Mathematical Biosciences and Engineering

MBE, 20(4): 6947−6965.
DOI: 10.3934/mbe.2023300
Received: 06 November 2022
Revised: 18 January 2023
Accepted: 26 January 2023
Published: 08 February 2023

http://www.aimspress.com/journal/MBE

Research article

Identification and verification of FN1, P4HA1 and CREBBP as potential biomarkers in human atrial fibrillation

Miao Zhu†, Tao Yan†, Shijie Zhu†, Fan Weng, Kai Zhu, Chunsheng Wang* and Changfa Guo*

Department of Cardiovascular Surgery, Zhongshan Hospital, Fudan University, 180 Fenglin Road, Shanghai 200032, China

* **Correspondence:** Email: guo.changfa@zs-hospital.sh.cn, wangchunsheng@fudan.edu.cn; Tel: +18616881228.

† These authors contributed equally to this work.

Abstract: *Background:* Atrial fibrillation (AF) is a common arrhythmia that can lead to cardiac complications. The mechanisms involved in AF remain elusive. We aimed to explore the potential biomarkers and mechanisms underpinning AF. *Methods:* An independent dataset, GSE2240, was obtained from the Gene Expression Omnibus database. The R package, "limma", was used to screen for differentially expressed genes (DEGs) in individuals with AF and normal sinus rhythm (SR). Weighted gene co-expression network analysis (WGCNA) was applied to cluster DEGs into different modules based on functional disparities. Enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery. A protein–protein interaction network was constructed, and hub genes were identified using cytoHubba. Quantitative reverse-transcription PCR was used to validate mRNA expression in individuals with AF and SR. *Results:* We identified 2,589 DEGs clustered into 10 modules using WGCNA. Gene Ontology analysis showed specific clustered genes significantly enriched in pathways associated with the extracellular matrix and collagen organization. Kyoto Encyclopedia of Genes and Genomes pathway analysis revealed that the target genes were mainly enriched for proteoglycans in cancer, extracellular matrix–receptor interaction, focal adhesion, and the PI3K-Akt signaling pathway. Three hub genes, *FN1*, *P4HA1* and *CREBBP*, were identified, which were highly correlated with AF endogenesis. mRNA expression of hub genes in patients with AF were higher than in individuals with normal SR, consistent with the results of bioinformatics analysis. *Conclusions: FN1*, *P4HA1*, and *CREBBP* may play critical roles in AF. Using bioinformatics, we found that expression of these genes was significantly elevated in patients with AF

than in individuals with normal SR. Furthermore, these genes were elevated at core positions in the mRNA interaction network. These genes should be further explored as novel biomarkers and target candidates for AF therapy.

Keywords: arrhythmia; atrial fibrillation; biomarkers; hub genes; weighted gene co-expression network analysis

Abbreviations: AF: atrial fibrillation; BPs: biological processes; CCs: cellular components; CREB: cyclic adenosine monophosphate response element-binding protein; DAVID: Database for Annotation, Visualization, and Integrated Discovery; DEGs: differentially expressed genes; ECM: extracellular matrix; FC: fold change; GEO: Gene Expression Omnibus; GO: gene ontology; GSEA: Gene Set Enrichment Analysis; KEGG: Kyoto Encyclopedia of Genes and Genomes; LAA: left atrial appendage; MFs: molecular functions; P4Hs: prolyl 4-hydroxylases; PPI: protein–protein interaction; qRT-PCR: quantitative real-time PCR; SR: sinus rhythm; STRING: Search Tool for the Retrieval of Interacting Genes; WGCNA: weighted gene co-expression network analysis

1. Introduction

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia [1], affecting approximately 33 million individuals worldwide [2]. Patients with AF are estimated to experience a two-fold increase in heart failure, myocardial infarction, and premature mortality [3]. AF is regarded as an independent risk factor for thromboembolism and stroke [4], leading to a higher occurrence of pulmonary embolism [5], and even death [6].

The Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo) is the largest public gene expression database, founded by the National Center for Biotechnology Information in 2000. The database contains a large amount of high-throughput gene expression and microarray data submitted by research institutions worldwide. With the improvement of GEO databases, bioinformatics analysis has become an essential tool for investigating the pathophysiological processes involved in various diseases.

Hemodynamic pressure overload stimulates increased secretion of natriuretic peptides from cardiomyocytes and can act as a biomarker for heart failure and AF [7−9]. C-reactive protein is widely used as a biomarker for inflammation; however, the protein has an essential correlation with AF incidence [8,10]. Myocardial fibrosis is a vital substrate for AF development [11]. In addition, biomarkers implicated in atrial fibrosis, such as soluble ST2 [12], MMP-9 [13], and TGF-β1 [14], are known to participate in AF development. However, most reported biomarkers are related to the extracellular matrix (ECM) and myocardial fibrosis, and only a few have been validated in human left atrial appendage (LAA) specimens. We aimed to employ the online databases and human tissues to explore the potential biomarkers and mechanisms underpinning AF. A large tissue bank containing human LAA specimens was used to validate our hypothesis and for subsequent experiments performed at our center. In addition, genes unrelated to fibrosis were specifically screened in our analysis, which may broaden our understanding of the mechanism behind AF development.

2. Materials and methods

2.1. Human sample collection

Fifteen patients with lone AF and fifteen with a normal sinus rhythm (SR) were included in this study. LAA tissues were collected from the participants during cardiac surgery. The specimens were immediately preserved in liquid nitrogen and transferred to a refrigerator at -80℃ for subsequent experiments. This study was conducted in full compliance with the Declaration of Helsinki and approved by the Medical Ethics Committee of Zhongshan Hospital, Fudan University (Shanghai, China). Written informed consent was obtained from all participants before surgery.

2.2. Data acquisition

Sequencing results of genes related to AF were obtained from the GEO database. The GEO dataset, GSE2240, based on the GPL96 platform (Affymetrix GeneChip Human Genome U133A Array), was downloaded for further analysis. Depending on the selected cohort, differentially expressed genes (DEGs) were screened between patients with AF and SR. An additional independent dataset, GSE41177, was downloaded for the validation of hub genes. Three LAA samples with SR and sixteen LAA samples with AF were included in analysis.

2.3. Data processing

The original expression profile of the GSE2240 dataset was processed using R software. The Linear Model for Microarray data (limma) package in R was used for dataset normalization, $\log_2$-transformation, and fold change (FC) calculation [15], and the SVA package was used to remove the batch effect. The *P*-values of genes were calculated using the Student's *t*-test, where those with *adj.P*-value < 0.05 were identified as DEGs and selected for further analyses to ensure integrity of the mRNA network.

2.4. Weighted gene co-expression network analysis

The weighted gene co-expression network analysis (WGCNA) [16] was constructed using genes with an *adj.P*-value < 0.05, as identified with the limma R package. The co-expression network was built based on a scale-free topology criterion of $R^2 > 0.80$ through an appropriate soft-thresholding power, β. All DEGs were clustered into modules of different colors according to the average linkage hierarchical clustering method. We applied a 0.25 threshold for module merging, and each module contained at least 20 genes. Finally, the correlation of each module with AF was calculated using the Pearson's correlation method.

2.5. Enrichment analyses

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of the target genes were performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, v6.8; https://david.ncifcrf.gov/) [17], which revealed

biological processes (BPs), molecular functions (MFs), cellular components (CCs) and pathways related to the DEGs. GO terms with an *adj.P*-value < 0.01 and KEGG maps with a *P*-value < 0.05 were considered to represent significant enrichment. Gene Set Enrichment Analysis [18] (GSEA) was used to validate significance of the GO-KEGG results, with *adj.P* < 0.05 and false discovery rate (FDR) < 0.25 as criteria. We applied DEGs and corresponding LogFC to GSEA at the level of gene sets, which are defined based on information about biological pathways. Enrichment scores were calculated to determine whether the selected gene set is at the top or bottom of the raked list, and, thus, whether the selected gene set is relevant for phenotypic class distinction.

