

# Classification patterns identification of immunogenic cell death-related genes in heart failure based on deep learning

Zhihui Ma<sup>1</sup>, Shixin Ma<sup>1</sup>, Bin Chen<sup>1</sup>, Yongjun Zhang<sup>1</sup>, Jinmei Zeng<sup>1</sup>, Jianping Tao<sup>1</sup>, Yu Hu<sup>1</sup>

<sup>1</sup>Department of Cardiology, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200233, People's Republic of China

**Correspondence to:** Shixin Ma; email: [shxma@sina.com](mailto:shxma@sina.com)

**Keywords:** heart failure, subtype identification, encoder, biomarkers, diagnosis

**Received:** June 21, 2023

**Accepted:** December 26, 2023

**Published:**

**Copyright:** © 2024 Ma et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Heart failure (HF) is a complex and prevalent disease, especially among the elderly population, characterized by symptoms like chest tightness, shortness of breath, and dyspnea. To address the need for improved classification and drug target identification in HF, we explored the potential role of Immunogenic Cell Death (ICD), a mode of cell death known for its significance in the tumor immune response but relatively uncharted in HF research. In recent years, deep learning models have exhibited remarkable performance in tasks such as classification, clustering, and regression. In this paper, we harnessed the power of deep learning by employing various encoder models to evaluate their effectiveness in clustering based on ICD-related genes. This novel approach allowed us to identify distinct subtypes within HF. Subsequently, we refined these subtypes by employing differentially expressed genes, leading to the discovery of significant variations in immune infiltration and functional enrichment across these subtypes. Moreover, we leveraged advanced machine learning techniques to identify diagnosis-related genes in HF. The AUC of the diagnostic model in the internal and external test sets could reach more than 0.99. These genes served as the foundation for constructing nomogram models and further exploration of their interactions with miRNA and transcription factors. In summary, our study uniquely combines the exploration of ICD in HF, the application of deep learning models, and the identification of diagnosis-related genes to provide a multifaceted understanding of HF subtypes and potential therapeutic targets.

## INTRODUCTION

Heart failure (HF) refers to the dysfunction of the cardiac function caused by many factors, which makes the stroke volume unable to meet the body's metabolic demands. Its clinical manifestations are mainly dyspnea, angina, and vertigo [1, 2]. The onset and progression of HF are accompanied by structural changes in cardiomyocytes and disrupted energy metabolism [3]. Advanced HF often occurs in the elderly and is challenging to diagnose. Therefore, it is urgent to develop biomarkers related to HF diagnosis, risk assessment, and therapeutic target identification [4, 5].

Immunogenic cell death (ICD) is a regulatory cell death mode that can trigger a variety of adaptive immune responses [6]. This reaction can present antigens to cytotoxic T cells through dendritic cells and then trigger an immune response [7]. For tumor diseases, chemotherapy drugs are inducers that can trigger ICD, promoting the presentation of tumor-related antigens and further eliminating the remaining tumor cells [8]. However, the role of ICD-related genes (ICDRGs) in HF is unclear.

Therefore, this paper systematically explored the role of ICDRGs in HF. Specifically, the gene expression

data of HF and its control group were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), and the differentially expressed genes were analyzed. Intersect the intersection gene with ICDRGs and extract the expression of the intersection gene. Based on three clustering methods, the HF subtypes were identified using DEICDRGs, and the differences in pathway enrichment and immune system among subtypes were explored. DEGs among clusters were used to identify further subtypes of HF, and immune cell infiltration and functional enrichment analysis were analyzed. In addition, we used two machine learning algorithms to screen some genes related to the diagnosis of HF and built a diagnostic nomogram model. Finally, the correlation between genes associated with diagnosis and immune cell types was discussed, and the miRNA-mRNA interaction network of genes related to diagnosis and the mRNA-transcription factor interaction network of genes associated with diagnosis were constructed. The subtypes and genes associated with diagnosis identified in this paper can provide a reference for individualized treatment and clinical diagnosis of HF.

