

RESEARCH

Open Access



Analysis of potential genetic biomarkers using machine learning methods and immune infiltration regulatory mechanisms underlying atrial fibrillation

Li-Da Wu[†], Feng Li[†], Jia-Yi Chen, Jie Zhang, Ling-Ling Qian^{*} and Ru-Xing Wang^{*}

Abstract

Objective: We aimed to screen out biomarkers for atrial fibrillation (AF) based on machine learning methods and evaluate the degree of immune infiltration in AF patients in detail.

Methods: Two datasets (GSE41177 and GSE79768) related to AF were downloaded from Gene expression omnibus (GEO) database and merged for further analysis. Differentially expressed genes (DEGs) were screened out using “*limma*” package in R software. Candidate biomarkers for AF were identified using machine learning methods of the LASSO regression algorithm and SVM-RFE algorithm. Receiver operating characteristic (ROC) curve was employed to assess the diagnostic effectiveness of biomarkers, which was further validated in another independent validation dataset of GSE14975. Moreover, we used CIBERSORT to study the proportion of infiltrating immune cells in each sample, and the Spearman method was used to explore the correlation between biomarkers and immune cells.

Results: 129 DEGs were identified, and *CYBB*, *CXCR2*, and *S100A4* were identified as key biomarkers of AF using LASSO regression and SVM-RFE algorithm. Both in the training dataset and the validation dataset, *CYBB*, *CXCR2*, and *S100A4* showed favorable diagnostic effectiveness. Immune infiltration analysis indicated that, compared with sinus rhythm (SR), the atrial samples of patients with AF contained a higher T cells gamma delta, neutrophils and mast cells resting, whereas T cells follicular helper were relatively lower. Correlation analysis demonstrated that *CYBB*, *CXCR2*, and *S100A4* were significantly correlated with the infiltrating immune cells.

Conclusions: In conclusion, this study suggested that *CYBB*, *CXCR2*, and *S100A4* are key biomarkers of AF correlated with infiltrating immune cells, and infiltrating immune cells play pivotal roles in AF.

Keywords: Atrial fibrillation, Immune infiltration, Biomarker, CIBERSORT, Diagnostic biomarker

Introduction

As the most common arrhythmia in clinic, patients with atrial fibrillation (AF) have high mortality and morbidity. It is reported that about 1–2% of the population are

troubled by AF, which contributes to heart failure and cardiogenic embolism [1]. AF could be divided into permanent AF, persistent AF (pAF), long standing pAF and paroxysmal AF based on its duration. Patients with hypertension, obesity, and diabetes etc. frequently develop AF, however, the molecular mechanisms underlying the development of AF remain unclear yet [2]. Inflammatory response plays an important role in the occurrence and development of AF. Studies have shown

*Correspondence: qll900@sina.com; ruxingw@aliyun.com

[†]Li-Da Wu and Feng Li contributed equally to this work

Department of Cardiology, Wuxi People's Hospital Affiliated to Nanjing Medical University, No. 299, Qingyang Road, Wuxi 214023, China



that TNF- α , CRP and IL -6 are significantly increased in atrial tissues of AF patients and related to outcomes of AF patients [3–5]. Moreover, studies demonstrated the anti-inflammatory therapies can significantly reduce AF episodes [6, 7]. In recent years, the role of immune cells infiltration in the inflammatory response of patients with AF has been widely concerned. Yamashita et al. confirmed that, in human AF, adhesion and recruitment of macrophages in heart endocardium promoted inflammatory responses [8]. Similarly, Hohmann et al. demonstrated that the number of CD3-positive T cells in left atrial appendage are significantly increased in patients with AF [9]. However, the more accurate association between infiltrating immune cells and AF still needs to further study.

Medicine is one of the early applications of artificial intelligence (AI), which is gradually changing the way many diseases are diagnosed and treated [10]. Machine learning is an important part of artificial intelligence that using algorithms to identify expression patterns of datasets. Machine learning has already been employed in prediction of myocardial infarction, pathological identification and surgical improvement [11]. Moreover, machine learning is also a research hotspot and cutting-edge technology in the field of arrhythmia diagnosis and treatment. Han et al. used machine learning algorithms to incorporate clinical signatures of AF patients, and their work provides prognostic value for risk stratification in stroke beyond CHA₂DS₂-VASc [12]. The diagnosis of atrial fibrillation requires electrocardiogram (ECG) test, but some asymptomatic patients are often difficult to find. Raghunath et al. collected 12 lead ECGs of 430,000 patients and predicted new onset atrial fibrillation within 1 year based on deep neural network. It was found that the accuracy of the algorithm reached 0.85 [13]. To the best of our knowledge, we firstly conducted a bioinformatics analysis to screen out key differentially expressed genes (DEGs) in AF as biomarkers based on machine learning algorithms. The CIBERSORT algorithm has been adopted to evaluate infiltrating immune cells based on gene expression profiles in various diseases [14–17]. We also performed a detailed analysis of immune cells infiltration in patients with AF using CIBERSORT algorithm.

Materials and methods

Microarray data

The workflow of this analysis is shown in Fig. 1. Three datasets related to AF (GSE41177, GSE79768 and GSE14975) were downloaded from Gene Expression Omnibus (GEO) database [18] via “*GEO query*” package [19]. The above three datasets were all based on GPL570 platform. GSE41177 contained 19 left atrial

tissue samples from 3 SR individuals and 16 AF patients [20]; GSE79768 consisted of 13 left atrial tissue samples from 7 AF patients and 6 SR individuals [21]; GSE14975 contained 10 left atrial tissue samples from 5 SR individuals and 5 AF patients [22]. The detailed characteristics of GSE41177, GSE79768 and GSE14975 is provided in Additional file 2: Table S1.

Data processing and DEGs screening

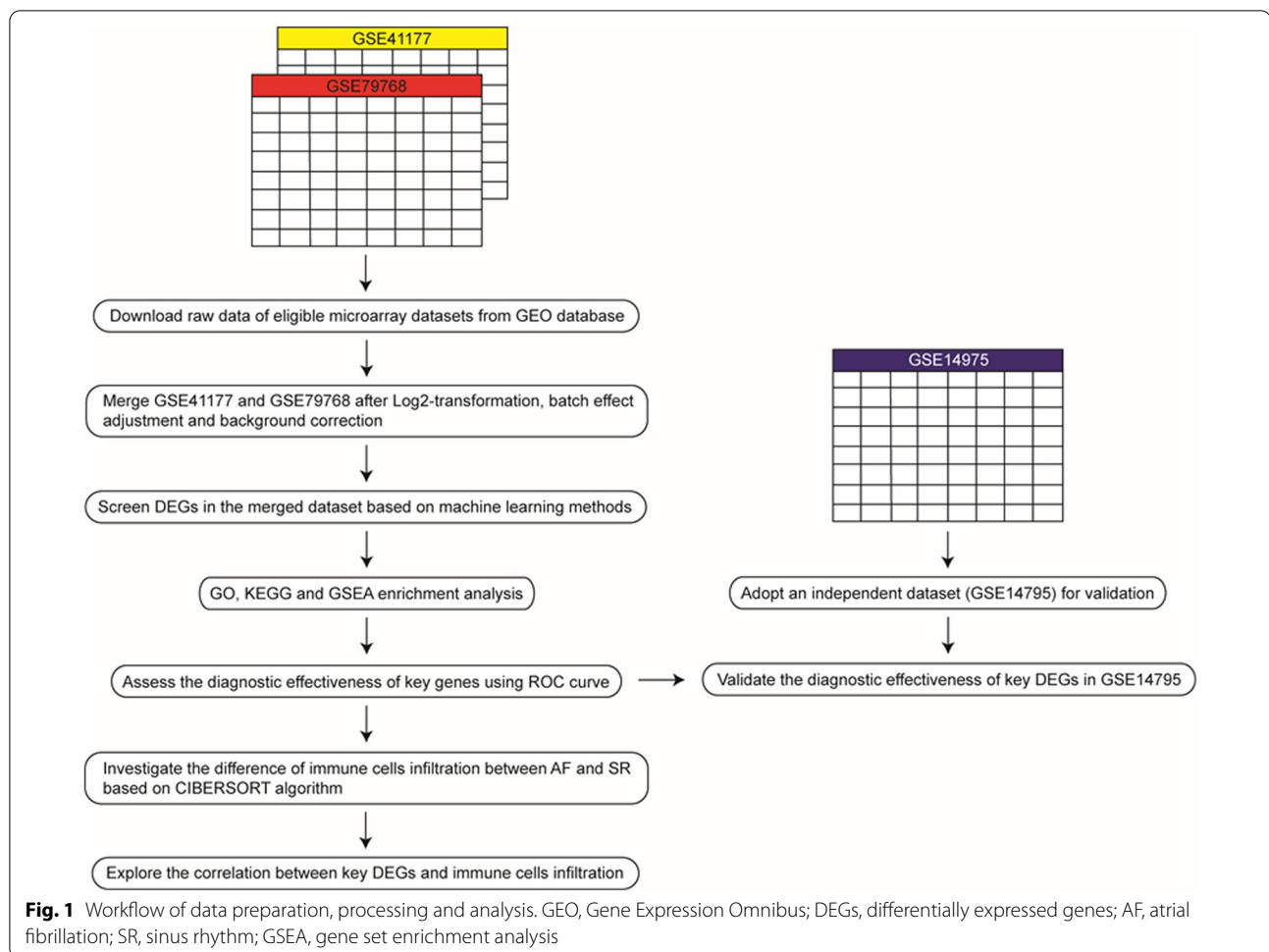
R software was employed to create gene expression matrices of GSE41177, GSE79768 and GSE14975. Log₂-transformation and background correction were performed on the expression profiles by the “*limma*” package [23]. Furthermore, “*SVA*” package was used for batch effects adjustment between the GSE41177 and GSE79768 [24]. GSE41177 and GSE79768 were merged for further analysis, and the GSE14975 was used as the validation cohort. “*heatmap*” package and “*ggplot2*” package was adopted to create to “heatmap” and “volcano plot” of DEGs.

Enrichment analysis

To understand the function of DEGs in AF patients, the “*clusterProfiler*” package was used to perform GO and KEGG pathway analysis [25]. Gene set enrichment analysis (GSEA) was also employed to identify pathways enriched in AF patients and SR individuals, respectively. “c2.cp.kegg.v7.0.symbols.gmt” from MSigDB database was adopted as the reference dataset [26].

Identification of key DEGs as biomarkers in AF using machine learning methods

Machine learning methods were adopted to screen out key DEGs as biomarkers in AF. LASSO algorithm, a regression analysis, often utilized to improve prediction accuracy. It belongs to linear regression model family and uses the default ten-fold cross validation. In recent years, LASSO regression analysis has been widely used in researches to screen out diagnostic or prognostic factors [27]. Jubair et al. found a meaningful way to identify subtype-specific biomarkers for the breast cancer survivability using LASSO regression analysis [28]. Ma et al. also identified key genes in blood of patients with intervertebral disc degeneration (IDD) as important biomarkers based on LASSO regression analysis [29]. To screen out key genes correlated with AF, “*glmnet*” package was used to perform LASSO regression algorithm. SVM-RFE is another machine learning algorithm, which has been widely used for classification and regression analysis. SVM-RFE model has nonlinear discrimination characteristics, which allows the results to be compared after modeling different numbers of variables, so as to screen the best combination of variables. Based on SVM-RFE algorithm, Zhang et al. screened ten discriminant features,



which provided a fast and effective diagnostic standard for Kashin–Beck disease [30]. We also used “*e107*” package to carry out SVM-RFE algorithm and identify key genes in occurrence and development of AF with discriminative power [31].

Diagnostic value of key DEGs as biomarkers in AF

Receiver operating characteristic (ROC) curve was established based on the meta-data cohort merged by GSE41177 and GSE79768 to evaluate the predictive value of biomarkers. We used the area under curve (AUC) value to determine the diagnostic effectiveness in discriminating AF from SR patients. Then, an independent dataset (GSE14975) was adopted to further validate the diagnostic effectiveness of biomarkers.

Evaluation of infiltrating immune cells

CIBERSORT algorithm was employed to evaluate infiltrating immune cells in patients with AF [17]. A large number of studies have used CIBERSORT to explore the function of immune cells in various diseases,

including osteoarthritis [14], high-grade serous ovarian cancer [15] and breast ductal and lobular carcinoma [16]. Proportions of infiltrating immune cells were visualized in R software using “*ggplot2*” package and “*pheatmap*” package. Correlation heatmap was created by “*corrplot*” package to visualize the correlation of infiltrating immune cells. The difference of immune cells infiltration between atrial tissue samples from AF patients and SR individuals were showed in the violin plot using the “*vioplot*” package. Then, “*ggplot2*” package was also adopted to perform principal components analysis (PCA) based on immune cells infiltration and draw a two dimensional PCA plot.

Correlation analysis of biomarkers and infiltrating immune cells

We used “Spearman” method to explore the correlation between biomarkers and immune infiltration, and then we used the “*ggplot2*” package to visualize results.

Results

Identification of DEGs in AF

Left atrial tissues from 23 AF patients and 9 normal individuals of GSE41177 and GSE79768 were analyzed. PCA plot clearly indicated that the batch effect between GSE41177 and GSE79768 was successfully removed (Fig. 2c). In the meta-data cohort merged by GSE41177 and GSE79768, 129 DEGs were identified using the “limma” package, including 71 upregulated genes and 58 downregulated genes (Fig. 2a, b).

Functional correlation analysis

We performed functional enrichment analysis of DEGs between AF and SR patients based on GO and KEGG databases (Fig. 3a). The biological processes were enriched in neutrophil activation involved in immune response, neutrophil mediated immunity, neutrophil degranulation, neutrophil activation and cell cellular defense response. The relationship between

biological processes terms and each DEG was shown in Fig. 3c. The enriched cellular components were mainly involved in collagen-containing extracellular matrix, secretory granule lumen, cytoplasmic vesicle lumen, vesicle lumen and NADPH oxidase complex. The molecular functions were mainly enriched in RAGE receptor binding, Toll-like receptor binding, calcium-dependent protein binding, superoxide-generating NADPH oxidase activity and long-chain fatty and binding oxidoreductase activity. KEGG pathway analysis shows that the osteoclast differentiation, staphylococcus aureus infection, leukocyte trans-endothelial migration, tight junction and cell adhesion molecules were mostly enriched (Fig. 3b). Moreover, GSEA results showed that Hedgehog signaling pathway and linoleic acid metabolism were mainly enriched in SR (Fig. 4a). The receptor signaling pathway, cell adhesion molecules, cytokine-cytokine receptor interaction, leukocyte trans-endothelial migration and natural killer

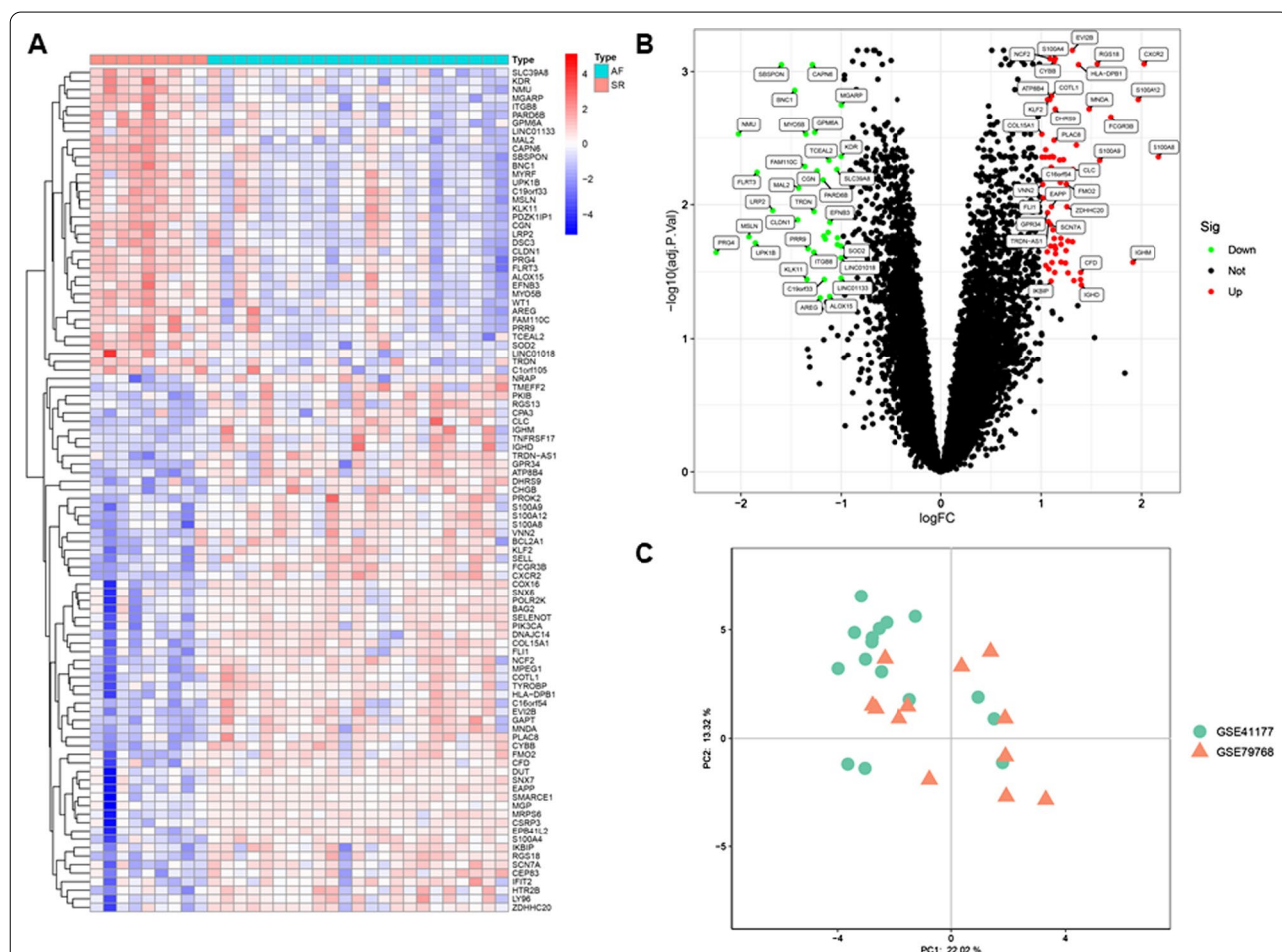
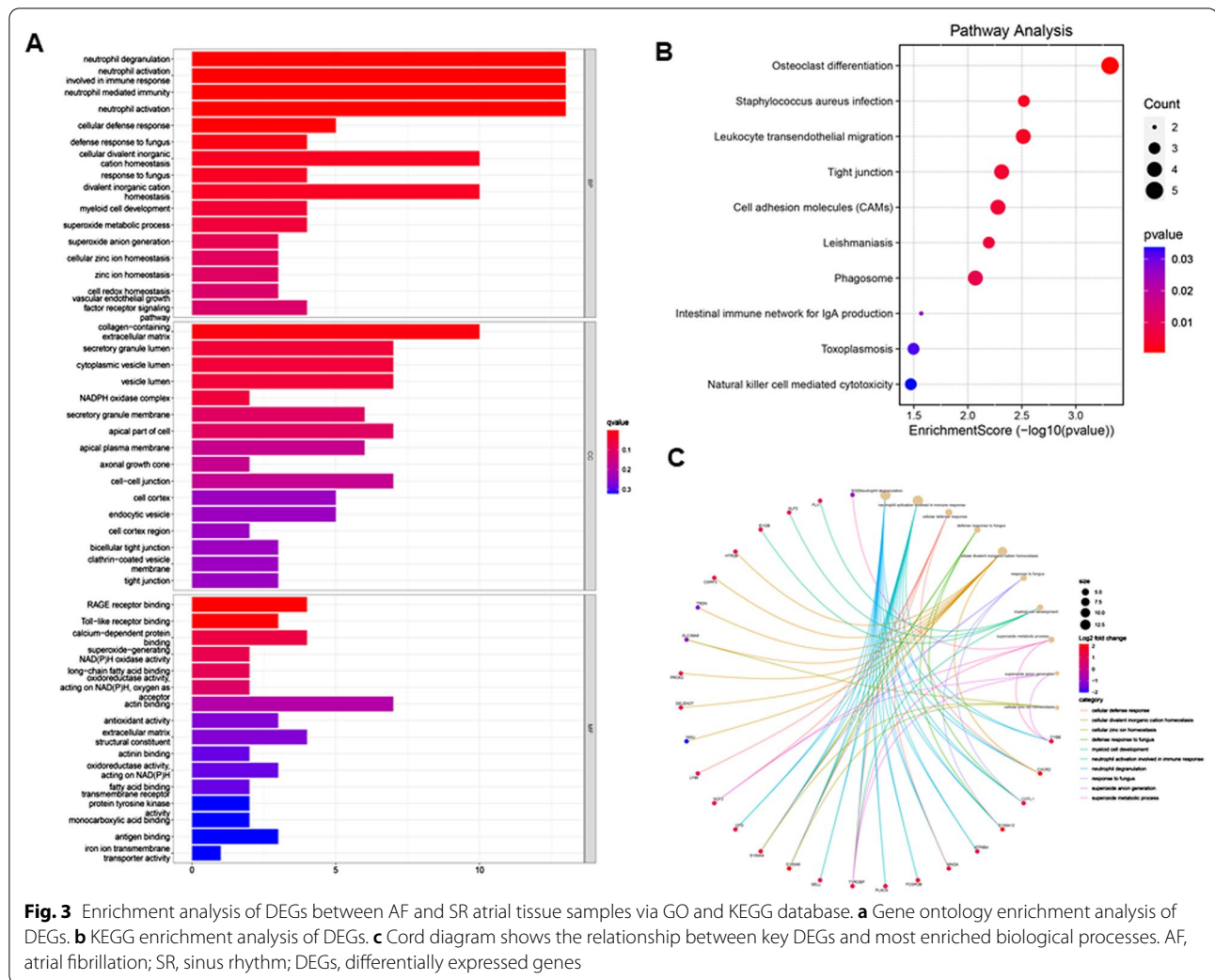


Fig. 2 Identification of DEGs between AF and SR atrial tissue samples. **a** Heatmap visualization of the top 100 DEGs between AF and SR atrial tissue samples. **b** Volcano plot visualization of DEGs between AF and SR atrial tissue samples. **c** PCA plot of AF and SR atrial tissue samples after removing batch effect between GSE41177 and GSE79768. AF, atrial fibrillation; SR, sinus rhythm; DEGs, differentially expressed genes



cell mediated cytotoxicity were mainly enriched in AF (Fig. 4b).

Identification key DEGs as biomarkers of AF based on machine learning algorithms

We used two different machine learning algorithms to screen key DEGs as biomarkers of AF. 9 key DEGs were identified using LASSO algorithm in the present study (Fig. 5a). Moreover, 40 DEGs was identified as biomarkers based on SVM-RFE algorithm (Fig. 5b). The three overlapping DEGs (*CXCR2*, *CYBB* and *S100A4*) were ultimately selected (Fig. 5c).

Diagnostic effectiveness of biomarkers in AF

Our results of ROC curves indicated that these three biomarkers screened out by machine learning algorithms also have a favorable diagnostic value in the meta-data cohort merged by GSE41177 and GSE79768, with an AUC of 0.942 (95% CI 0.845–1.000) in *CYBB*, AUC of

0.961 (95% CI 0.870–1.000) in *CXCR2*, and AUC of 0.932 (95% CI 0.768–1.000) in *S100A4* (Fig. 6a–c). Moreover, the diagnostic effectiveness of key DEGs was further validated in another independent dataset (GSE14795) with an AUC of 0.880 (95% CI 0.600–1.000) in *CYBB*, AUC of 0.760 (95% CI 0.400–1.000) in *CXCR2*, and AUC of 0.840 (95% CI 0.520–0.912) in *S100A4* (Fig. 6d–f).

Immune infiltration analysis

Based on CIBERSORT, we evaluated immune cells infiltration in patients with AF and normal individuals. Figure 7a, b illustrate the proportion of immune cells from 9 SR left atrial tissue samples and 23 AF left atrial tissue samples. As shown in Fig. 7c, compared with SR, left atrial tissue samples from AF patients contained higher neutrophils, mast cells resting and T cells gamma delta, whereas lower T cells follicular helper. Correlation analysis showed that dendritic cells activated and NK cells resting had the most intense positive relationship with

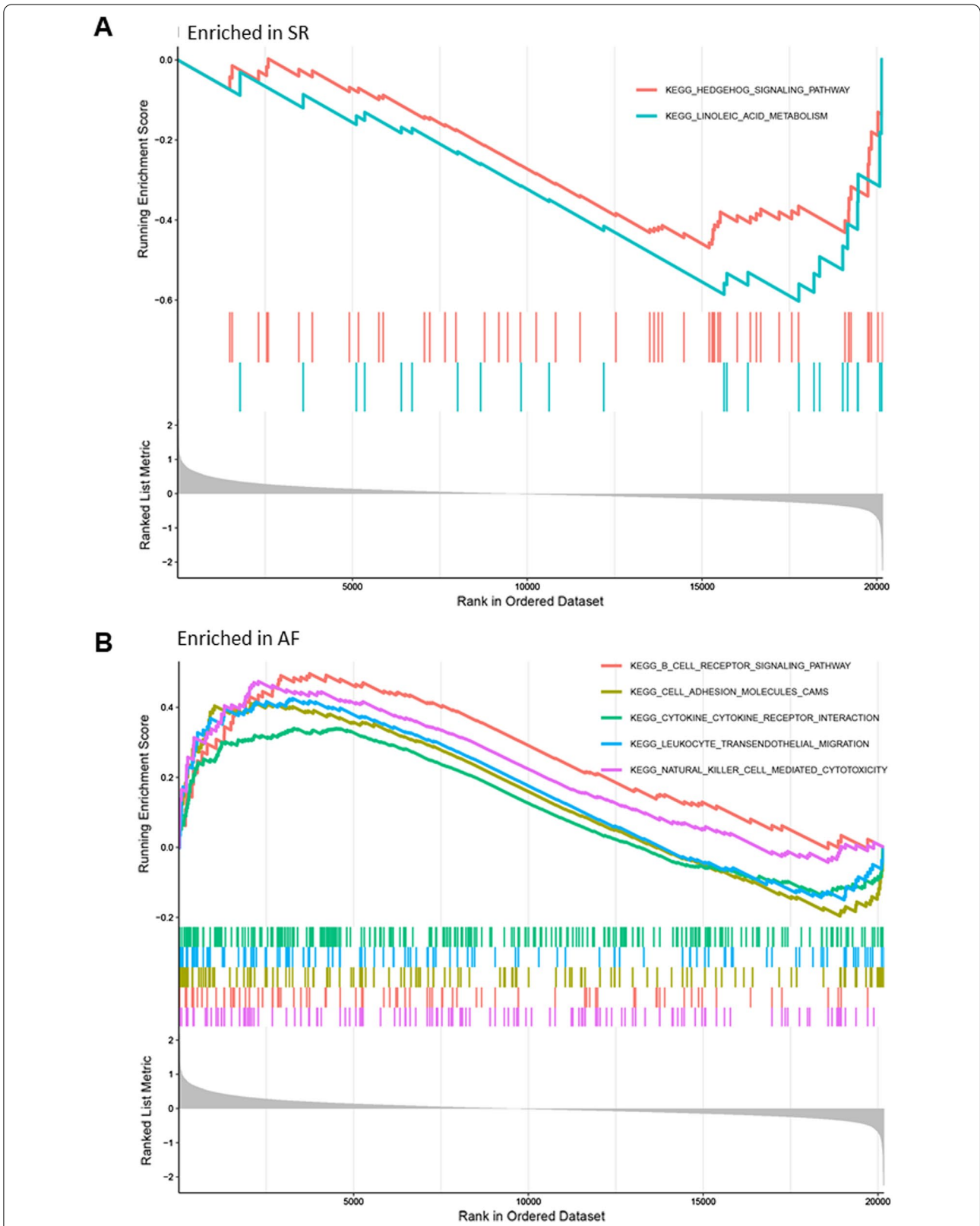
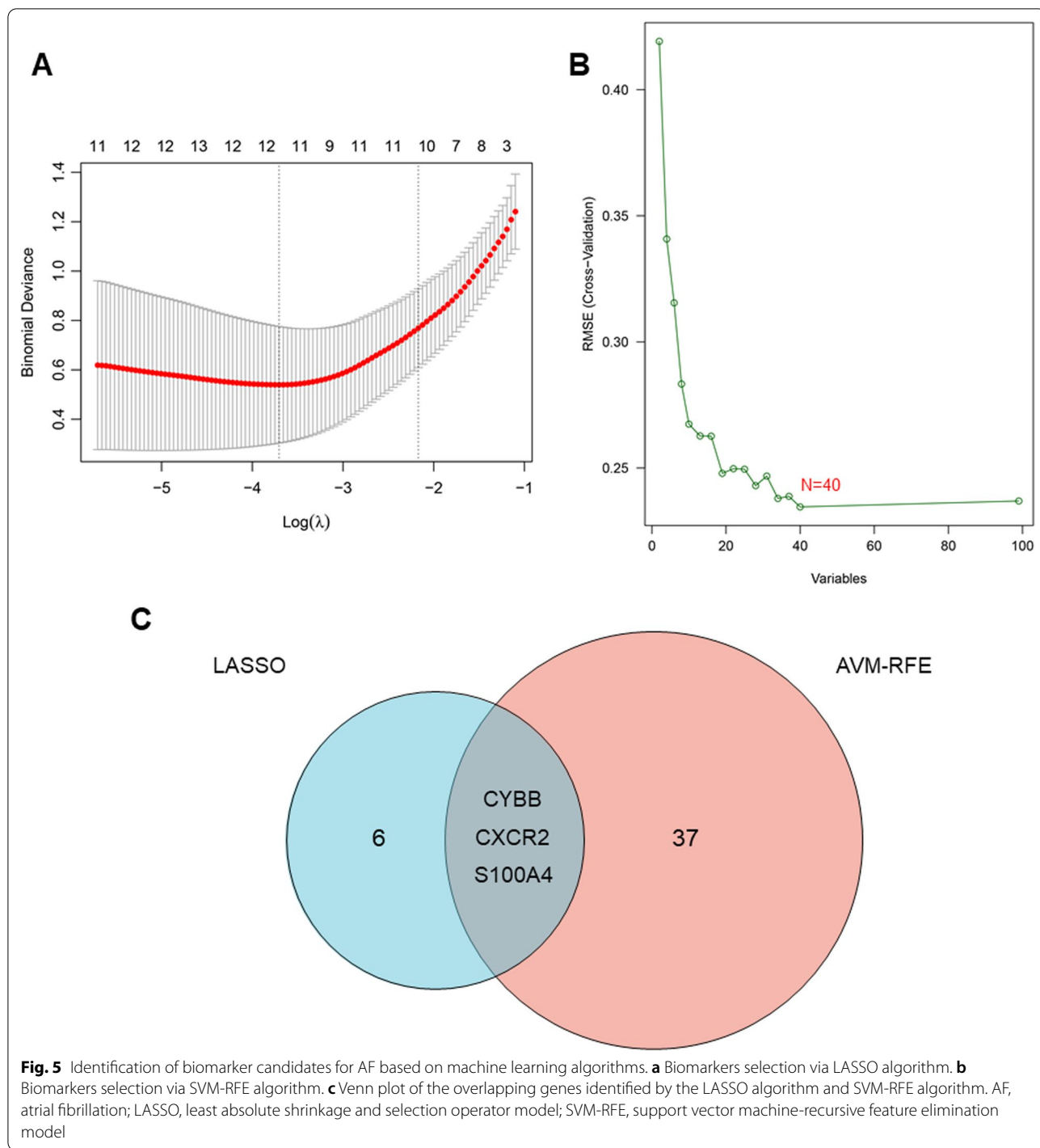


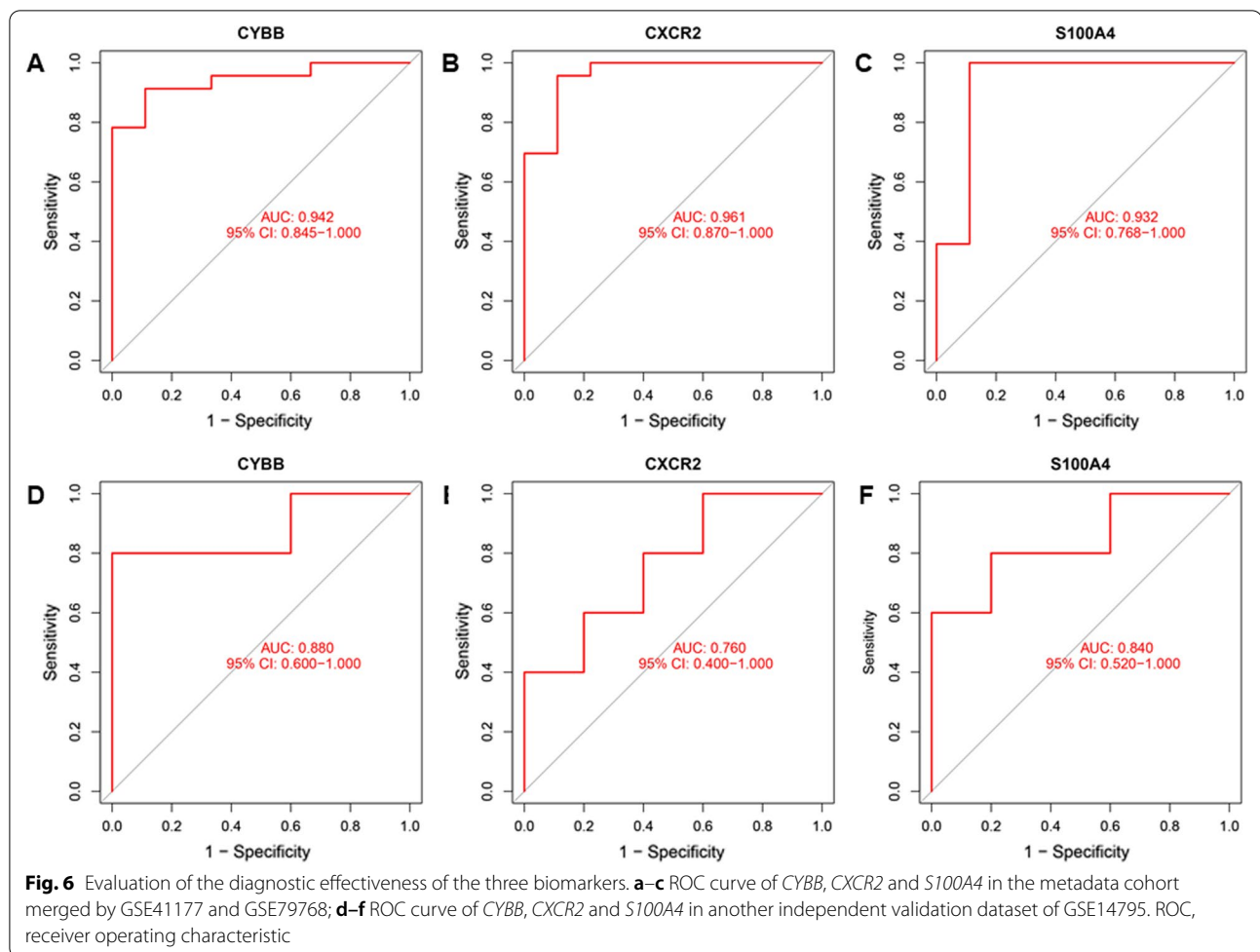
Fig. 4 GSEA enrichment analysis of DEGs between AF and SR atrial tissue samples. **a** GSEA enrichment analysis results in SR patients. **b** GSEA enrichment analysis results in AF patients. AF, atrial fibrillation; SR, sinus rhythm; GSEA, gene set enrichment analysis; DEGs, differentially expressed genes



$r=0.62$, B cells naïve and B cells memory had the most obvious negative correlation with $r=-0.51$ (Fig. 7d). PCA diagram revealed a distinct group bias and proved that the degree of immune cells infiltration is different between AF patients and SR individuals (Additional file 1: Fig. S1).

Correlation analysis between CXCR2, CYBB and S100A4 and infiltrating immune cells

In correlation analysis, we demonstrated that *CYBB* was positively correlated with T cells gamma delta ($r=0.28$, $P=0.029$) and negatively correlated with T cells CD8 ($r=-0.41$, $P=0.021$), T cells follicular helper

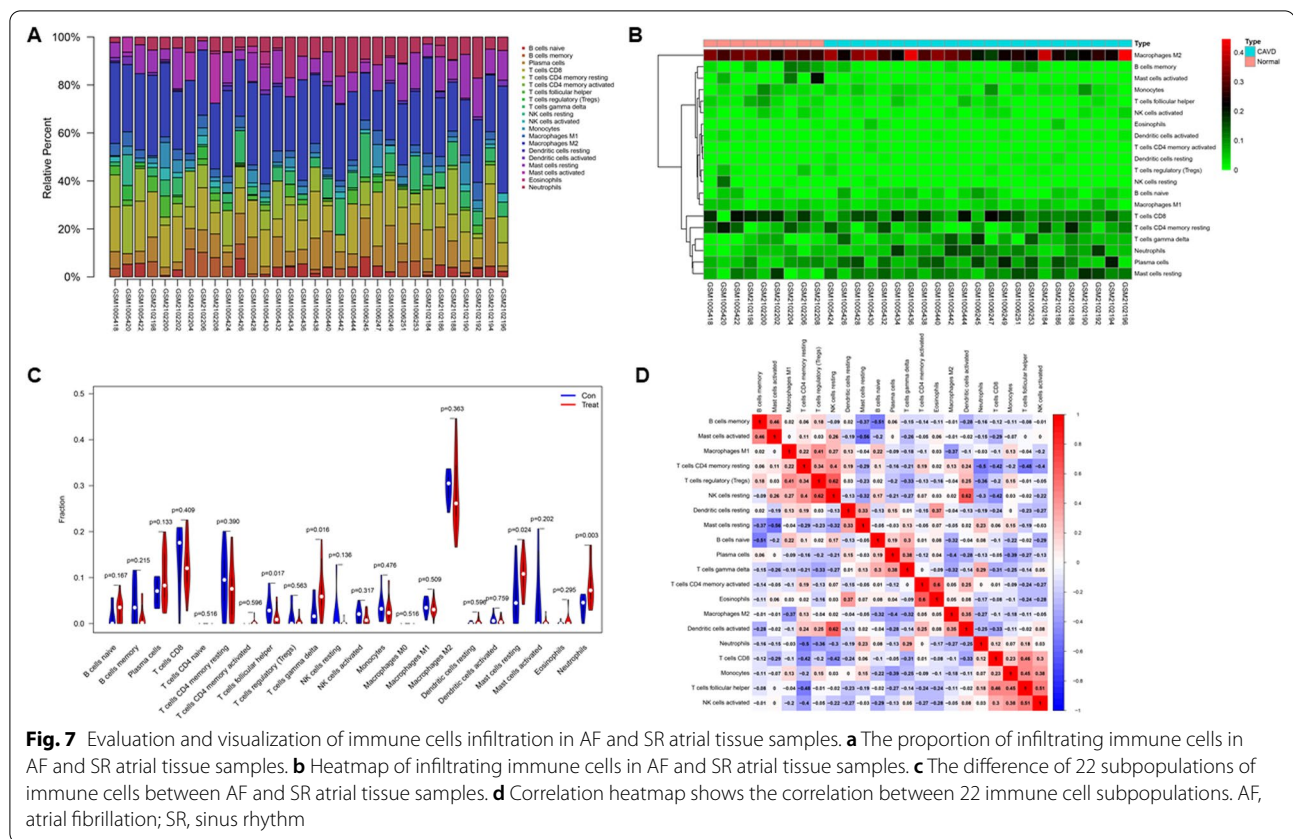


($r = -0.52$, $P = 0.002$) (Fig. 8a–d). *CXCR2* was positively correlated with T cells gamma delta ($r = 0.43$, $P = 0.014$), neutrophils ($r = 0.75$, $P < 0.001$) and negatively correlated with macrophages M2 ($r = -0.53$, $P = 0.002$) (Fig. 8e–h). *S100A4* was positively correlated with plasma cells ($r = 0.45$, $P = 0.01$) and mast cells resting ($r = 0.42$, $P = 0.017$) (Fig. 8i–k).

Discussion

AF is one of the most prevalent arrhythmias, however, the specific molecular mechanisms of AF still remain unclear. Despite the great improvement has been gained in the field of diagnosis and treatment, AF remains a leading cause of mortality and disability [32]. Drugs for rate control, oral anticoagulants for stroke prevention, antiarrhythmic drug and catheter ablation for conversion are main treatments for AF patients [33]. But the efficacy and safety of these treatment measures are still not well understood. Meanwhile, many asymptomatic patients, especially patients with paroxysmal AF, are difficult to find. This is the first study to identify biomarkers

of AF associated with immune cells infiltration. Two gene expression datasets from GEO database were merged and conducted an integrated analysis. 129 DEGs were detected using “*limma*” package. Enrichment analysis showed that these 129 DEGs were significantly correlated with immune and inflammatory responses. The relationship between inflammatory response and AF has been widely studied in the past decades and various regular anti-inflammatory biomarkers were found to be related to AF. It has been reported that patients with AF had increased IL-6, IL-8 and TNF- α , meanwhile, these inflammatory markers can also predict the outcome of AF ablation [34–38]. High-sensitive C-reactive protein (hsCRP) is also associated with increased risk of AF recurrence following successful electrical cardioversion and catheter ablation [39]. In addition, accumulating studies have demonstrated that the increase of TGF β 1 in AF patients promote atrial fibrosis, which plays a pivotal role in atrial structural remodeling in AF [40, 41]. Additionally, the critical role of various immune cells in the pathogenesis of AF has attracted more and more

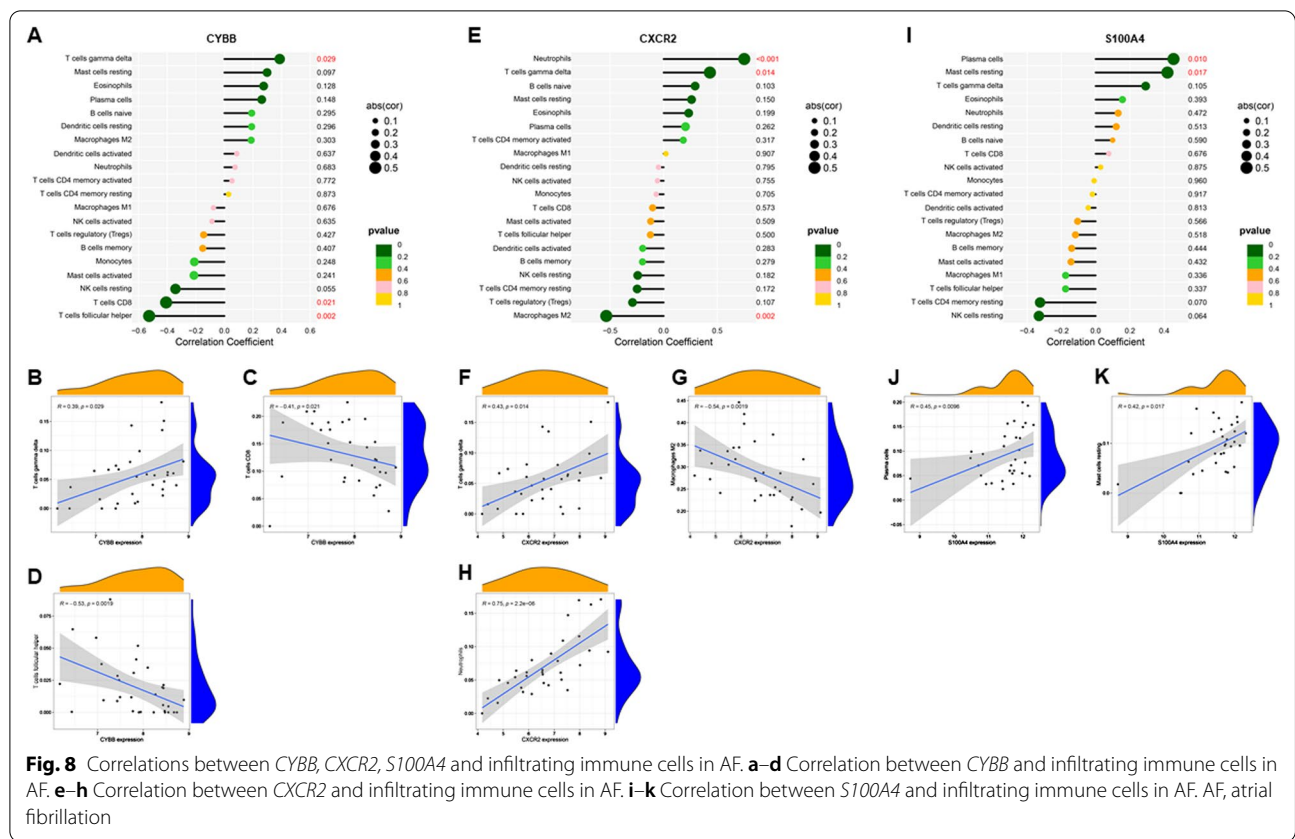


attention, including infiltrating in the atrium and secreting several chemokines and cytokines to regulate the microenvironment of the heart [42]. Our GSEA results are also in general agreement with the previous findings that immune cells infiltration participates in the pathogenesis of AF [8, 9]. Medicine is the earliest application field of AI. In the past few decades, AI technology, especially machine learning, has made great progress in the diagnosis and treatment of a variety of diseases, including cardiovascular diseases, nervous system diseases and genetic diseases [43]. Therefore, for the first time, we sought to screen out key DEGs between AF and SR patients as biomarkers based on machine learning methods and explore its relationship with immune cells infiltration in AF. Overlapping the results from two machine learning algorithms, *CYBB*, *CXCR2* and *S100A4* were identified as key DEGs and biomarkers of AF.

CYBB, also known as NOX2, has been implicated in oxidative stress in various cardiovascular diseases [44]. Pignatelli et al. reported that serum NOX2 can be used as one of the important indicators to predict vascular embolism events in [45]. In animal model, numerous studies have demonstrated that inhibition of NOX2-mediated production of reactive oxygen species (ROS) prevents atrial remodeling [46, 47]. In addition, atrial

electrical remodeling can also be alleviated by inhibiting NOX2 and oxidative stress [47, 48]. In human AF, NOX2 has also been demonstrated to participate in the atrial structural remodeling and electrophysiological remodeling, and up-regulation of NOX2 is associated with an enhanced risk of AF [49, 50].

The chemokine receptor *CXCR2*, encoding by *CXCR2*, belongs to chemokine receptors family, mediates cellular migration of immune cells [51]. The expression level of *CXCR2* is tightly regulated during infection and inflammation. It is worth noting that *CXCR2* is key stimulant of immune cells infiltration and recruitment, especially of neutrophils. Our results of evaluation of 22 subtypes immune cells infiltration showed that neutrophils are significantly elevated in AF patients compared with SR [52]. It is also reported that *CXCR2* was involved in atrial monocytes infiltration, which accelerates atrial fibrosis and promotes atrial remodeling. Therefore, blocking *CXCR2* may serve as a new therapeutic strategy for AF patients [53]. Moreover, *CXCR2* is also a crucial regulator of hypertension. In spontaneously hypertensive rats, Zhang et al. identified that inhibition of *CXCR2* could prevent the occurrence of hypertension-induced AF [53]. In angiotensin II-induced cardiac atrial fibrillation animal



model, *CXCR2* has also been proved to participate in immune cells infiltration and mediates cardiac hypertrophy and remodeling through regulation of monocyte [54].

S100A4, also known as fibroblast specific protein 1 (FSP1), is involved in various biological processes including cell survival, cell motility, and cell differentiation. Numerous studies have already revealed roles of *S100A4* in cancer progression, particularly the ability of enhancing metastasis. *S100A4* has also been linked to various diseases besides tumor, such as cardiac fibrosis and hypertrophy, kidney fibrosis and pulmonary disease. All of these diseases involve the inflammatory processes and rely heavily on tissue remodeling [55]. Studies revealed that *S100A4* expressed in normal human heart and increased in hypertrophic left ventricles [56–58]. In addition, *S100A4* is a key regulator of endothelial mesenchymal transformation (EMT), which is related to immune cells infiltration, making epithelial cells present mesenchymal cell phenotype and ultimately resulting in enhanced migration ability, enhanced anti apoptosis ability and production of a large number of extracellular matrix components. Recently, studies demonstrated that EMT occurs in the atrium of AF patients and contribute to fibroblast accumulation. Meanwhile, *S100A4* also had

significant correlations with left atrial dimension in AF patients [59].

We used CIBERSORT to evaluate the degree of infiltrating immune cells in the present study. We found reduced infiltration of T cells follicular helper, as well as increased neutrophils, mast cells resting and T cells gamma delta in AF. Neutrophils represent activated non-specific inflammation and have been found as markers of inflammation in various diseases. Correlation between inflammatory markers and cardiovascular diseases has been studied widely and the relationship between neutrophils and cardiovascular diseases has been confirmed in the past. It is reported that the level of neutrophils is an independent predictor for the prognosis of acute coronary syndrome [60]. Recently, accumulating studies have also reported that increased neutrophil/lymphocyte ratio is related to the increased risk of AF occurrence [61, 62]. Mast cells, tissue-specific innate immune cells, present in virtually all body tissues including the heart. Numerous inflammatory mediators secreted by mast cells including IL-1 β , IL-6 and TGF- β 1 participate in atrial structural remodeling and development of AF [63]. Liao et al. reported that the mast cells stabilization is associated with reduced atrial fibrosis and reduce AF incidence in animal model [64]. T follicular helper cells has also been

reported to function in AF by secreting IL-21 and ultimately promoting B cell proliferation and differentiation, which might be activated by Toll-like receptor 2 (TLR2) and TLR4 [65, 66]. We also studied the correlation between *CYBB*, *CXCR2*, *S100A4* and infiltrating immune cells. Based on the correlation analysis results, *CYBB*, *CXCR2* and *S100A4* appear to play key roles in regulation of immune cells infiltration.

However, a limitation of the present study should be noted. The occurrence and development of atrial fibrillation is a complex and dynamic process, and its pathogenesis includes atrial electrical remodeling, atrial structural remodeling and autonomic nervous dysfunction. Although a total of 42 participants were included, the input data might still be insufficient to identify and validate key genes in the atrial fibrillation development. Moreover, the 42 participants included in the study came from various regions with different diet, physical activity, genetic variation, susceptibility to cardiovascular diseases, and so on. All of these factors may have an impact on atrial fibrillation. Therefore, the diagnostic efficacy of *CYBB*, *CXCR2*, and *S100A4* in different populations and its role in the occurrence of atrial fibrillation still need more external validation.

Conclusions

We found that *CYBB*, *CXCR2* and *S100A4* may be key biomarkers of AF based on machine learning methods. The immune cells infiltration of patients with AF was measured in detail. Moreover, correlations between *CYBB*, *CXCR2* and *S100A4* and immune cells may play an important role in AF. Further researches for the specific molecular mechanism of these biomarkers and immune cells are required to study.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-022-01212-0>.

Additional file 1: Fig. S1. PCA plot based on infiltrating immune cells of AF and SR atrial tissue samples. PCA, principal component analysis. AF, atrial fibrillation; SR, sinus rhythm.

Additional file 2: Table S1. Characteristics of the datasets included in the analysis.

Acknowledgements

We acknowledge GEO database for providing their platforms and contributors for uploading their meaningful datasets. And we thank all participants involved in studies included in our present study.

Authors' contributions

R-XW was involved in the experiment design. L-DW and FL performed the experiments. L-LQ and J-YC analyzed the data. R-XW, L-DW and JZ wrote the manuscript. L-LQ, JZ and J-YC edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of China (81770331) and Research Foundation from Wuxi Health Commission for the Youth (Q202034).

Data availability

Publicly available datasets were analyzed in this study. All the raw data used in this study are derived from the public GEO data portal (<https://www.ncbi.nlm.nih.gov/geo/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 3 December 2021 Accepted: 14 March 2022

Published online: 19 March 2022

References

- Andrade J, Khairy P, Dobrev D, Nattel S. The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circ Res*. 2014;114:1453–68.
- Andrade JG, Aguilar M, Atzema C, Bell A, Cairns JA, Cheung CC, et al. The 2020 Canadian cardiovascular society/Canadian heart rhythm society comprehensive guidelines for the management of atrial fibrillation. *Can J Cardiol*. 2020;36:1847–948.
- Chung MK, Martin DO, Sprecher D, Wazni O, Kanderian A, Carnes CA, et al. C-reactive protein elevation in patients with atrial arrhythmias: inflammatory mechanisms and persistence of atrial fibrillation. *Circulation*. 2001;104:2886–91.
- Wu N, Xu B, Xiang Y, Wu L, Zhang Y, Ma X, et al. Association of inflammatory factors with occurrence and recurrence of atrial fibrillation: a meta-analysis. *Int J Cardiol*. 2013;169:62–72.
- Marcus GM, Smith LM, Ordovas K, Scheinman MM, Kim AM, Badhwar N, et al. Intracardiac and extracardiac markers of inflammation during atrial fibrillation. *Heart Rhythm*. 2010;7:149–54.
- Pfister R, Michels G, Brägelmann J, Sharp SJ, Luben R, Wareham NJ, et al. Plasma vitamin C and risk of hospitalisation with diagnosis of atrial fibrillation in men and women in EPIC-Norfolk prospective study. *Int J Cardiol*. 2014;177:830–5.
- Hemilä H, Suonsyrjä T. Vitamin C for preventing atrial fibrillation in high risk patients: a systematic review and meta-analysis. *BMC Cardiovasc Disord*. 2017;17:49.
- Yamashita T, Sekiguchi A, Iwasaki YK, Date T, Sagara K, Tanabe H, et al. Recruitment of immune cells across atrial endocardium in human atrial fibrillation. *Circ J*. 2010;74:262–70.
- Hohmann C, Pfister R, Mollenhauer M, Adler C, Kozlowski J, Wodarz A, et al. Inflammatory cell infiltration in left atrial appendageal tissues of patients with atrial fibrillation and sinus rhythm. *Sci Rep*. 2020;10:1685.
- Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2(10):719–31.
- Rajkumar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–58.
- Han L, Askari M, Altman RB, Schmitt SK, Fan J, Bentley JP, et al. Atrial fibrillation burden signature and near-term prediction of stroke: a machine learning analysis. *Circ Cardiovasc Qual Outcomes*. 2019;12(10):e005595.
- Raghunath S, Pfeifer JM, Ulloa-Cerna AE, Nemani A, Carbonati T, Jing L, et al. Deep neural networks can predict new-onset atrial fibrillation from the 12-lead ECG and help identify those at risk of atrial fibrillation-related stroke. *Circulation*. 2021;143(13):1287–98.