2.6. Construction of the protein–protein interaction network

Four algorithms were used to calculate the protein–protein interaction (PPI) network for the DEGs: MCC, Stress, Degree and EPC. Predicted target genes were imported into the Search Tool for the Retrieval of Interacting Genes database (STRING, v11.0; http://string-db.org) with the highest confidence (0.900) to identify interactions between genes in the PPI network [19]. Results downloaded from the STRING database were then used in the Cytoscape software (v3.8.1) [20] to visualize the PPI network and analyze node degrees for hub gene identification through cytoHubba, a plugin application [21].

2.7. Quantitative reverse-transcription PCR

Total cellular RNA was extracted from LAAs using a RNeasy Mini kit (QIAGEN, Frankfurt, Germany) following the manufacturer's instructions. Complementary DNA was synthesized using a PrimeScript RT reagent kit (Takara, Otsu, Japan) at 42℃ for 60 min, followed by an additional 5 min at 95℃. TB Green Premix Ex Taq II (Takara, Otsu, Japan) was used to perform quantitative reverse-transcription PCR (qRT-PCR) at 95℃ for 30 s, followed by 40 cycles at 95℃ for 5 s and 60℃ for 34 s on a QuantStudio 5 System (Thermo Fisher Science, Waltham, MA, USA). RNA expression was normalized using GAPDH as a standard, and the $2^{-\Delta\Delta CT}$ method was applied to each mRNA with three independent repeats. All RNA primers were obtained from Sangon Biotech (Shanghai, China), and their sequences are listed in Table 1.

Table 1. List of primer sequences used for quantitative real-time PCR.

MiRNA	Sequence
GAPDH	forward: GGAGCGAGATCCCTCCAAAAT
	reverse: GGCTGTTGTCATACTTCTCATGG
FN1	forward: CGGTGGCTGTCAGTCAAAG
	reverse: AAACCTCGGCTTCCTCCATAA
P4HA1	forward: AGTACAGCGACAAAAGATCCAG
	reverse: CTCCAACTCACTCCACTCAGTA
CREBBP	forward: CGGCTCTAGTATCAACCCAGG
	reverse: TTTTGTGCTTGCGGATTCAGT

2.8. Statistical analyses

Statistical analyses were performed using the SPSS software (v22.0; SPSS Inc., Chicago, IL, USA). The Student's *t*-test was performed to evaluate the mean values of quantitative variables, and the Mann–Whitney U test was used when data distribution was not normal. An *adj.P*-value < 0.05 was considered statistically significant.

3. Results

3.1. Identification of DEGs

Using our test criteria, genes with an *adj.P*-value < 0.05 were selected as DEGs. A set of 2589 genes, including 1244 upregulated and 1345 downregulated genes, was identified as DEGs in the GSE2240 dataset (Figure 1A). The DEGs with |logFC| > 1 were screened using the heat map shown in Figure 1B.

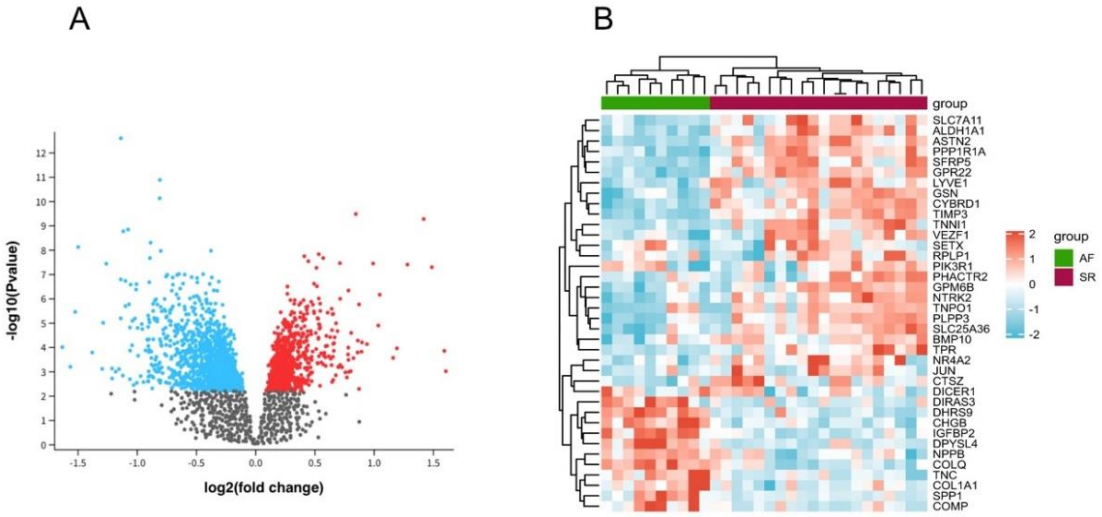

Figure 1. Identification of differentially expressed genes (DEGs). (A) Volcano plot of differentially expressed genes in GSE2240 (platform: gpl96); The red and blue dots represent up-regulated and down-regulated genes, respectively. (B) Heat map of DEGs with |logFC| ＞ 1.

3.2. Weighted gene co-expression network analysis

All identified DEGs were subjected to WGCNA to form modules based on the average linkage hierarchical clustering method. A scale-free co-expression network ($R^2 > 0.80$) was established with a soft-thresholding power of $\beta = 10$ (Figure 2D). Ten modules were eventually formed in tokens of different colors, including black, magenta, pink, green, midnight blue, yellow, blue, light green, red, and grey (Figure 2A). Only modules whose co-expression efficiency for AF was beyond 0.75 were chosen as target modules for further analysis. According to this criterion, the pink and green modules with 591 genes were considered significant. Figure 2B shows the clustering heatmap between modules,

where red means closer similarity and blue means less similarity. A topological overlap-based dissimilarity was also used to plot the clustering dendrogram of genes, together with the assigned merged module colors and the original module colors (Figure 2C).

