Identification of Potential Crucial Genes in Atrial Fibrillation: A Bioinformatic Analysis

Junguo Zhang
Guangdong Second Provincial General Hospital
Xin Huang
Guangdong Second Provincial General Hospital
Xiaojie Wang
Guangdong Second Provincial General Hospital
Yanhui Gao
Guangdong Pharmaceutical University
Li Liu
Guangdong Pharmaceutical University
Ziyi Li
Guangdong Second Provincial General Hospital
Xuejiao Chen
Guangdong Second Provincial General Hospital
Jie Zeng
Guangdong Second Provincial General Hospital
Zebing Ye
Guangdong Second Provincial General Hospital
Guowei Li (lig28@mcmaster.ca)

Research article

Keywords: Potential crucial genes, Atrial fibrillation, Bioinformatic gene analysis, Gene expression omnibus, Biomarkers

Posted Date: February 28th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-15436/v1

Version of Record: A version of this preprint was published on July 18th, 2020. See the published version at https://doi.org/10.1186/s12920-020-00754-5.

Abstract

Background Atrial fibrillation (AF) is clearly heritable, affecting 2-3% of the population in Europe and the USA. However, a substantial proportion of heritability is still lacking. In the present study, we aim to identify potential crucial genes associated with AF through bioinformatic analyses of public datasets.

Methods Microarray data sets of GSE115574, GSE31821, GSE79768, GSE41177 and GSE14975 from the Gene Expression Omnibus (GEO) database were enrolled. After merging all microarray data and adjusted batch effect, differentially expressed genes (DEGs) were identified. Functional enrichment analyses based on Gene Ontology (GO) resource, Kyoto Encyclopedia of Genes and Genomes (KEGG) resource, Gene Set Enrichment Analysis (GSEA), Reactome Pathway Database and Disease Ontology (DO) were carried out for DEGs. Protein-protein interaction (PPI) network was constructed using the STRING database. Combined with aforementioned significant bioinformatics information, potential crucial genes were subsequently selected. The potential crucial genes coupled with corresponding predicted microRNAs involved in AF were then assessed.

Result We identified 27 of DEGs with gene expression fold change (FC) ≥ 1.5 and 5 with FC ≥ 2 of AF patients compared with sinus rhythm controls. The most significantly enriched pathway was regulation of insulin-like growth factor transport and uptake by insulin-like growth factor binding proteins. IGFBP2, C1orf105, FHL2, CHGB, ATP1B4, IGFBP3, SLC26A9, CXCR4 and HTR2B were considered the potential crucial genes. Sixteen corresponding predicted microRNAs, of which 5 targeting IGFBP3 and 8 FHL2, might be associated with AF. The comparative toxicogenomics database (CTD) database showed CXCR4, IGFBP2, IGFBP3 and FHL2 had higher scores with AF.

Conclusions The 9 potential crucial genes, especially CXCR4, IGFBP2, IGFBP3 and FHL2, may be associated with risk of AF. MicroRNAs targeting IGFBP3 and FHL2 may be potential biomarkers or therapeutic targets for AF. Our study provided new insights of AF into genetics, molecular pathogenesis and new therapeutic targets.

Background

Atrial fibrillation (AF) is the most common sustained arrhythmia and is one of the major causes of stroke, heart failure, sudden death, and cardiovascular morbidity in the world [1]. The estimated number of AF patients is 34 million worldwide, and gradually increasing with the aging of the population [2]. However, the pathophysiologic mechanism underlying AF remains unclear, resulting in a lack of effective treatment [3]. Only a small number of AF patients can normalize heart rhythm by catheter ablation or cardiac surgery [4]. The higher prevalence and limited treatments of AF lead to substantial public health and economic burdens [5]. Therefore, it is necessary to improve our understanding of AF pathogenesis and to develop screening method for AF.

Epidemiological evidence over the last decade has identified that metabolic syndrome, pre-hypertension, obesity, obstructive sleep apnea, exercise, and dietary intake of stimulants are major risk factors for AF [4,

6]. However, only a small proportion of exposed individuals eventually developed AF, suggesting that a strong genetic component might be a risk factor contributing to the susceptibility of AF. From 2007 to 2017, genetic researches in European, Asian, and African-American ancestry groups have reported 17 independent signals at 14 genomic regions, such as PITX2, ZFHX3, and PRRX1, associated with AF [7, 8]. In 2018, a preliminary genome-wide association study meta-analysis including over 93,000 AF cases and more than 1 million referents identified at least 134 genetic loci significantly associated with risk of AF [9]. However, combined with the results of the genetic studies thus far, genetic variation only accounts for 42% of the heritability for AF [9]. Notably, a substantial proportion of heritability is still lacking. One potential interpretation is that unidentified gene may partially contribute to the missing heritability. Therefore, there are still many related genes to be identified, which will help us better understand the pathogenesis of AF and facilitate the discovery of novel diagnostic biomarkers or therapeutic target.

In this study, we aimed to identify the potential crucial genes for AF through Gene Expression Omnibus (GEO) database using bioinformatic methods and to analyze their expression, function and interaction.

Methods

Atrial fibrillation datasets

Raw files of 5 enrolled microarray data sets, including GSE115574, GSE31821, GSE79768, GSE41177 and GSE14975 (Table 1), were downloaded from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/). All of these datasets were obtained from the microarray platform of Affymetrix Human Genome U133 Plus 2.0 Array [HG-U133_Plus_2]. In each data set, only human left atrial appendage tissue (LAA) samples and sinus rhythm (SR) samples were selected, and finally 46 AF and 31 SR samples were included for subsequent analyses.

