Original Paper

Analysis of Genes Involved in Persistent Atrial Fibrillation: Comparisons of ‘Trigger’ and ‘Substrate’ Differences

Rongjun Zou^a Minglei Yang^b Wanting Shi^c Chengxi Zheng^d
Hui Zeng^a Xifeng Lin^a Dingwen Zhang^a Songran Yang^{e,f} Ping Hua^a

^aDepartment of Cardio-Vascular Surgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, ^bDepartment of Genetics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, ^cDepartment of Gastroenterology, Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, ^dDepartment of Dermatology, Henan No.2 Provincial People's Hospital, Zhengzhou, ^eThe Biobank of Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, ^fGuangdong Province Key Laboratory of Brain Function and Disease, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

Key Words

Atrial fibrillation • Pulmonary vein and the surrounding left atrial junction • Left atrial appendage • Bioinformatics analysis

Abstract

Background/Aims: Recent research has improved our understanding of the pulmonary vein and surrounding left atrial (LA-PV) junction and the left atrial appendage (LAA), which are considered the ‘trigger’ and ‘substrate’ in the development of atrial fibrillation (AF), respectively. Herein, with the aim of identifying the underlying potential genetic mechanisms, we compared differences in gene expression between LA-PV junction and LAA specimens via bioinformatic analysis. **Methods:** Microarray data of AF (GSE41177) were downloaded from the Gene Expression Omnibus database. In addition, linear models for microarray data limma powers differential expression analyses and weighted correlation network analysis (WGCNA) were applied. **Results:** From the differential expression analyses, 152 differentially expressed genes and hub genes, including *LEP*, *FOS*, *EDN1*, *NMU*, *CALB2*, *TAC1*, and *PPBP*, were identified. Our analysis revealed that the maps of extracellular matrix (ECM)-receptor interactions, PI3K-Akt and Wnt signaling pathways, and ventricular cardiac muscle tissue morphogenesis were significantly enriched. In addition, the WGCNA results showed high correlations between genes and related genetic clusters to external clinical characteristics. Maps of the ECM-receptor interactions, chemokine signaling pathways, and the cell cycle were significantly enriched in the genes of corresponding modules and closely associated with AF duration, left atrial diameter, and left ventricular ejection function, respectively. Similarly, mapping of the TNF signaling pathway indicated significant association with genetic traits of ischemic heart

R. Zou, M. Yang and W. Shi contributed equally to this work.

Ping Hua
and Songran Yang

Department of Cardio-Vascular Surgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University Guangzhou, 510120 (China)
E-Mail huaping88@sina.com, yangsongran@126.com

disease, hypertension, and diabetes comorbidity. **Conclusions:** The ECM-receptor interaction as a possible central node of comparison between LA-PV and LAA samples reflected the special functional roles of ‘triggers’ and ‘substrates’ and may be closely associated with AF duration. Furthermore, *LEP*, *FOS*, *EDN1*, *NMU*, *CALB2*, *TAC1*, and *PPBP* genes may be implicated in the occurrence and maintenance of AF through their interactions with each other.

© 2018 The Author(s)
Published by S. Karger AG, Basel

Introduction

Atrial fibrillation (AF), the most common dysrhythmia, is a major risk factor for cardiovascular (CV) mortality and morbidity resulting from thromboembolism and hemodynamic changes owing to the high-frequency fibrillation of the atrial wall and stagnation of blood flow [1]. Currently, there are 30 million cases of AF worldwide and approximately one-third of cases of CV-related mortality and morbidity are caused by AF [2, 3]. Recently, there has been an increasing focus on the relationships between subclinical episodes of AF and the risk of silent cerebral infarct (SCI) and stroke. Marfella's group reported findings related to silent episodes of AF and observed that subclinical episodes of AF may increase the risk of SCI and stroke, especially in patients with type 2 diabetes [4]. Emerging evidence has linked heart failure and comorbid AF to increased CV-related mortality [5]. Notably, AF also carries a greater risk of poor cardiovascular prognosis in patients with HF [5], indicating the importance of early AF detection, particularly to improve the quality of life of HF patients.

