Potential Target Genes in the Development of Atrial Fibrillation: A Comprehensive Bioinformatics Analysis

CE Liang Liu*
BC Yun Yu*
D Long-long Hu
D Quan-bin Dong
D Feng Hu
D Ling-juan Zhu
D Qian Liang
B Ling-ling Yu
A Hui-hui Bao**
A Xiao-shu Cheng**

* Liang Liu and Yun Yu contributed equally to this study
** Hui-hui Bao and Xiao-shu Cheng contributed equally to this study

Corresponding Authors: Hui-hui Bao, e-mail: huihui_bao77@126.com; Xiao-shu Cheng, e-mail: xiaoshumenfan@126.com
Source of support: Departmental sources

Background: Atrial fibrillation (AF) is the most prevalent arrhythmia worldwide. Although it is not life-threatening, the accompanying rapid and irregular ventricular rate can lead to hemodynamic deterioration and obvious symptoms, especially the risk of cerebrovascular embolism. Our study aimed to identify novel and promising genes that could explain the underlying mechanism of AF development.

Material/Methods: Expression profiles GSE41177, GSE79768, and GSE14975 were acquired from the Gene Expression Omnibus Database. R software was used for identifying differentially expressed genes (DEGs), and Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses were subsequently performed. A protein–protein interaction network was constructed in Cytoscape software. Next, a least absolute shrinkage and selection operator (LASSO) model was constructed and receiver-operating characteristic curve analysis was conducted to assess the specificity and sensitivity of the key genes.

Results: We obtained 204 DEGs from the datasets. The DEGs were mostly involved in immune response and cell communication. The primary pathways of the DEGs were related to the course or maintenance of autoimmune and chronic inflammatory diseases. The top 20 hub genes (high scores in cytoHubba) were selected in the PPI network. Finally, we identified 6 key genes (FCGR3B, CLEC10A, FPR2, IGSF6, S100A9, and S100A12) via the LASSO model.

Conclusions: We present 6 target genes that are potentially involved in the molecular mechanisms of AF development. In addition, these genes are likely to serve as potential therapeutic targets.

MeSH Keywords: Atrial Fibrillation • Biological Markers • Immunity, Active

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/928366

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

This work is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)
Background

Atrial fibrillation (AF) is the most prevalent (0.51%) and persistent heart rhythm disorder globally [1]. AF, increases the risk of blood clots causing thromboembolism, confers a nearly 5-fold increased risk of a debilitating stroke, and is a major risk factor for cardiovascular outcomes, including stroke [2,3]. Given its asymptomatic nature, many cases of AF go undetected until complications occur. Early detection of AF increases the chances of preventing stroke and other complications [4]. Therefore, it is imperative to explore the molecular mechanisms of AF pathogenesis as a means of improving the early diagnosis and treatment interventions of the disease.

Changes in genetic expression in AF are gradually gaining research attention [5,6]. Using the expression profiles of AF-associated target genes as a guide to examine the pathogenic sites of genetic variation and target gene regulation could be an effective approach [7,8]. Population studies show that a positive family history of AF increases the risk of developing AF among first-degree relatives by 30% [9,10]. The search for AF-associated genetic loci is propelled by the existing knowledge on the heritability of AF in the general population. Several genes, including some related to ion channel function, have been revealed to be associated with AF [11]. Ion channel-related genes associated with AF are primarily related to potassium channels. Previous studies have identified specific potassium channel genes or subunits that are associated with AF. Monogenic channels. Previous studies have identified specific potassium genes associated with AF are primarily related to potassium channels. Previous studies have identified specific potassium channel genes or subunits that are associated with AF. Monogenic AF pedigrees revealed that mutations leading to the gain or loss of function of channel genes or subunits that are associated AF. Monogenic AF pedigrees revealed that mutations leading to the gain or loss of function of channel genes or subunits that are associated AF.

Material and Methods

Data information and processing

The gene expression profiles of the GSE41177, GSE79768, and GSE14975 datasets were retrieved from the National Center for Biotechnology Information Gene Expression Omnibus database (NCBI GEO, http://www.ncbi.nlm.nih.gov/geo). The datasets contained data from AF patients with persistent AF for >3 months and SR patients without AF and not using any anti-arrhythmia drugs, with the latter serving as a control group. The GSE41177 dataset comprised 32 atrial appendage tissues from patients with persistent AF and 6 cases with SR. The GSE79768 and GSE14975 datasets consisted of atrial appendage tissues from 14 AF and 12 SR cases and 5 AF and 5 SR cases, respectively. The microarray data from GSE41177, GSE79768, and GSE14975 were based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array, HG-U133_Plus_2). The raw data were processed and analyzed in R (version 3.6.2). First, the raw data underwent background correction with the limma package (http://www.bioconductor.org/packages/limma/). Then, a robust mulitchip average was created and perfect matches from the raw data were log2 transformed. Afterwards, batch normalization and data merging for the 3 datasets was conducted using the sva package (http://www.bioconductor.org/packages/sva/).