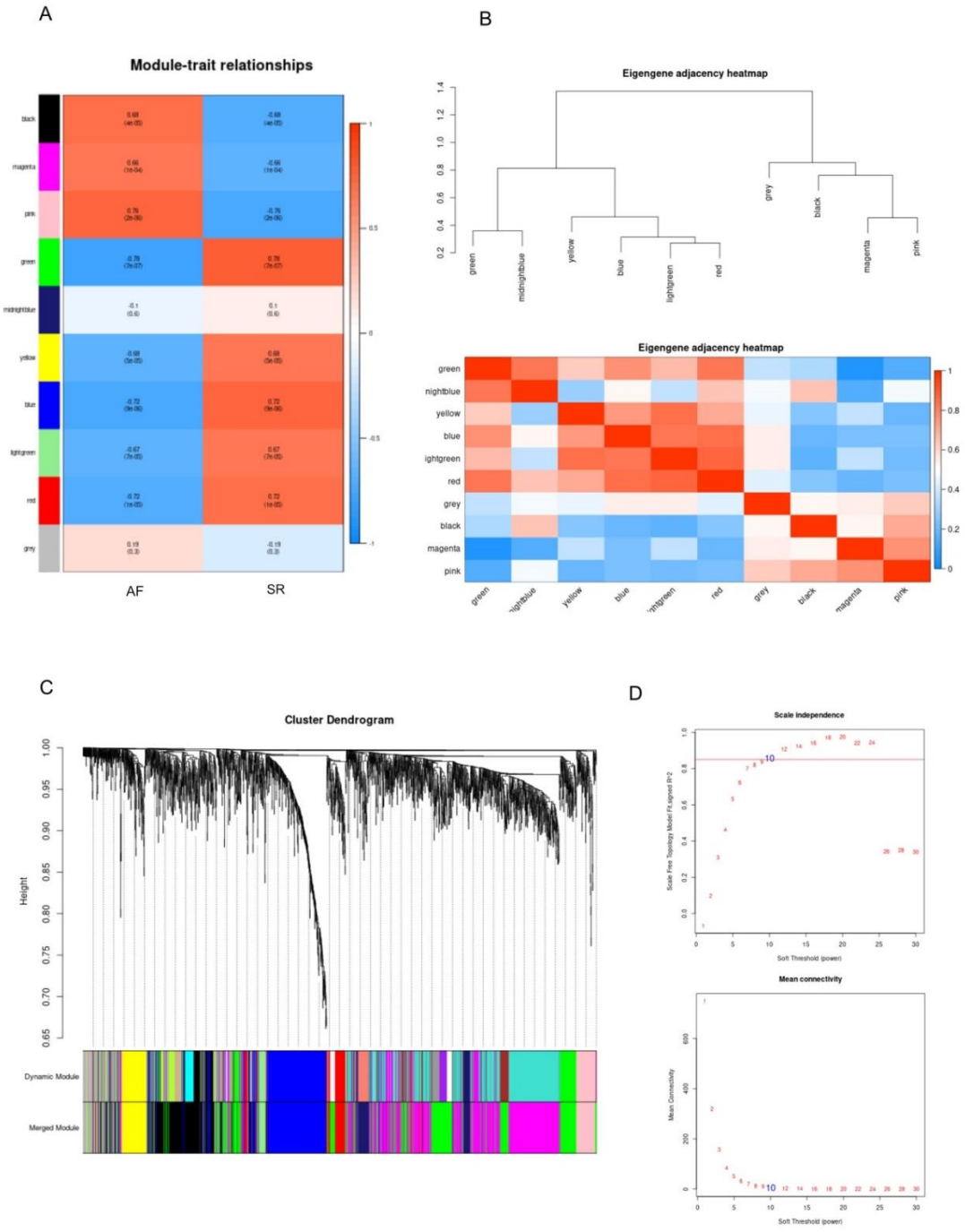

Figure 2. Weighted gene correlation network analysis. (A) The heatmap between modules. The closer similarity is in red, and the farther similarity is in blue. (B) The heatmap of module-trait correlations. Red means positive correlation, and blue means negative correlation. (C) The dendrogram of differentially expressed genes. (D) The result of soft-thresholding power β.

3.3. Enrichment analyses

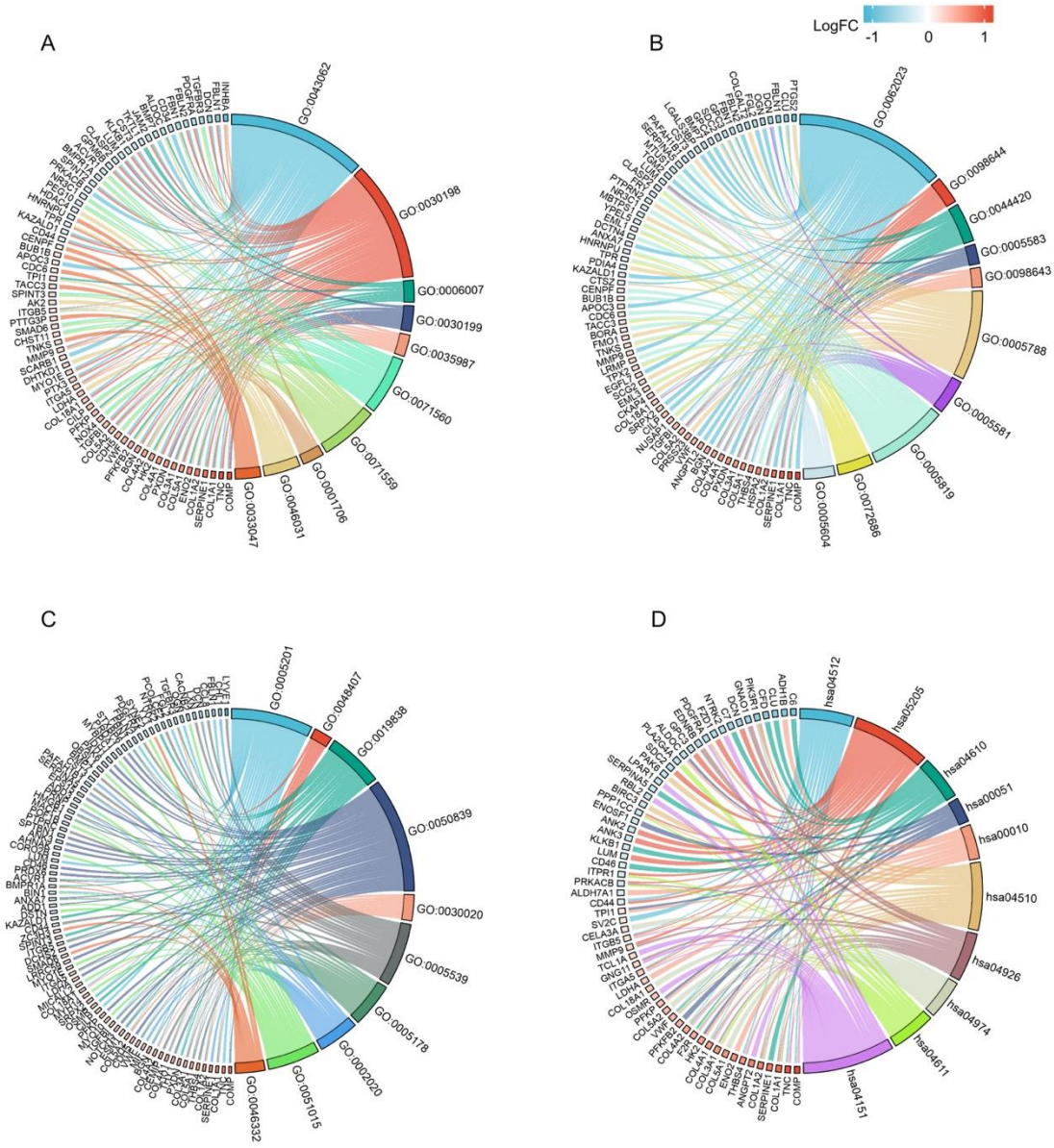

Figure 3. GO and KEGG enrichment analyses. (A) Biological process; (B) Cellular component; (C) Molecular function; (D) KEGG pathways.

GO and KEGG functional enrichment analyses were performed using the DAVID tool to investigate biological effects of the 591 genes clustered in the green and pink modules. The top 10 results for the BP, CC, MF and KEGG pathways are shown in Figure 3A–C. The KEGG pathway analysis suggested that the target genes were mainly enriched in protein digestion and absorption, ECM-receptor interaction, PI3K-Akt signaling pathway, and focal adhesion (Figure 3D). GSEA was performed using the clusterProfiler package [22]. A total of 19 gene sets were enriched with *adj.P* < 0.05 and FDR < 0.25. The top six identified gene sets are listed in Figure 4A–F.

