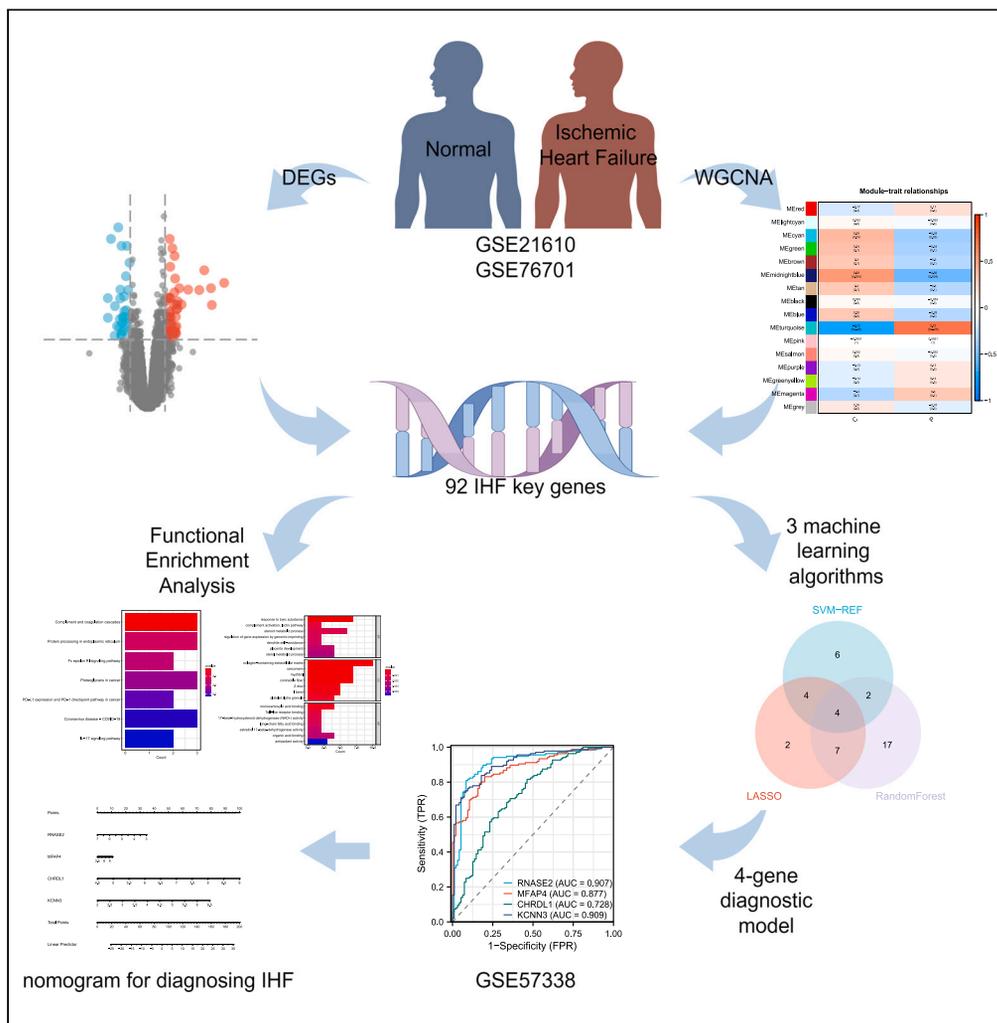**Article**

Identification of immune-related genes for the diagnosis of ischemic heart failure based on bioinformatics

Yiding Yu, Xiujuan Liu, Yitao Xue, Yan Li

xytsdzydfy@126.com (Y.X.)
liyan88130@163.com (Y.L.)

Highlights

We analyze the underlying molecular mechanisms of ischemic heart failure (IHF) and the immune cell infiltration environment within the failing myocardium

WGCNA combines three machine learning algorithms (LASSO, RF, and SVM-RFE) to make the results more accurate

RNASE2, MFAP4, CHRDL1, and KCNN3 could be potential biomarkers and therapeutic targets for IHF

Diagnostic models and nomogram tools providing a new understanding of IHF

Article

Identification of immune-related genes for the diagnosis of ischemic heart failure based on bioinformatics

Yiding Yu,[1] Xiujuan Liu,[2] Yitao Xue,[2,*] and Yan Li[2,3,*]

SUMMARY

The role of immune cells in the pathogenesis of ischemic heart failure (IHF) is well-established. However, identifying key genes in patients with IHF remains a challenge. We obtained two IHF datasets from the GEO database (GSE76701 and GSE21610), and identified four potential diagnostic candidate genes for IHF by using bioinformatics and machine learning algorithms, namely RNASE2, MFAP4, CHRDL1, and KCNN3. We constructed nomogram and validated the diagnostic value of these genes on additional GEO datasets (GSE57338). The results showed that these four genes had high diagnostic value (area under the curve value of 0.961). Furthermore, our immune infiltration analysis revealed the presence of three dysregulated immune cells in IHF, namely macrophages M2, monocytes, and T cells gamma delta. We also explored the potential molecular mechanisms of IHF. These findings provide new insights into the pathogenesis, diagnosis, and treatment of IHF.

INTRODUCTION

Cardiovascular disease is a leading cause of human mortality, encompassing hypertension, arrhythmias, coronary heart disease, and heart failure, among other heart-related conditions. Of these, heart failure is often the ultimate outcome of most cardiovascular diseases, primarily due to structural changes or functional impairments of the heart that hinder ventricular filling or ejection function.[1] With the advent of interventional techniques and drugs, survival rates for patients with ischemic heart disease and acute myocardial infarction are improving, leading to an increased number of patients at risk for heart failure.[2] Damaged and unrecoverable myocardium is the principal cause of heart failure and death in patients within five years of an infarction.[3] Consequently, identifying myocardial-specific genes for ischemic heart failure (IHF) at the transcriptome level could facilitate early diagnosis and targeted treatment of these patients.

Previous studies have demonstrated that immune cells play protective and destructive roles in cardiac remodeling after infarction.[4] Multipotent cells of the innate immune system, such as monocytes and macrophages, are critical for the initial inflammatory response to post-infarction myocardial injury and subsequent wound healing.[5] Clinical studies have aimed to target elements of the immune response in heart failure by modulating the inflammatory response. However, the results of these clinical studies have been unsatisfactory, and in some cases, have even exacerbated heart failure.[6] Therefore, examining the immune response and associated genes in the progression of IHF could help develop novel therapeutic strategies.

Rapid advances in high-throughput technologies and bioinformatics can aid in the screening of sensitive and specific diagnostic tools to diagnose and treat heart failure patients before they reach a refractory end stage. Additionally, with the development and maturation of machine learning in bioinformatics applications, multiple machine learning models can aid in uncovering potential mechanisms, prospective diagnostic tools, and therapeutic targets for IHF.[7]

In this study, we initiated by acquiring two IHF datasets from the GEO database, followed by identifying differentially expressed genes (DEGs) through the Limma package. Subsequently, we selected significant modular genes via weighted gene co-expression network analysis (WGCNA). We then performed functional enrichment analysis. We utilized three different machine learning models, namely, least absolute shrinkage and selection operator (LASSO), random forest (RF), and support vector machine recursive feature elimination (SVM-RFE), to determine diagnostic models. We further assessed these genes through nomogram and single-sample gene set enrichment analysis (ssGSEA). Additionally, we validated the diagnostic models on another GEO dataset. Finally, we conducted immune cell infiltration analysis on the dataset to identify immune-related genes in IHF and to elucidate the role of immune cells in the development of this condition. Figure 1 depicts the study flowchart.

[1]Shandong University of Traditional Chinese Medicine, Jinan 250014, China
[2]Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan 250014, China
[3]Lead contact
*Correspondence: xytsdzydfy@126.com (Y.X.), liyan88130@163.com (Y.L.)
https://doi.org/10.1016/j.isci.2023.108121
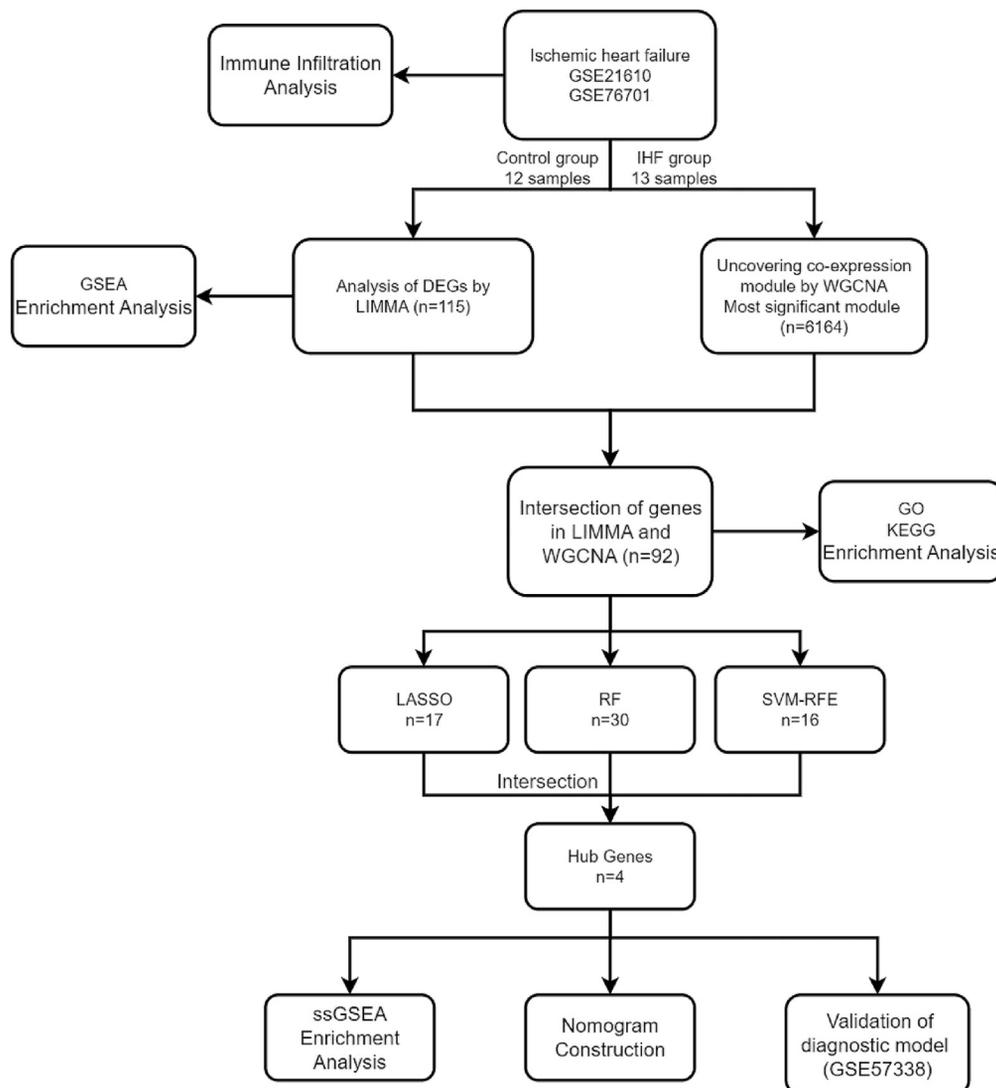
Figure 1. The study flowchart. This is a summary of our study as a whole