Data Preprocessing

Series matrix files were processed with ActivePerl 5.24.2 software (https://www.activestate.com/products/perl/) to convert the gene probe IDs to gene symbol codes. Because GSE14975 was extracted using the Affymetrix Microarray Suite 5 (MAS5) algorithm, it was log base 2 transformed. After merging all microarray data, batch effects were adjusted by the “combat” function of “sva” package of R software using empirical Bayes frameworks [10]. Finally, normalized expression values according to the “normalizeBetweenArrays” function of the package of “limma” in R software so that the expression values have similar distribution across a set of arrays [11].

Identification of differentially expressed genes (DEGs)

To assess differential expression, using the “limma” package of R software, a linear model was fitted and a simple empirical Bayes model was used to moderate standard errors [11]. A moderated t-statistic and a log-odds of differential expression was computed for each contrast for each gene. The Benjamini and Hochberg (BH) method was performed to adjust P value to reduce the false positive error. A gene was

defined as a DEG between the AF and SR sample, when the adjust P value was < 0.05 and the gene expression fold change (FC) value was > 1.5 (|log2 FC| ≥ 0.58), which were visualized as Volcano plots and heat map plots. “Ggplot” packages of R software was applied to generate box plots for genes which have the adjust P value < 0.05 and the gene expression FC value > 2 (log2 FC| ≥ 1).

Functional and pathway enrichment analyses of DEGs

The Gene Ontology (GO) Resource (http://geneontology.org/) is a bioinformatics tool providing a framework and set of concepts for describing the functions of gene products from all organisms [12]. Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.kegg.jp/) is a database resource integrated the information of genomes, biological pathways, diseases and chemicals [13]. Gene Set Enrichment Analysis (GSEA) (http://software.broadinstitute.org/gsea/index.jsp) is a computational method for interpreting gene expression data based on molecular signature database [14]. The Reactome Pathway Database (https://reactome.org/) is pathway annotation database collecting the biological pathways and processes in the human [15]. The Disease Ontology (DO) (http://disease-ontology.org) represents a comprehensive knowledge base of 8043 inherited, developmental and acquired human disease [16]. Before the preformation of enrichment analysis, human genome annotation package “org.Hs.eg.db” was used to convert gene symbol codes to Entrez ID. To better understand the biological function and characteristics, R software was used to perform enrichment analyse, with the “clusterProfiler” package for Go and KEGG enrichment analyses, the “GSEABase” package for GSEA analysis, the “ReactomePA” package for Reactome pathway analysis and the “DOSE” package for DO enrichment analysis. The “ggplot2”, “pathview” and “graphite” packages of R software were used to visualize the plots. GO terms and KEGG maps of biological functions associated with an adjusted P value < 0.05 and Q value < 0.05 was considered to be significantly enriched.

Protein-protein interaction (PPI) network and potential crucial genes analyses.

The STRING database (http:/ /string-db.org/) was performed to construct a PPI network to reveal the generic organization principles of functional cell systems and to predict protein-protein interactions [17]. The Molecular Complex Detection (MCODE) of Cytoscape was carried out to module analyze and visualize the result of PPI network. Default parameters (degree cutoff ≥2, node score cutoff ≥ 2, K-core ≥ 2, and maxi-mum depth =100) were used. To select potential crucial genes, we synthesized above significant bioinformatics information for subsequent analyses.

Potential microRNA targeting and international analyses

To predict which of the microRNAs could target potential crucial genes, online prediction tools including microRNA Data Integration Portal (mirDIP, http://ophid.utoronto.ca/mirDIP), miRDB (http://mirdb.org/), TargetScan (v7.1, http://www.targetscan.org/vert_71/), and DIANA Tools (http://diana.imis.athena-innovation.gr/DianaTools/) were used. Candidate microRNAs which had targeting evidence in more than 3 bioinformatics tools were selected. Subsequently, Diana-miRPath (v3.0;

http://www.microrna.gr/miRPathv3) was applied to interactions between candidate microRNAs and potential crucial genes involved in AF [18].

Identification of potential crucial genes associated with AF

The comparative toxicogenomics database (CTD, http://ctdbase.org/) integrated information including chemical-gene/protein interactions, chemical-disease and gene-disease relationships to develop hypotheses the mechanisms of disease [19]. The association between potential crucial genes and AF risk was analyzed using the data in CTD.

Results

Identification of DEGs

Gene expression levels of merged GEO series that have been adjusted batch effects were standardized and the results of pre- and post- standardized were presented in Supplementary Figure 1 and 2. The 54675 probes corresponding to 21654 genes in GSE115574, GSE31821, GSE79768, GSE41177 and GSE14975 datasets were identified and DEGs of AF were confirmed. Twenty-seven of DEGs with |log2 FC| ≥ 0.58 in LAA samples of AF patients compared with SR was identified, including 19 up-regulated genes and 8 down-regulated genes (Supplementary Table 1). Volcano plot and Heatmap plot of 27 DEGs enrolled in subsequent analyses was showed in Figure 1 and Supplementary Figure 3. Using a screening criteria of |log2 FC| ≥ 1, there were 5 genes identified, with 4 of these genes being up-regulated and 1 down-regulated (Table 2). Boxplots for the 5 selected genes were shown in Figure 2.

Functional enrichment analyses of DEGs

To further investigate the biological functions of the 27 DEGs, functional enrichment analyses were performed and results were shown in Table 3. The result of molecular function in GO revealed that two up-regulated DEGs (*IGFBP2* and *IGFBP3*) were enriched in insulin-like growth factor I binding process (adjusted P value = 0.0168 and Q value = 0.0120) and insulin-like growth factor binding process (adjusted P value = 0.0432 and Q value = 0.0310). Using screening criteria of adjusted P value < 0.05 and Q value < 0.05, no pathway was enriched in KEGG. The ‘mineral absorption’, ‘calcium signaling pathway’ and ‘proximal tubule bicarbonate reclamation’ pathways were enriched (P values = 0.0033, 0.0331 and 0.0343 respectively). Specifically, *SLC26A9* and *ATP1B4* genes were enriched in mineral absorption, while two up-regulated genes (*CXCR4* and *HTR2B*) were correlated with calcium signaling pathway. Only up-regulated *ATP1B4* gene was enriched in proximal tubule bicarbonate reclamation pathway. However, these enrichments did not remain significant after multiplicity adjustment by BH.