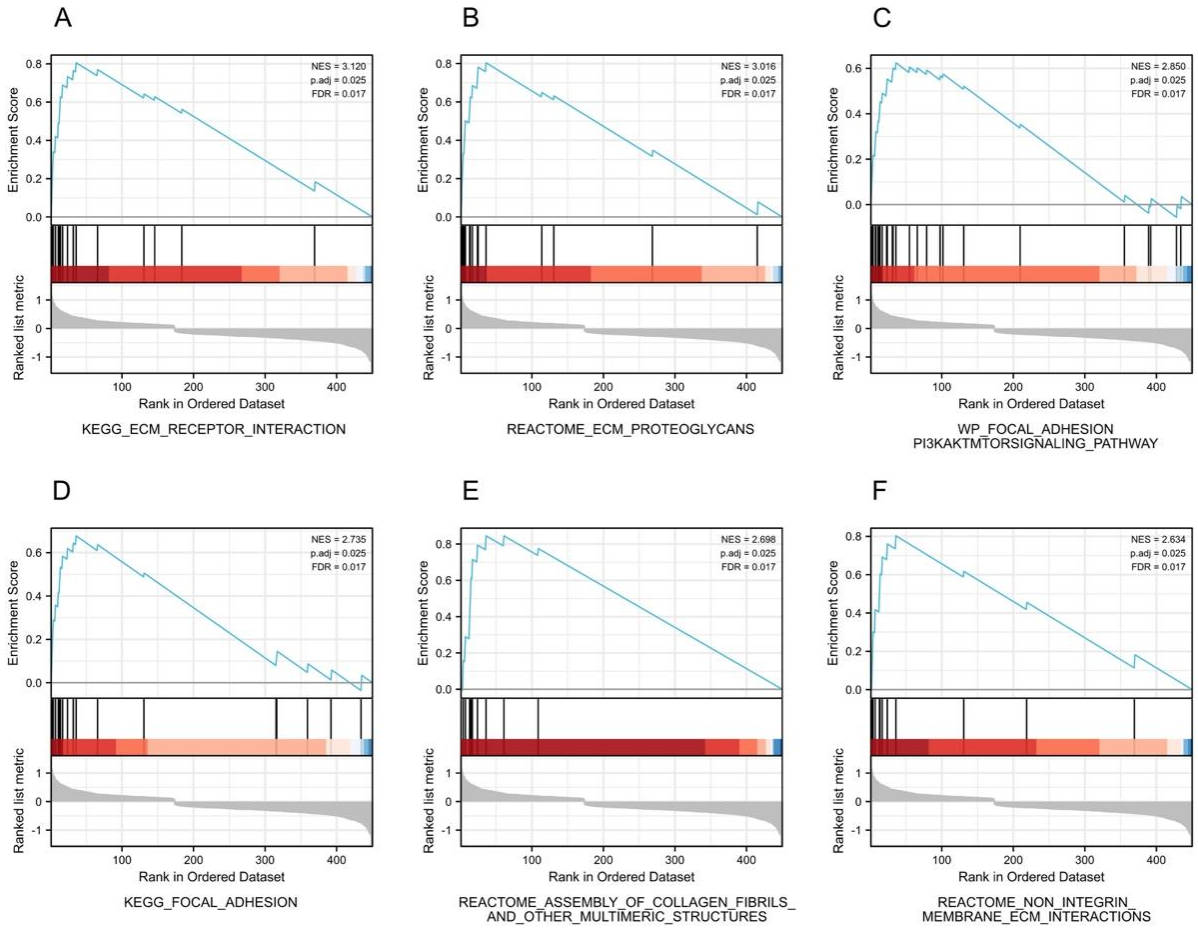

Figure 4. Gene set enrichment analysis. (A) The pathway of KEGG ECM RECEPTOR INTERACTION; (B) The pathway of REACTOME ECM PROTEOGLYCANS; (C) The pathway of WP FOCAL ADHESION-PI3K-AKTM-TOR-SIGNALING; (D) The pathway of KEGG FOCAL ADHESION; (E) The pathway of REACTOME ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES; (F) The pathway of REACTOME NON INTEGRIN MEMBRANE ECM INTERACTIONS.

3.4. Protein–protein interaction network and hub gene identification

The STRING online tool was used to observe interactions between genes clustered in the pink and green modules. The visualized PPI network is shown in Figure 5. CytoHubba [21] was then used to process the PPI network to identify the top 10 hub genes in the two modules. Results of the

PPI network calculations are shown in Table 2. Hub genes were identified using a Venn diagram (Figure 6). Five hub genes in the pink module, including collagen type I alpha 1 (*COL1A1*), *COL1A2*, *COL3A1*, prolyl 4-hydroxylase subunit alpha I (*P4HA1*), and fibronectin-1 (*FN1*; Figure 6A), and one in the green module, cyclic adenosine monophosphate response element-binding protein (CREB)-binding protein (*CREBBP*; Figure 6B), were selected. Since collagen family members are generally recognized as potential biomarkers of AF, *COL1A1*, *COL1A2*, and *COL3A1* were not verified in this study.

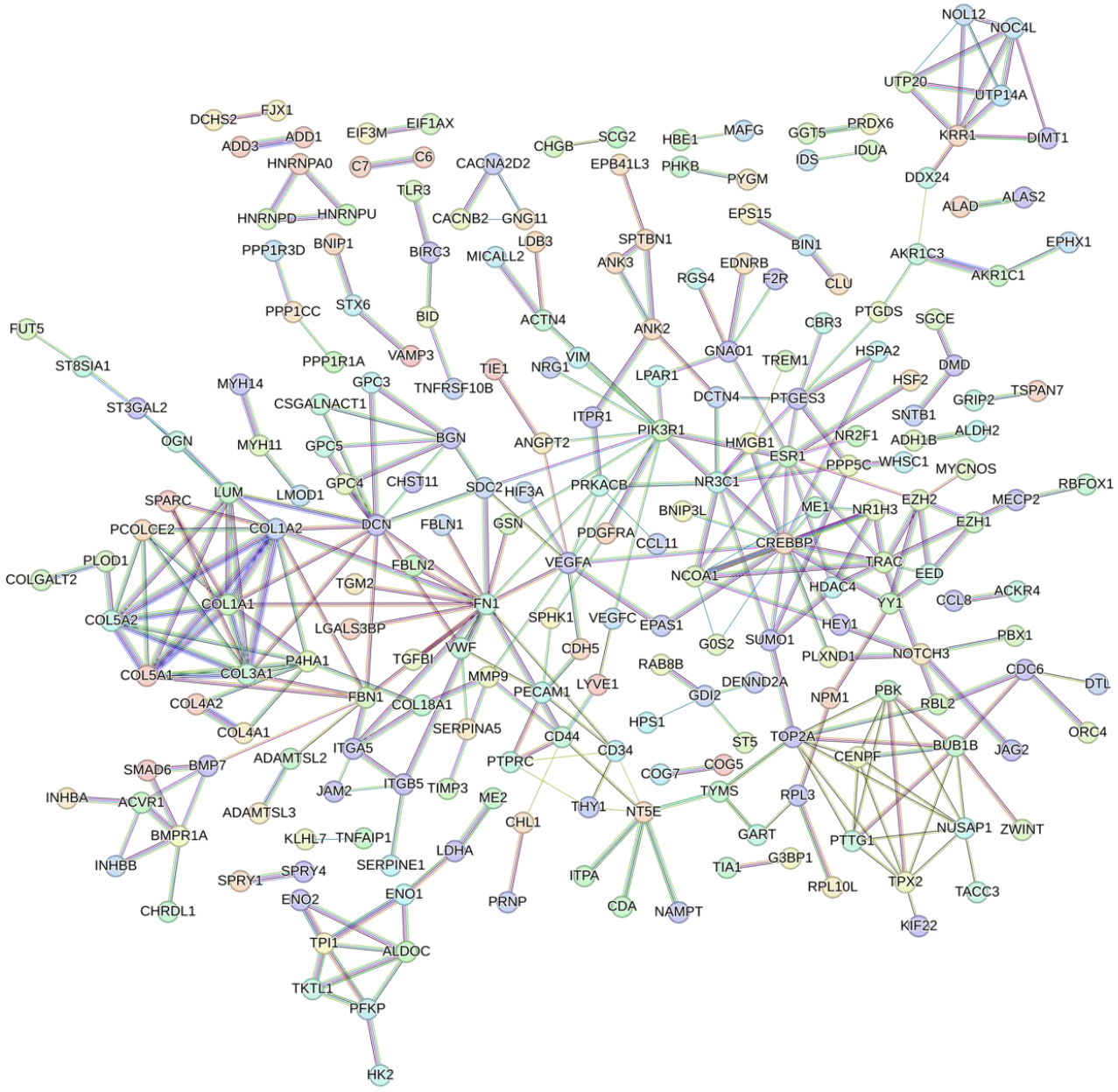

Figure 5. The PPI network of target genes, nodes represent proteins, while edges represent protein-protein associations.

Table 2. Rank of hub genes calculated by four algorithms.