## RESULTS

### ICD-related gene expression landscape

The overall flow chart of this paper is given in Figure 1. Firstly, the transcriptome data of the GSE141910 data set were analyzed for differential expression, and 8885 DEGs were obtained with  $\text{adj.P.Val} < 0.05$  as the standard. In addition, we obtained 20 ICDRGs from the previous literature. Fourteen DEICDRGs were obtained by crossing them (Figure 2A). Figure 2B, 2C show the differential expression thermogram and differential expression box diagram of 14 DEICDRGs in the sick and control groups, respectively. Figure 2D displays the chromosome position information of 14 DEICDRGs. Figure 2E shows the correlation thermogram among 14 DEICDRGs. It was obvious that most genes had a significant correlation. The details of DEGs, ICDRGs, and DEICDRGs can be found in Supplementary Table 1.

### Identification of subtypes of HF based on DEICDRGs

The expression levels of 14 DEICDRGs in diseased samples were extracted in the study. We set the number of clusters to 2-5, respectively. Figure 3 displays the t-sne dimensionality reduction graphs of AE, DAE, and K-means under different cluster numbers. Each point in the graph represented a sample. In order to evaluate the clustering performance of the three algorithms under different cluster numbers, this paper used three index columns: sample contour coefficient, Calinski-Harabasz Index, and Davies-Bouldin Index. We can observe the

histograms of the three algorithms in four clustering situations in Figure 4A–4C. From the figure, we found that DAE had the best comprehensive performance of the three indicators when the number of clusters was 2. Therefore, this paper would interpret it as the result of subtype identification in the future. Figure 4D, 4E are box graphs indicating the differences in infiltration abundance and immune function of immune cells between the two types. We also obtained the difference in gene expression related to immune inspection sites between the two subtypes (Figure 4F). According to Figure 4F, there were significant differences in the infiltration abundance of most immune cells, immune function, and expression of genes related to immune inspection sites between the two typing samples, which confirmed the typing ability of the DAE algorithm. In addition, this paper also analyzed the GSVA results for two types (Figure 4G). We would analyze the biological significance of these pathways in detail in the discussion section.

The expression levels of 14 DEICDRGs in diseased samples were extracted in the study. We set the number of clusters to range from 2 to 5. Figure 3 displays the t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction graphs of AE, DAE, and K-means clustering under different cluster numbers. Each point in the figure represented a sample. This paper used three index columns to evaluate the clustering performance of the three algorithms under other cluster numbers: sample silhouette coefficient, Calinski-Harabasz Index, and Davies-Bouldin Index. We can observe the histograms of the three algorithms in four clustering situations in Figure 4A–4C. From the figure, we found that DAE had the best overall performance of the three indicators when the number of clusters was 2. Therefore, this paper would interpret it as the outcome of subtype identification in the future. Figure 4D, 4E are box plots indicating the differences in infiltration abundance and immune function of immune cells between the two types. We also obtained the difference in gene expression related to immune inspection sites between the two subtypes (Figure 4F). According to Figure 4F, there were significant differences in the infiltration abundance of most immune cells, immune function, and expression of genes related to immune inspection sites between the two subtypes, which confirmed the classification ability of the DAE algorithm. This article introduces non-negative matrix factorization (NMF) and consensus clustering method (CC) to confirm the algorithm's performance further. Three clustering performance indicators were calculated when the number of clusters was 2. The dimensionality reduction results of these two algorithms and the performance comparison results with the research algorithms included in this article are shown in