Pathway enrichment using the REACTOME database identified that three up-regulated DEGs (*IGFBP2*, *IGFBP3* and *CHGB*) were enriched in Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs) (P value = 0.0008 and Q value < 0.0435, Supplementary Figure 4). DO enrichment analysis revealed that DEGs were enriched in 29 biological

processes (adjusted P value < 0.05 and Q value < 0.05), but not in AF associated process (Supplementary Figure 5). The result of GSEA analysis showed that no gene was enriched under the specific cutoff of P value.

PPI network construction and potential crucial genes selection

Using the STRING platform, PPI analysis of these DEGs identified 18 nodes and 26 interactions. In addition, one significant module with 5 nodes and 9 edges was screened out via MCODE (Supplementary Figure 6 and 7). *CXCR4*, *IGFBP2*, *IGFBP3*, *SNAI2* and *ANGPTL2* were hub nodes in module. Only *CXCR4*, *IGFBP2* and *IGFBP3* were selected for hub genes, all of which were involved in playing pivotal regulatory roles in PPI network, due to the high degree of connectivity (degree ≥ 5, Supplementary Figure 7). Furthermore, after combining with the results of differential expression, enrichment analyses and PPI, *IGFBP2*, *IGFBP3*, *CHGB*, *CXCR4*, *HTR2B*, *FHL2*, *C1orf105*, *ATP1B4* and *SLC26A9* were considered potential crucial genes for further analyses.

Identification of functional and pathway enrichment among predicted microRNAs and potential crucial genes

Bioinformatic prediction tools including mirDIP, miRDB, TargetScan, and DIANA were used to identify microRNAs targeting potential crucial genes involved in AF and these data were displayed in Table 4. There were 16 microRNAs met the inclusion criteria, of which 5 targeting *IGFBP3* and 8 *FHL2*.

In order to understand how predicted microRNAs are related to AF, functional and pathway enrichment analyses were performed using Diana-miRPath. GO analysis revealed hsa-miR-197-3p, hsa-miR-19a-3p, hsa-miR-19b-3p, hsa-miR-340-5p and hsa-miR-9-5p targeting *IGFBP3* were mainly enriched in insulin-like growth factor binding protein complex, protein tyrosine phosphatase activator activity, to mention a few. Hsa-miR-25-3p, hsa-miR-32-5p, hsa-miR-363-3p, hsa-miR-367-3p, hsa-miR-4325, hsa-miR-661, hsa-miR-92a-3p and hsa-miR-92b-3p targeting *FHL2* were mainly involved in atrial cardiac muscle cell development, ventricular cardiac muscle cell development and heart trabecula formation, among others. KEGG analysis result revealed that microRNAs targeting *IGFBP3* were enriched in p53 signaling pathway and transcriptional misregulation in cancer, while microRNAs targeting *FHL2* were involved in osteoclast differentiation.

Identification of potential crucial genes associated with AF

CTD database was employed to explore the interaction between potential crucial genes and AF. As shown in Supplementary Figure 8, potential crucial genes targeting AF, left ventricular dysfunction, heart diseases and cardiovascular diseases. Inference scores in CTD reflected the association between chemical, disease and genes. The interaction results showed that *CXCR4*, *IGFBP2*, *IGFBP3* and *FHL2* have a higher score with AF.

Discussion

In the present study, we integrated gene expression profiles of 46 AF samples and 31 SR samples from 5 GEO datasets and analyzed the data using bioinformatics tools. A total of 27 DEGs with |log2 FC| ≥ 0.58 and 5 with |log2 FC| ≥ 1 in AF compared with SR samples were selected. Furthermore, 9 potential crucial genes (IGFBP2, IGFBP3, CHGB, CXCR4, HTR2B, FHL2, C1orf105, ATP1B4 and SLC26A9), several important pathways and 16 microRNAs, of which 5 targeting IGFBP3 and 8 FHL2, which were associated with AF risk, were identified, suggesting these may play important role in the mechanism of AF.

IGFBP2, located in chromosome 2q35, which encoded the second most abundant circulating IGF-binding protein. IGFBP2, secreted by differentiating white adipocytes, regulate the functions of IGFs [20]. In the Framingham heart study, IGFBP2 was significantly associated with all-cause mortality [21]. Our study revealed that the IGFBP2 expression level was up-regulated in AF samples compared to SR samples. The enrichment analyses result of IGFBP2 was all correlated with IGFBP3.

IGFBP3 is known to have a function gene, encoded the primary carrier of IGFs (IGFBP3) in the circulation. IGFBP3 was found to be related to involve in oxidative stress, atherosclerosis and left ventricular hypertrophy [22, 23]. More important, previous study revealed low IGFBP3 serum level was an independent determinant of AF [22]. Enrichment analyses in the present study indicated that IGFBP2 and IGFBP3 were enriched in GO term of insulin-like growth factor binding and pathway of IGF transport and uptake by IGFBPs. Thus, IGFBP2 and IGFBP3 might bind and regulate the functions of IGFs through IGF transport and uptake by IGFBPs pathway, and then affected the susceptibility of AF [20]. To date, microRNAs have been regarded as possible targets for gene therapy in AF [24]. In the study, hsa-miR-197-3p, hsa-miR-19a-3p, hsa-miR-19b-3p, hsa-miR-340-5p and hsa-miR-9-5p targeting IGFBP3 were predicted and might be potential biomarkers of AF. Previous research showed that hsa-miR-19a-3p/19b-3p had lower expression in heart failure patients than in normal, suggesting these inhibited autophagy-mediated fibrogenesis which also related with AF [25]. Expression charts of microRNAs revealed miR-9-5p as a direct regulator may prevent atherosclerosis [26].