RESULTS

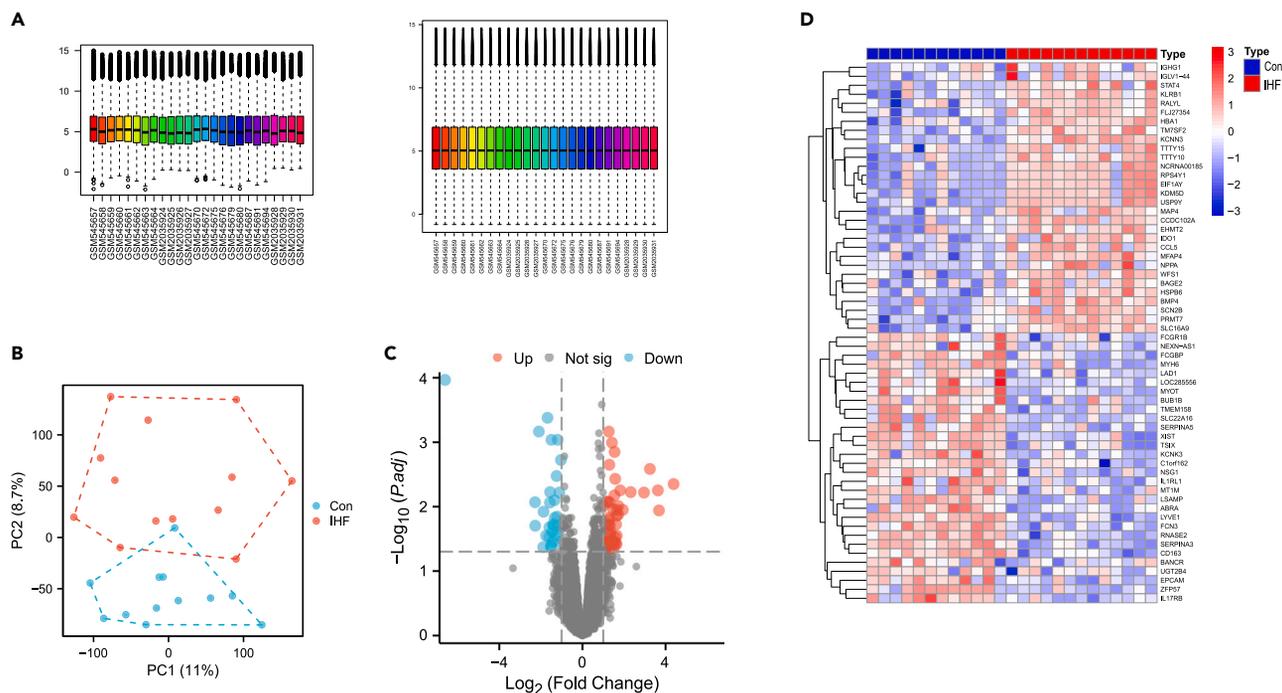### Identification of differentially expressed genes

After data merging and normalization, we identified a total of 115 differential genes in the merged dataset using the LIMMA package, of which there were 78 up-regulated and 37 down-regulated genes. The visualization of the associated results was shown in Figure 2.

Weighted gene co-expression network analysis and key module identification

We used WGCNA to identify the most relevant modules in IHF. The WGCNA package will automatically select the best β value for calculation at runtime. In this study, when the value of β = 6, the network is closer to the scale-free network. Ultimately, we obtained 16 gene co-expression modules. Among them, the turquoise module had the highest correlation with IHF (correlation coefficient = 0.72, p = 5e-05), containing a total of 6164 genes. Therefore, we selected the turquoise module as the key module for the subsequent analysis. The visualization of the results was shown in Figure 3.

Functional enrichment analysis of ischemic heart failure

To ensure that genes with significant functions but insignificant differential multiplicities were not overlooked, we conducted a GSEA on all the genes in the dataset. GSEA results indicate that genes related to IHF may be involved in biological functions such as programmed cell death, inflammation, and immunity. Among them, genes involved in antigen processing and presentation were generally up-regulated, while genes involved in apoptosis, proteasome, and Th1 and Th2 cell differentiation were generally down-regulated. Afterward, we intersected DEGs with

Figure 2. Results of differentially expressed genes

(A) Normalization of data. We have completed the normalization of samples, so that the data of each sample and parallel experiments are at the same level, making downstream analysis more accurate and reliable.

(B) Principal component analysis plot. The results show that the repeatability within the group is relatively good, the sample data are very similar, and there is good discrimination between groups.

(C) Volcano plot. We set adjust p values <0.05 and | log2(FC)| ≥ 1 as the difference genes. Red dots are upregulated genes, and blue dots are downregulated genes.

(D) Heatmap plot. We showed the top 30 upregulated and the top 30 downregulated genes in IHF and control groups.

the turquoise module genes to obtain 92 IHF key genes. We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis on the IHF key genes. The categories of GO analysis include biological process (BP), cellular component (CC), and molecular function (MF), and we will show the top five results, respectively. The KEGG enrichment analysis did not yield many pathways, so we will show all the enriched results. The visualization results were shown in Figure 4.

## Identification of hub genes via machine learning

We used three machine learning algorithms, LASSO, RF, and SVM-RFE, to further screen Hub genes for IHF. We identified 17 potential candidate biomarkers by the LASSO algorithm. The RF algorithm ranked the genes based on the importance calculation of each gene, and we selected the top 30 as potential candidates for IHF. The SVM-RFE algorithm exhibited the highest precision, identifying 13 genes with a constant precision score of 1 thereafter. To establish the optimal number of Hub genes, we selected the top 16 genes for the SVM-RFE algorithm results as candidate genes. By intersecting the results of all three algorithms, we identified four Hub genes for IHF: RNASE2, MFAP4, CHRDL1, and KCNN3. The visualization results were shown in Figure 5.

Diagnostic value assessment

We constructed nomograms based on the four Hub genes and built receiver operating characteristic (ROC) curves to assess the diagnostic specificity and sensitivity of each gene and nomogram. In addition, we plotted differential expression boxplots for the Hub genes. Finally, we completed the validation of Hub genes in GSE57338 using ROC curve analysis. The validation results showed that the area under the curve (AUC) of each gene was greater than 0.7, and the AUC of the 4-gene diagnostic model was 0.961, which had high diagnostic value. The visualization results were shown in Figure 6.

ssGSEA enrichment analysis

We performed ssGSEA enrichment analysis for RNASE2, MFAP4, CHRDL1, and KCNN3, respectively. The results showed that all these genes are involved in the development of IHF and the biological functions of immunity and inflammation in the disease process to varying degrees.
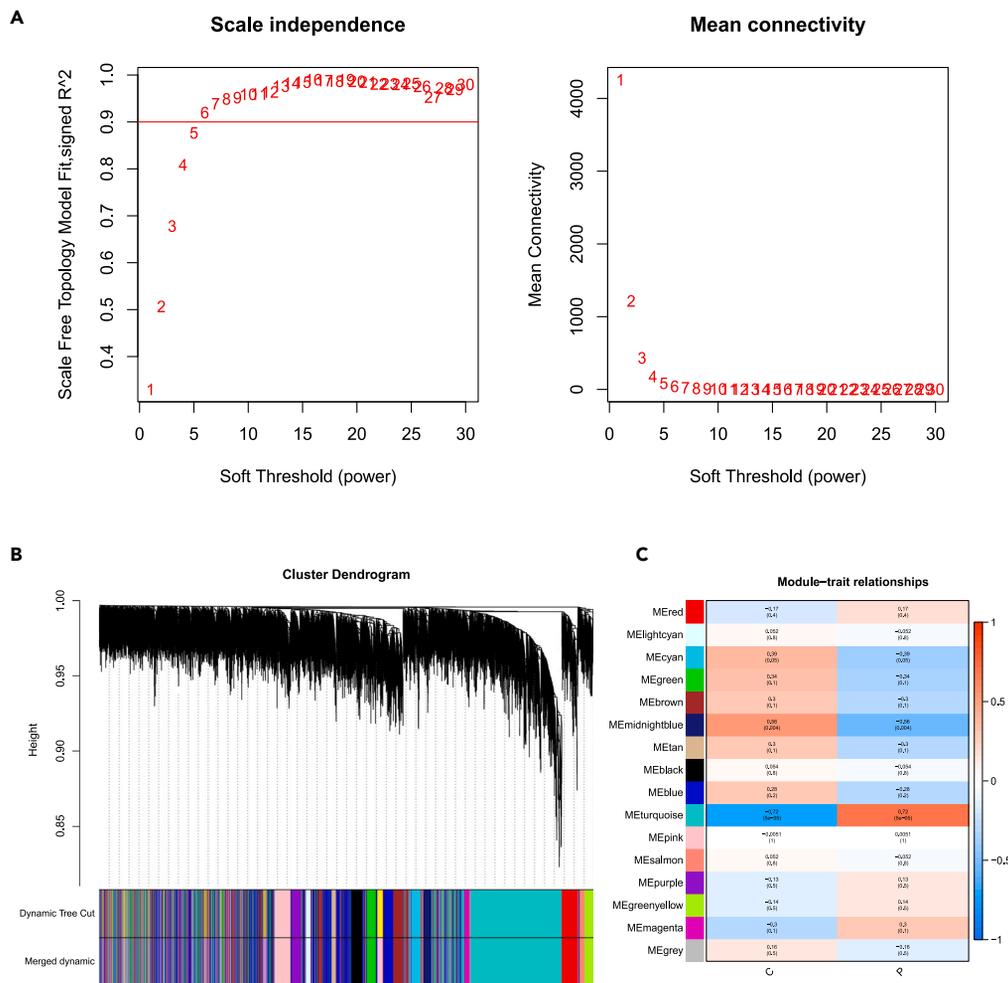
Figure 3. WGCNA analysis result

(A) The scale-free fit index for soft-thresholding powers and mean connectivity.