However, symptomatic patients have limited treatment options and currently available antiarrhythmic drugs are generally associated with potential adverse effects or toxicity. Furthermore, catheter ablation and surgical MAZE have limited efficacy in patients with persistent AF [6]. Surgical ablation is associated with a recurrence rate of ~30% and potential complications in 6% [6]. Similarly, in the synergy ablation system, the ABLATE trial demonstrated that the Cox Maze-IV procedure is effective in 76% and the rate of major adverse events is greater than 9% [7]. Preventing AF recurrence after catheter ablation is a major challenge, and a clinical trial conducted by Sardù's group illustrated that antioxidant treatment after ablation could reduce the serum levels of inflammatory factors, but regrettably it did not improve the poor prognosis in the first year after catheter ablation [8]. That study underscores the fact that the pathogenesis of AF and its underlying molecular mechanisms are not completely understood. Recently, a series of preclinical studies suggested that gene-based approaches could be used for the diagnosis and treatment of AF [9, 10]. Most studies have had encouraging results that indicate the potential of gene-based therapies to combat AF. Amit's group utilized an adenoviral-mediated delivery of KCNH2 (negative mutant of the IKr potassium channel) into a porcine AF model to enable epicardial atrial gene transfer and pacemaker implantation, as well as electrophysiological and molecular studies [9]. The authors found that KCNH2 transfer prolonged atrial action potential duration and prevented AF [9]. Similarly, Zhang's group found that miR-206 modulated intrinsic cardiac autonomic nerve remodeling in AF via right atrial tachypacing via the regulation of superoxide dismutase 1 (SOD1) *in vivo*, and this may be a potential therapeutic target for AF because it shortens the atrial effective refractory period [10]. Furthermore, many loci or target genes associated with ion channels, autonomic innervation, gap junctions, and substrate modifiers have been identified in preclinical studies to attenuate fibrosis, autonomic nerve sprouting, and AF duration [11-13]. However, inflammation and the immune response, oxidative stress, and genetic mechanisms do not fully explain the development of AF, and autonomic system dysfunction may exert direct effects on the initiation and maintenance of AF due to heterogeneous changes in the atrial electrophysiology [14, 15]. In a large clinical study, Rizzo's group offered compelling evidence of the relationships between autonomic dysfunction and silent AF in type 2 diabetes, thus revealing a potential prognostic factor and principle for target therapies that potentially offer greater specificity and efficacy in the control of AF [15]. In sum, AF may be a final common endpoint of atrial remodeling or

electrical, structural, cardiac autonomic nerves, and ion channel changes that cause irregular atrial rhythms and unique spatial heterogeneous conduction [16]; thus it is essential to develop a clinical scheme for the early and accurate detection of AF and appropriate interventions [13]. At present, gene-based transformation therapy and tailored treatments may be promising solutions because of the advent of genome-wide association studies.

Materials and Methods

Gene Expression Omnibus (GEO) dataset

We downloaded the expression microarray data (CEL data) from the GSE41177 dataset from Gene Expression Omnibus (GEO) (<http://ncbi.nlm.nih.gov/geo/>) [17]. The GSE41177 dataset included 16 pairs of pulmonary veins and surrounding left atrial (LA-PV) junctions and left atrial appendage (LAA) specimens obtained from persistent AF patients after valvular surgery. We measured gene expression using GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix Inc., Santa Clara, CA). In addition, platform information, probe annotation, and clinical information for all samples were pooled for further analysis.

Data processing

We processed RAW data (CEL file) using the Affy R package (<http://www.bioconductor.org/packages/release/bioc/html/affy.html>), provided by the Bioconductor project. We used linear models for microarray data (LIMMA) package in R language, Robust multi-array analysis method, and k-nearest neighbors algorithm to identify differentially expressed (DE) genes between the LA-PV and LAA specimens [18]. We also assessed background correction, quantile normalization, and probe summarization. Averages were considered as gene expression values if the multiple probes were mapped to the same gene symbol. We used a false discovery rate to calculate fold-changes (FC) and also the Benjamini-Hochberg method [19]. Gene expression values of $\log_2\text{FC} > 1$ and adjusted $p < 0.05$ were chosen as thresholds. Hierarchical clustering analysis was used to build a hierarchy of clusters of the DE genes [20]. Subsequently, we performed a visualization analysis of chromosome loci and applied it to the initiation site information to map gene regions and obtain biological information on the DE genes. The database of hg19_refGene was referenced in the gene mapping analysis [21].

Weighted correlation network analysis (WGCNA), using the Eigengene network methodology, was applied to study the biological networks and select clinical features based on weighted correlations between quantitative variables (normalized gene expression data) [22, 23]. Here, we chose the first 25% least mean square deviation of the expression variable to guarantee the stability and accuracy of the bioinformatics statistics to a certain extent to perform further co-expression network analysis. Soft thresholding, based on the scale-free topology criterion, was employed for further analysis to preserve the continuous nature of co-expression information [22, 23]. The dynamic branch cutting approach was applied to identify the gene modules, and the Eigengenes method was used to draw characteristic values based on more complex predictive models, including decision trees and Bayesian networks [23]. Here, the smallest value of the soft thresholding parameter was determined by a condition of high-scale independence. Heat maps were generated to evaluate the connection strengths and module Eigengene of the correlation network. The analysis strategy is presented in Fig. 1.