Our study revealed that CHGB was involved in IGF transport and uptake by IGFBPs pathway including IGFBP2 and IGFBP3, and the gene expression of CHGB was higher in AF patients. Chromogranin B (CHGB) was an emerging cardiovascular biomarker, which was encoded by CHGB [27]. CHGB can regulate B-type natriuretic peptide (BNP) production through polycystin 2(PC2)-CGB-BNP signaling axis in cardiomyocyte, and integrate information from myocardial stress and neuro-endocrine activation [27, 28]. CHGB level was significantly increased in heart failure patients [27]. These findings suggested that IGF transport and uptake by IGFBPs pathway with IGFBP2, IGFBP3 and CHGB may participate in the occurrence and development of AF. A more thorough understanding of IGF transport and uptake by IGFBPs pathway as well as its microRNAs and proteins in AF is necessary.

With PPI analysis, IGFBP2, IGFBP3 and CXCR4 were divided into a group according to protein-protein interactions. There was evidence that CXCR4 was overexpressed in chronic AF patients, and might contribute to the process of AF through regulating atrial fibrosis and structural remodeling [29]. In the present study, CXCR4 was also found to be potential crucial gene related to AF. KEGG pathway

enrichment analysis showed that the CXCR4 and HTR2B were enriched in calcium signaling pathway, which had been extensively characterized in the role in cardiac hypertrophy and remodeling processes [30]. HTR2B is located in chromosome 2q37.1. 5-HT_{2B} (5-hydroxytryptamine receptor 2B) receptor coded by HTR2B, is presented in the cardiovascular system, and may indirectly produce life-threatening arrhythmias and cardiodepression [31, 32]. In auricular myocytes of newborn rat, the activation of 5-HT_{2B} enhances gap junctional intercellular communication (GJIC) in a receptor subtype-specific manner, and prolongs 5-HT exposure to alter the Cx expression pattern which associated with AF [32].

Increasing evidence demonstrated that FHL2 and its proteins products had a function in cardiovascular disease [33, 34]. FHL2, located at the sarcomere, interacted with extracellular signal regulated kinase (ERK) and regulated cardiac growth, suggesting FHL2 a protective role in adrenergic-mediated cardiac hypertrophy [35, 36]. We found that FHL2 was up-regulated in AF samples compared to SR in this study. Predicted microRNAs targeting FHL2 were mainly enriched in GO terms of atrial cardiac muscle cell development, ventricular cardiac muscle cell development and heart trabecula formation. The expressions of hsa-miR-25-3p and hsa-miR-92a-3p in patients with cardiovascular disease were lower than controls [37, 38]. These results indicated that FHL2 and its microRNAs might be potential biomarkers of AF.

In this study, C1orf105 and ATP1B4 had 2 folds gene expression in AF patients than SR control. SLC26A9 and ATP1B4 were enriched in KEGG pathways of mineral absorption and proximal tubule bicarbonate reclamation. Previous study showed that a SNP on C1orf105 was associated with remodeling response to atherosclerosis [39]. Slc26a9 encoded transporters with diverse functional attributes and RT-PCR showed that Slc26a9 is detectable in heart [40]. Above evidence revealed that C1orf105, SLC26A9 and ATP1B4 were related with cardiovascular disease and might have a function in AF.

In current study, we have discussed that 9 potential crucial genes is involved in the occurrence and development of AF, suggesting these genes may server as potential biomarkers and therapeutic targets for AF. However, the limitations of this study should be considered. Firstly, it is difficult to consider some important factors such as regions, races and age. Considering that the development of AF resulted from various environmental and genetic factors, some unmeasured factors including region, family history and risk factors of AF should be evaluated in the further research. In addition, the potential crucial genes need further validation by RT-qPCR in clinical samples. Finally, the mechanisms of genes were not clear. More evidences is required to find out the biological foundation.

Conclusion

Our study integrated data with relative larger sample size from multiple GEO datasets and identified 9 potential crucial genes (IGFBP2, IGFBP3, CHGB, CXCR4, HTR2B, FHL2, C1orf105, ATP1B4 and SLC26A9), pathways and microRNAs using bioinformatic analyses. The exploration of potential crucial genes of AF may provide some potential aid in further identification of new biomarkers and targets for the susceptible of AF.

Abbreviations

AF Atrial Fibrillation

DEGs Differentially Expressed Genes

GEO Gene Expression Omnibus

FC Fold Change

LAA Left Atrial Appendage Tissue

SR Sinus Rhythm

MAS5 Affymetrix Microarray Suite 5

BH Benjamini and Hochberg method

GO Gene Ontology

KEGG Kyoto Encyclopedia of Genes and Genomes

GSEA Gene Set Enrichment Analysis

DO Disease Ontology

PPI Protein-Protein Interaction

MCODE Molecular Complex Detection

CTD Comparative Toxicogenomics Database

IGF Insulin-like Growth Factor

IGFBPs Insulin-like Growth Factor Binding Proteins

Declarations

Acknowledgments

We acknowledge the research grant from the Science Foundation of Guangdong Second Provincial General Hospital.