	MCC		Stress		Degree		EPC	
Rank	Pink	Green	Pink	Green	Pink	Green	Pink	Green
1	COL1A1	CREBBP	FN1	VEGFA	P4HA1	CREBBP	COL1A1	CREBBP
2	COL1A2	KRR1	P4HA1	PIK3R1	COL1A1	ESR1	COL1A2	ESR1
3	COL3A1	NCOA1	COL1A1	CREBBP	COL1A2	PIK3R1	P4HA1	NR3C1
4	P4HA1	TOP2A	COL1A2	PTGES3	COL3A1	NR3C1	COL3A1	NCOR2
5	COL5A2	NOC4L	VWF	ESR1	FN1	NCOR2	COL5A1	NCOA1
6	COL5A1	BUB1B	PECAM1	DCN	COL5A2	DCN	COL5A2	NR1H3
7	SPARC	TPX2	COL3A1	SDC2	COL5A1	NCOA1	SPARC	SUMO1
8	FN1	UTP20	MYH11	NR3C1	SPARC	PTGES3	FN1	HDAC4
9	COL4A2	PTTG1	COL18A1	PTGDS	COL4A2	VEGFA	COL4A1	PIK3R1
10	COL4A1	UTP14A	HK2	MMP9	COL4A1	KRR1	COL4A2	PTGES3

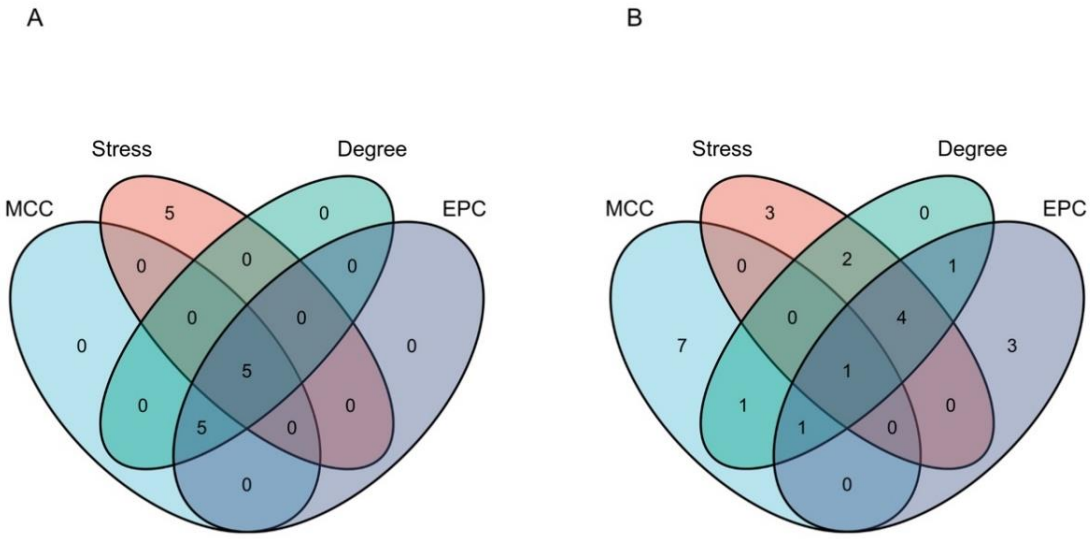

Figure 6. The Venn diagram of hub genes calculated by four algorithms (MCC, Stress, Degree, EPC). (A) Pink module; (B) Green module.

3.5. Validation of target hub genes

The selected biomarkers, *FN1*, *P4HA1* and *CREBBP*, were identified in 15 LAAs of patients with AF and SR using qRT-PCR. However, expression of the hub genes in patients with AF were higher than those in patients with SR (Figure 7, Table 3). In addition, one dataset, GSE41177, was downloaded from the GEO database to validate our hypothesis (Figure 8, Table 4). The expression of the three verified genes was significantly higher in the AF group than in the SR group, and the qRT-PCR and GSE41177 validation results showed that they were statistically significant.

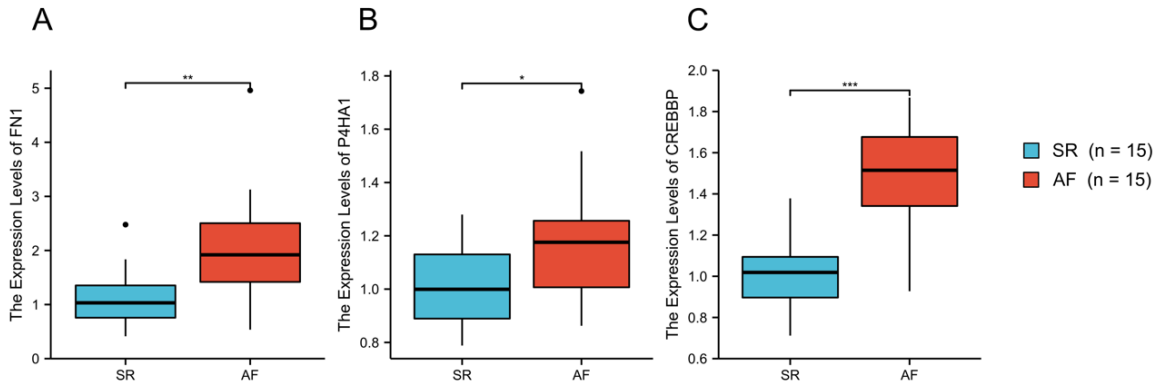

Figure 7. The relative expression of hub genes between patients with atrial fibrillation (AF, n = 15) and sinus rhythm (SR, n = 15). (A) FN1; (B) P4HA1; (C) CREBBP. The error bar represents standard deviations.

Table 3. Relative expression of three hub genes utilizing qRT-PCR.

	The difference [AF-SR]	Confidence Interval [95%]	P-value
FN1	1.000	0.372−1.628	0.003
P4HA1	0.155	0.005−0.305	0.043
CREBBP	0.472	0.292−0.652	< 0.001

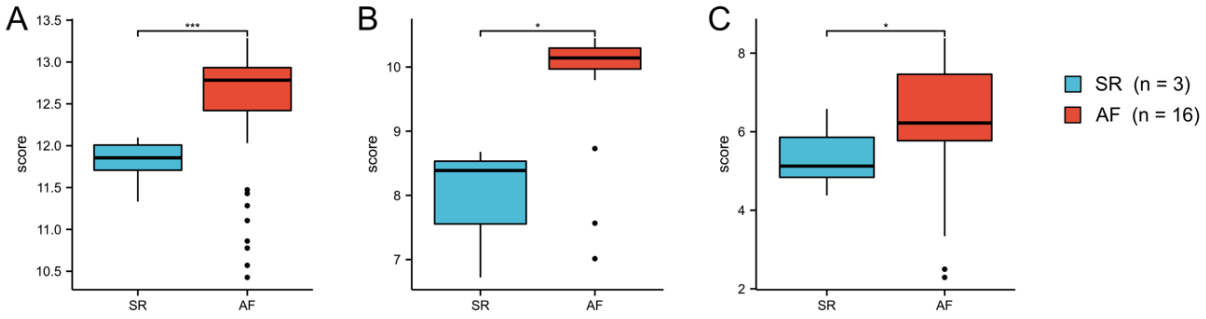

Figure 8. Boxplot of hub genes compared in GSE41177 between the AF group (n = 16) and SR group (n = 3). (A) FN1; (B) P4HA1; (C) CREBBP. The error bar represents the standard deviation.

Table 4. Disparity validation of three hub genes in GSE41177.

	The difference [AF-SR]	Confidence Interval [95%]	P-value
FN1	0.865	0.688−1.108	< 0.001
P4HA1	1.754	0.342−3.489	0.023
CREBBP	1.088	0.131−1.673	0.031

3.6. Clinical information of selected specimens

Clinical data for the SR group was not made available to protect donor privacy because the LA specimens were taken from healthy donors during heart transplant surgery. However, we were able to provide clinical information on patients with AF who underwent qRT-PCR. Information of fifteen AF patients was collected: eleven male and four female patients were enrolled. The mean age was 64.333 (7.659) years. Five patients had preoperative diabetes, and were on treatment with long-term hypoglycemic drugs or injected insulin. Four patients had a history of chronic smoking, including two who had smoked for more than three decades, one for four decades and one for five decades. Hypertension was observed in several of the patients, with four patients having grade I hypertension, five having grade II hypertension, and one having grade III hypertension. The mean pre-operative CHA_2DS_2-VASc score was 2.267 (1.223). The pre-operative echocardiogram showed a mean left atrium diameter of 44.733 (5.946) mm, left ventricular end-diastolic dimension of 50.467 (5.125) mm, pulmonary artery systolic pressure of 33.800 (5.199) mmHg and left ventricular ejection fraction of 63.800 (7.930)%. The results of the serological test were Cr 83.400 (15.565) μmol/L, cTnT 0.010 (0.004) ng/ml and BNP 917.160 (907.249) pg/ml preoperatively.