(B) Gene and trait clustering dendrograms. Gene clustering trees (dendrograms) obtained by hierarchical clustering of neighbor-based differences.

(C) 16 gene co-expression modules. The numbers in each cell means the correlation coefficient and p value. C: Control group; P: IHF group.

For example, RNASE2 was involved in adrenergic signaling in cardiomyocytes, cell cycle, cellular senescence, IL-17 signaling pathway, and proteasome. MFAP4 was involved in biosynthesis of unsaturated fatty acids and histidine metabolism. KCNN3 participates in the TGF-beta signaling pathway. Chordin-like 1(CHRDL1) participates in proteasome, primary immunodeficiency, and circadian rhythm, etc. Detailed results and visualization were shown in Figure 7.

Immune cell infiltration analysis

Due to the important role of immune cells in the development of IHF, we also performed an immune infiltration analysis of the dataset by means of the CIBERSORT algorithm. The bar charts clearly show the content of the different subpopulations in each sample. We assessed the heterogeneity of cell composition between the heart failure samples and the healthy samples and the results showed that there were three immune cell infiltrates that were significantly different. Macrophages M2 and monocytes were higher in normal samples than in heart failure samples, and T cells gamma delta was lower than in heart failure samples. The differential infiltration of these 3 types of immune cells may provide potential regulatory points for the treatment of IHF. The visualization results were shown in Figure 8.

DISCUSSION

HF is always a major public health challenge. Among them, left ventricular systolic dysfunction due to obstructive coronary artery disease is the most common cause of heart failure worldwide. While new treatments such as mechanical unloading and modulation of the inflammatory response look promising, understanding the mechanisms of IHF is a key step in finding new ways to address the risk of heart failure.[8]
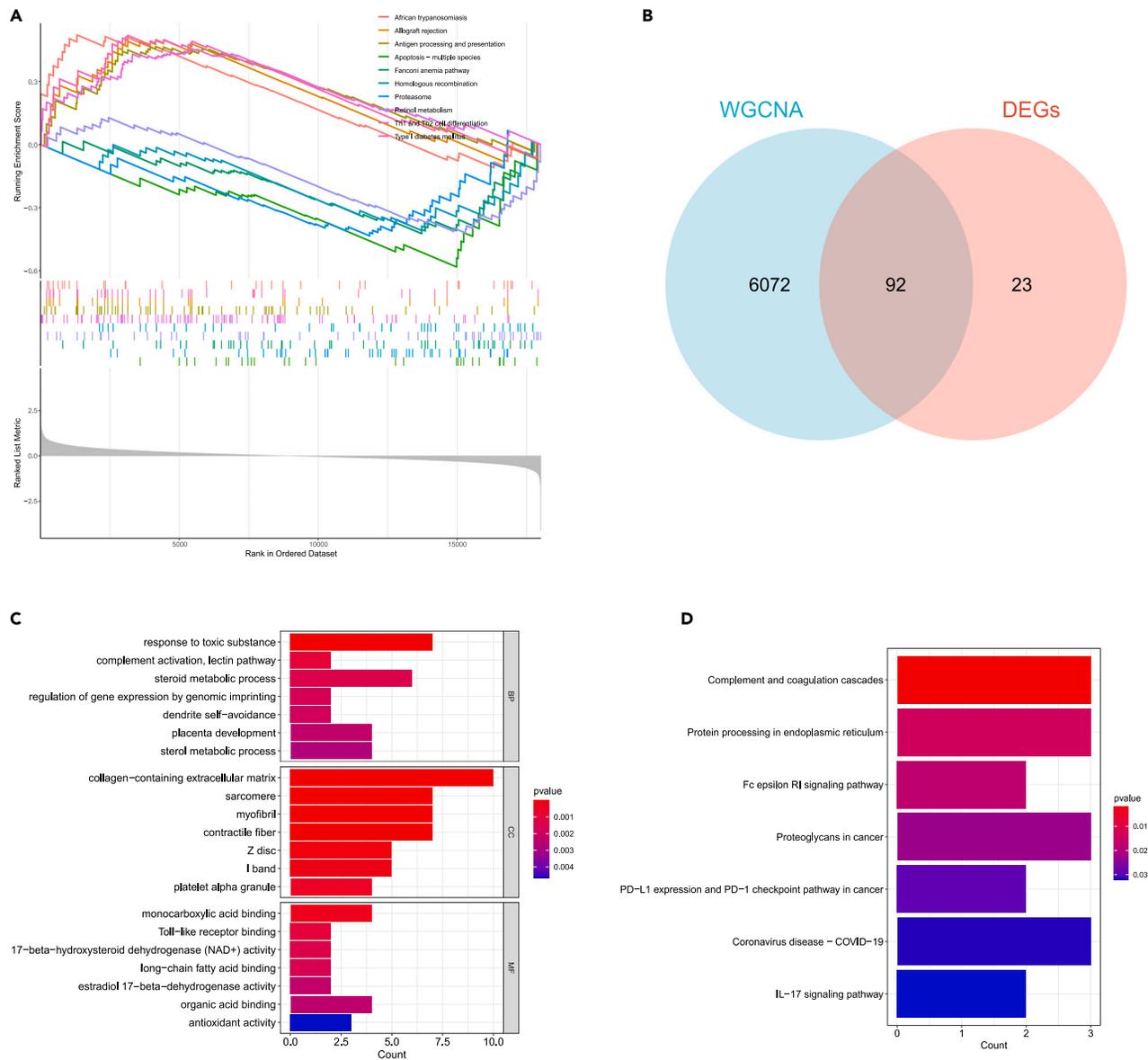
Figure 4. Functional enrichment analysis of IHF

(A) GSEA enrichment analysis results. Include upregulated gene pathway and downregulated gene pathway.

(B) The intersection of DEGs and the turquoise module. We intersected DEGs with the turquoise module genes to obtain 92 IHF key genes.

(C) GO enrichment analysis results.

(D) KEGG enrichment analysis results.

In this study, we explored the potential mechanisms of IHF through three enrichment analyses. The results suggest that the underlying mechanisms of IHF are mostly related to inflammation, immunity, and metabolism. For example, antigen processing and presentation and Th1 and Th2 cell differentiation in the enrichment results are involved in cellular immune processes. It has been demonstrated that cardiac fibroblasts take up and process antigens and promote cardiac fibrosis and dysfunction through IFNγ.[9] Inflammatory cytokines such as IFNγ, TNF-α, IL-2, and IL-6 secreted by Th1 and Th2 cells have also been important players in the development of heart failure.[10,11] The proteasome involves the ubiquitin-proteasome system, which plays a key role in maintaining protein homeostasis and cardiac function.[12] On the one hand, the ubiquitin-proteasome system is the major protein degradation system involved in the regulation of inflammation and selective mitochondrial autophagy during heart failure.[13] On the other hand, this system is also involved in the development of heart failure in terms of cardiac energy metabolism.[14] In addition, how to mitigate the cardiotoxicity induced by proteasome inhibitors of anticancer drugs has been a popular research in recent years.[15]

We also used bioinformatics tools and machine learning algorithms to screen for Hub genes for IHF, namely RNASE2, MFAP4, CHRDL1, and KCNN3, and validated them in a larger dataset. The AUC value of this 4-gene diagnostic model in the validation set was 0.961, implying
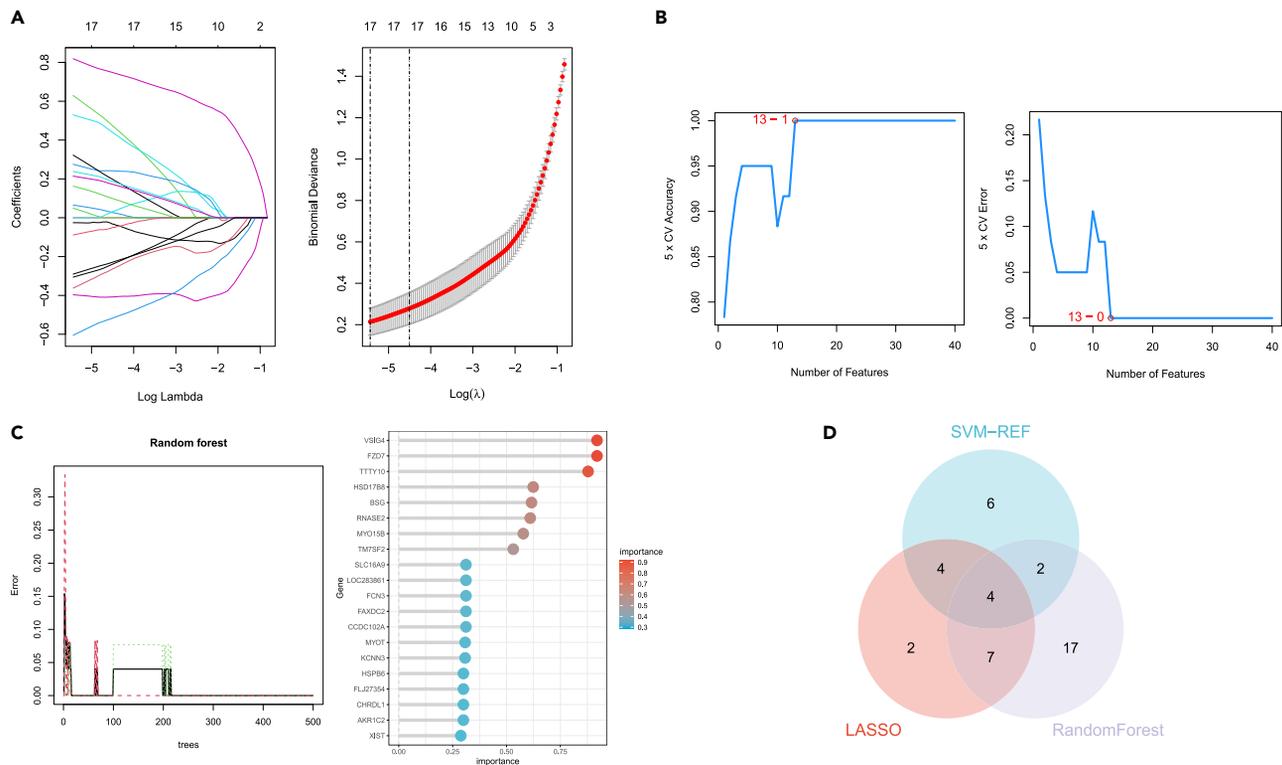
Figure 5. Machine learning in screening candidate diagnostic biomarkers for IHF

(A) Biomarkers screening in the Lasso model. LASSO coefficient profiles of the candidate optimal feature genes and the optimal lambda was determined when the partial likelihood deviance reached the minimum value. Each coefficient curve in the left picture represents a single gene. The solid vertical lines in right picture represent the partial likelihood deviance, and the number of genes (n = 17) corresponding to the lowest point of the cure is the most suitable for LASSO.