Funding

This study was funded by research grant from the Science Foundation of Guangdong Second Provincial General Hospital (YY2018-002)

Availability of data and materials

Microarray datasets (GSE115574, GSE31821, GSE79768, GSE41177 and GSE14975) for this study are openly available in Gene Expression Omnibus database at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115574, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31821, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79768, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41177 and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14975, respectively (last accessed on 27 Dec 2019).

Authors' Contributions

JZ and GL were responsible for study conception, design of the study, data acquisition, and analysis and interpretation of results. JZ, XH and ZY were responsible for data acquisition. XW, YG, LL took part in the discussion of the paper. JZ wrote the manuscript that was reviewed and revised by GL, ZY, XH, XW, YG, LL, ZL, XC and JZ. All authors have read and approved the manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests, and all authors should confirm its accuracy.

References

1. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B *et al*: 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Europace. 2016;18:1609-78.
2. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ *et al*: Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation. 2014;129:837-47.
3. Schotten U, Verheule S, Kirchhof P, Goette A: Pathophysiological mechanisms of atrial fibrillation: a translational appraisal. Physiol Rev. 2011;91:265-325.
4. Lau DH, Schotten U, Mahajan R, Antic NA, Hatem SN, Pathak RK *et al*: Novel mechanisms in the pathogenesis of atrial fibrillation: practical applications. Eur Heart J. 2016;37:1573-81.

5. Kirchhof P: The future of atrial fibrillation management: integrated care and stratified therapy. Lancet. 2017;390:1873-87.
6. Turagam MK, Velagapudi P, Kocheril AG, Alpert MA: Commonly Consumed Beverages in Daily Life: Do They Cause Atrial Fibrillation? Clinical Cardiology. 2015;38:317-22.
7. Ellinor PT, Lunetta KL, Albert CM, Glazer NL, Ritchie MD, Smith AV *et al*: Meta-analysis identifies six new susceptibility loci for atrial fibrillation. Nat Genet. 2012;44:670-5.
8. Tucker NR, Ellinor PT: Emerging directions in the genetics of atrial fibrillation. Circulation Research. 2014;114:1469-82.
9. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM *et al*: Multi-ethnic genome-wide association study for atrial fibrillation. Nat Genet. 2018;50:1225-33.
10. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882-3.
11. Diboun I, Wernisch L, Orengo CA, Koltzenburg M: Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. Bmc Genomics. 2006;7:252.
12. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM *et al*: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25-9.
13. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M: New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019;47:D590-D5.
14. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-50.
15. Joshi-Tope G, Vastrik I, Gopinath GR, Matthews L, Schmidt E, Gillespie M *et al*: The Genome Knowledgebase: a resource for biologists and bioinformaticists. Cold Spring Harb Symp Quant Biol. 2003;68:237-43.
16. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L *et al*: Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Research. 2019;47:D955-D62.
17. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J *et al*: STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research. 2019;47:D607-D13.
18. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M *et al*: DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res. 2013;41:W169-73.
19. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J *et al*: The Comparative Toxicogenomics Database: update 2019. Nucleic Acids Res. 2019;47:D948-D54.
20. Dunger D, Yuen K, Ong K: Insulin-like growth factor I and impaired glucose tolerance. Horm Res. 2004;62 Suppl 1:101-7.

21. Ho JE, Lyass A, Courchesne P, Chen G, Liu C, Yin X *et al*: Protein Biomarkers of Cardiovascular Disease and Mortality in the Community. J Am Heart Assoc. 2018;7.
22. Duron E, Vidal JS, Funalot B, Brunel N, Viollet C, Seux ML *et al*: Insulin-like growth factor I, insulin-like growth factor binding protein 3, and atrial fibrillation in the elderly. The journals of gerontology Series A, Biological sciences and medical sciences. 2014;69:1025-32.
23. Friedrich N, Aumann N, D Rr M, Felix SB, Nauck M, Wallaschofski H *et al*: Lack of association between insulin-like growth factor-1 or insulin-like growth factor-binding protein-3 and left ventricular hypertrophy: results of the Study of Health in Pomerania. Journal of Hypertension. 2010;28:856-64.
24. Hucker WJ, Hanley A, Ellinor PT: Improving Atrial Fibrillation Therapy: Is There a Gene for That? J Am Coll Cardiol. 2017;69:2088-95.
25. Zou M, Wang F, Gao R, Wu J, Ou Y, Chen X *et al*: Autophagy inhibition of hsa-miR-19a-3p/19b-3p by targeting TGF-β R II during TGF-β1-induced fibrogenesis in human cardiac fibroblasts. Scientific Reports. 2016;6:24747.
26. D'Amore S, Hardfeldt J, Cariello M, Graziano G, Copetti M, Di Tullio G *et al*: Identification of miR-9-5p as direct regulator of ABCA1 and HDL-driven reverse cholesterol transport in circulating CD14+ cells of patients with metabolic syndrome. Cardiovasc Res. 2018;114:1154-64.
27. Heidrich FM, Melz C, Buechau MS, Pfluecke C, Quick S, Speiser U *et al*: Regulation of circulating chromogranin B levels in heart failure. Biomarkers. 2018;23:78-87.
28. Heidrich FM, Zhang K, Estrada M, Huang Y, Giordano FJ, Ehrlich BE: Chromogranin B regulates calcium signaling, nuclear factor kappaB activity, and brain natriuretic peptide production in cardiomyocytes. Circ Res. 2008;102:1230-8.
29. Wang XX, Zhang FR, Zhu JH, Xie XD, Chen JZ: Up-Regulation of CXC Chemokine Receptor 4 Expression in Chronic Atrial Fibrillation Patients with Mitral Valve Disease May Be Attenuated by Renin–Angiotensin System Blockers. Journal of International Medical Research. 2009;37:1145-51.
30. Dewenter M, von der Lieth A, Katus HA, Backs J: Calcium Signaling and Transcriptional Regulation in Cardiomyocytes. Circ Res. 2017;121:1000-20.
31. Kaumann AJ, Levy FO: 5-hydroxytryptamine receptors in the human cardiovascular system. Pharmacol Ther. 2006;111:674-706.
32. Derangeon M, Bozon V, Defamie N, Peineau N, Bourmeyster N, Sarrouilhe D *et al*: 5-HT4 and 5-HT2 receptors antagonistically influence gap junctional coupling between rat auricular myocytes. J Mol Cell Cardiol. 2010;48:220-9.
33. Radke MH, Polack C, Methawasin M, Fink C, Granzier HL, Gotthardt M: Deleting Full Length Titin Versus the Titin M-Band Region Leads to Differential Mechanosignaling and Cardiac Phenotypes. Circulation. 2019;139:1813-27.
34. Liang Y, Bradford WH, Zhang J, Sheikh F: Four and a half LIM domain protein signaling and cardiomyopathy. Biophys Rev. 2018;10:1073-85.
35. Purcell NH, Darwis D, Bueno OF, Muller JM, Schule R, Molkentin JD: Extracellular Signal-Regulated Kinase 2 Interacts with and Is Negatively Regulated by the LIM-Only Protein FHL2 in Cardiomyocytes.