4. Discussion

AF is the most prevalent tachyarrhythmia in the clinic that increases the risk of ischemic stroke and systemic arterial embolism, possibly leading to high morbidity and mortality [23]. Heart failure, myocardial infarction, dementia and kidney damage are associated with AF [24−27]. Previous research has provided profound insights into the mechanisms underpinning AF, especially the biomarkers of hemodynamic stress, inflammation and tissue fibrosis. However, these biomarkers were discovered mainly in serum and have seldom been validated in human LAA specimens. The strength of our study was the combined use of online databases and human appendage specimens to identify potential biomarkers and possible mechanisms involved in AF.

Structural remodeling caused by the degeneration of atrial myocytes and proliferation of interstitial fibers is a possible mechanism behind AF development [28,29]. The proliferation and differentiation of cardiac fibroblasts into cardiac myofibroblasts caused by various pathological irritants promote atrial fibrosis and cardiac structural remodeling [30]. Immoderate deposition of the ECM and aberrant proliferation of atrial fibroblasts are critical characteristics of atrial fibrosis [31].

FN1 encodes fibronectin, a member of the glycoprotein family, which is abundant in the ECM [32]. The upregulation of *FN1* promotes TGF-β, c-caspase-9/t-caspase-9, and NF-κB p65 protein expression, and reduces the p-PI3K/PI3K and p-AKT/AKT ratio, indicating that *FN1* has a positive correlation with the TGF-β/PI3K/Akt signaling pathway [33]. Fibronectin can bind to ECM proteins through multiple domains, such as the gelatin domain combined with type II collagen, which acts as an adhesion molecule. In addition, FN1 induces the expression of COL1A1, which is involved in ECM synthesis [34]. COL1A1 constitutes a major component of type I collagen [35]. Two chains of COL1A1 and one of COL1A2 comprise type I collagen [36,37], which accounts for 85% of the ECM synthesized by cardiac fibroblasts [38]. Studies have shown that ECM composition and volume are closely related to AF [39], with evidence of collagen I and collagen III elevation in the left atrial tissue of patients with AF [40]. Owing to the increase in fibrosis and extracellular volume, cardiac and atrial fibrosis are presumed to interfere with the mechanical transduction of electricity between

cardiomyocytes by disturbing cell connections [41,42].

The involvement of abnormal COL1A1 and COL1A2 expression in tumor metastasis and proliferation has been implicated [43,44]; however, their functions in the pathological mechanisms underpinning AF have seldom been reported. COL1A1, COL1A2 and FN1 jointly participate in regulating the PI3K/Akt pathway, which is associated with various BPs, including cell proliferation and differentiation, apoptosis, tumorigenesis and glycogen synthesis. The role of the PI3K/Akt pathway in AF has been widely reported. The administration of H_2S has been indicated to significantly mitigate atrial fibrosis and reduce the incidence of AF via the activation of the PI3K/Akt pathway [45]. Another study showed that IGF1 plays a fibrotic role in AF by activating the PI3K-Akt pathway [46]. Moreover, the Wnt/β-catenin-induced PI3K/ATK-activated protein C system can be regulated by edoxaban to improve thromboembolism in AF [47].

Prolyl 4-hydroxylases (P4Hs) are key enzymes in collagen synthesis, and P4HA1 is the main isoform, exerting a rate-limiting function during collagen maturation and secretion [48,49]. P4HA1 is correlated with pan-cancer prognosis, indicating an immunosuppressive microenvironment, and can be used as a potential target for immunotherapy [50]. Research has indicated that P4HA1 might play a role in myocardial fibrosis in diabetic cardiomyopathy through the CD36-JNK-AP1 pathway [51]. Blocking the p38 MAPK and ASK1-JNK-NonO pathways can also reduce vascular remodeling in patients with hypertension caused by elevated P4HA1 levels [48]. An E-box-like sequence (CACGGG) located at -135 bp of the *P4HA1* promoter is responsible for 80% of the basal transcriptional activity, and appears to be upregulated by TGF-β1 [52], which may contribute to the development of atrial fibrosis and fibrillation by mediating CD44/STAT3 signaling [53]. However, research on its direct effect on triggering and maintenance of AF is scarce.

CREBBP is a lysine acetyltransferase with a myriad of cellular functions. CREBBP exerts its influence primarily through its role as a transcriptional regulator, with the non-transcriptional effect of DNA replication and metabolism inside and outside the nucleus. Binding of CREBBP to CREB is proposed to trigger apoptosis via the CREBBP/p53/p21/Bax pathway [54]. Studies have indicated that transgenic mice with directional expression of CREB repressors in cardiomyocytes may develop atrial dilation and spontaneous AF [55,56]. Tachypacing activates cAMP/PKA/CREB signaling and promotes CREB/CD44 linkage, thus resulting in the downregulation of genes encoding L-type calcium channels and shortening of action potential duration and refractoriness [57]. This may hinder conduction and promote atrial reentry, contributing significantly to the onset of AF [57−60].

Various molecules participate in the pathophysiological process of AF by activating the PI3K/Akt pathway and ECM production or organization, indicating that molecules present in this pathway, such as FN1, P4HA1 and CREBBP, may be potential therapeutic targets for patients with AF.

However, our study had some limitations. First, this study was based on the data downloaded from the GEO database that included only a few patients. Further studies with more patients should be conducted to support our results. Second, the clinical characteristics were not used in our study. Third, although we validated the expression of target hub genes using qRT-PCR, further research is required to explore their potential mechanisms in AF development.

5. Conclusions

In this study, we identified three hub genes, *FN1*, *P4HA1* and *CREBBP*, as potential biomarkers of AF through integrated bioinformatics analyses of three datasets from the GEO database. The

expression of these three genes in the myocardium of patients with AF was higher than that in individuals with normal SR. In-depth investigation of these biomarkers may provide insights into the potential mechanism behind AF endogenesis.

6. Data availability

The GSE2240 dataset is available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse2240, and the GSE41177dataset is available through the following links: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse41177.

7. Contributions

Changfa Guo and Chunsheng Wang conceived and designed this study; Miao Zhu, Tao Yan, Shijie Zhu, and Fan Weng contributed to the data analysis and prepared the main manuscript. All authors reviewed the final manuscript.

Acknowledgments

We thank Dr. Zhengyang Hu and Qi Wu for their proofreading and valuable advice. This study was supported by the General Program of the National Natural Science Foundation of China (No. 81770408).

Conflict of interest

We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. C. T. January, L. S. Wann, J. S. Alpert, H. Calkins, J. E. Cigarroa, J. C. ClevelandJr, et al., 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society, *J. Am. Coll. Cardiol.*, **130** (2014), e199–e267. https://doi.org/10.1161/CIR.0000000000000041
2. S. S. Chugh, R. Havmoeller, K. Narayanan, D. Singh, M. Rienstra, E. J. Benjamin, et al., Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study, *Circulation*, **129** (2014), 837−847. https://doi.org/10.1161/CIRCULATIONAHA.113.005119
3. R. S. Wijesurendra, B. Casadei, Mechanisms of atrial fibrillation, *Heart*, **105** (2019), 1860−1867. https://doi.org/10.1136/heartjnl-2018-314267
4. S. Jame, G. Barnes, Stroke and thromboembolism prevention in atrial fibrillation, *Heart*, **106** (2020), 10−17. https://doi.org/10.1136/heartjnl-2019-314898
5. C. C. Wang, C. L. Lin, G. J. Wang, C. T. Chang, F. C. Sung, C. H. Kao, Atrial fibrillation associated with increased risk of venous thromboembolism−A population-based cohort study, *Thromb. Haemost.*, **113** (2015), 185−192. https://doi.org/10.1160/TH14-05-0405