(B) Biomarkers screening in the SVM-RFE model. The SVM-RFE algorithm was used to further candidate optimal feature genes with the highest accuracy and lowest error obtained in the curves. The x axis shows the number of feature selections, and the y axis shows the prediction. accuracy.

(C) Biomarkers screening in the RF model. Relative importance of overlapping candidate genes calculated in random forest. We showed the results for the top 20 genes.

(D) Venn diagram shows that four candidate diagnostic genes are identified via the previous three algorithms.

that this diagnostic model has a high diagnostic value. Nomograms were also plotted to aid the application of this diagnostic model. We completed the ssGSEA analysis of these 4 genes in order to provide new ideas for our understanding of the molecular mechanism of IHF.

Ribonuclease A family member 2 (RNASE2) is a non-secretory ribonuclease that belongs to the RNaseA superfamily. Also known as an eosinophil-derived neurotoxin (EDN), RNASE2 is involved in immune and inflammatory related pathways.[16] Its broad antiviral activity is primarily directed against single-stranded RNA viruses, such as human immunodeficiency virus.[17] Studies have shown that RNASE2 acts as a catalyst for human dendritic cells and promotes the secretion of a plethora of cytokines and chemokines.[18] RNASE2 also functions as an endogenous ligand for toll-like receptor (TLR) 2. On the one hand, downstream pathway stimulation via TLR may activate myeloid differentiation factor 88 (MyD88) and mitogen-activated protein kinase (MAPK) to promote the production of pro-inflammatory factors such as IL-10.[19,20] On the other hand, TLR immunosensing against live pathogens may also allow RNASE2 to act as a bridge between innate and adaptive immunity.[21] In the cardiovascular context, TLR2 has been reported to regulate myocardial ischemia, and thus RNASE2 may be involved in innate immune responses in the pathogenesis of IHF via TLR2.[22]

Microfibrillar associated protein 4 (MFAP4) is an extracellular matrix protein that belongs to the fibrinogen-associated protein superfamily. Vascular smooth muscle cells produce MFAP4, which is highly enriched in the vessels of the heart and lungs, and is believed to contribute to the structure and function of elastic fibers.[23] Thus, MFAP4 holds significant research value in arterial vascular-related diseases. Studies have shown that MFAP4 can induce the proliferation and migration of vascular smooth muscle cells and promote monocyte chemotaxis, which can accelerate neointimal proliferation after vascular injury.[24] Additionally, two clinical studies have demonstrated the potential of MFAP4 as a biomarker of atherosclerotic disease. In patients with stable coronary artery disease, serum MFAP4 levels were lower compared to patients with acute infarction.[25] Moreover, MFAP4 has the potential to serve as a biomarker for assessing the degree of coronary artery stenosis in acute heart attacks.[26] Concerning heart failure, MFAP4 plays a crucial role in macrophage infiltration, inflammation, and myocardial fibrosis.[27] MFAP4 knockout experiments have shown that MFAP4 deficiency can lead to dysregulated integration of G protein-coupled receptors and
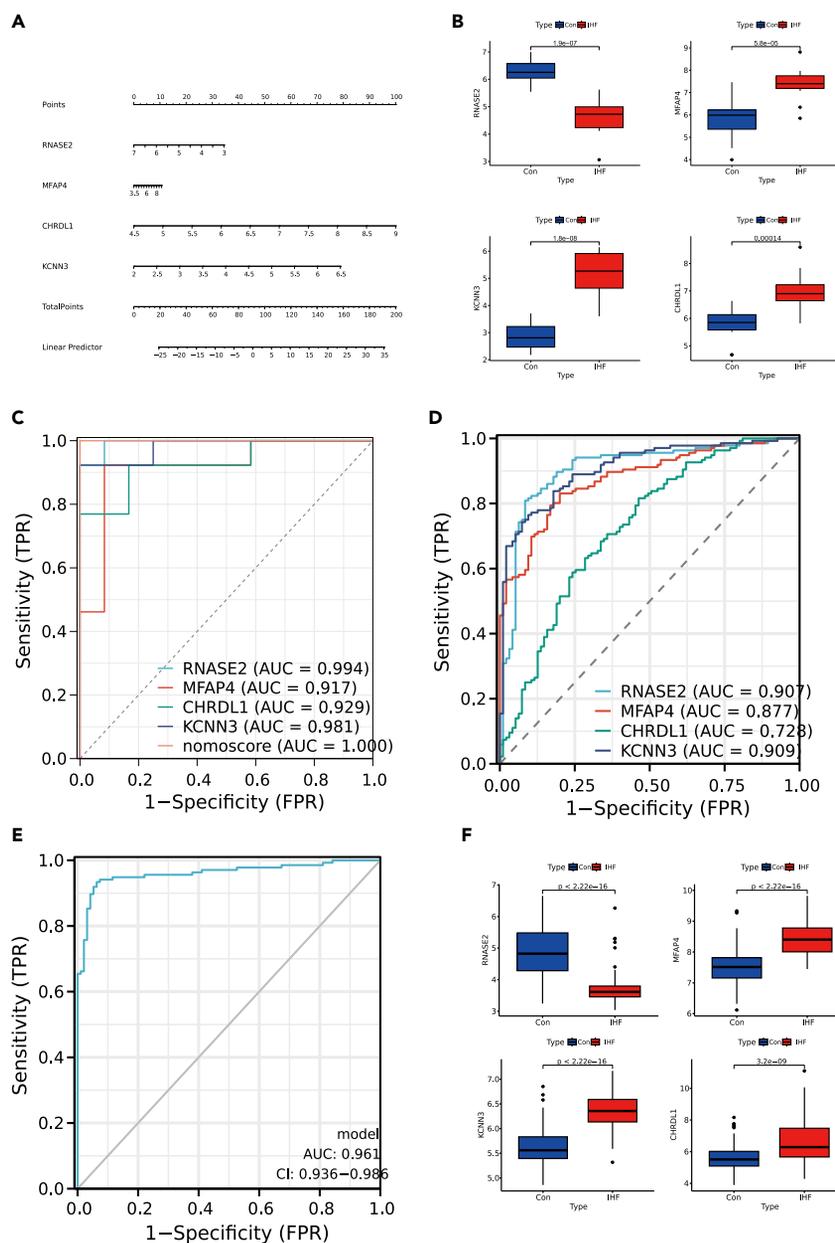
Figure 6. Results of diagnostic value assessment

(A) The visible nomogram for diagnosing IHF.

(B) Expression of Hub genes in IHF patients compared to normal controls in GSE21610 and GSE76701.

(C) The ROC curve of nomogram and each candidate genes showed the IHF diagnostic value.

(D) The ROC curve of each candidate genes in GSE57338.

(E) The ROC curve of the 4-gene diagnostic model in GSE57338.

(F) Expression of Hub genes in IHF patients compared to normal controls in GSE57338.

integrin signaling in the heart, which exacerbates cardiomyocyte hypertrophy.[28] Therefore, the decreased ejection fraction and myocardial fibrosis in IHF patients may be related to the dysregulation of MFAP4.