Molecular & Cellular Biology. 2004;24:1081-95.

36. Lorenz K, Schmitt JP, Vidal M, Lohse MJ: Cardiac hypertrophy: targeting Raf/MEK/ERK1/2-signaling. The international journal of biochemistry & cell biology. 2009;41:2351-5.
37. Gumus G, Giray D, Bobusoglu O, Tamer L, Karpuz D, Hallioglu O: MicroRNA values in children with rheumatic carditis: a preliminary study. Rheumatol Int. 2018;38:1199-205.
38. Bulent VM, Kalayci YA, Akdemir R, Tarik AM, Akif CM, Aksoy M *et al*: Altered Plasma MicroRNA Expression in Patients with Mitral Chordae Tendineae Rupture. J Heart Valve Dis. 2016;25:580-8.
39. Harrison SC, Zabaneh D, Asselbergs FW, Drenos F, Jones GT, Shah S *et al*: A gene-centric study of common carotid artery remodelling. Atherosclerosis. 2013;226:440-6.
40. Chang MH, Plata C, Zandi-Nejad K, Sindic A, Sussman CR, Mercado A *et al*: Slc26a9–anion exchanger, channel and Na+ transporter. The Journal of membrane biology. 2009;228:125-40.

Tables

Table 1 Characteristics of datasets in this study

GSE series	Platform	Total	AF	SR	Country	Contributors
GSE115574	GPL570	29	14	15	Turkey	Deniz GC et.al
GSE31821	GPL570	6	4	2	France	Morel E et.al
GSE79768	GPL570	13	7	6	Taiwan	Tsai F et.al
GSE41177	GPL570	19	16	3	Taiwan	Yeh Y et.al
GSE14975	GPL570	10	5	5	Germany	Adam O et.al

GSE, Gene Expression Omnibus; AF, Atrial Fibrillation; SR, Sinus Rhythm

Table 2 The DEGs of merged data set with the use of criteria of adjust P value <0.05 and |log2FC|≥1

Gene	Log2FC	AveExpr	t	P value	adjust P value	B
IGFBP2	1.1746	9.5881	-6.0548	<0.0001	0.0006	8.0172
C1orf105	-1.0560	7.2935	5.6106	<0.0001	0.0011	6.3668
FHL2	1.1255	9.9882	-5.4755	<0.0001	0.0016	5.8742
CHGB	1.0836	8.5851	-5.3090	<0.0001	0.0025	5.2745
ATP1B4	1.0469	5.0220	-4.3592	<0.0001	0.0283	2.0308

Log2FC, log2 Fold Change; AveExpr, Average Expression

Table 3 Significant enriched GO terms and pathways of DEGs

	Term	Count	Genes	P value	adjust P value	Q value
GO terms						
GO:0031994	insulin-like growth factor I binding	2	*IGFBP2/IGFBP3*	0.0001	0.0168	0.0121
GO:0005520	insulin-like growth factor binding	2	*IGFBP2/IGFBP3*	0.0006	0.0433	0.0311
KEGG Pathway						
hsa04978	Mineral absorption	2	*SLC26A9/ATP1B4*	0.0033	0.2088	0.2088
hsa04020	Calcium signaling pathway	2	*CXCR4/ HTR2B*	0.0332	0.2741	0.2741
hsa04964	Proximal tubule bicarbonate reclamation	1	*ATP1B4*	0.0343	0.2741	0.2741
Reactome Pathway						
R-HSA-381426	Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	3	*IGFBP2/ IGFBP3/ CHGB*	0.0008	0.0632	0.0435

GO, Gene Ontology; DEGs, Differentially Expressed Genes

Table 4 The GO terms and KEGG pathways enrichment among predicted miRNAs and potential crucial genes

