM1, MMP9, ITGB2, RAC2, and vWF had a high degree of interaction in the PPI network and were key genes in CAVD. These key genes were significantly enriched in pathway of leukocyte transendothelial migration. The gene cluster containing HLA-DPA1, HLA-DRA, HLA-DMA, HLA-DMB, and CD74 detected by MCODE also suggested the role of immune response in CAVD. Moreover, the down-regulated gene ALDH2 might play a protective role in CAVD.

Inflammation is involved in the development and progression of CAVD (20), and calcification of the aortic valve...
was considered an inflammation-dependent process (21). T lymphocytes (22) and macrophages (5) had been identified in early aortic valve lesions. In the present study, 4 hub genes (VCAM1, MMP9, ITGB2, and RAC2) involved in leukocyte transendothelial migration, which was the first step of the immune cells infiltration. Aortic valve endothelial cells cover the surface of the aortic valve, and provide a protective function on aortic valve (23,24). High expression of VCAM1 had been found in the endothelium of calcified human aortic valves (24). VCAM1 on the endothelial cells are involved in the firm adhesion of leukocytes (25), and it also activates intracellular calcium release and NADPH oxidase Nox2, which further promotes the migration of leukocytes (26). Increase in MMP-9 had been found in aortic stenosis (27). MMP9 is synthesized by macrophages, which regulates the activation of T lymphocytes (28). MMP9 also can regulate leukocyte migration by processing chemokines and cytokines (29). ITGB2 normally expressed only in leukocytes, which promote leukocyte adhesion to endothelium (30). RAC2 is a member of GTPase signaling, which had been proven involved in regulation of leukocyte transendothelial migration (31). However, the role of ITGB2 and RAC2 in CAVD has not been illuminated yet. HLA-DPA1, HLA-DRA, HLA-DMA, HLA-DMB, and CD74 are MHC class II molecules, which mainly expressed in antigen presenting cells such as macrophages and dendritic cells (32). It had been previously reported that the proportion of circulating CD8+ T cell expressing HLA-DR was elevated during CAVD, which suggested that besides leukocytes infiltration of the aortic valve, a systemic adaptive immunity also play a role during CAVD (33).

vWF plays a major role in platelet adhesion and hemostasis (34). An in vitro study conducted by Balaoing et al. showed that vWF stimulated by histamine from porcine aortic valve endothelial cells significantly increased valvular interstitial cell calcification in vitro, which suggested a possible role of vWF in progression of CAVD (35).

ALDH2 is a detoxifying enzyme for the removal of toxic acetaldehyde such as malondialdehyde (MDA) and 4-hydroxynonenal (4-HNE) (36). More than 40% of the East Asians population carries a common ALDH2*2 mutant allele (Glu504Lys), which results in reduction of the enzymatic activity (37). Recently, various studies had highlighted the relationship between ALDH2 and

**Figure 3** PPI network analysis. (A) Protein-protein interaction network of DEGs. Red nodes represent up-regulated genes and green nodes represent down-regulated genes. The size of each node is positively correlated with the degree of interaction. The width of each edge is positively correlated with the combined score; (B) pathway enrichment analysis of the top 5 hub genes. DEG, differentially expressed gene; PPI, protein-protein interaction.
cardiovascular disease. Epidemiological studies suggested that the inactive East-Asian variant of ALDH2 (ALDH2*2) increased the risk of coronary artery disease (38). Animal study suggested that Alda-1, an activator of ALDH2, inhibited atherosclerosis in apolipoprotein E-knockout mice (39). In vitro study also suggested that activation of ALDH2 significantly attenuated the oxygenized LDL induced endoplasmic reticulum stress and apoptosis in smooth muscle cells (40). The present study suggested that the expression of ALDH2 was down-regulated in CAVD, however, the role of ALDH2 in CAVD had not been reported yet. It has been demonstrated that dysregulation of antioxidant mechanisms led to increased oxidative stress, which contributed to the progression of CAVD (41). We speculated that ALDH2-mediated detoxification effect reduced the oxidative stress in aortic valve, which might play a protective role in CAVD.

There are several limitations of the present study. The sample sizes of the microarray datasets used in the integrated analysis were relatively small. Besides, due to the very small sample sizes, subgroup analysis was not able to be performed to distinguish the possible different mechanisms involved between bicuspid and tricuspid valves. Although the integrated bioinformatics approach obtained a steadier result than single microarray, further studies with larger sample sizes covering wider range of races are still needed.

In conclusion, the present study identified VCAM1, MMP9, ITGB2, RAC2, vWF and ALDH2 as key genes in the progression of CAVD by integrated analysis of microarray datasets. Further bioinformatics analysis suggested that immune cells infiltration might play a key role in the progression of CAVD, while ALDH2-mediated detoxification effect might play a protective role in CAVD. The study provides a set of useful targets for further investigation into the molecular mechanisms, biomarkers and treatment targets of CAVD. Further molecular biological experiments are needed to confirm the function of the identified genes associated with CAVD.

**Acknowledgments**

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No. 81670350) and
Natural Science Foundation of Zhejiang Province, China (Grant No. LQ20H020003).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References