CHRDL1 is a secreted protein that acts as an antagonist of bone morphogenetic protein (BMP).[29] There has been little research on the association of CHRDL1 with heart disease, but recently Mauro Giacca and his team have identified the cardioprotective effects of CHRDL1 by a method called cardiac FunSel.[30] Its cardioprotective effect stems from the maintenance of cardiomyocyte viability by blocking the negative effect of BMP4 on cardiomyocyte autophagy. Also, CHRDL1 inhibits post-infarction cardiac fibrosis and enables post-infarction remodeling by inhibiting the negative effects of TGF-β on cardiac fibroblasts. In addition, a chromosomal genetic analysis showed that
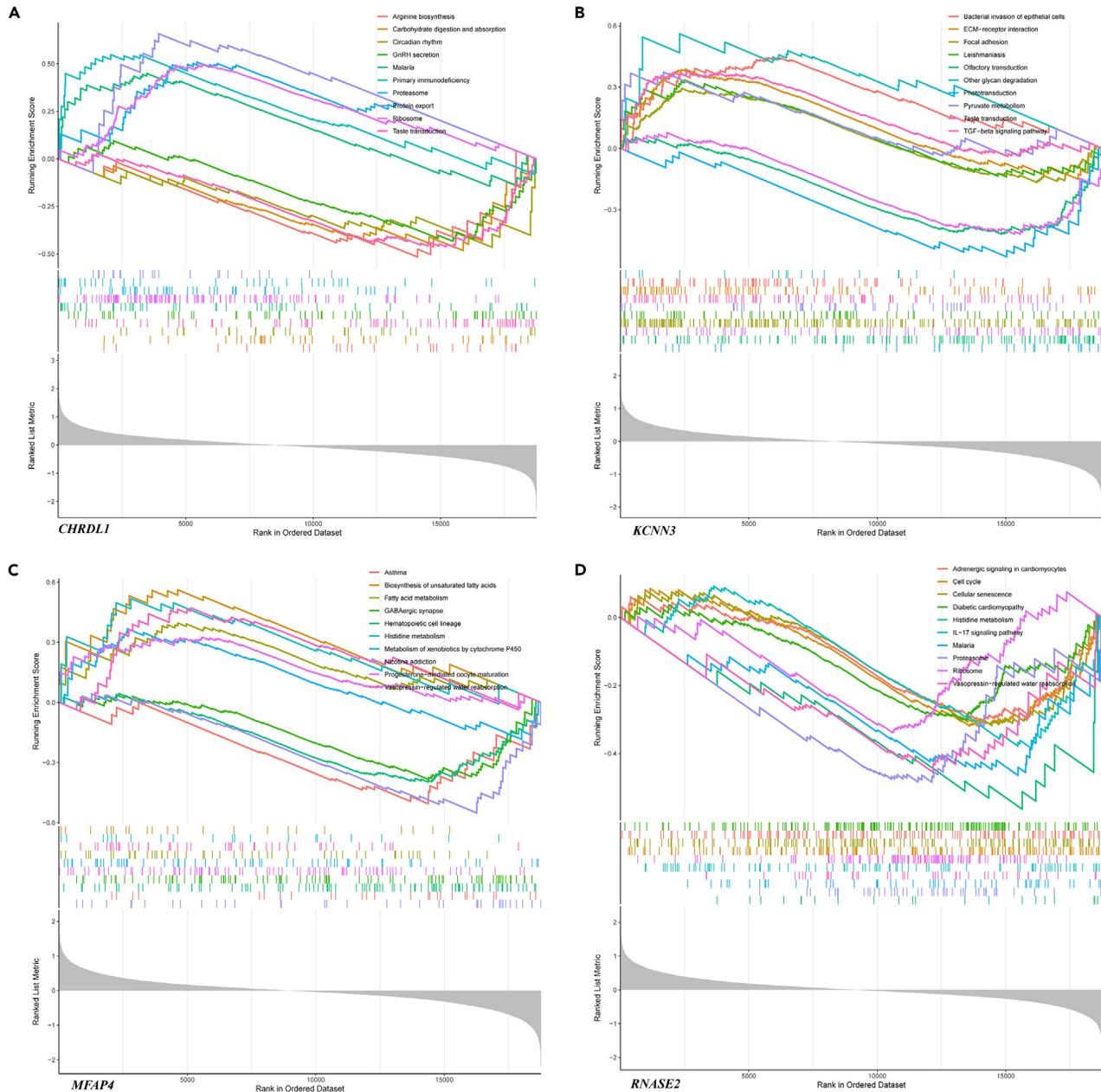
Figure 7. The results of ssGSEA enrichment analysis. Include upregulated gene pathway and downregulated gene pathway

(A) The results of CHRDL1.

(B) The results of KCNN3.

(C) The results of MFAP4.

(D) The results of RNASE2.

CHRDL1 was strongly associated with lowered lipids, suggesting that CHRDL1 has potential in coronary plaque control and may improve myocardial supply and treat IHF by lowering LDL and reversing plaque.[31] Therefore, targeting CHRDL1 therapy may improve exercise tolerance and cardiac output in IHF patients.

Potassium calcium-activated channel subfamily N member 3(KCNN3) belongs to the KCNN family of potassium channels. It encodes an integral membrane protein that forms a voltage-independent calcium-activated channel, which is thought to regulate neuronal excitability by contributing to the slow component of synaptic AHP.[32] Thus, KCNN3 has potential in the treatment of arrhythmias, especially since GWAS evidence suggests that variants in the KCNN3 gene are associated with atrial fibrillation.[33] In the field of heart failure, it has been shown that ventricular expression of KCNN3 is significantly increased in heart failure when ventricular tissue from patients with heart failure is compared

A

B

C
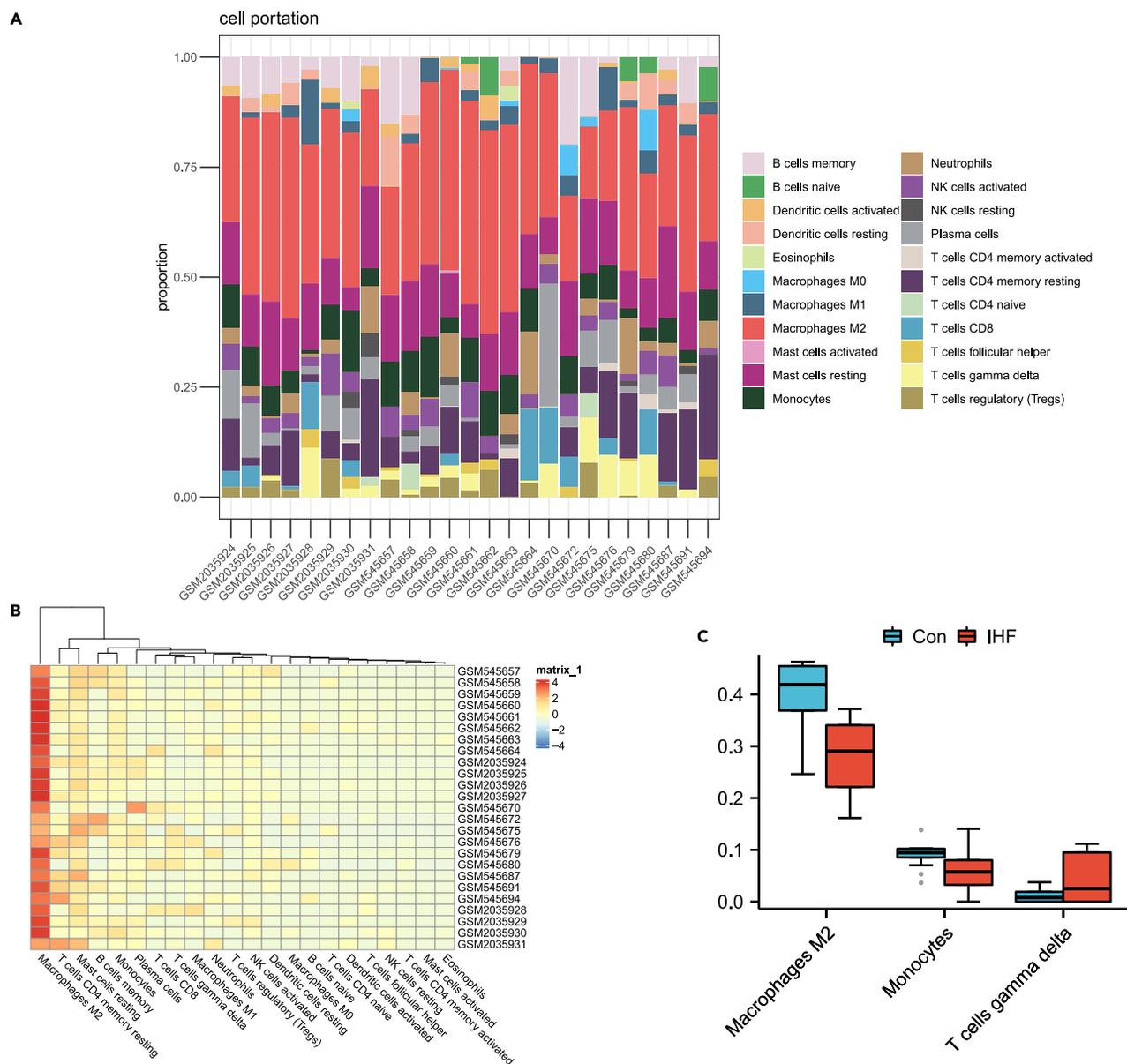
Figure 8. Immune cell infiltration analysis between IHF and control
(A) The proportion of 22 kinds of immune cells in different samples visualized from the bar plot.
(B) The expression of the 22 immune cells in the different samples can be seen in the heatmap.
(C) Expression of the 3 dysregulated immune cells in the IHF and controls as seen in the boxplot.

to healthy samples, in line with our findings.[34] It has also been shown that upregulation of KCNN3 leads to deterioration of ventricular function, which may be related to the involvement of this gene in the induction of ventricular tachycardia, but the specific mechanisms behind this need further investigation.[35] In general, the overexpression of KCNN3 in IHF patients may increase the risk of arrhythmia and sudden cardiac death, and aggravate clinical symptoms such as palpitations after activity in IHF patients.[36]

The results of the immune infiltration analysis showed that there were three types of immune cell infiltration that were significantly different. Macrophages were divided into two subpopulations based on their function and level of inflammatory factor secretion: macrophages M1 and macrophages M2. Macrophages M2 have an anti-inflammatory effect and are mainly activated by IL-4 inflammatory factor, inhibiting M1 macrophages mainly by secreting anti-inflammatory cytokines such as IL-10, which play a role in processes such as wound healing and tissue repair.[37] Macrophages M2 are lower in heart failure samples than in normal samples, and imbalanced M1/M2 macrophages may exacerbate inflammatory damage to cardiomyocytes, exacerbating myocardial dysfunction, and fibrosis.[38] Monocytes play a key role in orchestrating the inflammatory cascade response and the pathogenesis of HF. However, the highly differentiated nature of monocytes complicates their

function. On the one hand, monocytes are one of the major cellular targets of pro-inflammatory cytokines, and TNF-α induces monocytes to promote NO synthase production, thereby inducing apoptosis in cardiomyocytes, which in turn leads to further activation of the cascade by monocytes, resulting in a vicious cycle in the failing myocardium.[39] On the other hand, it has been suggested that activated monocytes can infiltrate the myocardium in order to exert phagocytic and reparative effects.[40] This may explain the general lack of clinical success in treating HF by modulating the cytokine system. The main biological effect of T cells gamma delta (Tgd cells) is cytotoxicity, which may account for its elevation in heart failure samples. Studies have shown that modulation of the IL-17A/Tgd cells axis can effectively modulate inflammation levels and slow down the process of myocardial fibrosis, thus exerting an anti-heart failure effect.[41,42]

The 4 Hub genes screened out by us also seem to be related to the immune infiltration of IHF samples. Studies have shown that RNASE2 is the most abundant member of the RNASEA family in macrophages and is also mainly expressed in monocytes, which may be one of the reasons why the expressions of RNASE2, macrophages M2, and monocytes in IHF samples are lower than those in normal samples.[43,44] In addition, the overexpression of MFAP4 and KCNN3 can promote the migration of monocytes, which may be another reason why the monocytes of IHF samples are lower than those of normal samples.[45,46]

In addition to our study, there are currently several bioinformatic studies on IHF. However, due to the different datasets and analysis methods selected, the Hub genes obtained by each study are also different. For example, Wang C and Kong X, although their researches were all based on the dataset GSE57338, Wang C obtained the three Hub genes of ASB14, CD163, and CCL5 through the traditional PPI algorithm, and Kong X obtained seven Hub genes through WGCNA and machine learning algorithms.[47,48] This does not mean that these results are in conflict. The occurrence and development of the disease involves multiple genes and multiple pathways, and it is difficult to explain clearly with a few genes. Our results and theirs can complement each other and provide more ideas for further understanding the mechanism and therapeutic targets of IHF.