Identification of protein-protein interaction (PPI) networks

We used the online Search Tool for the Retrieval of Interacting Genes (STRING database,

Fig. 1. Flowchart of the analysis strategy.

Version 10.5; <http://string-db.org/>) to construct a PPI network based on uniquely comprehensive coverage and predictive function of genome-wide data [24]. We used a reliability threshold of a combined score of > 0.4 to construct the PPI network and used Cytoscape software (Version 3.5.1; <http://cytoscape.org/>) to visualize and analyze the biological networks [24]. Thus, the degree of each node was used to screen “hub” genes that interacted with other genes or proteins.

Bio-function enrichment analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) is a web-based analytical tool used to extract comprehensive biological information about candidate genes that can then provide functional interpretation of large lists of genes derived from genomic studies [25]. We used DAVID resources to perform Gene Ontology (GO) functions and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. We assessed GO terms in biological processes, molecular functions, and cellular components. GO terms and KEGG maps of biological functions were enriched with a threshold p -value of < 0.05 .

Results

Data processing and identification of DE genes and co-expression modules

After normalization, 152 genes were found to be differentially expressed between LA-PVs samples and LAA samples. Among these DE genes, 104 were down-regulated and 48 were up-regulated. A heat map was generated to assess the expressed values of the DE genes to distinguish LA-PV samples and LAA samples in AF patients. Interestingly, the DE genes included more down-regulated genes in LA-PV than in LAA (Fig. 2A). Additionally, genetic mapping information showed that more DE genes were associated with the negative log of the corresponding p -value between 10×10^{-8} and 15×10^{-8} , and 7 hub genes were at more than 10×10^{-8} (Fig. 2C).

To confirm the WGCNA study, we performed cluster analysis among the top 5121 genes (first 25%), and thus GSM1005428, GSM1005444, GSM1006247, and GSM1006249 were excluded from the sample cluster analysis (Fig. 4A). In addition, the power of $\beta = 12$ was chosen to ensure that the scale independence was up to 0.78 with higher mean connectivity (Fig. 4B). As such, we identified 16 gene clusters/modules of co-expressed genes based on the dynamic branch cutting approach (the merge cut height was 0.25) (Fig. 4D). In addition, following the module-trait relationships analysis, the grey module was primarily associated with AF duration ($p = 0.005$), the cyan module was mainly associated with left ventricular ejection fraction (LVEF) ($p = 0.002$), the blue module was significantly associated with left atrium diameter (LAD) ($p = 0.001$), and the magenta module was particularly important for ischemic heart disease, hypertension, and diabetes comorbidities (the p -values were 0.005, 0.02, and 0.05, respectively) (Fig. 4C). Interaction analysis of the co-expression module demonstrated non-statistically significant results with a brightness area among different modules and a higher scale of independence among these clusters (Fig. 4F). Furthermore, the Eigengene adjacency plot also illustrated a significant difference in connectivity effects (Fig. 4E). The sample clinical information is depicted in Fig. 4A.

KEGG Pathway analysis

KEGG pathway maps of biological functions associated with up- and down-regulated DE genes were obtained from the DAVID web-based search tool, respectively. Table 1 shows the up- and down-regulated genes and the pathways in which they interact. We assessed the co-pathways of the DE genes and identified hsa04512: extracellular matrix (ECM)-receptor interaction (up-regulated: *COMP*, *THBS2*, *THBS4*; down-regulated: *COL6A6*, *RELN*, *TNN*) as possible points of similarity between the LA-PV and LAA samples, reflecting the special functional roles of ‘trigger’ and ‘substrate’ (these data are shown in in Table 1). The expression of hsa04512-related DE genes is depicted in Fig. 5A.

In the WGCNA study, the grey module genes were mainly enriched in hsa04512: ECM-receptor interaction (including 10 genes with p -values of $1.59\text{E-}04$) and further revealed

Fig. 2. Differentially expressed (DE) gene expression, DE gene chromosome locus analysis, and protein-protein interaction (PPI) network. (A) The results of hierarchical clustering analysis of the DE genes for a comparison of LA-PV and LAA samples. Red, higher expression of DE genes. Blue, lower expression of DE genes. (B) The expression of DE genes. (C) Manhattan plots presenting a visualization analysis of chromosome locus via mapping the initiation site information. There were 55 DE genes located in the association with negative logs of the corresponding p-values between 10×10^{-8} and 15×10^{-8} , and 7 hub genes at over 10×10^{-8} . (D) The PPI network constructed via the STRING database for the DE genes with a threshold value > 0.4 . The red nodes represent up-regulated genes. The blue nodes represent down-regulated genes. The sizes of the nodes represent the degree of gene interaction, and the thickness of the blue lines indicates the combined score between the DE genes.