Identification of DEGs

Gene annotation was conducted before the analysis of the DEGs between AF and SR samples. The probe identification
numbers of the merged data were matched with the gene symbols using the affy package (http://www.bioconductor.org/packages/affy/). If multiple probes were mapped to the same gene, the median of the expression value was selected as the gene expression value. Next, we utilized the limma package to identify the DEGs, and the cutoff criteria were set at an adjusted $P$-value (adj. $P$) $<0.05$ and an absolute value of log fold change ($|\log_2 FC|$) $>1$. We apply adj. $P$ to correct false positives. The data for the listed DEGs were processed and plotted in a heatmap and volcano plots using the pheatmap and ggplot2 R packages, respectively.

**Gene ontology and pathway enrichment analysis**

Based on the analysis of DEGs, added potential functional annotations in Functional Enrichment analysis tool (Funrich) were performed. Gene Ontology (GO) term enrichment analysis, included biological process (BP), cellular component (CC), and molecular function (MF). In addition, the Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8, https://david.ncifcrf.gov/) was applied to analyze the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. DAVID is a biological information database with gene annotation, gene symbol conversion, visualization, and integrated discovery function, and it provides comprehensive and systematic biological function annotation information of genes. Subsequently, we visualized the KEGG pathway analyses using the ggplot2 R package. The cutoff criterion was set at adj. $P<0.05$.

**Analysis of protein–protein interaction networks of DEGs**

The protein–protein interaction (PPI) networks of the DEGs were analyzed using the STRING online database (version 10.5; http://string-db.org/) to predict PPIs and protein functional associations. A confidence score of $\geq 0.4$ was set as the threshold. After that, Cytoscape software (version 3.7.2; http://cytoscape.org/) was used to analyze and visualize the biological networks and node degrees of DEGs. The cytoHubba plugin (version 0.1) was utilized to identify the PPI networks. Based on the 4 filtering algorithms (Stress, Closeness, EcCentricity, and MCC), the top 20 hub genes obtained by each algorithm were used. Subsequently, co-expression of the hub genes and enrichment genes in the pathways was identified and presented in a Venn diagram.

Figure 1. (A) Volcano plot of potential differentially expressed genes (DEGs). Red dots denote upregulated genes; green dots denote downregulated genes. (B) Heat maps for the DEGs between atrial fibrillation (AF) and sinus rhythm (SR).
DATASET ANALYSIS

A. Percentage of genes – Log10(p-value)

- Immune response: 12.6%, p = 0.001
- Cell communication: 5.5%, p = 0.001
- Signal transduction: 4.3%, p = 0.001
- Invasive immune response: 5.5%, p = 0.001
- Nucleic acid metabolism: 5.5%, p = 0.001
- Protein modification: 5.5%, p = 0.001
- Regulation of immune response: 8.2%, p = 0.001
- Cell surface receptor activity: 8.2%, p = 0.001
- Signaling transduction: 8.2%, p = 0.001
- Regulation of gene expression: 8.2%, p = 0.001

B. Biological process

- Plasma membrane: 45.7%, p = 0.001
- Exosomes: 27.7%, p = 0.001
- Complement C1: 22.5%, p = 0.001
- Extracellular: 13.3%, p = 0.001
- Secretory granule: 7.5%, p = 0.001
- Extracellular region: 7.5%, p = 0.001
- Adipocyte: 7.5%, p = 0.001
- Gp120/130: 7.5%, p = 0.001
- Small ribosomal subunit: 7.5%, p = 0.001

C. Molecular function

- Repression activity: 9.3%, p = 0.001
- MHC class I receptor activity: 2.7%, p = 0.008
- Membrane receptor activity: 2.7%, p = 0.008
- Growth factor binding: 1.1%, p = 0.008
- Complement activity: 1.5%, p = 0.008
- Adaptor binding: 1.5%, p = 0.008
- Calcium ion binding: 1.5%, p = 0.008
- Transmembrane receptor activity: 1.5%, p = 0.008

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)
Percentage of genes were well clustered between the AF and control cases. Compared with the control SR differential genes in the data sets, AF-related genes were mostly upregulated. The expression data of the hub genes, based on the ratio 7: 3, were randomly allocated to the training and testing sets. The receiver-operating characteristic (ROC) curve was plotted using the pROC package in R. The value of ROC area under the curve (AUC) was obtained to assess the sensitivity and specificity of the key genes screened by the LASSO model.