The novel aspects of our study are as follows. First, in the selection of datasets, we selected 2 IHF datasets and merged them. Currently, there are no similar studies using these two datasets. Secondly, we identified RNASE2, MFAP4, CHRDL1, and KCNN3 as potential biomarkers and therapeutic targets for IHF through bioinformatics and three machine learning approaches. Thirdly, we validated these four genes in other dataset, and the validation results showed that the diagnostic model composed of these four genes has high diagnostic value, which provides new ideas for our future research on the molecular mechanisms of IHF. In addition, our enrichment analysis and immune infiltration analysis of the dataset showed that the molecular mechanisms of IHF are related to immunity and inflammation, which provides us with ideas for developing new therapeutic modalities for IHF.

In summary, we conducted a bioinformatic analysis of the GEO dataset to investigate the underlying molecular mechanisms of IHF and the immune cell infiltration environment within the failing myocardium. Through the implementation of three machine learning algorithms (LASSO, RF, and SVM-RFE), we have identified RNASE2, MFAP4, CHRDL1, and KCNN3 as potential biomarkers and therapeutic targets for the treatment of IHF. Of particular significance, we have developed diagnostic models and nomogram tools based on these four genes, providing a novel understanding of the pathogenesis of IHF and offering exciting prospects for future in-depth studies.

Limitations of the study

Nevertheless, there are some shortcomings in this study. Firstly, it is difficult to establish a causal relationship between gene expression differences and the pathophysiological mechanisms of heart failure. Secondly, the sample lacked information on gender and race, which may affect the generalizability of the results. Thirdly, our dataset was derived from myocardial tissue and the lack of validation of the peripheral blood dataset may limit the application of diagnostic models. Therefore, although our results were validated in external datasets, further clinical trials are needed to affirm the value of our diagnostic model. Whether interfering with the Hub gene can treat IHF needs to be further verified in animal experiments.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data processing and differentially expressed gene screening
  - Weighted Gene Co-Expression Network Analysis and module gene selection
  - Functional enrichment analysis
  - Machine learning
  - Nomogram construction and validation of diagnostic model
  - ssGSEA enrichment analysis
  - Immune infiltration analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108121.

REFERENCES

1. Hunt, S.A., Abraham, W.T., Chin, M.H., Feldman, A.M., Francis, G.S., Ganiats, T.G., Jessup, M., Konstam, M.A., Mancini, D.M., Michl, K., et al. (2005). Acc/aha 2005 guideline update for the diagnosis and management of chronic heart failure in the adult summary article. J. Am. Coll. Cardiol. *46*, 1116–1143. https://doi.org/10.1016/j.jacc2005.08023.

2. Heidenreich, P.A., Bozkurt, B., Aguilar, D., Allen, L.A., Byun, J.J., Colvin, M.M., Deswal, A., Drazner, M.H., Dunlay, S.M., Evers, L.R., et al. (2022). 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation *145*, e895–e1032. https://doi.org/10.1161/CIR.0000000000001063.

3. Peng, H., and Abdel-Latif, A. (2019). Cellular Therapy for Ischemic Heart Disease: An Update. Adv. Exp. Med. Biol. *1201*, 195–213. https://doi.org/10.1007/978-3-030-31206-0_10.

4. Kologrivova, I., Shtatolkina, M., Suslova, T., and Ryabov, V. (2021). Cells of the Immune System in Cardiac Remodeling: Main Players in Resolution of Inflammation and Repair After Myocardial Infarction. Front. Immunol. *12*, 664457. https://doi.org/10.3389/fimmu.2021.664457.

5. Peet, C., Ivetic, A., Bromage, D.I., and Shah, A.M. (2020). Cardiac monocytes and macrophages after myocardial infarction. Cardiovasc. Res. *116*, 1101–1112. https://doi.org/10.1093/cvr/cvz336.

6. Zhang, Y., Bauersachs, J., and Langer, H.F. (2017). Immune mechanisms in heart failure. Eur. J. Heart Fail. *19*, 1379–1389. https://doi.org/10.1002/ejhf.942.

7. Kumar, N., Narayan Das, N., Gupta, D., Gupta, K., and Bindra, J. (2021). Efficient Automated Disease Diagnosis Using Machine Learning Models. J. Healthc. Eng. *2021*, 9983652. https://doi.org/10.1155/2021/9983652.

8. Del Buono, M.G., Moroni, F., Montone, R.A., Azzalini, L., Sanna, T., and Abbate, A. (2022). Ischemic Cardiomyopathy and Heart Failure After Acute Myocardial Infarction. Curr. Cardiol. Rep. *24*, 1505–1515. https://doi.org/10.1007/s11886-022-01766-6.

9. Ngwenyama, N., Kaur, K., Bugg, D., Theall, B., Aronovitz, M., Berland, R., Panagiotidou, S., Genco, C., Perrin, M.A., Davis, J., and Alcaide, P. (2022). Antigen presentation by cardiac fibroblasts promotes cardiac dysfunction. Nat. Cardiovasc. Res. *1*, 761–774. https://doi.org/10.1038/s44161-022-00116-7.

10. Kumar, V., Prabhu, S.D., and Bansal, S.S. (2022). CD4+ T-lymphocytes exhibit biphasic kinetics post-myocardial infarction. Front. Cardiovasc. Med. *9*, 992653. https://doi.org/10.3389/fcvm.2022.992653.

11. Gage, J.R., Fonarow, G., Hamilton, M., Widawski, M., Martínez-Maza, O., and Vredevoe, D.L. (2004). Beta blocker and angiotensin-converting enzyme inhibitor therapy is associated with decreased Th1/Th2 cytokine ratios and inflammatory cytokine production in patients with chronic heart failure. Neuroimmunomodulation *11*, 173–180. https://doi.org/10.1159/000076766.

12. Bi, H.L., Zhang, X.L., Zhang, Y.L., Xie, X., Xia, Y.L., Du, J., and Li, H.H. (2020). The deubiquitinase UCHL1 regulates cardiac hypertrophy by stabilizing epidermal growth factor receptor. Sci. Adv. *6*, eaax4826. https://doi.org/10.1126/sciadv.aax4826.

13. Nishida, K., and Otsu, K. (2017). Sterile Inflammation and Degradation Systems in Heart Failure. Circ. J. *81*, 622–628. https://doi.org/10.1253/circj.CJ-17-0261.

14. Brown, D.I., Parry, T.L., and Willis, M.S. (2017). Ubiquitin Ligases and Posttranslational Regulation of Energy in the Heart: The Hand that Feeds. Compr. Physiol. *7*, 841–862. https://doi.org/10.1002/cphy.c160024.

15. Pokorna, Z., Jirkovsky, E., Hlavackova, M., Jansova, H., Jirkovska, A., Lencova-Popelova, O., Brazdova, P., Kubes, J., Sotakova-Kasparova, D., Mazurova, Y., et al. (2019). In vitro and in vivo investigation of cardiotoxicity associated with anticancer proteasome inhibitors and their combination with anthracycline. Clin. Sci. *133*, 1827–1844. https://doi.org/10.1042/CS20190139.

16. Gupta, S.K., Haigh, B.J., Griffin, F.J., and Wheeler, T.T. (2013). The mammalian secreted RNases: mechanisms of action in host defence. Innate Immun. *19*, 86–97. https://doi.org/10.1177/1753425912446955.

17. Bedoya, V.I., Boasso, A., Hardy, A.W., Rybak, S., Shearer, G.M., and Rugeles, M.T. (2006). Ribonucleases in HIV type 1 inhibition: effect of recombinant RNases on infection of primary T cells and immune activation-induced RNase gene and protein expression. AIDS Res. Hum. Retroviruses *22*, 897–907. https://doi.org/10.1089/aid.2006.22.897.

18. Yang, D., Chen, Q., Rosenberg, H.F., Rybak, S.M., Newton, D.L., Wang, Z.Y., Fu, Q., Tchernev, V.T., Wang, M., Schweitzer, B., et al. (2004). Human ribonuclease A superfamily members, eosinophil-derived neurotoxin and pancreatic ribonuclease, induce dendritic cell maturation and activation. J. Immunol. *173*, 6134–6142. https://doi.org/10.4049/jimmunol.173.10.6134.

19. Yang, D., Chen, Q., Su, S.B., Zhang, P., Kurosaka, K., Caspi, R.R., Michalek, S.M., Rosenberg, H.F., Zhang, N., and Oppenheim, J.J. (2008). Eosinophil-derived neurotoxin acts as an alarmin to activate the TLR2-MyD88 signal pathway in dendritic cells and enhances Th2 immune responses. J. Exp. Med. *205*, 79–90. https://doi.org/10.1084/jem.20062027.

20. Yang, D., Rosenberg, H.F., Chen, Q., Dyer, K.D., Kurosaka, K., and Oppenheim, J.J. (2003). Eosinophil-derived neurotoxin (EDN), an antimicrobial protein with chemotactic activities for dendritic cells. Blood *102*, 3396–3403. https://doi.org/10.1182/blood-2003-01-0151.