6. E. J. Benjamin, P. A. Wolf, R. B. D'Agostino, H. Silbershatz, W. B. Kannel, D. Levy, Impact of atrial fibrillation on the risk of death: the Framingham Heart Study, *Circulation*, **98** (1998), 946−952. https://doi.org/10.1161/01.CIR.98.10.946
7. T. J. Wang, M. G. Larson, D. Levy, E. J. Benjamin, E. P. Leip, T. Omland, et al., Plasma natriuretic peptide levels and the risk of cardiovascular events and death, *N. Engl. J. Med.*, **350** (2004), 655−663. https://doi.org/10.1056/NEJMoa031994
8. R. B. Schnabel, M. G. Larson, J. F. Yamamoto, L. M. Sullivan, M. J. Pencina, J. B. Meigs, et al., Relations of biomarkers of distinct pathophysiological pathways and atrial fibrillation incidence in the community, *Circulation*, **121** (2010), 200−207. https://doi.org/10.1161/CIRCULATIONAHA.109.882241
9. K. K. Patton, P. T. Ellinor, S. R. Heckbert, R. H. Christenson, C. DeFilippi, J. S. Gottdiener, et al., N-terminal pro-B-type natriuretic peptide is a major predictor of the development of atrial fibrillation: the Cardiovascular Health Study, *Circulation*, **120** (2009), 1768−1774. https://doi.org/10.1161/CIRCULATIONAHA.109.873265
10. R. J. Aviles, D. O. Martin, C. Apperson-Hansen, P. L. Houghtaling, P. Rautaharju, R. A. Kronmal, et al., Inflammation as a risk factor for atrial fibrillation, *Circulation*, **108** (2003), 3006−3010. https://doi.org/10.1161/01.CIR.0000103131.70301.4F
11. K. W. Lee, T. H. EverettIV, D. Rahmutula, J. M. Guerra, E. Wilson, C. Ding, et al., Pirfenidone prevents the development of a vulnerable substrate for atrial fibrillation in a canine model of heart failure, *Circulation*, **114** (2006), 1703−1712. https://doi.org/10.1161/CIRCULATIONAHA.106.624320
12. M. Rienstra, X. Yin, M. G. Larson, J. D. Fontes, J. W. Magnani, D. D. McManus, et al., Relation between soluble ST2, growth differentiation factor-15, and high-sensitivity troponin I and incident atrial fibrillation, *Am. Heart J.*, **167** (2014) 109−115. https://doi.org/10.1016/j.ahj.2013.10.003
13. Y. Nakano, S. Niida, K. Dote, S. Takenaka, H. Hirao, F. Miura, et al., Matrix metalloproteinase-9 contributes to human atrial remodeling during atrial fibrillation, *J. Am. Coll. Cardiol.*, **43** (2004), 818−825. https://doi.org/10.1016/j.jacc.2003.08.060
14. F. Gramley, J. Lorenzen, E. Koellensperger, K. Kettering, C. Weiss, T. Munzel, Atrial fibrosis and atrial fibrillation: the role of the TGF-beta1 signaling pathway, *Int. J. Cardiol.*, **143** (2010), 405−413. https://doi.org/10.1016/j.ijcard.2009.03.110
15. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, et al., *limma* powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.*, **43** (2015), e47. https://doi.org/10.1093/nar/gkv007
16. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, *Stat. Appl. Genet. Mol. Biol.*, **4** (2005). https://doi.org/10.2202/1544-6115.1128
17. D. W. Huang, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res.*, **37** (2009), 1−13. https://doi.org/10.1093/nar/gkn923
18. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.*, **102** (2005), 15545−15550. https://doi.org/10.1073/pnas.0506580102

19. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Res.*, **47** (2019), D607−D613. https://doi.org/10.1093/nar/gky1131
20. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13** (2003), 2498−2504. https://doi.org/10.1101/gr.1239303
21. C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, C. Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, *BMC Syst. Biol.*, **8** (2014). https://doi.org/10.1186/1752-0509-8-S4-S11
22. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS*, **16** (2012), 284−287. https://doi.org/10.1089/omi.2011.0118
23. S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, et al., Heart disease and stroke statistics-2020 update: a report from the American heart association, *Circulation*, **141** (2020), e139−e596. https://doi.org/10.1161/CIR.0000000000000757
24. M. Böhm, M. D. Ezekowitz, S. J. Connolly, J. W. Eikelboom, S. H. Hohnloser, P. A. Reilly, et al., Changes in renal function in patients with atrial fibrillation: an analysis from the RE-LY trial, *J. Am. Coll. Cardiol.*, **65** (2015), 2481−2493. https://doi.org/10.1016/j.jacc.2015.03.577
25. E. Z. Soliman, M. M. Safford, P. Muntner, Y. Khodneva, F. Z. Dawood, N. A. Zakai, et al., Atrial fibrillation and the risk of myocardial infarction, *JAMA Intern. Med.*, **174** (2014), 107−114. https://doi.org/10.1001/jamainternmed.2013.11912
26. D. P. Morin, M. L. Bernard, C. Madias, P. A. Rogers, S. Thihalolipavan, N. A. M. Estes III, The state of the art: atrial fibrillation epidemiology, prevention, and treatment, *Mayo Clin. Proc.*, **91** (2016), 1778−1810. https://doi.org/10.1016/j.mayocp.2016.08.022
27. L. Staerk, J. A. Sherer, D. Ko, E. J. Benjamin, R. H. Helm, Atrial fibrillation: epidemiology, pathophysiology, and clinical outcomes, *Circ. Res.*, **120** (2017), 1501−1517. https://doi.org/10.1161/CIRCRESAHA.117.309732
28. S. Nattel, Molecular and cellular mechanisms of atrial fibrosis in atrial fibrillation, *JACC Clin. Electrophysiol.*, **3** (2017), 425−435. https://doi.org/10.1016/j.jacep.2017.03.002
29. Y. Iwasaki, K. Nishida, T. Kato, S. Nattel, Atrial fibrillation pathophysiology: implications for management, *Circulation*, **124** (2011), 2264−2274. https://doi.org/10.1161/CIRCULATIONAHA.111.019893
30. C. Zhang, Y. Zhang, H. Zhu, J. Hu, Z. Xie, MiR-34a/miR-93 target c-Ski to modulate the proliferaton of rat cardiac fibroblasts and extracellular matrix deposition in vivo and in vitro, *Cell. Signalling*, **46** (2018), 145−153. https://doi.org/10.1016/j.cellsig.2018.03.005
31. Q. Wang, Y. Yu, P. Zhang, Y. Chen, C. Li, J. Chen, et al., The crucial role of activin A/ALK4 pathway in the pathogenesis of Ang-II-induced atrial fibrosis and vulnerability to atrial fibrillation, *Basic Res. Cardiol.*, **112** (2017), 47. https://doi.org/10.1007/s00395-017-0634-1
32. B. Li, W. Shen, H. Peng, Y. Li, F. Chen, L. Zheng, et al., Fibronectin 1 promotes melanoma proliferation and metastasis by inhibiting apoptosis and regulating EMT, *Onco Targets Ther.*, **12** (2019), 3207−3221. https://doi.org/10.2147/OTT.S195703