Fig. 3. Gene Ontology (GO) terms enrichment: The functional enrichment analysis was performed using the Database for Annotation, Visualization and Integration Discovery (DAVID database). GO terms enrichment for the selected up- and down-regulated DE genes are shown in Fig. 3. in relation to biological process, molecular function, and cellular component. The sizes of the dots represent the counts of enriched DE genes, and the dot color represents the P-value.

the biological mechanisms in relation to AF duration. The blue module genes were mainly enriched in hsa04062: chemokine signaling pathway (including 27 genes with p-values of 2.13×10^{-7}) and closely associated with LAD. In addition, the cyan module genes were significantly associated with cfa04110: cell cycle (including 9 genes with p-values of 3.51×10^{-8}).

Fig. 4. Construction and interaction analysis of co-expression gene modules. (A) Cluster analysis of samples and the corresponding clinical information. (B) The effect of different power values on the scale independence degree and mean connectivity degree of co-expression modules. (C) The relationships of the co-expression modules with clinical traits. The value at the top of each square and the p value in parentheses reflect the correlation coefficient between the module Eigengene and the clinical characteristics. Red represents a positive correlation, and green represents a negative correlation. (D) The construction of co-expression modules based on a dynamic branch-cutting method. The merge cut height is 0.25. (E & F) The interactions and connectivity of Eigengenes among different gene co-expression modules. The colored squares and numbers on the right represent the different modules and the gene numbers, respectively.

10) related to LVEF, and the magenta module genes were primarily associated with ssc04668: TNF signaling pathway (including 9 genes with p -values of 2.68E-07) may be significantly associated with genetic traits of ischemic heart disease, hypertension, and diabetes (Fig. 5C).

GO functional analysis

We performed GO enrichment analysis of the up-regulated and down-regulated DE genes following the LIMMA study. For the up-regulated genes, we obtained several clusters of interest based on official gene symbol information. Among the biological processes, the first cluster was significantly associated with the Wnt signaling pathway, including the non-canonical Wnt signaling pathway, positive regulation of the canonical Wnt signaling pathway, and negative regulation of the canonical pathway. Among the molecular functions, our results showed that the up-regulated DE genes were significantly associated with the binding of heparin. In cellular components enrichment analysis, the terms of the extracellular region and extracellular space were also enriched. Additionally, different terms were significantly enriched among the down-regulated genes, including regulation of blood pressure, notch signaling path, and PI3K signaling. However, following molecular function analysis, calcium ion binding was significantly enriched in the

Table 1. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment for the AF DE genes in a comparison of LA-PV and LAA specimens

Term	Count	Genes	P-Value
Upregulated DEGs			
hsa04510: Focal adhesion	4	MYL2, COMP, THBS2, THBS4	1.1E-2
hsa04512: ECM-receptor interaction	3	COMP, THBS2, THBS4	1.7E-2
hsa04145: Phagosome	3	COMP, THBS2, THBS4	4.8E-2
hsa04062: Chemokine signaling pathway	3	PPBP, CXCL14, PF4	6.8E-2
hsa04060: Cytokine-cytokine receptor interaction	3	PPBP, CXCL14, PF4	9.8E-2
Downregulated DEGs			
hsa04640: Hematopoietic cell lineage	3	CD3D, IL7, IL7R	4.6E-2
hsa04512: ECM-receptor interaction	3	COL6A6, RELN, TNN	4.8E-2
hsa04151: PI3K-Akt signaling pathway	5	COL6A6, IL7, RELN, TNN, IL7R	4.9E-2

Fig. 5. The expression of ECM-related genes and the Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment among the different genetic modules. (A) The expression of ECM-related genes. (B) The GO term enrichment of the interesting modules, and (C) the KEGG pathway enrichment.

down-regulated DE genes. Furthermore, the cellular component terms of the extracellular regions, extracellular space, and integral components of plasma membranes were significantly enriched in our study (Fig. 3).

In the WGCNA study, the genes in the grey module, which were associated with AF duration, were mainly enriched in association with the positive regulation of inflammatory responses. Similarly, the genes in the blue module were significantly associated with immune responses, illustrating the potential molecular mechanism in relation to LAD. In addition, we also found that the term mitotic cytokinesis was significantly enriched in the cyan module genes, which were closely associated with LVEF. Nevertheless, regarding the genetic traits of comorbid ischemic heart disease, hypertension, and diabetes, the terms immune response and fat cell differentiation were primarily identified among the genes in the magenta module (Fig. 5B).