21. Ostendorf, T., Zillinger, T., Andryka, K., Schlee-Guimaraes, T.M., Schmitz, S., Marx, S., Bayrak, K., Linke, R., Salgert, S., Wegner, J.,

et al. (2020). Immune Sensing of Synthetic, Bacterial, and Protozoan RNA by Toll-like Receptor 8 Requires Coordinated Processing by RNase T2 and RNase 2. Immunity *52*, 591–605.e6. https://doi.org/10.1016/j.immuni.2020.03.009.

22. Zhang, X.Y., Huang, Z., Li, Q.J., Zhong, G.Q., Meng, J.J., Wang, D.X., and Tu, R.H. (2020). Ischemic postconditioning attenuates the inflammatory response in ischemia/reperfusion myocardium by upregulating miR-499 and inhibiting TLR2 activation. Mol. Med. Rep. *22*, 209–218. https://doi.org/10.3892/mmr.2020.11104.

23. Ong, S.L.M., de Vos, I.J.H.M., Meroshini, M., Poobalan, Y., and Dunn, N.R. (2020). Microfibril-associated glycoprotein 4 (Mfap4) regulates haematopoiesis in zebrafish. Sci. Rep. *10*, 11801. https://doi.org/10.1038/s41598-020-68792-8.

24. Schlosser, A., Pilecki, B., Hemstra, L.E., Kejling, K., Kristmannsdottir, G.B., Wulf-Johansson, H., Moeller, J.B., Füchtbauer, E.M., Nielsen, O., Kirketerp-Møller, K., et al. (2016). MFAP4 Promotes Vascular Smooth Muscle Migration, Proliferation and Accelerates Neointima Formation. Arterioscler. Thromb. Vasc. Biol. *36*, 122–133. https://doi.org/10.1161/ATVBAHA.115.306672.

25. Wulf-Johansson, H., Lock Johansson, S., Schlosser, A., Trommelholt Holm, A., Rasmussen, L.M., Mickley, H., Diederichsen, A.C.P., Munkholm, H., Poulsen, T.S., Tornøe, I., et al. (2013). Localization of microfibrillar-associated protein 4 (MFAP4) in human tissues: clinical evaluation of serum MFAP4 and its association with various cardiovascular conditions. PLoS One *8*, e82243. https://doi.org/10.1371/journal.pone.0082243.

26. Han, C., Peng, Y., Yang, X., Guo, Z., Yang, X., Su, P., Guo, S., and Zhao, L. (2023). Declined plasma microfibrillar-associated protein 4 levels in acute coronary syndrome. Eur. J. Med. Res. *28*, 32. https://doi.org/10.1186/s40001-023-01002-z.

27. Kanaan, R., Medlej-Hashim, M., Jounblat, R., Pilecki, B., and Sorensen, G.L. (2022). Microfibrillar-associated protein 4 in health and disease. Matrix Biol. *111*, 1–25. https://doi.org/10.1016/j.matbio.2022.05.008.

28. Dorn, L.E., Lawrence, W., Petrosino, J.M., Xu, X., Hund, T.J., Whitson, B.A., Stratton, M.S., Janssen, P.M.L., Mohler, P.J., Schlosser, A., et al. (2021). Microfibrillar-Associated Protein 4 Regulates Stress-Induced Cardiac Remodeling. Circ. Res. *128*, 723–737. https://doi.org/10.1161/CIRCRESAHA.120.317146.

29. Pei, Y.F., Zhang, Y.J., Lei, Y., Wu, W.D., Ma, T.H., and Liu, X.Q. (2017). Hypermethylation of the CHRDL1 promoter induces proliferation and metastasis by activating Akt and Erk in gastric cancer. Oncotarget *8*, 23155–23166. https://doi.org/10.18632/oncotarget.15513.

30. Ruozi, G., Bortolotti, F., Mura, A., Tomczyk, M., Falcione, A., Martinelli, V., Vodret, S., Braga, L., Dal Ferro, M., Cannatà, A., et al. (2022). Cardioprotective factors against myocardial infarction selected in vivo from an AAV secretome library. Sci. Transl. Med. *14*, eabo0699. https://doi.org/10.1126/scitranslmed.abo0699.

31. Natarajan, P., Pampana, A., Graham, S.E., Ruotsalainen, S.E., Perry, J.A., de Vries, P.S., Broome, J.G., Pirruccello, J.P., Honigberg, M.C., Aragam, K., et al. (2021). Chromosome Xq23 is associated with lower atherogenic lipid concentrations and favorable cardiometabolic indices. Nat. Commun. *12*, 2182. https://doi.org/10.1038/s41467-021-22339-1.

32. National Center for Biotechnology Information. PubChem Gene Summary for Gene 3782. https://pubchem.ncbi.nlm.nih.gov/gene/KCNN3/human.

33. Ellinor, P.T., Lunetta, K.L., Glazer, N.L., Pfeufer, A., Alonso, A., Chung, M.K., Sinner, M.F., de Bakker, P.I.W., Mueller, M., Lubitz, S.A., et al. (2010). Common variants in KCNN3 are associated with lone atrial fibrillation. Nat. Genet. *42*, 240–244. https://doi.org/10.1038/ng.537.

34. Darkow, E., Nguyen, T.T., Stolina, M., Kari, F.A., Schmidt, C., Wiedmann, F., Baczkó, I., Kohl, P., Rajamani, S., Ravens, U., and Peyronnet, R. (2021). Small Conductance Ca2 +-Activated K+ (SK) Channel mRNA Expression in Human Atrial and Ventricular Tissue: Comparison Between Donor, Atrial Fibrillation and Heart Failure Tissue. Front. Physiol. *12*, 650964. https://doi.org/10.3389/fphys.2021.650964.

35. Ortega, A., Tarazón, E., Roselló-Lletí, E., Gil-Cayuela, C., Lago, F., González-Juanatey, J.R., Cinca, J., Jorge, E., Martínez-Dolz, L., Portolés, M., and Rivera, M. (2015). Patients with Dilated Cardiomyopathy and Sustained Monomorphic Ventricular Tachycardia Show Up-Regulation of KCNN3 and KCNJ2 Genes and CACNG8-Linked Left Ventricular Dysfunction. PLoS One *10*, e0145518. https://doi.org/10.1371/journal.pone.0145518.

36. Mahida, S., Mills, R.W., Tucker, N.R., Simonson, B., Macri, V., Lemoine, M.D., Das, S., Milan, D.J., and Ellinor, P.T. (2014). Overexpression of KCNN3 results in sudden cardiac death. Cardiovasc. Res. *101*, 326–334. https://doi.org/10.1093/cvr/cvt269.

37. Mouton, A.J., Li, X., Hall, M.E., and Hall, J.E. (2020). Obesity, Hypertension, and Cardiac Dysfunction: Novel Roles of Immunometabolism in Macrophage Activation and Inflammation. Circ. Res. *126*, 789–806. https://doi.org/10.1161/CIRCRESAHA.119.312321.

38. Zhang, L., Chen, J., Yan, L., He, Q., Xie, H., and Chen, M. (2021). Resveratrol Ameliorates Cardiac Remodeling in a Murine Model of Heart Failure With Preserved Ejection Fraction. Front. Pharmacol. *12*, 646240. https://doi.org/10.3389/fphar.2021.646240.

39. Balligand, J.L., Ungureanu, D., Kelly, R.A., Kobzik, L., Pimental, D., Michel, T., and Smith, T.W. (1993). Abnormal contractile function due to induction of nitric oxide synthesis in rat cardiac myocytes follows exposure to activated macrophage-conditioned medium. J. Clin. Invest. *91*, 2314–2319. https://doi.org/10.1172/JCI116461.

40. Wrigley, B.J., Lip, G.Y.H., and Shantsila, E. (2011). The role of monocytes and inflammation in the pathophysiology of heart failure. Eur. J. Heart Fail. *13*, 1161–1171. https://doi.org/10.1093/eurjhf/hfr122.

41. Blanco-Domínguez, R., de la Fuente, H., Rodríguez, C., Martín-Aguado, L., Sánchez-Díaz, R., Jiménez-Alejandre, R., Rodríguez-Arabaolaza, I., Curtabbi, A., García-Guimaraes, M.M., Vera, A., et al. (2022). CD69 expression on regulatory T cells protects from immune damage after myocardial infarction. J. Clin. Invest. *132*, e152418. https://doi.org/10.1172/JCI152418.

42. Yan, X., Shichita, T., Katsumata, Y., Matsuhashi, T., Ito, H., Ito, K., Anzai, A., Endo, J., Tamura, Y., Kimura, K., et al. (2012). Deleterious effect of the IL-23/IL-17A axis and

γδT cells on left ventricular remodeling after myocardial infarction. J. Am. Heart Assoc. *1*, e004408. https://doi.org/10.1161/JAHA.112.004408.

43. Zhu, Y., Tang, X., Xu, Y., Wu, S., Liu, W., Geng, L., Ma, X., Tsao, B.P., Feng, X., and Sun, L. (2022). RNASE2 Mediates Age-Associated B Cell Expansion Through Monocyte Derived IL-10 in Patients With Systemic Lupus Erythematosus. Front. Immunol. *13*, 752189. https://doi.org/10.3389/fimmu.2022.752189.

44. Lu, L., Li, J., Wei, R., Guidi, I., Cozzuto, L., Ponomarenko, J., Prats-Ejarque, G., and Boix, E. (2022). Selective cleavage of ncRNA and antiviral activity by RNase2/EDN in THP1-induced macrophages. Cell. Mol. Life Sci. *79*, 209. https://doi.org/10.1007/s00018-022-04229-x.

45. Pilecki, B., de Carvalho, P.V.S.D., Kirketerp-Møller, K.L., Schlosser, A., Kejling, K., Dubik, M., Madsen, N.P., Stubbe, J., Hansen, P.B.L., Andersen, T.L., et al. (2021). MFAP4 Deficiency Attenuates Angiotensin II-Induced Abdominal Aortic Aneurysm Formation Through Regulation of Macrophage Infiltration and Activity. Front. Cardiovasc. Med. *8*, 764337. https://doi.org/10.3389/fcvm.2021.764337.

46. Zhang, S.X., Wang, X.P., Gao, C.Y., Ju, C.H., Zhu, L.J., and DU, Y.M. (2015). Sheng Li Xue Bao *67*, 505–512.

47. Wang, C., Yang, H., and Gao, C. (2019). Potential biomarkers for heart failure. J. Cell. Physiol. *234*, 9467–9474. https://doi.org/10.1002/jcp.27632.