Results

Construction of the LASSO model and ROC curve analysis

Based on the hub genes, we constructed a least absolute shrinkage and selection operator (LASSO) model because of its strong predictive value using the glmnet package (http://www.bioconductor.org/packages/glmnet/). Then, the expression data of the hub genes, based on the ratio 7: 3, were randomly allocated to the training and testing sets. The receiver-operating characteristic (ROC) curve was plotted using the pROC package in R. The value of ROC area under the curve (AUC) was obtained to assess the sensitivity and specificity of the key genes screened by the LASSO model.

GO term enrichment analysis of DEGs

We identified the DEGs for the GO analysis using the Funrich software, and the DEGs were examined based on 3 categories: BP, CC, and MF. In the BP group, the DEGs were mainly enriched in immune response, cell communication, and signal transduction; in the CC group, the DEGs were primarily enriched in plasma membrane and exosomes; and in the MF group, the enrichment of the DEGs was mostly in the receptor activity, MHC class I receptor activity, and MHC class II receptor activity (Figure 2A–2C, Table 1).

Signaling pathway enrichment analysis

The KEGG pathway enrichment analysis was performed for the DEGs, and the results are shown in Figure 2D and Table 2. Enrichment of the DEGs was mainly in the Staphylococcus aureus infection, intestinal immune network for IgA production, systemic lupus erythematosus, asthma, and viral myocarditis.

PPI network and hub genes identification

Using the STRING database and Cytoscape software, we constructed 4 PPI networks for the DEGs as shown in Figure 3A–3E. Furthermore, the top 20 hub genes of each dataset were identified using the 4 algorithms (Stress, Closeness, Eccentricity, and MCC) of the cytoHubba plugin. Among these hub genes,
SPI1, HCK, TLR8, FPR2, RAC2, IGF6, CSF3R, CXCR4, S100A8, S100A9, FCGR3B, TYROBP, S100A12, and CLEC10A exhibited the highest degree of association with AF and are listed in the Venn diagram (Figure 3F).

Validation of the diagnostic value of key genes

The construction of a LASSO model was based on the A gene expression profile of the 14 hub genes (Figure 4A). Based on the model, 6 genes were identified according to the regression coefficients that were not equal to zero. ROC curve analysis (Figure 4B, 4C) identified the AUC for FCGR3B, CLEC10A, FPR2, IGF6, S100A9, and S100A12 in AF patients and normal controls; in the training set, the AUC values were 0.9981 (95% CI, 0.993–1), 0.9944 (95% CI, 0.9821–1), 0.987 (95% CI, 0.9655–1), 0.9481 (95% CI, 0.8566–1), 0.9852 (95% CI, 0.9616–1), and 0.9981 (95% CI, 0.993–1); in the test set, the AUC values were 0.8762 (95% CI, 0.7271–1), 0.9714 (95% CI, 0.9079–1), 0.981 (95% CI, 0.9358–1), 0.8857 (95% CI, 0.7092–1), 0.8952 (95% CI, 0.7517–1), and 0.9429 (95% CI, 0.825–1).

Discussion

The initiation and development of AF are regulated by the expression of many genes. Here, we systematically investigated gene expression profiles obtained from AF microarray studies. Gene expression data were downloaded from 3 GEO datasets in the GCBI. To better explore the DEGs, we identified the potential and remarkable DEGs, biological pathways, and processes based on comparisons between AF and normal SR samples.
GO analysis revealed that a considerable number of the co-expressed DEGs were primarily associated with immune response, cell communication, signal transduction, plasma membrane, exosomes, receptor activity, MHC class I receptor activity, and MHC class II receptor activity. Previous studies indicated that the immune response can be a complex and powerful factor in the pathophysiological process of AF and its concomitant complications [23,24]. The major immune cells found in the left atrial appendages of patients with AF were immunologically active monocytes and macrophages, suggesting a pro-inflammatory state characterized by increased infiltration of iNOS-positive, but Arg1-negative immune cells [25,26]. AF-induced cardiac injury processes of the myocardium initiate a cardiac immune response as shown by damage-associated molecular patterns (DAMPs), such as extracellular matrix and reactive oxygen species [27]. The pattern recognition receptor then initiates recognition of DAMPs, before the activation of innate and adaptive immune cells. This immune process forms a long-term feedback loop and promotes the formation of permanent forms of AF [28].