Genes	Predicted miRNAs	Category		P value
IGFBP3	hsa-miR-197-3p	KEGG pathway	P53 signaling pathway	0.0000
	hsa-miR-19a-3p		Transcriptional misregulation in cancer	0.0201
	hsa-miR-19b-3p	GO terms	Insulin-like growth factor binding protein complex	0.0043
	hsa-miR-340-5p		Protein tyrosine phosphatase activator activity	0.0046
	hsa-miR-9-5p		Type B pancreatic cell proliferation	0.0051
			Insulin-like growth factor I binding	0.0051
			Negative regulation of smooth muscle cell migration	0.0051
			Positive regulation of insulin-like growth factor receptor signaling pathway	0.0051
			Positive regulation of myoblast differentiation	0.0055
			Regulation of glucose metabolic process	0.0061
			Fibronectin binding	0.0061
			Negative regulation of signal transduction	0.0061
			Negative regulation of smooth muscle cell proliferation	0.0069
			Insulin-like growth factor binding	0.0072
			Regulation of growth	0.0098
			Positive regulation of catalytic activity	0.0100
			Negative regulation of protein phosphorylation	0.0100
			Regulation of cell growth	0.0101
			Positive regulation of MAPK cascade	0.0126
			Osteoblast differentiation	0.0150
			Cellular protein metabolic process	0.0248
			Positive regulation of apoptotic process	0.0360
			Negative regulation of cell proliferation	0.0443
FHL2	hsa-miR-25-3p	KEGG pathway	Osteoclast differentiation	0.0040
	hsa-miR-32-5p	GO terms	Atrial cardiac muscle cell development	0.0069
	hsa-miR-363-3p		Ventricular cardiac muscle cell development	0.0119
	hsa-miR-367-3p		Heart trabecula formation	0.0119
	hsa-miR-4325		M band	0.0119
	hsa-miR-661		Androgen receptor signaling pathway	0.0119
	hsa-miR-92a-3p		Androgen receptor binding	0.0119
	hsa-miR-92b-3p		Cellular lipid metabolic process	0.0171
			Response to hormone	0.0171
			Z disc	0.0262
			Osteoblast differentiation	0.0262
			Actin cytoskeleton	0.0413

	Transcription coactivator activity	0.0413
	Protein binding transcription factor activity	0.0428

KEGG, Kyoto Encyclopedia of Genes and Genomes

Supplemental Information Note

Supplementary Figure 1. Data standardization. Pre-standardization gene expression levels of each data set are presented as blue boxplots.

Supplementary Figure 2. Data standardization. Post-standardization gene expression levels of each data set are presented as red boxplots.

Supplementary Figure 3. Heat map of DEGs in AF samples and SR samples. Each row represents a sample, and each column represents a single gene. Purple color represented AF samples, and blue color represented SR samples. The color scale shows the relative genes expression level in certain slide: green indicates low relative expression levels; red indicates high relative expression levels.

Supplementary Figure 4. Pathway view of Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs) using the REACTOME database. *IGFBP2*, *IGFBP3* and *CHGB* were enriched in the pathway.

Supplementary Figure 5. Bar plot of DO enrichment of DEGs.

Supplementary Figure 6. PPI network of the DEGs and modular analysis. Yellow nodes represent DEGs in the same module.

Supplementary Figure 7. Bar plot for the interaction numbers of each gene in PPI network.

Supplementary Figure 8. Relationship to AF diseases related to potential crucial genes based on the CTD database.