48. Kong, X., Sun, H., Wei, K., Meng, L., Lv, X., Liu, C., Lin, F., and Gu, X. (2023). WGCNA combined with machine learning algorithms for analyzing key genes and immune cell infiltration in heart failure due to ischemic cardiomyopathy. Front. Cardiovasc. Med. *10*, 1058834. https://doi.org/10.3389/fcvm.2023.1058834.

49. Schwientek, P., Ellinghaus, P., Steppan, S., D'Urso, D., Seewald, M., Kassner, A., Cebulla, R., Schulte-Eistrup, S., Morshuis, M., Röfe, D., et al. (2010). Global gene expression analysis in nonfailing and failing myocardium pre- and postpulsatile and nonpulsatile ventricular assist device support. Physiol. Genomics *42*, 397–405. https://doi.org/10.1152/physiolgenomics.00030.2010.

50. Kim, E.H., Galchev, V.I., Kim, J.Y., Misek, S.A., Stevenson, T.K., Campbell, M.D., Pagani, F.D., Day, S.M., Johnson, T.C., Washburn, J.G., et al. (2016). Differential protein expression and basal lamina remodeling in human heart failure. Proteomics. Clin. Appl. *10*, 585–596. https://doi.org/10.1002/prca.201500099.

51. Liu, Y., Morley, M., Brandimarto, J., Hannenhalli, S., Hu, Y., Ashley, E.A., Tang, W.H.W., Moravec, C.S., Margulies, K.B., Cappola, T.P., et al. (2015). RNA-Seq identifies novel myocardial gene expression signatures of heart failure. Genomics *105*, 83–89. https://doi.org/10.1016/j.ygeno.2014.12.002.

52. Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics *23*, 1846–1847. https://doi.org/10.1093/bioinformatics/btm254.

53. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.

https://doi.org/10.1093/bioinformatics/bts034.

54. Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., and Dunnington, D. (2016). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. R Package Version 3.2.1 (Stata Software Package).

55. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

56. Smyth, G.K. (2005). Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor (Springer), pp. 397–420.

57. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 9, 559. https://doi.org/10.1186/1471-2105-9-559.

58. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287. https://doi.org/10.1089/omi.2011.0118.

59. Engebretsen, S., and Bohlin, J. (2019). Statistical predictions with glmnet. Clin. Epigenetics 11, 123. https://doi.org/10.1186/s13148-019-0730-1.

60. Yi, M., Nissley, D.V., McCormick, F., and Stephens, R.M. (2020). ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical implications. Sci. Rep. 10, 10258. https://doi.org/10.1038/s41598-020-66986-8.

61. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453–457. https://doi.org/10.1038/nmeth.3337.

62. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 41, D991–D995. https://doi.org/10.1093/nar/gks1193.

63. The Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47, D330–D338. https://doi.org/10.1093/nar/gky1055.

64. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30. https://doi.org/10.1093/nar/28.1.27.

65. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

66. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. 33, 1–22.

67. Petralia, F., Wang, P., Yang, J., and Tu, Z. (2015). Integrative random forest for gene regulatory network inference. Bioinformatics 31, i197–i205. https://doi.org/10.1093/bioinformatics/btv268.

68. Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., and Xu, W. (2018). Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics 15, 41–51. https://doi.org/10.21873/cgp.20063.

69. Gou, M., Qian, N., Zhang, Y., Wei, L., Fan, Q., Wang, Z., and Dai, G. (2022). Construction of a nomogram to predict the survival of metastatic gastric cancer patients that received immunotherapy. Front. Immunol. 13, 950868. https://doi.org/10.3389/fimmu.2022.950868.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
IHF dataset 1	(Schwientek P et al.)[49]	GEO: GSE21610
IHF dataset 2	(Kim EH et al.)[50]	GEO: GSE76701
IHF verification set	(Liu Y et al.)[51]	GEO: GSE57338
Software and algorithms		
R (v4.2.0)	The R Project	https://www.r-project.org
GEOquery (R package)	(Davis S et al.)[52]	https://bioconductor.org/packages/release/bioc/html/GEOquery.html
sva (R package)	(Leek JT et al.)[53]	https://bioconductor.org/packages/release/bioc/html/sva.html
ggplot2 (R package)	(Wickham H et al.)[54]	https://cran.r-project.org/web/packages/ggplot2/index.html
ComplexHeatmap (R package)	(Gu Z et al.)[55]	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
Limma (R package)	(Ritchie et al.)[56]	https://bioconductor.org/packages/release/bioc/html/limma.html
WGCNA (R package)	(Langfelder et al.)[57]	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/
clusterProfiler (R package)	(Yu G et al.)[58]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
glmnet (R package)	(Engebretsen S et al.)[59]	https://github.com/cran/glmnet
ssGSEA (R package)	(Yi et al.)[60]	https://github.com/broadinstitute/ssGSEA2.0
CIBERSORT (R package)	(Newman AM et al.)[61]	https://github.com/zomithex/CIBERSORT

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yan Li (liyan88130@163.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data

This study analyzes existing, publicly available data. The sources for the datasets are listed in the key resources table.

Code

This study does not report original code. All codes were used in this study in alignment with recommendations made by authors of R packages in their respective user's guide, which can be accessed at https://bioconductor.org. All code used in the analyses is deposited on https://github.com/YidingYu96/ML.

Additional information requests

Any additional information required to reanalyze the data used in this study is available from the lead contact upon request.

Availability of data and materials

Publicly available datasets were analyzed in this study. These data can be found here: GSE76701; GSE21610; GSE57338.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We used "ischemic heart failure" as the keyword to search in the GEO database (as of March 31, 2023). Select the "Homo sapiens" dataset and experiment type should be "Expression profiling by array". In addition, the disease must be clearly defined as ischemic heart failure in the dataset description. We finally obtained 8 IHF-related datasets (GSE57345 and GSE57338 are the same dataset). The details of the datasets are shown in Table S1. We select datasets based on the following two principles: 1. The sample size of the control group and the disease group are close. 2. Select as many datasets as possible for merging. Since only datasets of the same platform can be combined, we finally selected GSE21610 and GSE76701 of the GPL570 platform as the dataset for ischemic heart failure, and GSE57338 as the verification set.[49–51,62]

METHOD DETAILS

Data processing and differentially expressed gene screening

To collate and analyze our data, we utilized R software (version 4.2.0) and accessed the GEO database through the GEOquery package to download the GSE21610 and GSE76701 datasets.[52] To ensure accuracy and consistency, we removed probes corresponding to multiple molecules and retained only the probe with the highest signal value for each molecule. We utilized the ComBat function of the sva package to eliminate batch effect from the filtered data.[53] The essence of difference analysis is a generalized linear model. The limma package can fit a linear equation to the expression of each gene. Therefore, using the limma package for difference analysis is currently the most recommended way. To gain insights into the differences between heart failure and normal samples, we employed the limma package and identified genes with adjust p values <0.05 and $|\log2(FC)| \geq 1$ as the difference genes. We further visualized our findings using the ggplot2 package and the pheatmap package.[54–56] To assess normalization, we utilized a boxplot, and for clustering between sample subgroups, we used a PCA plot.

Weighted Gene Co-Expression Network Analysis and module gene selection

In addition to obtaining differential genes with the limma package, we also used the WGCNA package to explore gene modules with relatedness.[57] A scale-free co-expression network was created after removing ineligible genes and samples by the goodSamplesGenes function with a filtering criterion of 0.5. Subsequently, adjacency was calculated by default using $\beta = 30$ and scale-free R2 = 0.9 as a soft threshold, and adjacency was converted to a topological overlap matrix (TOM), which was used to determine gene ratios and dissimilarity. Genes with the same expression profile were grouped into gene modules using average linkage hierarchical clustering. We preferred larger modules, so we set the minimum module size to 200. Finally, we calculate the similarity of the modules' signature genes, select the cut lines of the module dendrogram to combine several modules for the next step of the study, and complete the visualization of the signature gene network. WGCNA analysis was used to identify important modules in IHF.

Functional enrichment analysis

We utilized the clusterProfiler package to conduct Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) enrichment analyses and visualizations.[58,63–65] Specifically, we performed GSEA enrichment analysis on all genes within our dataset. Subsequently, we identified the key genes of IHF by intersecting the DEGs with the important module genes of WGCNA. Finally, we performed GO and KEGG enrichment analyses on the key genes of IHF.

Machine learning

We will use three machine learning algorithms, LASSO, RF and SVM-RFE, to further screen candidate genes for IHF diagnosis.[66–68] We use the glmnet package to perform the LASSO algorithm, choosing 10-fold cross-validation to select the prominent genes.[59] We used the randomForest package to complete the RF algorithm, selecting the top 30 genes as alternative genes. We used the e1071 package to complete the SVM-RFE algorithm to select the number of genes with the highest precision as candidate genes. After completing the calculation, we selected the intersection of the three as the diagnostic genes for IHF.

Nomogram construction and validation of diagnostic model

We utilized the rms package to develop a nomogram for the identification of IHF diagnostic genes.[69] Points represent the scores of candidate genes and Total Points represent the sum of all the above gene scores. Subsequently, boxplots of gene expression were generated (The differences in gene expression between IHF and normal groups were compared by t-test) and receiver operating characteristic (ROC) curves were constructed to determine the diagnostic value of the candidate genes. The area under the curve (AUC) was calculated to quantify their value, with an AUC value greater than 0.7 being considered as the ideal diagnostic threshold. To further validate our findings, we conducted an analysis of individual and combined genes using the GSE57338 dataset. We assessed the discriminatory ability of the diagnostic model by ROC curves once again.

ssGSEA enrichment analysis

We used the clusterProfiler package to complete ssGSEA enrichment analysis of genes from the IHF diagnostic model to explore the function of these genes in the IHF process.[60]

Immune infiltration analysis

We used the CIBERSORT package to assess the content of immune cells and stromal cells in IHF myocardial samples to depict a cellular heterogeneous landscape of myocardial expression profiles and to complete the immune cell infiltration analysis.[61] Bar charts were used to visualize the proportion of each type of immune cell in the different samples. Differences in cell distribution between the IHF and normal groups were compared by t-test, with the cut-off value set at $p < 0.05$.