PPI network analysis of DE genes

Based on the STRING database, we constructed a PPI network containing 152 DE genes with a combined score of > 0.4 (Fig. 2D). There were 66 nodes with network density and heterogeneity of 0.04 and 0.88, respectively. After calculating the degree of each DE gene, 7 genes were considered central to the maintenance of AF with a degree of > 6. In addition, leptin (*LEP*), FBJ murine osteosarcoma viral oncogene homolog (*FOS*), endothelin 1 (*EDN1*), neuromedin U (*NMU*), calbindin 2 (*CALB2*), tachykinin (*TAC1*), and pro-platelet basic protein (chemokine (C-X-C motif) ligand 7) (*PPBP*) could interact with each other in the PPI network (as shown in Fig. 2D). The expression levels of these hub genes are shown in Fig. 2B.

Discussion

The LA-PV junction is thought to be an important ‘driver’ or ‘trigger’ region for the maintenance and immediate recurrence of AF and short bursts of LA-PV tachycardias. LAA may act as a “substrate” or “tract” in AF. Pison’s group reviewed the potential mechanisms of sustained AF triggered by ectopic beats arising from muscle sleeves of the LA-PV to drive focal activity or local re-entry emanating from spiral waves from a curved central core. The mechanisms behind the maintenance of AF are increasingly being attributed to these re-entry substrates [26]. Romero’s group reported that the LA-PV is the key to AF maintenance—they used PV isolation to treat AF [27]—and further alterations of atrial substrates related to

atrial inflammation and fibrosis may reduce the atrial potential duration (APD) and cause electrical heterogeneity and greater vulnerability of single or multiple re-entry beats and random-wave front propagation [28-30]. However, LAA is considered the primary source of thromboemboli in patients with persistent or permanent AF [31].

To develop novel strategies or individualized treatment, it is essential to discriminate region-specific gene expression and understand the pathophysiological mechanisms associated with AF maintenance. Recently, Yeh's group used a genome-wide approach to identify region-specific gene expression in paired LA-PV junction and LAA specimens of patients with AF [32], and found that the LA-PV junction had a higher expression of arrhythmia-related genes; conversely, several thrombosis-related genes were up-regulated in LAA samples [32]. In our LIMMA study, 152 DE genes (48 up-regulated and 104 down-regulated genes) and 7 hub genes were identified, and the ECM-receptor interaction and chemokine signaling pathway were both significantly enriched in the up-regulated DE genes, as were both the ECM-receptor interaction and PI3K-Akt signaling pathway in the down-regulated DE genes. Moreover, to assess the additional relationships between the gene modules and external clinical information, we carried out a WGCNA study and found that ECM-receptor interactions, chemokine signaling pathways, and the cell cycle may be potentially targeted for AF duration, LAD, and LVEF, respectively, at least with a soft threshold cutoff of 12.

In this study, both differential gene expression and co-expression network analyses powerfully revealed an important regulatory role for the ECM-receptor interaction pathway in the comparison of 'trigger' and 'substrate', which turned out to be closely related to AF duration. Recently, Husser's group performed clinical and genetic predictor analysis based on pathway enrichment analysis of GWAS (Genome-wide Association Study) data, and demonstrated that the ECM-receptor interaction pathways were significantly associated with left atrial enlargement and AF type and response to AF catheter ablation [33]. Presumably, ECM remodeling revealed complex relationships between cell-to-cell interactions and communication, and was validated as a central mediator that tightly links ion channels with the development of an ectopic pacemaker [33]. Although the previous studies and our results have provided evidence to show that ECM interaction is a potential target, the precise molecular mechanisms of the ECM in the progressive nature of AF are still unknown.

LEP is a hormone synthesized by adipocytes that has signaling ability and regulates food intake and energy metabolism. Abd El-Aziz's group compared circulating LEP and genotyped the LEP gene and leptin receptor (LEPR) gene and their gene polymorphisms among 100 coronary artery disease (CAD) patients and 100 healthy subjects [34]. The group reported greater LEP expression in CAD patients, and gene polymorphisms were more frequent in the LEP gene (AA genotype) and the leptin receptor gene (RR genotype) in CAD patients who experienced heart failure [34]. In addition, Li's group demonstrated that *LEP*-related obesity may be related to cardiac contraction defects resulting from damage to nicotinamide adenine dinucleotide phosphate (NADPH) oxidase activation, oxidative modification of sarco(endo)plasmic reticulum Ca^{2+} -ATPase, and myosin heavy chains [35]. Similarly, based on the rapidly growing data from experimental studies, the overexpression of *FOS* may play a central role in heart disease because of its association with inflammation and cardiac remodeling. Zhang's group illustrated that the expression of c-fos in cardiomyocytes and serum is rapidly increased at days 1, 3, 7, and 10 after coronary ligation [36]. Drosos's group assessed elevations in the expression of LEP protein in aortic root and coronary arterial perivascular adipose tissue (C-PVAT) and internal mammary arterial perivascular adipose tissue (IMA-PVAT) and found higher expression of hypoxia-inducible factor-1 α (HIF-1 α) and Fos-like antigen (FOSL)2 in hypoxia C-PVAT compared with IMA-PVAT via regulation of the hypoxia-inflammation-fibrosis axis in atherosclerosis, which reflected an increased risk of atherosclerotic plaque and CV-related death [37]. In addition, *EDN1* is an important intermediate for guiding the intrinsic ability of these progenitors to differentiate and self-organize into a functional vasculature [38], and has a pivotal role in the maintenance of normal contractile function via the superoxide-metalloproteinase 9-(Mmp9) cascade. Moreover, less than 35% decrease