Inflammation and immune response are dependent on the onset and maintenance of AF, which is in turn dependent on inflammation and immune response. In addition, a clinical phenomenon that AF promotes inflammation supports a vicious cycle of “AF begets AF” [29]. The results of our pathway analysis showed that the DEGs identified were mainly associated with the course or maintenance of autoimmune diseases and chronic inflammatory diseases. Multiple systems of systemic lupus erythematosus (SLE), a chronic inflammatory autoimmune disease, may be an independent risk factor for AF since SLE has cardiac manifestations. Recent studies have shown that the disease status of SLE is independently associated with AF after adjusting for age, sex, race, and coronary artery disease [30]. Moreover, systemic inflammation and disease pathways of chronic inflammatory diseases are considered as pathogenic contributors to the initiation and development of AF [31,32]. Chemokines guide the migration and activation of systemic leukocytes and may influence AF development [33]. Some studies have suggested that the level of inflammatory cytokines/chemokines is related to the progression of AF to myocardial fibrosis [34].
expression of cytokines in AF promotes the degree of the recruitment of immune cells [35]. As a classic way related in immune reaction, transforming growth factor (TGF)-β and interleukin (IL)-6 were expressed more frequently in the atria of macrophages after migration in AF [36].

Further analysis of the co-expression genes in the PPI network and LASSO model identified 6 key genes (FCGR3B, CLEC10A, FPR2, IGSF6, S100A9, and S100A12) as the most significance. Among the key genes, FCGR3B showed a closer association with AF. FCGR3B encodes a low-affinity receptor for the Fc portion of gamma immunoglobulins (IgG) known as FcyRIIb (CD16b). The FcyRIIb can bind either the monomeric or aggregated IgG, and it may function to capture dedicated immune complexes in the peripheral circulation [37]. On neutrophils, FcyRIIb regulates immune responses [38]. Our findings provide strong evidence that recruitment of neutrophils, particularly polymorphonuclear neutrophils, promotes the development of AF [39,40]. Human FcyRIIb lacks intracellular signaling motifs and is anchored in the cell membrane via a glycosphatidylinositol anchor [41]. In addition, intracellular signaling promotes phagocytosis of antibody-opsonized microbes by human neutrophils through FcyRIIb in collaboration with other associated FcyR [42]. In the rapid adhesion of neutrophils to endothelial cells, FcyRIIb binding immune complexes remains an important link [43]. Thus, this gene is involved in the recruitment and activation of polymorphonuclear neutrophils at the site of inflammation [44].

Although the potential role of FCGR3B in autoimmunity has been extensively investigated, its role in AF development remains unclear [45]. The protein interacts with the Fc portion of the IgG molecule involved in immune regulation. Moreover, the variation of FCGR3B copy numbers is associated with susceptibility to several autoimmune diseases [46]. Importantly, FCGR3B could be linked to cardiovascular disorder of the myocardium and coronary heart disease [47]. In addition, CD16+ monocytes have been shown to be involved in atrial remodeling in the pathophysiology of AF [48]. Therefore, the FCGR3B gene could be a potential biomarker for AF.

CLEC10A is one of the members of the c-type lectin domain family 12, which serves as a characteristic galactose lectin on macrophages and dendritic cells. CLEC10A can be activated through Toll-like receptor signaling and increase the secretion of various cytokines, including tumor necrosis factor-α, IL-8, and IL-10 [49]. In addition, lectin has been linked to the development of AF, and it can particularly predict thrombosis and left atrial appendage remodeling in patients with AF [50]. A synergistic interaction exists between lectin and TGF-β1, which can induce AF by activating the TGF-β1/Smad pathway in patients with AF [51]. Formyl peptide receptors (FPRs) belong to a G-protein-coupled chemokine receptor family and include 3 subtypes (FPR1, FPR2, and FPR3) that play important roles in host defense and inflammation [52]. FPR2 is involved in the initiation and resolution of inflammation. Similarly, the FPR2 signaling has important vascular effects through the inflammatory response [53,54].

Figure 3. (A) Protein–protein interaction (PPI) networks of the differentially expressed genes (DEGs). The top 20 genes in cytoHubba plugins: (B) Stress, (C) Closeness, (D) EcCentricity, and (E) MCC. (F) Venn diagrams of the top 20 genes.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

DATABASE ANALYSIS

APPROVED GALLEY PROOF
Conclusions

In summary, we used bioinformatics analysis tools to delineate the pathomechanism of AF based on DEGs. Autoimmune and chronic inflammatory pathways, neutrophil chemotaxis, and immune response were confirmed to play roles in the development of AF. In addition, 6 genes, including FCGR3B, CLEC10A, FPR2, IGSF6, S100A9, and S100A12 were identified as potential AF biomarkers. Our findings pave the way for further research on AF occurrence and development.

There are some limitations of our study. First, the clinical atrial tissue samples were not analyzed in the laboratory. Therefore, future studies on the mechanism of AF occurrence and development should incorporate such samples. In addition, given the current limited datasets, the sample sizes were relatively small and a multicenter study involving a larger sample size is needed. Further investigation is required to determine whether the 6 key genes (FCGR3B, CLEC10A, FPR2, IGSF6, S100A9, and S100A12) activate any of the pathways or other mechanisms known to induce AF in humans.

Conflict of interest

None.
References:


