



Research article

Predicting potential biomarkers and immune infiltration characteristics in heart failure

Xuesi Chen, Qijun Zhang and Qin Zhang*

Cardiovascular Department, the Affiliated People's Hospital of Ningbo University, Ningbo, Zhejiang, China

* **Correspondence:** Email: zhangqin1993102@163.com; Tel: +86-15895915928.

Abstract: *Background:* Studies have demonstrated that immune cell activation and their infiltration in the myocardium can have adverse effects on the heart, contributing to the pathogenesis of heart failure (HF). The purpose of this study is used by bioinformatics analysis to determine the potential diagnostic markers of heart failure and establish an applicable model to predict the association between heart failure and immune cell infiltration. *Methods:* Firstly, gene expression profiles of dilated heart disease GSE3585 and GSE120895 were obtained in Gene Expression Omnibus (GEO) database. This study then selected differentially expressed genes (DEGs) in 54 patients with HF and 13 healthy controls. In this study, biomarkers were identified using Least Absolute Shrinkage and Selector Operation (LASSO) and Support Vector Machine-Recursive Feature Elimination (SVM-RFE). Additionally, we evaluated the prognostic discrimination performance by the receiver operating characteristic (ROC) curve. Cell type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) was used for analyzing immune cell infiltration in HF tissues. Lastly, immune biomarkers were correlated with each other. *Result:* After 24 DEGs were analyzed using a combinatorial model of LASSO regression and SVM-RFE analysis, four key genes were obtained, namely NSG1, NPPB, PHLDA1, and SERPINE2. The area under the curve (AUC) of these four genes were greater than 0.8. Subsequently, using CIBERPORT, we also found that compared with normal people, the proportion of M1 macrophages and activated mast cells in heart failure tissues decreased. In addition, correlation analysis showed that NPPB, PHLDA1 and SERPINE2 were associated with immune cell infiltration. *Conclusion:* NSG1, NPPB, PHLDA1 and SERPINE2 were identified as potential biomarkers of heart failure. It reveals the comprehensive role of relevant central genes in immune infiltration, which provides a new research idea for the treatment and early detection in heart failure.

Keywords: immune infiltrate; heart failure; machine learning; LASSO; CIBERSORT; SVM-RFE

1. Introduction

Heart failure (HF) is a complicated syndrome associated with cardiac tissue injury, which is usually accompanied by systemic inflammation, arrhythmia and cardiac conduction defects [1]. As a global public health problem with increasing and rapid development, heart failure is the main cause of incidence rate, mortality rate and hospitalization in the world, and is also a major burden of clinical, social and economic burden [2–4].

The heart is also immunoactive and contains all major leukocyte populations, either located in the heart or waiting to penetrate into the heart [5,6]. When the heart is ischemic, the death of damaged and necrotic cardiomyocytes will lead to the activation of immune cells in the heart tissue. Macrophages resident in myocardial tissue undergo a series of obvious phenotypic and functional changes to remove dead cells and matrix fragments, lead to the release of cytokines and growth factors, and stimulate the formation of connective tissue and new blood vessels [7].

Recently, microarray technology and integrated bioinformatics analysis have been applied to identify potential key genes associated with different illnesses, which can be further used as biological markers for diagnosis and prognosis [8–10]. Machine learning method is widely used to find biomarkers. For example, the team used regularized logistic regression to predict biomarkers of spontaneous preterm birth from gene expression data [11]. CIBERPORT is a common computational method for immune cell identification. CIBERSORT contributes to evaluate the relative subpopulations for RNA transcripts and also to quantify cell components [12]. Further, previous articles suggest that CIBERSORT can precisely estimate immune cell landscape in several forms of malignant tumors, such as hepatocellular carcinoma [13] and bladder cancer [14].

In addition to cancer, more and more studies began to focus on the role of immune cell infiltration in heart failure. In this study, two machine-learning methods were employed to explore and identify the key genes of patients associated with heart failure and to analyze the immune cell infiltration. Then further evaluate the correlation between immune cell infiltration and dominant gene in heart failure tissues, so as to provide new research ideas for the treatment and early detection of heart failure.

2. Materials and methods

2.1. Data download and identification of differentially expressed genes (DEGs)

The series of matrix files of the GSE3585 and GSE120895 datasets were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). The GSE3585 was based on the GPL96 platform of Affymetrix Human Genome U133A Array. The GSE120895 was based on the GPL570 platform of Affymetrix Human Genome U133 Plus 2.0 Array. The GSE3585 dataset included 7 HFs and 5 controls collected from left ventricular tissue samples, whereas the GSE120895 dataset included 47 HFs and 8 controls collected from the heart tissues. GSE9800 dataset was used for verification, and also from GEO database. 12 HFs and 11 controls were obtained from GSE9800. The GSE9800 was based on the GPL887 platform of Agilent-012097 Human 1A Microarray (V2) G4110B.

According to the corresponding probe annotation files, the probe IDs of every dataset were

converted to gene symbols. We evaluated probe average to be the ultimate gene level as to multiple probes in line with the same gene symbols. In addition to this, subsequent integration analyses were conducted with the two datasets, which were integrated as one metadata cohort. Moreover, this work adopted R software “SVA” package for removing batch effect [15]. To establish DEGs between the HF and the control group, the “limma” package was selected for R software [16]. Samples with an adjusted false discovery rate $P < 0.05$ and $|\log \text{fold change (FC)}| \geq 1$ were considered as the threshold points for differentially expressed genes. Volcano map and their corresponding heatmaps were plotted with the ggplot2 R package (<https://ggplot2.tidyverse.org>) in combination with heatmap R package (<https://CRAN.R-project.org/package=heatmap>).

2.2. Functional enrichment of DEGs

R software and the “clusterProfiler” package [17] were used to conduct Gene Ontology (GO) analysis. First symbol codes were converted to Entrez ID using Human genome annotation package “org.Hs.eg.db”. Disease ontology (DO) enrichment analyses were performed on DEGs using the “clusterProfiler” and DOSE packages in R [17,18]. We used the “ggplot2” [19] of R software to visualize the plots. Adjusted P-value < 0.05 and Q value < 0.05 were considered significant for GO analysis and DO enrichment analyses.

2.3. Candidate diagnostic biomarker screening and diagnostic value of feature biomarkers

The Least Absolute Shrinkage and Selector Operation (LASSO) applies regularization in improving prediction accuracy. Besides, the “glmnet” package was adopted for analyzing LASSO regression [20]. Support vector machine (SVM) represents the supervised machine-learning technology commonly used in regression and classification. For avoiding overfitting, we selected RFE algorithm for choosing those best genes in meta-data cohort. Thus, Support Vector Machine - Recursive Feature Elimination (SVM-RFE) was conducted for analyzing proper characteristics aiming to recognize the gene set having the maximum discriminative power [21]. The “e1071” package was applied in performing the SVM algorithm [22]. We defined those overlapped genes of these two algorithms as hub genes.

This study produced the receiver operating characteristic (ROC) curve with hub gene expression data for investigating whether those discovered biomarkers were of prediction value. We used the area under curve (AUC) for determining the efficiency in differentiating between HFs and control samples and further validated in the GSE9800 dataset.

2.4. Discovery of immune cell subtypes

We applied CIBERSORT (<https://cibersortx.stanford.edu/>) in immune cell infiltration calculations. We forecasted the possible levels of immune cells with the reference set containing 1000 permutations and 22 immune cell subtypes. We created violin plots with R software “vioplot” package for visualizing different infiltration degrees of immune cells. Meanwhile, CorrPlot package was used to produce a correlation heat map. Adjusted-P value < 0.05 suggested the threshold value for statistical significance of the relevant cell type.

2.5. Correlation analysis of immune cells with diagnostic biomarker

The association of immune cells with featured genes of R software was analyzed by Spearman's rank correlation analysis and "ggplot2" package was used for data visualization. P-value < 0.05 suggested the threshold value for being statistically significant.

2.6. Construction of heart failure model and quantitative real-time polymerase chain reaction (qRT-PCR)

The animal experiment had passed the animal ethics Association of Ningbo University, and the ethics number is 11118. C57BL/6 mice were induced to heart failure through ligation of left anterior descending coronary artery. The control group was the sham operation group, that is, treated with the same surgery without tying the left anterior descending coronary artery. The pale color of the affected myocardium during the procedure confirmed the success of left anterior descending coronary artery ligation. The concrete method was performed as previously reported [23,24]. After 4 weeks, the mice were sacrificed and RNA were extracted from heart tissues by using Trizol reagent (America Invitrogen).

RNA was reverse transcribed into cDNA using primescript one step RT-PCR Kit (Japan Takara). Then cDNA was generated and quantitative PCR was performed by using the SYBR Green (America sigma). All primers used for quantitative PCR were showed as Table 1. The expression levels of heart failure markers ANP, BNP and β -MHC were measured by qRT-PCR. When these markers increase, it proves that our heart failure model is successful [25]. GAPDH served as internal control. Expression levels of all genes were analyzed by using $2^{-\Delta\Delta Ct}$ method.

Table1. Primers used in Quantitative Real-Time PCR.

Primers	Sequence (5'→ 3')	
GAPDH	Forward	5'- AGGTCGGTGTGAACGGATTTG -3'
	Reverse	5'- GGGGTCGTTGATGGCAACA -3'
ANP	Forward	5'- GAGAAGATGCCCGGTAGAAGA-3'
	Reverse	5'- AAGCACTGCCGTCTCTCAGA -3'
BNP	Forward	5'- CTGCTGGAGCTGATAAGAGA -3'
	Reverse	5'- TGCCCAAAGCAGCTTGAGAT -3'
β -MHC	Forward	5'- GTGCCAAGGGCCTGAATGAG-3'
	Reverse	5'- GCAAAGGCTCCAG-GTCTGA-3'
NSG1	Forward	5'- GTGAAGTTGGGGAATAATTTTCGC -3'
	Reverse	5'- CACGACCTTATCTGGGGGTG -3'
PHLDA1	Forward	5'- GGGCTACTGCTCATACCGC -3'
	Reverse	5'- AAAAGTGCAATTCCTTCAGCTTG -3'
NPPB	Forward	5'- GAGGTCACCTCCTATCCTCTGG -3'
	Reverse	5'- GCCATTCCTCCGACTTTTCTC -3'
SERPINE2	Forward	5'- CACATGGGATCGCGTCCATC -3'
	Reverse	5'- CAGCACTTTACCAACTCCGTTTA -3'

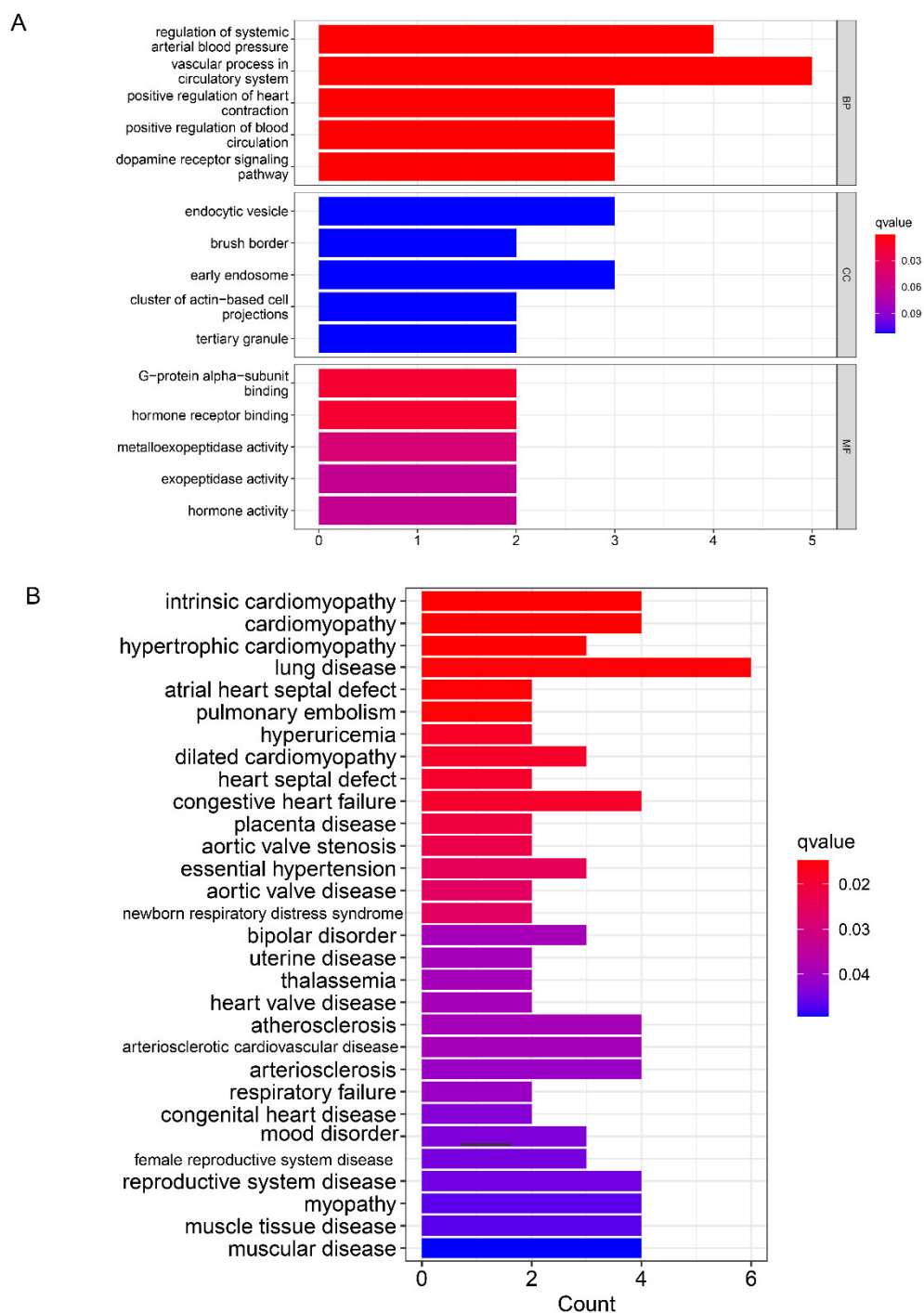


Figure 2. The functional enrichment analysis of DEGs. **(A)** The GO enrichment; **(B)** The DO enrichment.

3.3. Diagnostic biomarker discovery

Then, we further look for diagnostic biomarkers from the 24 DEGs obtained by the following two methods. Using the LASSO regression algorithm and SVM-RFE algorithm, 10 variables were found as diagnostic biomarkers of HF by using the LASSO regression algorithm (Figure 3A), and 6 features

of the DEGs were recognized (Figure 3B) by SVM-RFE algorithm.

Finally, four overlapping features (NSG1, NPPB, PHLDA1, and SERPINE2) of the two algorithms were selected (Figure 3C).

Additionally, ROC analysis was adopted for further evaluation of the predictive value for four characteristic genes. The AUC values were 0.91 for NPPB (Figure 4A), 0.927 for NSG1 (Figure 4B), 0.925 for PHLDA1 (Figure 4C) and 0.899 for SERPINE2 (Figure 4D). It showed that the characteristic biomarkers have a high diagnostic ability.

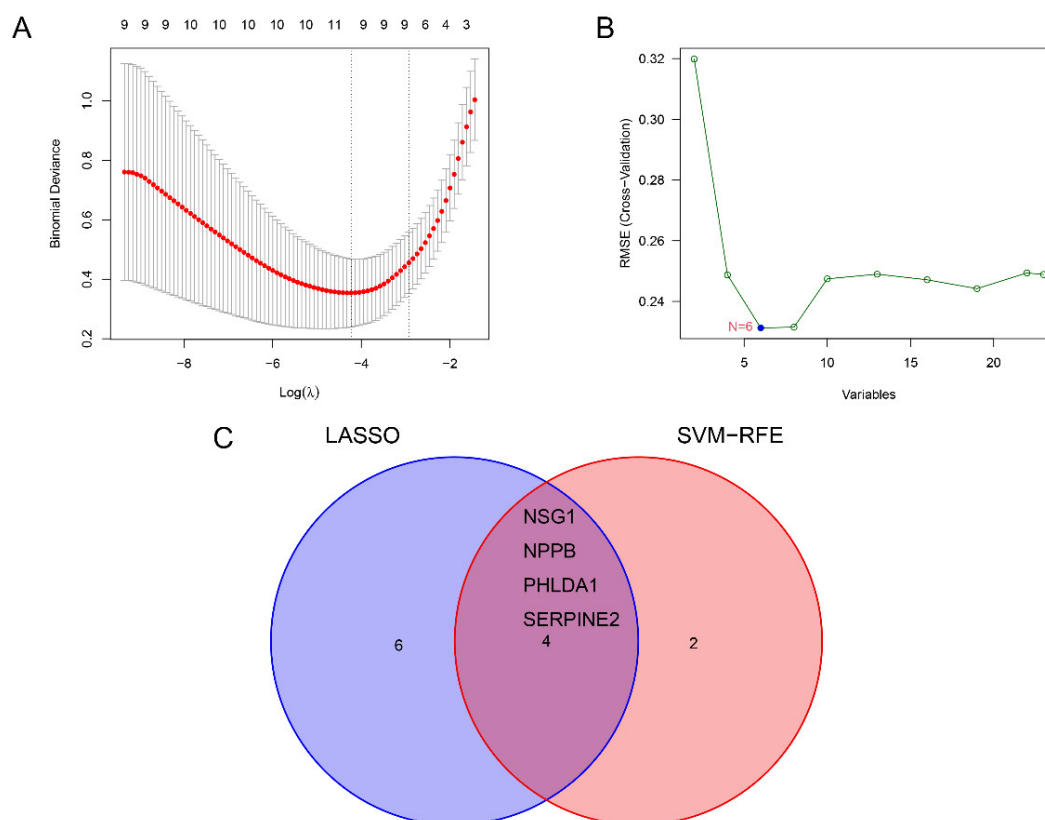


Figure 3. Establishment of prognostic genes. **(A)** Feature screening in LASSO model. **(B)** A plot showing biomarker screening through the SVM-RFE algorithm. **(C)** Venn plot showing those intersected feature genes between SVM-RFE and LASSO regression.

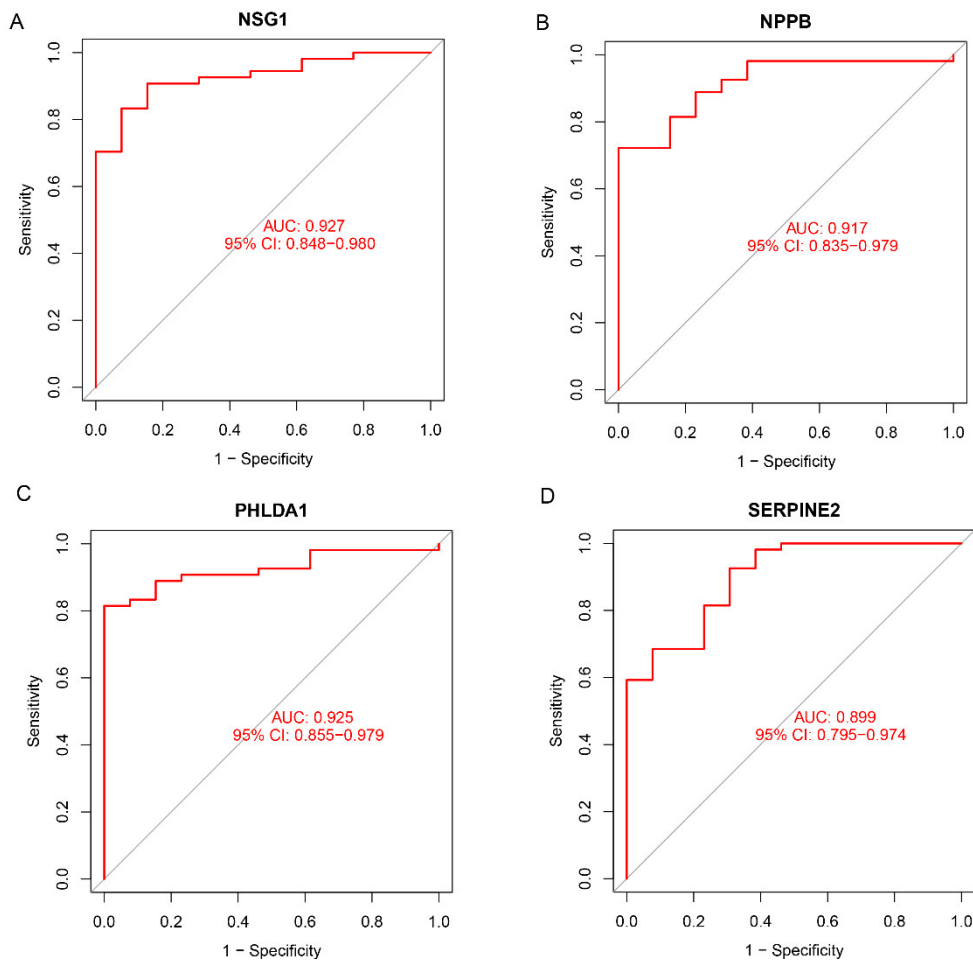


Figure 4. ROC curve showing diagnostic ability of feature genes. **(A–D)** ROC curve for NSG1, NPPB, PHLDA1 and SERPINE2.

3.4. Immune infiltration analysis

Differences between controls and HF tissues concerning immune cell infiltration were analyzed with 22 subpopulations of immune cells based on the CIBERSORT algorithm. As shown in Figure 5A, the percentage for 22 immune cells has been displayed. The proportion of macrophages M1 and activated mast cells in heart failure tissues decreased, as shown in the vioplot (Figure 5B).

Then we analyzed the correlation of 22 immune cells. The correlation heatmap of immune cells also revealed the relationship between different immune cells (Figure 6). The Figure 6 showed M0 Macrophages were observed to be in a positive correlation with activated memory CD4T cells ($R = 0.96$) and M1 macrophages ($R = 0.58$), while naive B cells were negatively correlated with memory B cells ($R = -0.82$).

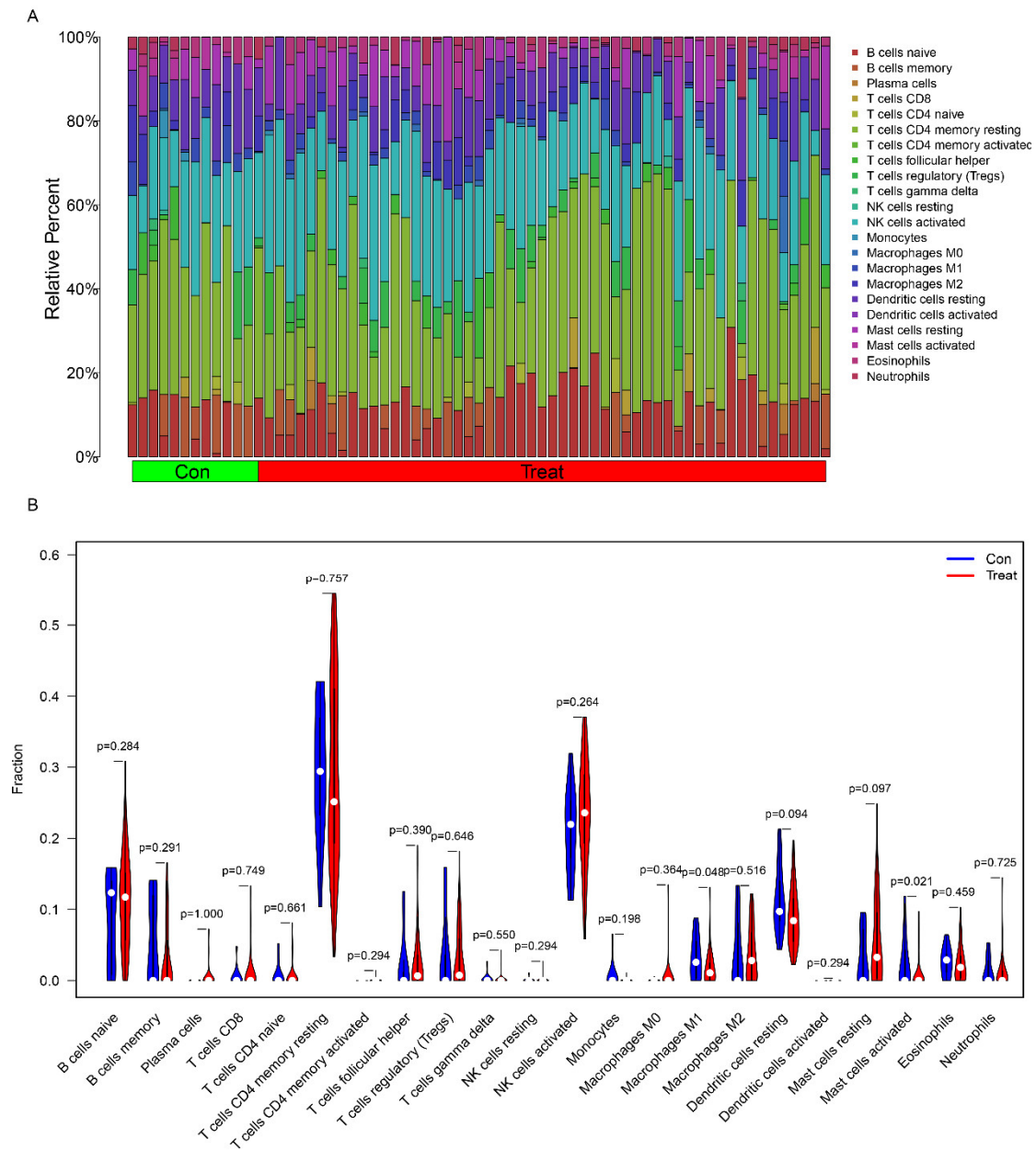


Figure 5. Visualization and evaluation of infiltration of immune cells. **(A)** Relative percentage of 22 types of immune cells; **(B)** Violin plot showing diverse immune cells. The red and blue bars stand for peripheral blood from HF cases and normal samples, separately.

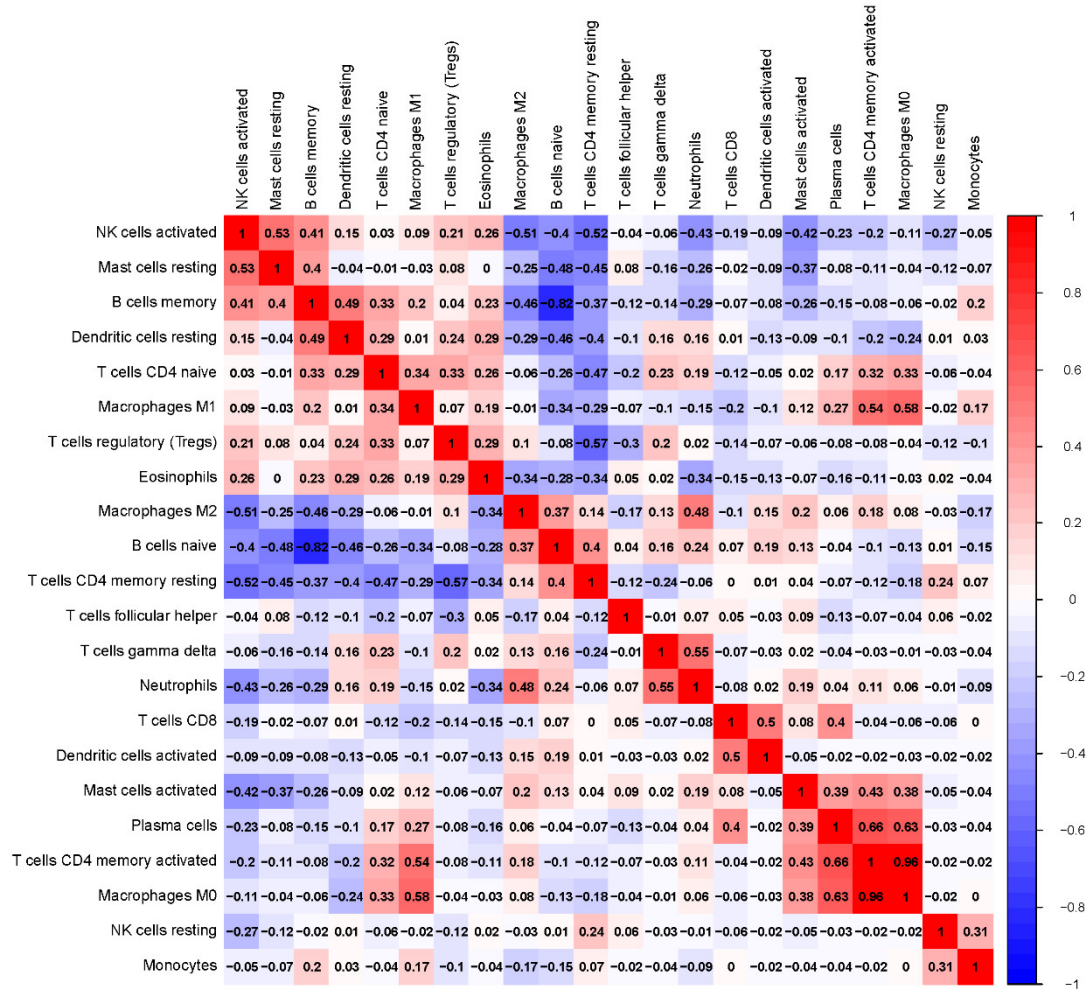


Figure 6. Correlation heatmap for immune cells. The blue and red colors are presented to negative and positive relations of 22 immune cells, separately. The darker color indicates the closer relation.

3.5. Association between diagnostic biomarker and immune cells

According to Figure 7A, NPPB showed positive relation to T follicular helper cells ($R = 0.28$, $p = 0.022$) but a negative correlation was found with resting dendritic cells ($R = -0.33$, $p = 0.006$). PHLDA1 showed positive relation to T follicular helper cells ($R = 0.26$, $p = 0.031$) and M0 Macrophages ($R = 0.26$, $p = 0.032$), but a negative correlation with the remaining Dendritic cells ($R = -0.27$, $p = 0.028$) (Figure 7B) was observed. SERPINE2 and Plasma cells were positively associated with each other (Figure 7C). No correlation was seen between NSG1 and immune cell infiltration ($R = 0.34$, $p = 0.0055$) (Figure 7D).

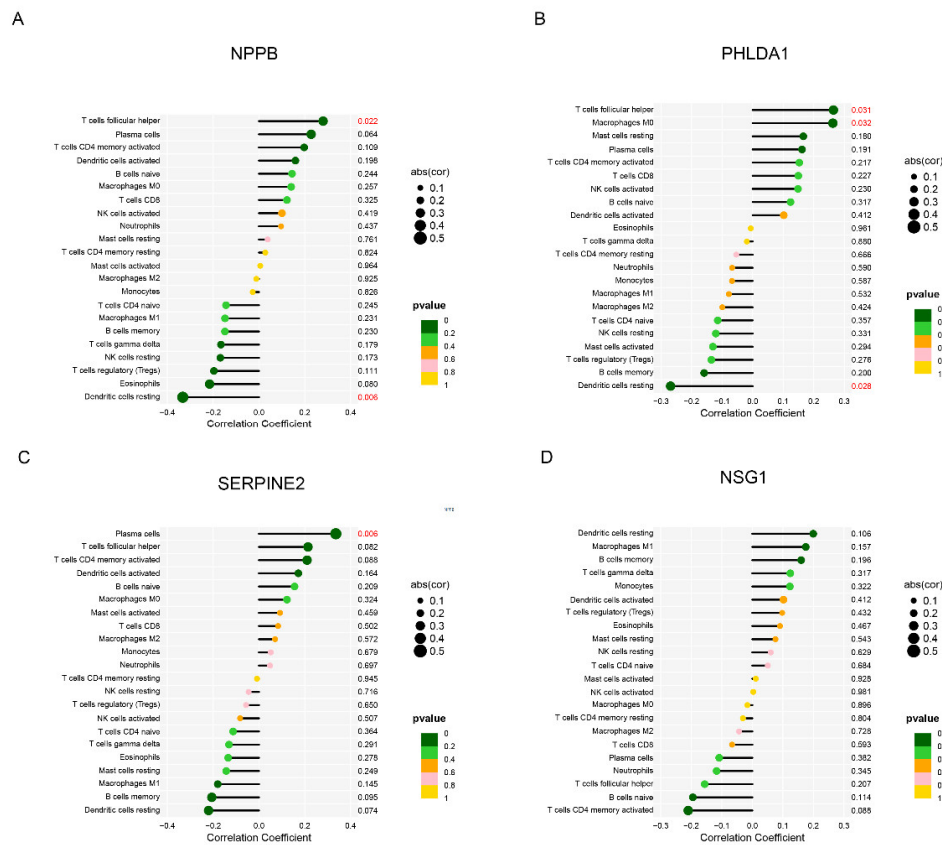


Figure 7. Correlation of NPPB, PHLDA1, SERPINE2 and NSG1 with immune cells. **(A)** Association of NPPB with immune cells. **(B)** Association of PHLDA1 with immune cells. **(C)** Association of SERPINE2 with immune cells. **(D)** Association of NSG1 with immune cells. Dot size stands for association strength of gene biomarkers with immune cells. Dot color stands for P -value.

3.6. Validation of diagnostic biomarkers

Finally, in order to produce more accurate and reliable results, GSE9800 dataset is used to verify the expression of these four characteristic genes. The expression levels of NSG1 in heart failure tissues were significantly lower than those in the control group (Figures 8A,B; $P < 0.05$), and the expression levels of NPPB and PHLDA1 in heart failure tissues were significantly higher than those in the control group (Figure 8C; $P < 0.05$). SERPINE2 was up-regulated in the heart failure group, but there was no significant difference in SERPINE2 (Figure 8D; $P > 0.05$).

The diagnostic ability of other three biomarkers in discriminating HF from the control samples in GSE9800 dataset demonstrated a favorable diagnostic value. The AUC of NSG1 was 0.848 (95% CI 0.667–0.970) (Figure 8E), NPPB was 0.784 (95% CI 0.572–0.981) (Figure 8F), PHLDA1 was 0.833 (95% CI 0.621–0.985) (Figure 8G), and the AUC of SERPINE2 was 0.720 (95% CI 0.492–0.909) (Figure 8H).

Then we constructed the heart failure model in mice, and obtained the heart tissue of heart failure mice. The heart failure markers ANP, BNP and β -MHC increased significantly in the heart failure

model group (Figures 9A–C; $P < 0.05$), indicating the success of our heart failure model. Compared with the control group, we observed a significant increase in the expression levels of NSG1, NPPB, PHLDA1 and SERPINE2 in heart failure tissues (Figures 10A–D, $P < 0.05$).

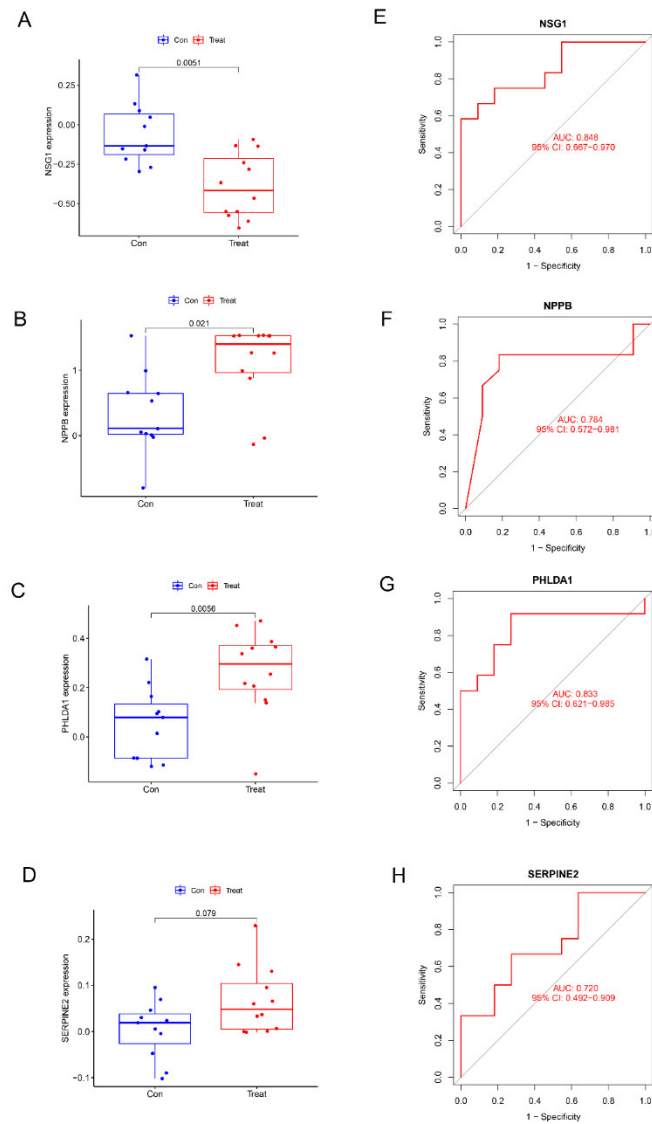


Figure 8. Validation of the expression of diagnostic biomarkers in the GSE9800 dataset. (A–D) The expression level of NSG1, NPPB, PHLDA1 and SERPINE2. (E–H) ROC curve for NSG1, NPPB, PHLDA1 and SERPINE2.

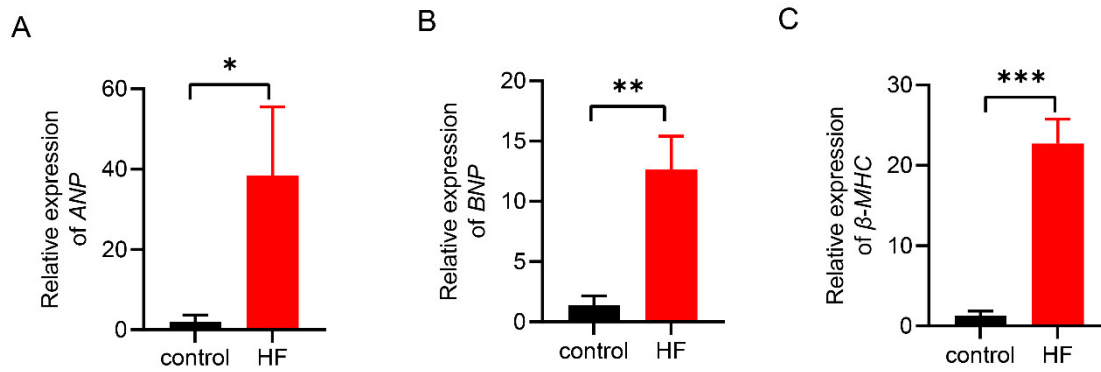


Figure 9. Expression of heart failure markers. (A–C) The expression level of ANP, BNP and β -MHC. * represents $P < 0.05$, ** represents $P < 0.01$. *** represents $P < 0.001$.

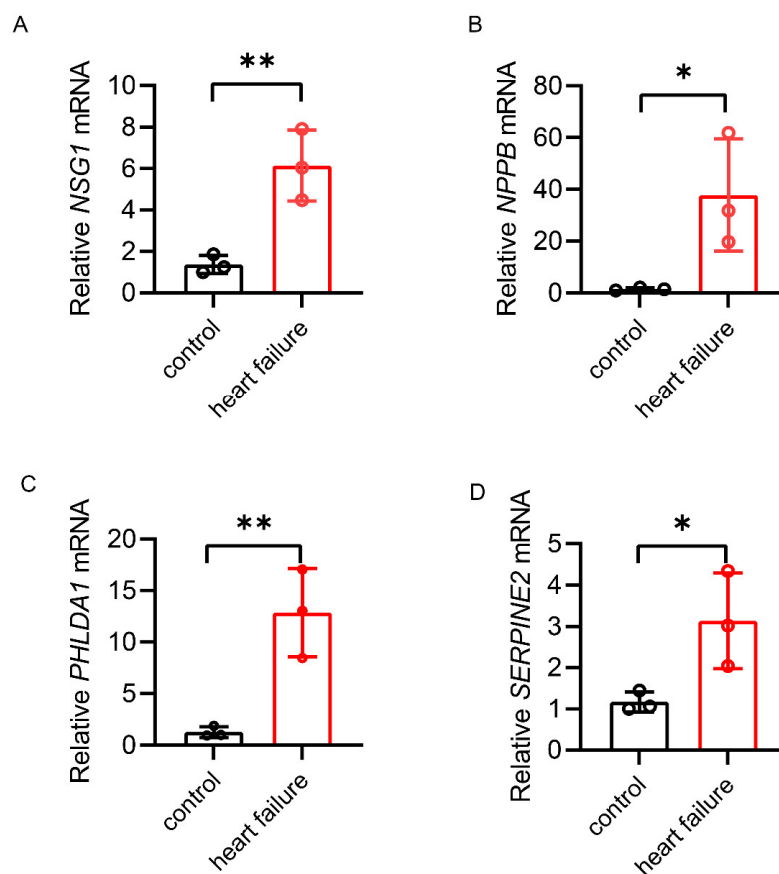


Figure 10. Validation of the expression of diagnostic biomarkers in the mice heart failure model. (A–D) The expression level of NSG1, NPPB, PHLDA1 and SERPINE2. * represents $P < 0.05$, ** represents $P < 0.01$.

4. Discussion

Gene expression differences exert significant effects both on heart failure diagnosis and therapy. During the previous decades, microarray technology and bioinformatics analysis have been extensively used to screen gene changes at the genomic level, to propose novel personalized therapies, and to explore the functions of associated biomolecules.

Support vector machine is one of the important tools of machine-learning [22]. It has excellent performance in classification and prediction, and is widely used in disease diagnosis or medical assistance. However, at first, the application of SVM in analyzing biomedical data was limited. The proposal of SVM-RFE enables us to accurately select variables and explain the direction and intensity of the correlation between predictors and results when analyzing biomedical data [26]. LASSO is a regression-based method, which allows the use of a large number of covariates in the model and has the unique characteristics of punishing the absolute value of regression coefficient [27]. The traditional regression technology is widely limited in analyzing and synthesizing a large number of covariates. Therefore, LASSO has been more and more widely used in biomedicine [28].

Here, a combinatorial assay of the lasso regression model and SVM-RFE algorithm was adopted to select candidate biomarkers with the highest value for diagnosis: NSG1, NPPB, PHLDA1, and SERPINE2.

In addition to NSG1, almost all genes had been well verified. There are two probable reasons for the increase of NSG1 in the heart failure mice model. One is that the species are different. In the GEO database, we choose human samples, but we use mice to simulated heart failure model. Second, we use the heart failure model caused by myocardial infarction in mice, but there are many reasons for heart failure. For example, studies have used single-cell RNA-sequencing to reveal the cell types involved in heart failure, and researchers have found the potential difference between ischemic cardiomyopathy and dilated cardiomyopathy in heart failure. For example, ischemic cardiomyopathy has more specific immune cell types than dilated cardiomyopathy, such as macrophages and dendritic cells [29]. But in general, NPPB, PHLDA1 and SERPINE2 have been repeatedly certified in the experimental results, which further verifies the reliability of our model.

SERPINE2 is a serine protease inhibitor, a 44 kDa to 50 kDa glycoprotein, and a key regulator of biological processes inside and outside cells [30]. According to the current research, SERPINE2 is expressed in many organs, including the heart [31]. A recent observation shows that SERPINE2 is closely related to cardiac tissue fibrosis. The researchers observed the increased expression of SERPINE2 in pathological cardiac fibrosis and found that SERPINE2 caused cardiac fibrosis mainly by increasing collagen deposition, and ERK1/2 signaling pathway played an important role in this process [32]. Another study also found that aspirin can reduce cardiac interstitial fibrosis by inhibiting ERK1/2-SERPINE2 signaling pathway [33]. PHLDA1, known as Pleckstrin homology-like domain, family A, member 1, was first identified in T cell receptor activation induced apoptosis [34]. Up to now, many studies have found that *phlda1* is closely related to a variety of heart diseases, such as; Ischemic cardiomyopathy [35], dilated cardiomyopathy [36], myocardial damage [37]. Similarly, through bioinformatics studies, we observed that PHLDA1 was highly expressed in heart failure tissues and was significantly correlated with T follicular helper cells. PHLDA1 is expected to become a new marker of heart failure.

In order to further explore the infiltration of immune cells in heart failure tissues and the role of these four genes in the pathogenesis of heart failure. Using CIBERSORT, we observed that M1 macrophages and activated mast cells infiltrated differently in heart failure tissues and normal tissues.

In addition, NPPB and PHLDA1 showed positive relation to T follicular helper cells. SERPINE2 was positively associated with plasma cells.

As far as we know, mast cells are one of the resident cell groups in heart tissue, which may play an important role in the progression of myocardial infarction. Stimulated by a series of inflammatory factors, mast cells release TNF and histamine, triggering a series of inflammatory reactions after myocardial infarction [38].

Macrophage population is also one of the resident cell groups in myocardial tissue [39]. Heart failure is generally accompanied by cardiac inflammation and subsequent tissue damage. Macrophages are activated and infiltrated into the myocardium, coordinate with other immune cells, and become a key regulator of tissue repair or regeneration [40]. However, when infiltrating macrophages are over stimulated by inflammatory cytokines, the macrophages will have functional disorder, which can then lead to abnormal repair and sustained injury of cardiomyocytes, and finally progress to heart failure [40].

CIBERSORT results showed that when heart failure occurs, various types of immune cells are infiltrating into myocardial tissue, and the immune cells infiltrating into cardiac tissue play a complex mechanism. However, this study had some limitations and requires further experimental data, such as confirmation of immunophenotype by immunohistochemistry.

5. Conclusions

In short, our study identified four key genes through two machine-learning methods. In the future, it is possible to distinguish patients with heart failure from patients without heart failure by detecting the expression of these four genes. At the same time, combined with CIBERPORT, the identified immune cells may be developed as the target of immunotherapy for patients with heart failure, providing new insights for the diagnosis and treatment of patients with heart failure.

Acknowledgments

This study was supported by the Natural Science Foundation of Zhejiang Province (LY20H020001). Thanks for the support of “Yinzhou district science and technology project in the field of agricultural and social development” with NO.20201YZQ010102, NO.20211YZQ010088 and 20221YZQ070021.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. D. Snipelisky, S. P. Chaudhry, G. C. Stewart, The many faces of heart failure, *Card. Electrophysiol. Clin.*, **11** (2019), 11–20. <https://doi.org/10.1016/j.ccep.2018.11.001>
2. A. L. Bui, T. B. Horwich, G. C. Fonarow, Epidemiology and risk profile of heart failure, *Nat. Rev. Cardiol.*, **8** (2011), 30–41. <https://doi.org/10.1038/nrcardio.2010.165>
3. J. B. Young, The global epidemiology of heart failure, *Med. Clin. North Am.*, **88** (2004), 1135–1143. <https://doi.org/10.1016/j.mcna.2004.06.001>

4. L. Klein, M. Gheorghiadu, Coronary artery disease and prevention of heart failure, *Med. Clin. North Am.*, **88** (2004), 1209–1235. <https://doi.org/10.1016/j.mcna.2004.03.002>
5. M. A. Evans, N. Smart, K. N. Dubé, S. Bollini, J. E. Clark, H. G. Evans, et al., Thymosin β 4-sulfoxide attenuates inflammatory cell infiltration and promotes cardiac wound healing, *Nat. Commun.*, **4** (2013), 2081. <https://doi.org/10.1038/ncomms3081>
6. D. P. Nelson, E. Setser, D. G. Hall, S. M. Schwartz, T. Hewitt, R. Klevitsky, et al., Proinflammatory consequences of transgenic fas ligand expression in the heart, *J. Clin. Invest.*, **105** (2000), 1199–1208. <https://doi.org/10.1172/JCI8212>
7. N. G. Frangogiannis, The mechanistic basis of infarct healing, *Antioxid. Redox Signaling*, **8** (2006), 1907–1939. <https://doi.org/10.1089/ars.2006.8.1907>
8. X. Ma, Q. Zhang, H. Zhu, K. Huang, W. Pang, Q. Zhang, Establishment and analysis of the lncRNA-miRNA-mRNA network based on competitive endogenous RNA identifies functional genes in heart failure, *Math. Biosci. Eng.*, **18** (2021), 4011–4026. <https://doi.org/10.3934/mbe.2021201>
9. Y. Shah, A. Verma, A. R. Marderstein, J. White, B. Bhinder, J. S. Garcia Medina, et al., Pan-cancer analysis reveals molecular patterns associated with age, *Cell Rep.*, **37** (2021), 110100. <https://doi.org/10.1016/j.celrep.2021.110100>
10. Y. Y. He, X. M. Xie, H. D. Zhang, J. Ye, S. Gencer, E. P. C. van der Vorst, et al., Identification of hypoxia induced metabolism associated genes in pulmonary hypertension, *Front. Pharmacol.*, **12** (2021), 753727. <https://doi.org/10.3389/fphar.2021.753727>
11. L. Li, Z. P. Liu, Biomarker discovery for predicting spontaneous preterm birth from gene expression data by regularized logistic regression, *Comput. Struct. Biotechnol. J.*, **18** (2020), 3434–3446. <https://doi.org/10.1016/j.csbj.2020.10.028>
12. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, et al., Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods*, **12** (2015), 453–457. <https://doi.org/10.1038/nmeth.3337>
13. C. Huang, C. Zhang, J. Sheng, D. Wang, Y. Zhao, L. Qian, et al., Identification and validation of a tumor microenvironment-related gene signature in hepatocellular carcinoma prognosis, *Front. Genet.*, **12** (2021), 717319. <https://doi.org/10.3389/fgene.2021.717319>
14. X. Zheng, X. Zhou, H. Xu, D. Jin, L. Yang, B. Shen, et al., A novel immune-gene pair signature revealing the tumor microenvironment features and immunotherapy prognosis of muscle-invasive bladder cancer, *Front. Genet.*, **12** (2021), 764184. <https://doi.org/10.3389/fgene.2021.764184>
15. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, *Bioinformatics*, **28** (2012), 882–883. <https://doi.org/10.1093/bioinformatics/bts034>
16. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.*, **43** (2015), e47. <https://doi.org/10.1093/nar/gkv007>
17. G. Yu, L. G. Wang, Y. Han, Q. Y. He, ClusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS: J. Integr. Biol.*, **16** (2012), 284–287. <https://doi.org/10.1089/omi.2011.0118>
18. G. Yu, L. G. Wang, G. R. Yan, Q. Y. He, DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis, *Bioinformatics*, **31** (2015), 608–609. <https://doi.org/10.1093/bioinformatics/btu684>

19. L. Myint, A. Hadavand, L. Jager, J. Leek, Comparison of beginning R students' perceptions of peer-made plots created in two plotting systems: a randomized experiment, *J. Stat. Educ.*, **28** (2020), 98–108. <https://doi.org/10.1080/10691898.2019.1695554>
20. J. H. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J. Stat. Software*, **33** (2010), 1–22. <https://doi.org/10.18637/jss.v033.i01>
21. S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, W. Xu, Applications of support vector machine (SVM) learning in cancer genomics, *Cancer Genomics Proteomics*, **15** (2018), 41–51. <https://doi.org/10.21873/cgp.20063>
22. M. L. Huang, Y. H. Hung, W. M. Lee, R. K. Li, B. R. Jiang, SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier, *Sci. World J.*, **2014** (2014), 795624. <https://doi.org/10.1155/2014/795624>
23. Q. Zhou, J. Deng, X. Pan, D. Meng, Y. Zhu, Y. Bai, et al., Gut microbiome mediates the protective effects of exercise after myocardial infarction, *Microbiome*, **10** (2022), 82. <https://doi.org/10.1186/s40168-022-01271-6>
24. K. Zhao, Y. Li, Z. Zhou, Y. Mao, X. Wu, D. Hua, et al., Ginkgolide A alleviates cardiac remodeling in mice with myocardial infarction via binding to matrix metalloproteinase-9 to attenuate inflammation, *Eur. J. Pharmacol.*, **923** (2022), 174932. <https://doi.org/10.1016/j.ejphar.2022.174932>
25. C. Kong, D. Lyu, C. He, R. Li, Q. Lu, Dioscin elevates lncRNA MANTIS in therapeutic angiogenesis for heart diseases, *Aging Cell*, **20** (2021), e13392. <https://doi.org/10.1111/accel.13392>
26. H. Sanz, C. Valim, E. Vegas, J. M. Oller, F. Reverter, SVM-RFE: selection and visualization of the most relevant features through non-linear kernels, *BMC Bioinf.*, **19** (2018), 432. <https://doi.org/10.1186/s12859-018-2451-4>
27. R. Tibshirani, The lasso method for variable selection in the Cox model, *Stat. Med.*, **16** (1997), 385–395. [https://doi.org/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3)
28. A. J. McEligot, V. Poynor, R. Sharma, A. Panangadan, Logistic LASSO regression for dietary intakes and breast cancer, *Nutrients*, **12** (2020), 2652. <https://doi.org/10.3390/nu12092652>
29. X. Shi, L. Zhang, Y. Li, J. Xue, F. Liang, H. W. Ni, et al., Integrative analysis of bulk and single-cell RNA sequencing data reveals cell types involved in heart failure, *Front. Bioeng. Biotechnol.*, **9** (2021), 779225. <https://doi.org/10.3389/fbioe.2021.779225>
30. J. Sommer, S. M. Gloor, G. F. Rovelli, J. Hofsteenge, H. Nick, R. Meier, et al., cDNA sequence coding for a rat glia-derived nexin and its homology to members of the serpin superfamily, *Biochemistry*, **26** (1987), 6407–6410. <https://doi.org/10.1021/bi00394a016>
31. R. Vidal, J. U. G. Wagner, C. Braeuning, C. Fischer, R. Patrick, L. Tombor, et al., Transcriptional heterogeneity of fibroblasts is a hallmark of the aging heart, *JCI Insight*, **4** (2019), e131092. <https://doi.org/10.1172/jci.insight.131092>
32. X. Li, D. Zhao, Z. Guo, T. Li, M. Qili, B. Xu, et al., Overexpression of serpinE2/protease nexin-1 contribute to pathological cardiac fibrosis via increasing collagen deposition, *Sci. Rep.*, **6** (2016), 37635. <https://doi.org/10.1038/srep37635>
33. X. Li, G. Wang, M. QiLi, H. Liang, T. Li, X. E, et al., Aspirin reduces cardiac interstitial fibrosis by inhibiting erk1/2-serpine2 and P-Akt signalling pathways, *Cell. Physiol. Biochem.*, **45** (2018), 1955–1965. <https://doi.org/10.1159/000487972>

34. C. G. Park, S. Y. Lee, G. Kandala, S. Y. Lee, Y. Choi, A novel gene product that couples TCR signaling to Fas(CD95) expression in activation-induced cell death, *Immunity*, **4** (1996), 583–591. [https://doi.org/10.1016/S1074-7613\(00\)80484-7](https://doi.org/10.1016/S1074-7613(00)80484-7)
35. J. Wang, F. Wang, J. Zhu, M. Song, J. An, W. Li, Transcriptome profiling reveals PHLDA1 as a novel molecular marker for ischemic cardiomyopathy, *J. Mol. Neurosci.*, **65** (2018), 102–109. <https://doi.org/10.1007/s12031-018-1066-6>
36. L. Liu, J. Huang, Y. Liu, X. Pan, Z. Li, L. Zhou, et al., Multiomics analysis of transcriptome, epigenome, and genome uncovers putative mechanisms for dilated cardiomyopathy, *Biomed. Res. Int.*, **2021** (2021), 6653802. <https://doi.org/10.1155/2021/6653802>
37. Y. Guo, P. Jia, Y. Chen, H. Yu, X. Xin, Y. Bao, et al., PHLDA1 is a new therapeutic target of oxidative stress and ischemia reperfusion-induced myocardial injury, *Life Sci.*, **245** (2020), 117347. <https://doi.org/10.1016/j.lfs.2020.117347>
38. N. G. Frangogiannis, M. L. Lindsey, L. H. Michael, K. A. Youker, R. B. Bressler, L. H. Mendoza, et al., Resident cardiac mast cells degranulate and release preformed TNF-alpha, initiating the cytokine cascade in experimental canine myocardial ischemia/reperfusion, *Circulation*, **98** (1998), 699–710. <https://doi.org/10.1161/01.CIR.98.7.699>
39. S. D. Prabhu, N. G. Frangogiannis, The biological basis for cardiac repair after myocardial infarction: from inflammation to fibrosis, *Circ. Res.*, **119** (2016), 91–112. <https://doi.org/10.1161/CIRCRESAHA.116.303577>
40. W. P. Lafuse, D. J. Wozniak, M. V. S. Rajaram, Role of cardiac macrophages on cardiac inflammation, fibrosis and tissue repair, *Cells*, **10** (2020), 51. <https://doi.org/10.3390/cells10010051>



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)