in *EDN1* expression is sufficient to cause significant malignant cardiovascular changes [39]. *NMU* produces a brain-gut peptide with potent contractile effects that is significantly expressed in human plasma, epicardial adipose tissue, and the vasculature, including the left ventricle, coronary artery, and the saphenous vein. Several studies have shown that *NMU* precursor mRNA is significantly up-regulated in the left ventricle in dilated cardiomyopathy and ischemic heart disease patients [40, 41].

In conclusion, we combined the powerful LIMMA differential gene expression analysis with WGCNA to identify the hub genes and pathways related to AF perpetuation, recurrence, and maintenance. Of note, 152 DE genes and 7 hub genes, including *LEP*, *FOS*, *EDN1*, *NMU*, *CALB2*, *TAC1*, and *PPBP* were identified and may contribute to the pathogenic and pathophysiological mechanisms. Moreover, the pathway of ECM-receptor interaction was primarily enriched and closely linked to the initiation and maintenance of AF.

Limitations

Although gene-based analysis results are encouraging, we currently do not understand how candidate genes contribute to AF or whether these pathways will translate into improved clinical outcomes to prevent AF. It is worth noting that while the hub genes and pathways in the comparison between AF ‘triggers’ and ‘substrates’ were found, these findings need to be verified in animal models. Moreover, the PPI network information, GO terms, and KEGG pathway maps from the STRING and DAVID databases are not available in similar databases.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC No.81771165), the Special Program for Applied Research on Super Computation from the NSFC-Guangdong Joint Fund (the second phase), the Natural Science Foundation Project in Guangdong province, China (Grant No. 2016A030313295), the Major Project Development and Emerging, Interdisciplinary Funding Projects of Sun Yat-sen University (Grant No. 15ykjc17b) and the Guangzhou science and technology project of Major Special Research Topics on International Collaborative Innovation (Grant No. 201704030032).

Disclosure Statement

The authors have no conflict of interests to disclose.

References

- 1 Koizumi R, Funamoto K, Hayase T, Kanke Y, Shibata M, Shiraishi Y, Yambe T: Numerical analysis of hemodynamic changes in the left atrium due to atrial fibrillation. *J Biomech* 2015;48:472-478.
- 2 Piccini JP, Hammill BG, Sinner MF, Jensen PN, Hernandez AF, Heckbert SR, Benjamin EJ, Curtis LH: Incidence and prevalence of atrial fibrillation and associated mortality among Medicare beneficiaries, 1993-2007. *Circ Cardiovasc Qual Outcomes* 2012;5:85-93.
- 3 Freeman JV, Wang Y, Akar J, Desai N, Krumholz H: National Trends in Atrial Fibrillation Hospitalization, Readmission, and Mortality for Medicare Beneficiaries, 1999-2013. *Circulation* 2017;135:1227-1239.
- 4 Marfella R, Sasso FC, Siniscalchi M, Cirillo M, Paolisso P, Sardu C, Barbieri M, Rizzo MR, Mauro C, Paolisso G: Brief episodes of silent atrial fibrillation predict clinical vascular brain disease in type 2 diabetic patients. *J Am Coll Cardiol* 2013;62:525-530.
- 5 Sardu C, Santamaria M, Rizzo MR, Barbieri M, di Marino M, Paolisso G, Santulli G, Marfella R: Telemonitoring in heart failure patients treated by cardiac resynchronisation therapy with defibrillator (CRT-D): the TELECARD Study. *Int J Clin Pract* 2016;70:569-576.