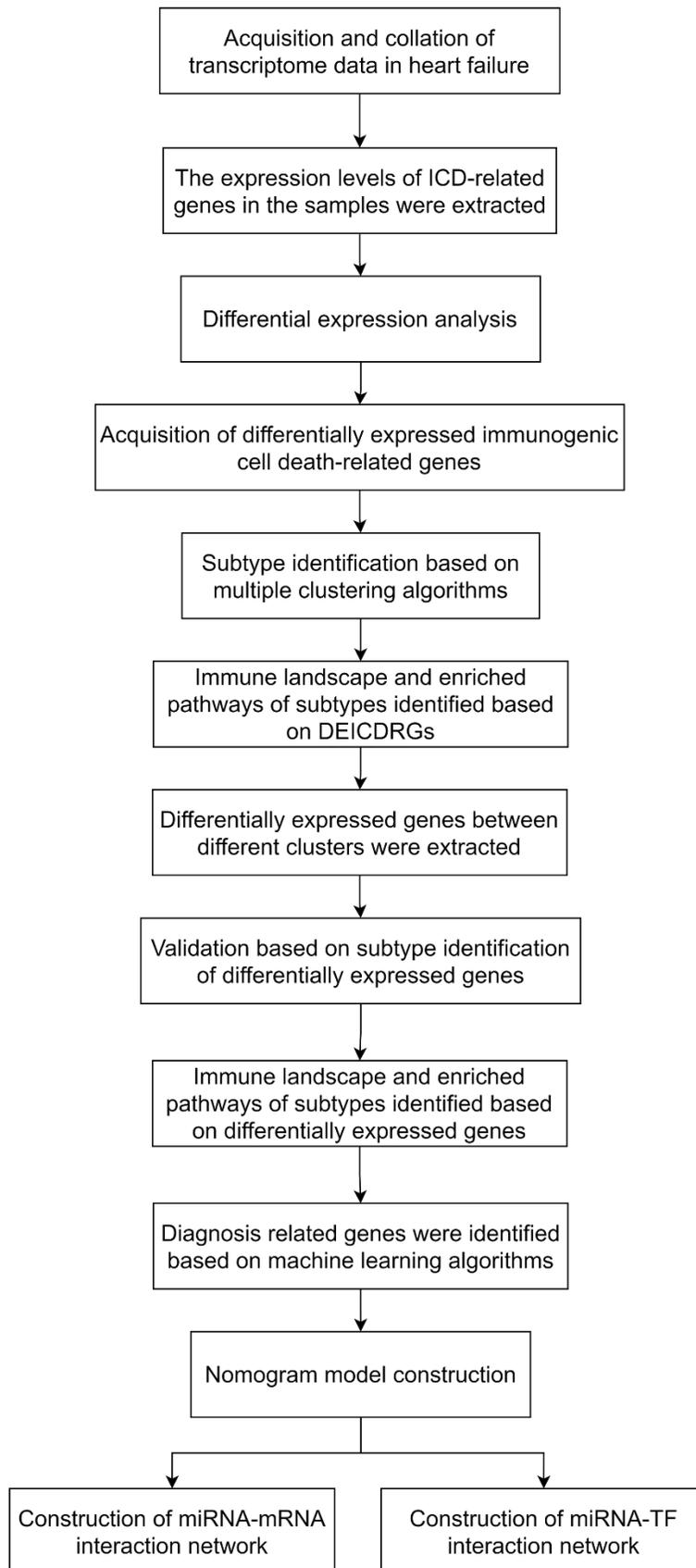


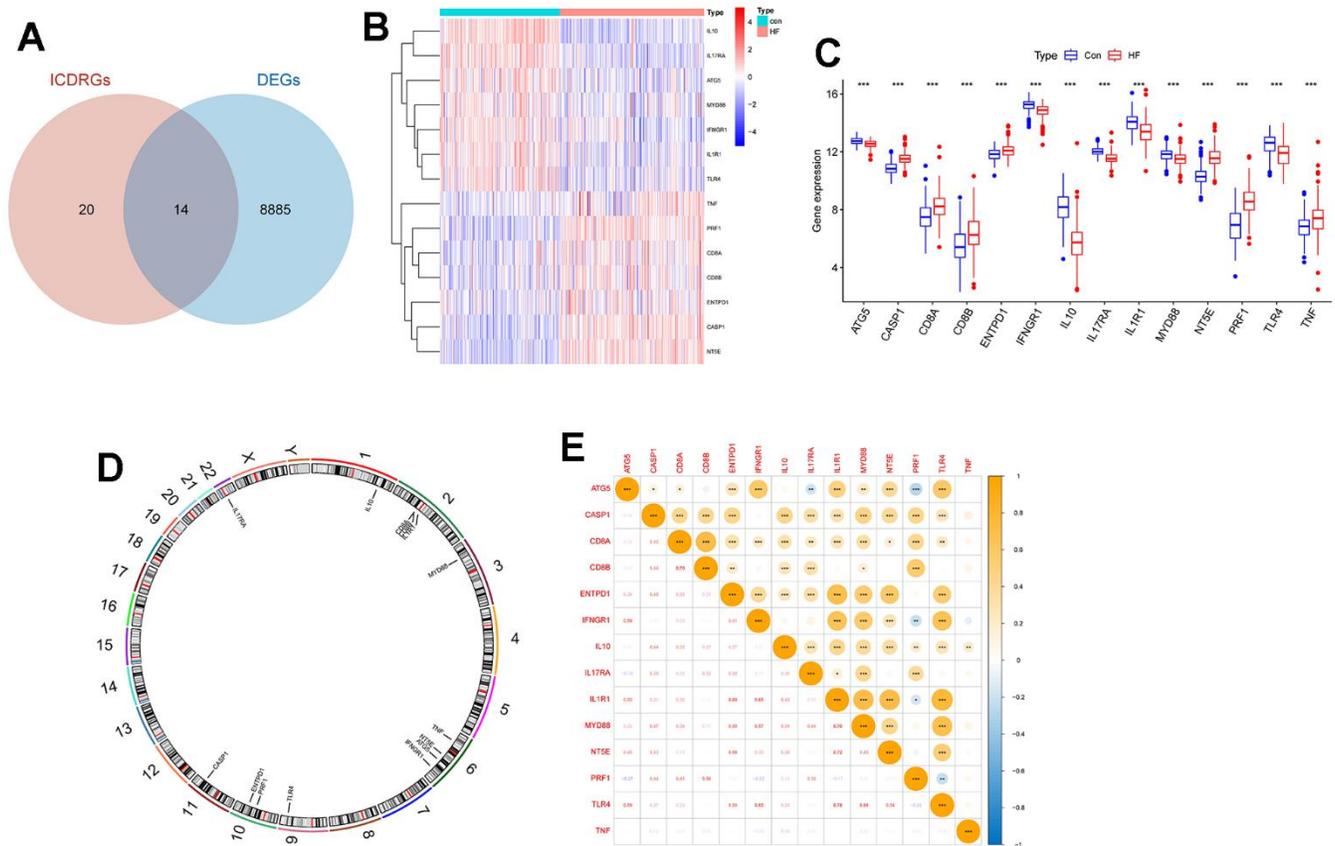
Figure 1. The technical roadmap of the article.

Supplementary Figures 1, 2 in the Supplementary Table 2.

In addition, this paper also analyzed the GSVA results for two types (Figure 4G). Through the GSVA analysis of the two subtypes, we found that there were significant differences in several biological processes between the two subtypes (Figure 4G). Almost all of these pathways are associated with heart failure—for example, melanoma, apoptosis, and the JAK-STAT signal pathway. Early scholars reported a case where metastatic malignant melanoma could result in rapid occlusion of the right ventricle, thus leading to congestive HF [9]. Melanoma often involves the heart, resulting in cardiac issues such as HF and myocardial infarction [10, 11]. HF is also intricately linked to the complex pathophysiology of apoptosis [12, 13]. Researchers have confirmed that isoproterenol (ISO) can induce apoptosis, improve heart function, and relieve and treat ISO-induced HF models and cellular HF in rats [14]. The JAK-STAT signaling pathway has been proven to play an essential role in the pathophysiology of HF [15, 16].

### Verification of subtypes of HF based on DEGs cluster

To further validate the rationality of the subtype identification mentioned above, we conducted differential expression analysis on the two subtypes, identifying 43 DEGs with  $p < 0.05$ . Utilizing these DEGs, this study employed three algorithms to generate t-SNE dimensionality reduction plots under four clustering numbers (Figure 5). The evaluation results of three metrics in Figure 6A–6C are presented. It can be observed that compared to Figure 5, the clustering effect has been significantly improved. This confirms the rationality of using DAE to classify patients into two subtypes. Similarly, we provided identification results of the immune microenvironment and functional biological characteristics of different gene clusters in Figure 6D–6G. There were significant differences between the two subtypes in terms of immune cell infiltration abundance, immune functions, immune checkpoint-related gene expression, GSVA analysis pathways, and others.