33. H. Zhang, X. Chen, P. Xue, X. Ma, J. Li, J. Zhang, FN1 promotes chondrocyte differentiation and collagen production via TGF-beta/PI3K/Akt pathway in mice with femoral fracture, *Gene*, **769** (2021), 145253. https://doi.org/10.1016/j.gene.2020.145253
34. Y. X. Liao, Z. P. Zhang, J. Zhao, J. P. Liu, Effects of fibronectin 1 on cell proliferation, senescence and apoptosis of human glioma cells through the PI3K/AKT signaling pathway, *Cell. Physiol. Biochem.*, **48** (2018), 1382−1396. https://doi.org/10.1159/000492096
35. H. P. Ma, H. L. Chang, O. A. Bamodu, V. K. Yadav, T. Y. Huang, A. T. H. Wu, et al., Collagen 1A1 (COL1A1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis, *Cancers (Basel)*, **11** (2019), 786. https://doi.org/10.3390/cancers11060786.
36. K. Gelse, E. Pöschl, T. Aigner, Collagens−structure, function, and biosynthesis, *Adv. Drug Delivery Rev.*, **55** (2003), 1531−1546. https://doi.org/10.1016/j.addr.2003.08.002
37. J. Y. Exposito, U. Valcourt, C. Cluzel, C. Lethias, The fibrillar collagen family, *Int. J. Mol. Sci.*, **11** (2010), 407−426. https://doi.org/10.3390/ijms11020407
38. K. T. Weber, Cardiac interstitium in health and disease: the fibrillar collagen network, *J. Am. Coll. Cardiol.*, **13** (1989), 1637−1652. https://doi.org/10.1016/0735-1097(89)90360-4
39. J. Xu, G. Cui, F. Esmailian, M. Plunkett, D. Marelli, A. Ardehali, et al., Atrial extracellular matrix remodeling and the maintenance of atrial fibrillation, *Circulation*, **109** (2004), 363−368. https://doi.org/10.1161/01.CIR.0000109495.02213.52
40. A. Boldt, U. Wetzel, J. Lauschke, J. Weigl, J. Gummert, G. Hindricks, et al., Fibrosis in left atrial tissue of patients with atrial fibrillation with and without underlying mitral valve disease, *Heart*, **90** (2004), 400−405. https://doi.org/10.1136/hrt.2003.015347
41. F. G. Akar, R. D. Nass, S. Hahn, E. Cingolani, M. Shah, G. G. Hesketh, et al., Dynamic changes in conduction velocity and gap junction properties during development of pacing-induced heart failure, *Am. J. Physiol. Heart Circ. Physiol.*, **293** (2007), H1223−H1230. https://doi.org/10.1152/ajpheart.00079.2007
42. C. Rucker-Martin, P. Milliez, S. Tan, X. Decrouy, M. Recouvreur, R. Vranckx, et al., Chronic hemodynamic overload of the atria is an important factor for gap junction remodeling in human and rat hearts, *Cardiovasc. Res.*, **72** (2006), 69−79. https://doi.org/10.1016/j.cardiores.2006.06.016
43. I. I. de Caceres, E. Dulaimi, A. M. Hoffman, T. Al-Saleem, R. G. Uzzo, P. Cairns, Identification of novel target genes by an epigenetic reactivation screen of renal cancer, *Cancer Res.*, **66** (2006), 5021−5028. https://doi.org/10.1158/0008-5472.CAN-05-3365
44. V. F. Bonazzi, D. J. Nancarrow, M. S. Stark, R. J. Moser, G. M. Boyle, L. G. Aoude, et al., Cross-platform array screening identifies COL1A2, THBS1, TNFRSF10D and UCHL1 as genes frequently silenced by methylation in melanoma, *PLoS One*, **6** (2011), e26121. https://doi.org/10.1371/journal.pone.0026121
45. X. Xue, X. Ling, W. Xi, P. Wang, J. Sun, Q. Yang, et al., Exogenous hydrogen sulfide reduces atrial remodeling and atrial fibrillation induced by diabetes mellitus via activation of the PI3K/Akt/eNOS pathway, *Mol. Med. Rep.*, **22** (2020), 1759−1766. https://doi.org/10.3892/mmr.2020.11291
46. J. Wang, Z. Li, J. Du, J. Li, Y. Zhang, J. Liu, et al., The expression profile analysis of atrial mRNA in rats with atrial fibrillation: the role of IGF1 in atrial fibrosis, *BMC Cardiovasc. Disord.*, **19** (2019), 40. https://doi.org/10.1186/s12872-019-1013-7

47. X. Shan, Z. Liu, M. Wulasihan, S. Ma, Edoxaban improves atrial fibrillation and thromboembolism through regulation of the Wnt-beta-induced PI3K/ATK-activated protein C system, *Exp. Ther. Med.*, **17** (2019), 3509−3517. https://doi.org/10.3892/etm.2019.7379
48. X. Liu, X. Huang, L. Chen, Y. Zhang, M. Li, L. Wang, et al. Mechanical stretch promotes matrix metalloproteinase-2 and prolyl-4-hydroxylase alpha1 production in human aortic smooth muscle cells via Akt-p38 MAPK-JNK signaling, *Int. J. Biochem. Cell Biol.*, **62** (2015), 15−23. https://doi.org/10.1016/j.biocel.2015.02.009
49. K. I. Kivirikko, T. Pihlajaniemi, Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases, *Adv. Enzymol. Relat. Areas Mol. Biol.*, **72** (1998), 325−398. https://doi.org/10.1002/9780470123188.ch9
50. Q. Zhao, J. Liu, P4HA1, a prognostic biomarker that correlates with immune infiltrates in lung adenocarcinoma and pan-cancer, *Front. Cell Dev. Biol.*, **9** (2021), 754580. https://doi.org/10.3389/fcell.2021.754580
51. T. Zhao, H. Chen, C. Cheng, J. Zhang, Z. Yan, J. Kuang, et al., Liraglutide protects high-glucose-stimulated fibroblasts by activating the CD36-JNK-AP1 pathway to downregulate P4HA1, *Biomed. Pharmacother.*, **118** (2019), 109224. https://doi.org/10.1016/j.biopha.2019.109224
52. L. Chen, Y. H. Shen, X. Wang, J. Wang, Y. Gan, N. Chen, et al., Human prolyl-4-hydroxylase alpha(I) transcription is mediated by upstream stimulatory factors, *J. Biol. Chem.*, **281** (2006), 10849−10855. https://doi.org/10.1074/jbc.M511237200
53. S. H. Chang, Y. H. Yeh, J. L. Lee, Y. J. Hsu, C. T. Kuo, W. J. Chen, Transforming growth factor-beta-mediated CD44/STAT3 signaling contributes to the development of atrial fibrosis and fibrillation, *Basic Res. Cardiol.*, **112** (2017), 58. https://doi.org/10.1007/s00395-017-0647-9
54. J. Yang, L. Chen, J. Yang, J. Ding, S. Li, H. Wu, et al., MicroRNA-22 targeting CBP protects against myocardial ischemia-reperfusion injury through anti-apoptosis in rats, *Mol. Biol. Rep.*, **41** (2014), 555−561. https://doi.org/10.1007/s11033-013-2891-x
55. P. Kirchhof, E. Marijon, L. Fabritz, N. Li, W. Wang, T. Wang, et al. Overexpression of cAMP-response element modulator causes abnormal growth and development of the atrial myocardium resulting in a substrate for sustained atrial fibrillation in mice, *Int. J. Cardiol.*, **166** (2013), 366−374. https://doi.org/10.1016/j.ijcard.2011.10.057
56. N. Li, D. Y. Chiang, S. Wang, Q. Wang, L. Sun, N. Voigt, et al., Ryanodine receptor-mediated calcium leak drives progressive development of an atrial fibrillation substrate in a transgenic mouse model, *Circulation*, **129** (2014), 1276−1285. https://doi.org/10.1161/CIRCULATIONAHA.113.006611
57. S. H. Chang, Y. H. Chan, W. J. Chen, G. J. Chang, J. L. Lee, Y. H. Yeh, et al., Tachypacing-induced CREB/CD44 signaling contributes to the suppression of L-type calcium channel expression and the development of atrial remodeling, *Heart Rhythm*, **18** (2021), 1760−1771. https://doi.org/10.1016/j.hrthm.2021.05.021
58. R. F. Bosch, X. Zeng, J. B. Grammer, K. Popovic, C. Mewis, V. Kühlkamp, Ionic mechanisms of electrical remodeling in human atrial fibrillation, *Cardiovasc. Res.*, **44** (1999), 121−131. https://doi.org/10.1016/S0008-6363(99)00178-9
59. L. Yue, P. Melnyk, R. Gaspo, Z. Wang, S. Nattel, Molecular mechanisms underlying ionic remodeling in a dog model of atrial fibrillation, *Circ. Res.*, **84** (1999), 776−784. https://doi.org/10.1161/01.RES.84.7.776

60. X. Y. Qi, Y. H. Yeh, L. Xiao, B. Burstein, A. Maguy, D. Chartier, et al., Cellular signaling underlying atrial tachycardia remodeling of L-type calcium current, *Circ. Res.*, **103** (2008), 845−854. https://doi.org/10.1161/CIRCRESAHA.108.175463

AIMS Press