- 6 Bhargava M, Di Biase L, Mohanty P, Prasad S, Martin DO, Williams-Andrews M, Wazni OM, Burkhardt JD, Cummings JE, Khaykin Y, Verma A, Hao S, Beheiry S, Hongo R, Rossillo A, Raviele A, Bonso A, Themistoclakis S, Stewart K, Saliba WJ, Schweikert RA, Natale A: Impact of type of atrial fibrillation and repeat catheter ablation on long-term freedom from atrial fibrillation: results from a multicenter study. *Heart Rhythm* 2009;6:1403-1412.
- 7 Philpott JM, Zemlin CW, Cox JL, Stirling M, Mack M, Hooker RL, Morris A, Heimansohn DA, Longoria J, Gandhi DB, McCarthy PM: The ABLATE Trial: Safety and Efficacy of Cox Maze-IV Using a Bipolar Radiofrequency Ablation System. *Ann Thorac Surg* 2015;100:1541-1546; discussion 1547-1548.
- 8 Sardu C, Santulli G, Santamaria M, Barbieri M, Sacra C, Paolisso P, D'Amico F, Testa N, Caporaso I, Paolisso G, Marfella R, Rizzo MR: Effects of Alpha Lipoic Acid on Multiple Cytokines and Biomarkers and Recurrence of Atrial Fibrillation Within 1 Year of Catheter Ablation. *Am J Cardiol* 2017;119:1382-1386.
- 9 Amit G, Kikuchi K, Greener ID, Yang L, Novack V, Donahue JK: Selective molecular potassium channel blockade prevents atrial fibrillation. *Circulation* 2010;121:2263-2270.
- 10 Zhang Y, Zheng S, Geng Y, Xue J, Wang Z, Xie X, Wang J, Zhang S, Hou Y: MicroRNA profiling of atrial fibrillation in canines: miR-206 modulates intrinsic cardiac autonomic nerve remodeling by regulating SOD1. *PLoS One* 2015;10:e0122674.
- 11 Kunamalla A, Ng J, Parini V, Yoo S, McGee KA, Tomson TT, Gordon D, Thorp EB, Lomasney J, Zhang Q, Shah S, Browne S, Knight BP, Passman R, Goldberger JJ, Aistrup G, Arora R: Constitutive Expression of a Dominant-Negative TGF-beta Type II Receptor in the Posterior Left Atrium Leads to Beneficial Remodeling of Atrial Fibrillation Substrate. *Circ Res* 2016;119:69-82.
- 12 Aistrup GL, Cokic I, Ng J, Gordon D, Koduri H, Browne S, Arapi D, Segon Y, Goldstein J, Angulo A, Wasserstrom JA, Goldberger JJ, Kadish AH, Arora R: Targeted nonviral gene-based inhibition of Galpha(i/o)-mediated vagal signaling in the posterior left atrium decreases vagal-induced atrial fibrillation. *Heart Rhythm* 2011;8:1722-1729.
- 13 Soucek R, Thomas D, Kelemen K, Bikou O, Seyler C, Voss F, Becker R, Koenen M, Katus HA, Bauer A: Genetic suppression of atrial fibrillation using a dominant-negative ether-a-go-go-related gene mutant. *Heart Rhythm* 2012;9:265-272.
- 14 Chen PS, Chen LS, Fishbein MC, Lin SF, Nattel S: Role of the autonomic nervous system in atrial fibrillation: pathophysiology and therapy. *Circ Res* 2014;114:1500-1515.
- 15 Rizzo MR, Sasso FC, Marfella R, Siniscalchi M, Paolisso P, Carbonara O, Capoluongo MC, Lascar N, Pace C, Sardu C, Passavanti B, Barbieri M, Mauro C, Paolisso G: Autonomic dysfunction is associated with brief episodes of atrial fibrillation in type 2 diabetes. *J Diabetes Complications* 2015;29:88-92.
- 16 Nattel S, Harada M: Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. *J Am Coll Cardiol* 2014;63:2335-2345.
- 17 Clough E, Barrett T: The Gene Expression Omnibus Database. *Methods Mol Biol* 2016;1418:93-110.
- 18 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249-264.
- 19 Fu G, Saunders G, Stevens J: Holm multiple correction for large-scale gene-shape association mapping. *BMC Genet* 2014;15 Suppl 1:S5.
- 20 Serban N, Jiang H: Multilevel functional clustering analysis. *Biometrics* 2012;68:805-814.
- 21 Amirkia VD, Qiubao P: The Antimicrobial Index: a comprehensive literature-based antimicrobial database and reference work. *Bioinformatics* 2011;5:365-366.
- 22 Zhang B, Horvath S: A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005;4:Article17.
- 23 Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- 24 Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C: The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. 2017;45:D362-d368.
- 25 Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, Lempicki RA: DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* 2007;35:W169-175.