14. Deng YJ, Ren EH, Yuan WH, Zhang GZ, Wu ZL, Xie QQ, et al. GRB10 and E2F3 as diagnostic markers of osteoarthritis and their correlation with immune infiltration. *Diagnostics (Basel)*. 2020;10(3):171.
15. Liu R, Hu R, Zeng Y, Zhang W, Zhou H-H. Tumour immune cell infiltration and survival after platinum-based chemotherapy in high-grade serous ovarian cancer subtypes: a gene expression-based computational study. *EBioMedicine*. 2020;51:102602.
16. Zhang B, Wang Q, Fu C, Jiang C, Ma S. Exploration of the immune-related signature and immune infiltration analysis for breast ductal and lobular carcinoma. *Ann Transl Med*. 2019;7(23):730.
17. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453–7.
18. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41:991–5.
19. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846–7.
20. Yeh YH, Kuo CT, Lee YS, Lin YM, Nattel S, Tsai FC, et al. Region-specific gene expression profiles in the left atria of patients with valvular atrial fibrillation. *Heart Rhythm*. 2013;10(3):383–91.
21. Tsai FC, Lin YC, Chang SH, Chang GJ, Hsu YJ, Lin YM, et al. Differential left-to-right atria gene expression ratio in human sinus rhythm and atrial fibrillation: Implications for arrhythmogenesis and thrombogenesis. *Int J Cardiol*. 2016;222:104–12.
22. Adam O, Lavall D, Theobald K, Hohl M, Grube M, Ameling S, et al. Rac1-induced connective tissue growth factor regulates connexin 43 and N-cadherin expression in atrial fibrillation. *J Am Coll Cardiol*. 2010;55(5):469–80.
23. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
24. 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.
25. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (N Y)*. 2021;2(3):100141.
26. Powers RK, Goodspeed A, Pielke-Lombardo H, Tan AC, Costello JC. GSEA-InContext: identifying novel and common patterns in expression experiments. *Bioinformatics*. 2018;34(13):i555–64.
27. Wang H, Lengerich BJ, Aragao B, Xing EP. Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. *Bioinformatics*. 2019;35(7):1181–7.
28. Jubair S, Alkhateeb A, Tabl AA, et al. A novel approach to identify subtype-specific network biomarkers of breast cancer survivability. *Netw Model Anal Health Inform Bioinform*. 2020;9:43.
29. Ma X, Su J, Wang B, et al. Identification of characteristic genes in whole blood of intervertebral disc degeneration patients by weighted gene co-expression network analysis (WGCNA). *Comput Math Methods Med*. 2022;2022:6609901.
30. Zhang Y, Wei X, Cao C, Yu F, Li W, Zhao G, et al. Identifying discriminative features for diagnosis of Kashin-Beck disease among adolescents. *BMC Musculoskelet Disord*. 2021;22(1):801.
31. Tang J, Wang Y, Luo Y, Fu J, Zhang Y, Li Y, et al. Computational advances of tumor marker selection and sample classification in cancer proteomics. *Comput Struct Biotechnol J*. 2020;18:2012–25.
32. 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.
33. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC Jr, et al. 2019 AHA/ACC/HRS focused update of the 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 clinical practice guidelines and the heart rhythm society in collaboration with the society of thoracic surgeons. *Circulation*. 2019;140:e125–151.
34. Luan Y, Guo Y, Li S, et al. Interleukin-18 among atrial fibrillation patients in the absence of structural heart disease. *Europace*. 2010;12:1713–8.
35. Qu YC, Du YM, Wu SL, et al. Activated nuclear factor kappaB and increased tumor necrosis factor-alpha in atrial tissue of atrial fibrillation. *Scand Cardiovasc J*. 2009;43:292–7.
36. Psychari SN, Apostolou TS, Sinos L, et al. Relation of elevated C-reactive protein and interleukin-6 levels to left atrial size and duration of episodes in patients with atrial fibrillation. *Am J Cardiol*. 2005;95:764–7.
37. Liuba I, Ahlmroth H, Jonasson L, et al. Source of inflammatory markers in patients with atrial fibrillation. *Europace*. 2008;10:848–53.
38. Marcus GM, Whooley MA, Glidden DV, et al. Interleukin-6 and atrial fibrillation in patients with coronary artery disease: data from the Heart and Soul Study. *Am Heart J*. 2008;155:303–9.
39. Kornej J, Reinhardt C, Kosiuk J, Arya A, Hindricks G, Adams V, et al. Response of high-sensitive C-reactive protein to catheter ablation of atrial fibrillation and its relation with rhythm outcome. *PLoS ONE*. 2012;7:e44165.
40. Kunamalla A, Ng J, Parini V, Yoo S, McGee KA, Tomson TT, et al. Constitutive expression of a dominant-negative TGF- β type II receptor in the posterior left TGF- β type II receptor in the posterior left atrium leads to beneficial remodeling of atrial fibrillation substrate. *Circ Res*. 2016;119:69–82.
41. Teodorescu P, Pasca S, Jurj A, Gafencu G, Joelsson JP, Selicean S, et al. Transforming growth factor- β -mediated CD44/STAT3 signaling contributes to the development of atrial fibrosis and fibrillation. *Basic Res Cardiol*. 2017;112:58.
42. Guo Y, Lip GY, Apostolakis S. Inflammation in atrial fibrillation. *J Am Coll Cardiol*. 2012;60:2263–70.
43. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–30.
44. Zhang M, Perino A, Ghigo A, Hirsch E, Shah AM. NADPH oxidases in heart failure: poachers or gamekeepers? *Antioxid Redox Signal*. 2013;18(9):1024–41.
45. Pignatelli P, Pastorì D, Carnevale R, Farcomeni A, Cangemi R, Nocella C, et al. Serum NOX2 and urinary isoprostanes predict vascular events in patients with atrial fibrillation. *Thromb Haemost*. 2015;113(3):617–24.
46. Saito S, Teshima Y, Fukui A, Kondo H, Nishio S, Nakagawa M, et al. Glucose fluctuations increase the incidence of atrial fibrillation in diabetic rats. *Cardiovasc Res*. 2014;104(1):5–14.
47. Chen WJ, Chang SH, Chan YH, Lee JL, Lai YJ, Chang GJ, et al. Tachycardia-induced CD44/NOX4 signaling is involved in the development of atrial remodeling. *J Mol Cell Cardiol*. 2019;135:67–78.
48. Lu G, Li J, Zhai Y, Li Q, Xie D, Zhang J, et al. Spironolactone suppresses aldosterone-induced Kv1.5 expression by attenuating mineralocorticoid receptor-Nox1/2/4-mediated ROS generation in neonatal rat atrial myocytes. *Biochem Biophys Res Commun*. 2019;520(2):379–84.
49. Violi F, Carnevale R, Calvieri C, Nocella C, Falcone M, Farcomeni A, et al. Nox2 up-regulation is associated with an enhanced risk of atrial fibrillation in patients with pneumonia. *Thorax*. 2015;70(10):961–6.
50. Kim YM, Guzik TJ, Zhang YH, Zhang MH, Kattach H, Ratnatunga C, et al. A myocardial Nox2 containing NAD(P)H oxidase contributes to oxidative stress in human atrial fibrillation. *Circ Res*. 2005;97(7):629–36.
51. Planagumà A, Domènech T, Pont M, Calama E, García-González V, López R, et al. Combined anti CXC receptors 1 and 2 therapy is a promising anti-inflammatory treatment for respiratory diseases by reducing neutrophil migration and activation. *Pulm Pharmacol Ther*. 2015;34:37–45.
52. Zhang X, Guo R, Kambara H, Ma F, Luo HR. The role of CXCR2 in acute inflammatory responses and its antagonists as anti-inflammatory therapeutics. *Curr Opin Hematol*. 2019;26(1):28–33.
53. Zhang YL, Cao HJ, Han X, Teng F, Chen C, Yang J, et al. Chemokine receptor CXCR-2 initiates atrial fibrillation by triggering monocyte mobilization in mice. *Hypertension*. 2020;76(2):381–92.
54. Wang L, Zhang YL, Lin QY, Liu Y, Guan XM, Ma XL, et al. CXCL1-CXCR2 axis mediates angiotensin II-induced cardiac hypertrophy and remodeling through regulation of monocyte infiltration. *Eur Heart J*. 2018;39:1818–31.
55. Schneider M, Hansen JL, Sheikh SP. S100A4: a common mediator of epithelial-mesenchymal transition, fibrosis and regeneration in diseases? *J Mol Med (Berl)*. 2008;86(5):507–22.
56. Lawson WE, Polosukhin VV, Zoia O, Stathopoulos GT, Han W, Plieth D, Loyd JE, Neilson EG, Blackwell TS. Characterization of fibroblast-specific protein 1 in pulmonary fibrosis. *Am J Respir Crit Care Med*. 2005;171:899–907.
57. Schneider M, Kostin S, Strom CC, Aplin M, Lyngbaek S, Theilade J, et al. S100A4 is upregulated in injured myocardium and promotes growth and survival of cardiac myocytes. *Cardiovasc Res*. 2007;75:40–50.
58. Strøm CC, Kruhøffer M, Knudsen S, Stensgaard-Hansen F, Jonassen TEN, Ørntoft TF, Haunsø S, et al. Identification of a core set of genes that signifies pathways underlying cardiac hypertrophy. *Comp Funct Genomics*. 2004;5:459–70.

59. Kato T, Sekiguchi A, Sagara K, Tanabe H, Takamura M, Kaneko S, et al. Endothelial-mesenchymal transition in human atrial fibrillation. *J Cardiol*. 2017;69(5):706–11.
60. Bhat T, Teli S, Rijal J, Bhat H, Raza M, Khoueiry G. Neutrophil to lymphocyte ratio and cardiovascular diseases: a review. *Expert Rev Cardiovasc Ther*. 2013;11(1):55–9. <https://doi.org/10.1586/erc.12.159>.
61. Shao Q, Chen K, Rha SW, Lim HE, Li G, Liu T. Usefulness of neutrophil/lymphocyte ratio as a predictor of atrial fibrillation: a meta-analysis. *Arch Med Res*. 2015;46(3):199–206.
62. Paquissi FC. The predictive role of inflammatory biomarkers in atrial fibrillation as seen through neutrophil-lymphocyte ratio mirror. *J Biomark*. 2016;2016:8160393.
63. Legere SA, Haidl ID, Légaré JF, Marshall JS. Mast cells in cardiac fibrosis: new insights suggest opportunities for intervention. *Front Immunol*. 2019;28(10):580.
64. Liao C, Akazawa H, Tamagawa M, Ito K, Yasuda N, Kudo Y, et al. Cardiac mast cells cause atrial fibrillation through PDGF-A-mediated fibrosis in pressure-overloaded mouse hearts. *J Clin Invest*. 2010;120:242–53.
65. Hu YF, Chen YJ, Lin YJ, Chen SA. Inflammation and the pathogenesis of atrial fibrillation. *Nat Rev Cardiol*. 2015;12(4):230–43.
66. Kared H, Camous X, Larbi A. T cells and their cytokines in persistent stimulation of the immune system. *Curr Opin Immunol*. 2014;29:79–85.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