Figures

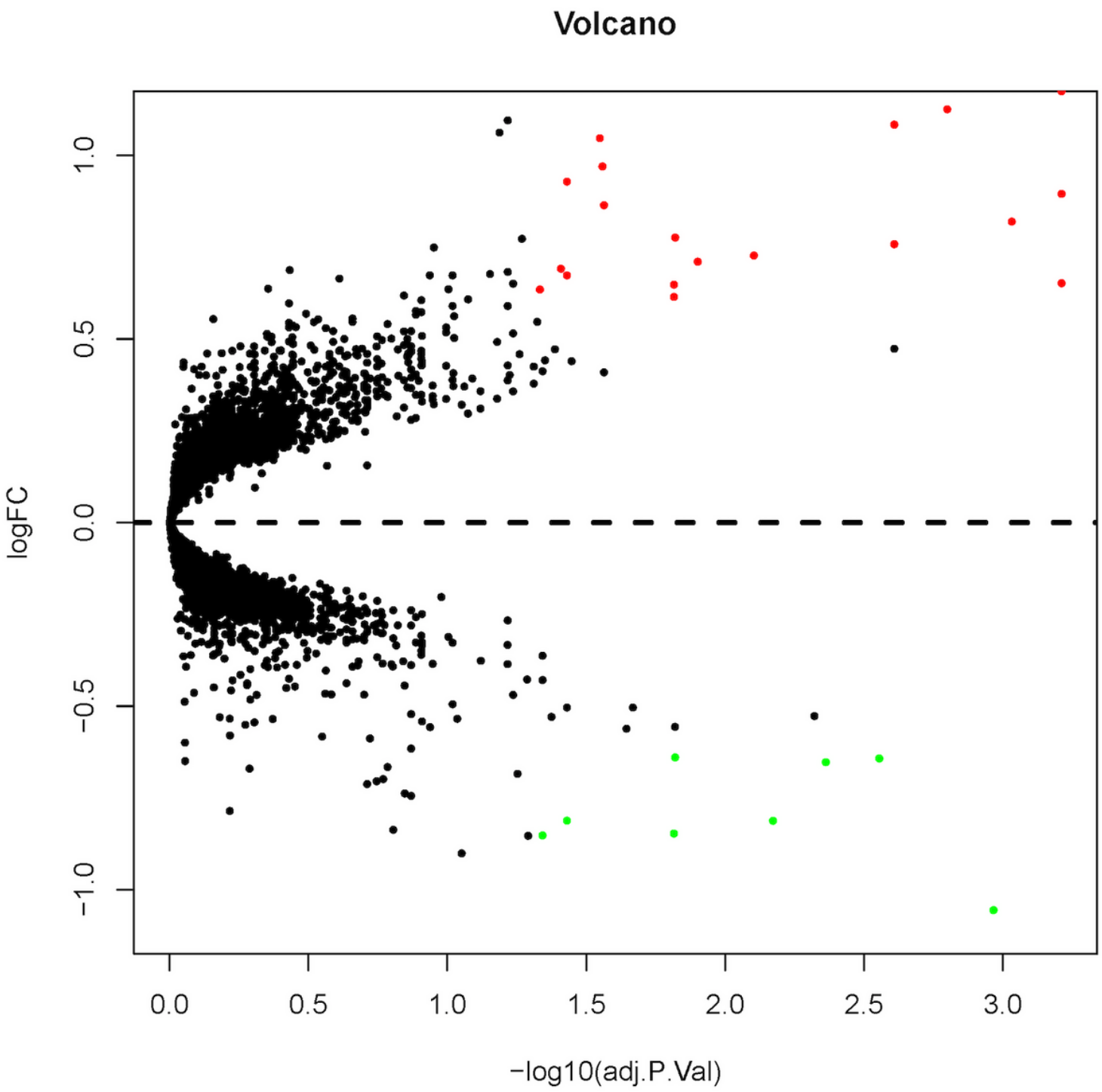

Figure 1

Volcano plot of DEGs in AF samples compared to SR samples. Red indicates the gene expression was up-regulated in AF samples compared to primary samples (adjust P value < 0.05 and |log2 FC| ≥ 0.58); Green indicates the gene expression was down-regulated in AF samples compared with primary samples (adjust P value < 0.05 and |log2 FC| ≥ 0.58); Black indicates the adjust P value was >0.05.

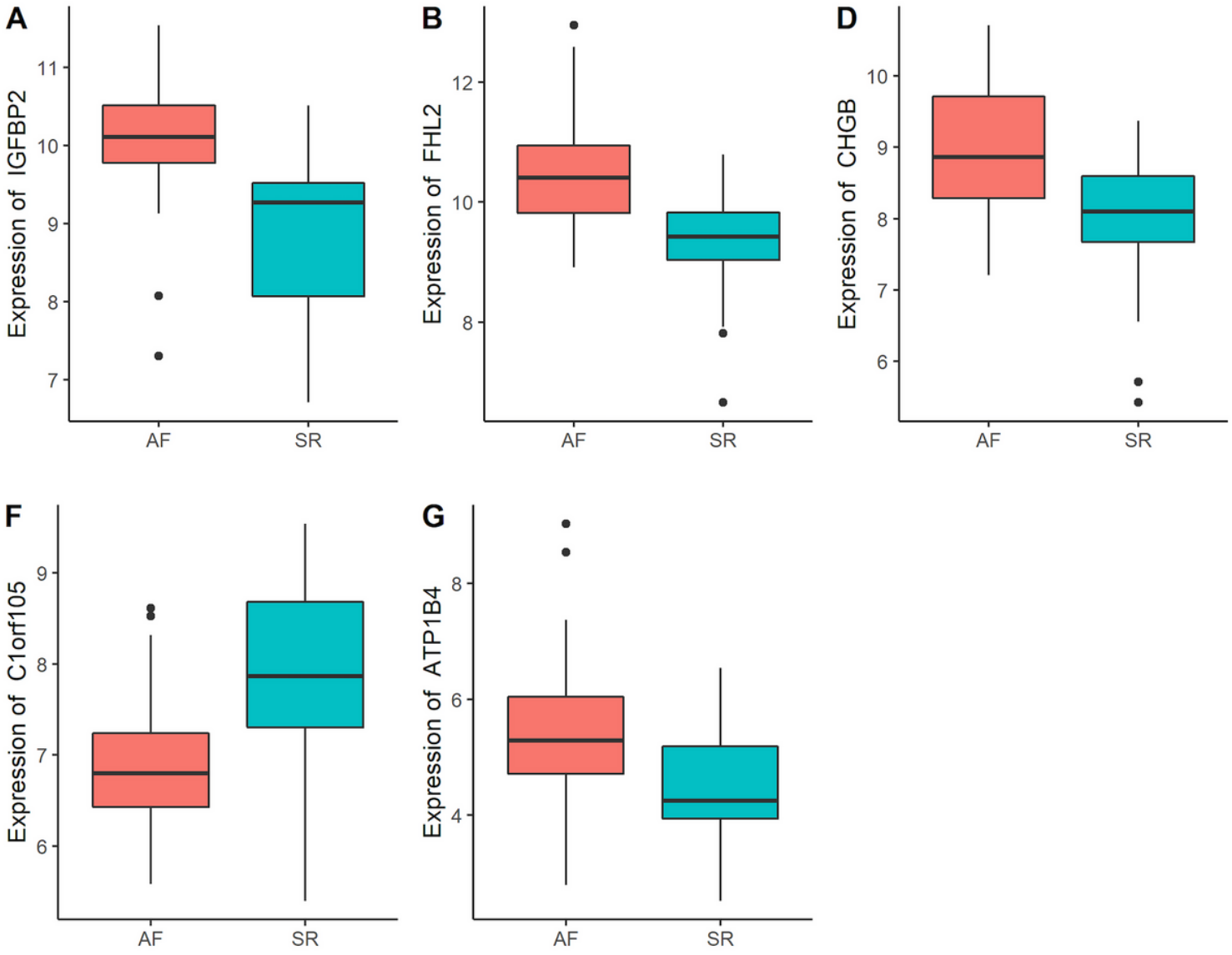

Figure 2

Boxplots of gene expressions for 5 selected genes with a screening criteria of |log2 FC| ≥ 1 and adjust P value < 0.05.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.tiff
- SupplementaryFigure8.tiff
- SupplementaryTable.docx
- SupplementaryFigure3.tiff
- SupplementaryFigure4.tiff

- SupplementaryFigure5.tiff
- SupplementaryFigure6.tif
- SupplementaryFigure7.tiff
- SupplementaryFigure2.tiff