- 26 Pison L, Tilz R, Jalife J, Haissaguerre M: Pulmonary vein triggers, focal sources, rotors and atrial cardiomyopathy: implications for the choice of the most effective ablation therapy. *J Intern Med* 2016;279:449-456.
- 27 Romero J, Gianni C, Natale A, Di Biase L: What Is the Appropriate Lesion Set for Ablation in Patients with Persistent Atrial Fibrillation? *Curr Treat Options Cardiovasc Med* 2017;19:35.
- 28 Sardu C, Santamaria M, Paolisso G, Marfella R: microRNA expression changes after atrial fibrillation catheter ablation. *Pharmacogenomics* 2015;16:1863-1877.
- 29 Rinne S, Kiper AK, Schmidt C, Ortiz-Bonnin B, Zwiener S, Seebohm G, Decher N: Stress-Kinase Regulation of TASK-1 and TASK-3. *Cell Physiol Biochem* 2017;44:1024-1037.
- 30 Petric S, Klein S, Dannenberg L, Lahres T, Clasen L, Schmidt KG, Ding Z, Donner BC: Pannexin-1 Deficient Mice Have an Increased Susceptibility for Atrial Fibrillation and Show a QT-Prolongation Phenotype. *Cell Physiol Biochem* 2016;38:487-501.
- 31 Caliskan E, Cox JL, Holmes DR, Jr., Meier B, Lakkireddy DR, Falk V, Salzberg SP, Emmert MY: Interventional and surgical occlusion of the left atrial appendage. *Nat Rev Cardiol* 2017;14:727-743.
- 32 Yeh YH, Kuo CT, Lee YS, Lin YM, Nattel S, Tsai FC, Chen WJ: Region-specific gene expression profiles in the left atria of patients with valvular atrial fibrillation. *Heart Rhythm* 2013;10:383-391.
- 33 Husser D, Buttner P, Ueberham L, Dinov B, Sommer P, Arya A, Hindricks G, Bollmann A: Genomic Contributors to Rhythm Outcome of Atrial Fibrillation Catheter Ablation - Pathway Enrichment Analysis of GWAS Data. *PLoS One* 2016;11:e0167008.
- 34 Abd El-Aziz TA, Mohamed RH, Mohamed RH, Pasha HF: Leptin, leptin gene and leptin receptor gene polymorphism in heart failure with preserved ejection fraction. *Heart Vessels* 2012;27:271-279.
- 35 Li SY, Yang X, Ceylan-Isik AF, Du M, Sreejayan N, Ren J: Cardiac contractile dysfunction in Lep/Lep obesity is accompanied by NADPH oxidase activation, oxidative modification of sarco(endo)plasmic reticulum Ca²⁺-ATPase and myosin heavy chain isozyme switch. *Diabetologia* 2006;49:1434-1446.
- 36 Zhang S, Zhang M, Goldstein S, Li Y, Ge J, He B, Ruiz G: The effect of c-fos on acute myocardial infarction and the significance of metoprolol intervention in a rat model. *Cell Biochem Biophys* 2013;65:249-255.
- 37 Drosos I, Chalikias G, Pavlaki M, Kareli D, Epitropou G, Bougioukas G, Mikroulis D, Konstantinou F, Giatromanolaki A, Ritis K, Munzel T, Tziakas D, Konstantinides S, Schafer K: Differences between perivascular adipose tissue surrounding the heart and the internal mammary artery: possible role for the leptin-inflammation-fibrosis-hypoxia axis. *Clin Res Cardiol* 2016;105:887-900.
- 38 Soh BS, Ng SY, Wu H, Buac K, Park JH, Lian X, Xu J, Foo KS, Felldin U, He X, Nichane M, Yang H, Bu L, Li RA, Lim B, Chien KR: Endothelin-1 supports clonal derivation and expansion of cardiovascular progenitors derived from human embryonic stem cells. *Nat Commun* 2016;7:10774.
- 39 Hathaway CK, Grant R, Hagaman JR, Hiller S, Li F, Xu L, Chang AS, Madden VJ, Bagnell CR, Rojas M, Kim HS, Wu B, Zhou B, Smithies O, Kakoki M: Endothelin-1 critically influences cardiac function via superoxide-MMP9 cascade. *Proc Natl Acad Sci U S A* 2015;112:5141-5146.
- 40 Mitchell JD, Maguire JJ, Kuc RE, Davenport AP: Expression and vasoconstrictor function of anorexigenic peptides neuromedin U-25 and S in the human cardiovascular system. *Cardiovasc Res* 2009;81:353-361.
- 41 Brighton PJ, Szekeres PG, Willars GB: Neuromedin U and its receptors: structure, function, and physiological roles. *Pharmacol Rev* 2004;56:231-248.