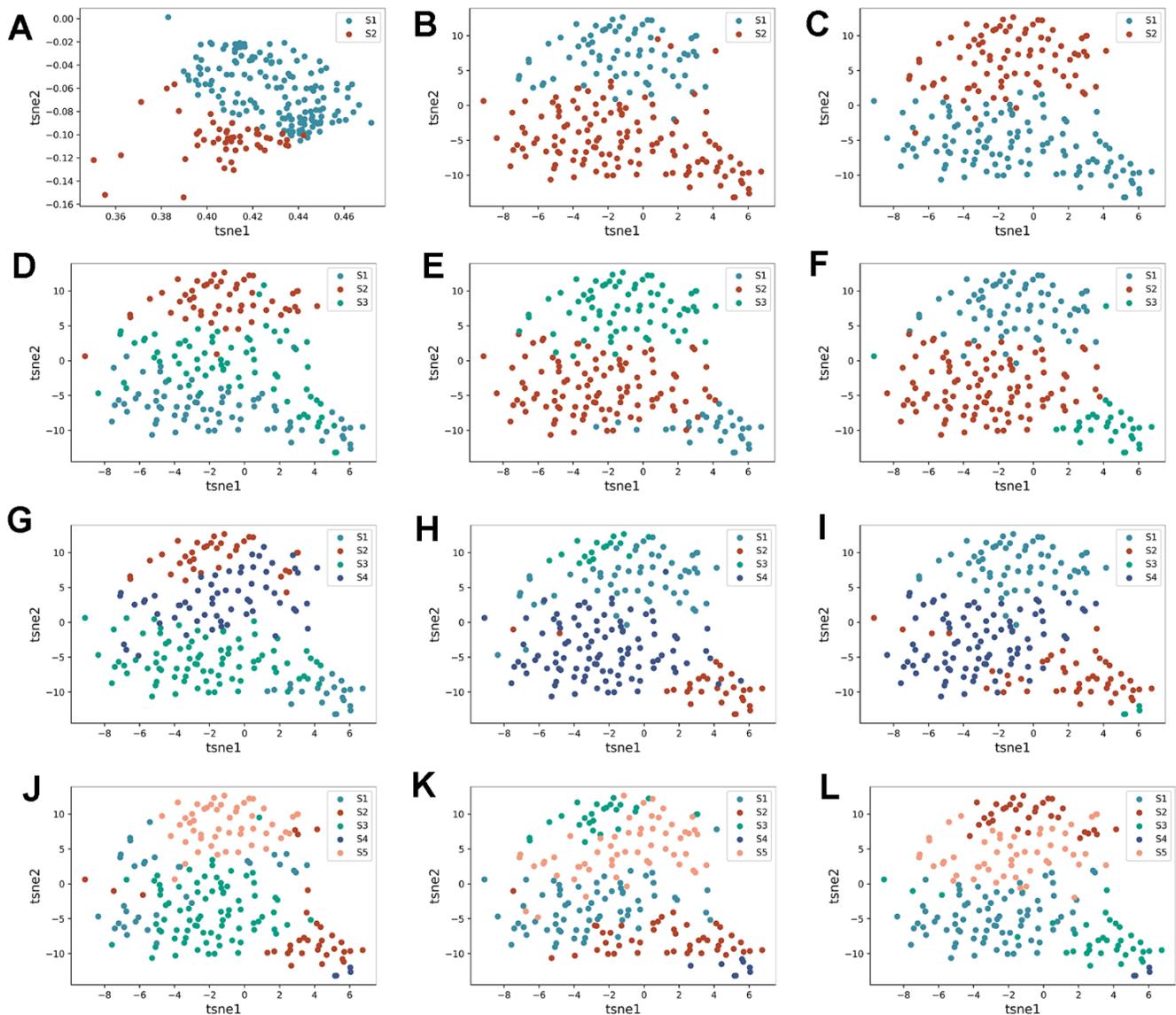


**Figure 2. The expression landscape of DEICDRGs. (A)** The intersection Wayne diagram of DEGs and ICDRGs ( $p=3.409506e-52$ ). **(B)** The expression heat map of DEICDRGs obtained by differential expression analysis. **(C)** The box diagram of the differential expression of DEICDRGs in HF and its control group. **(D)** The chromosome map of DEICDRGs. **(E)** The correlation analysis heat map of DEICDRGs.

### Construction and verification of lasso model and SVM model

This paper extracted diagnosis-related genes from 14 DEICDRGs using the LASSO and SVM-RFE algorithms to identify genes relevant to HF diagnosis. Figure 7A, 7B display the curves depicting the relationship and cross-validation results of the L1 norm and coefficients obtained by applying the LASSO algorithm. Figure 7C is the result of feature selection using SVM regression. When the number of DEICDRGs was 12, the AUC of 10-fold cross-validation was 0.978. Figure 7D is the

Venn diagram of the gene intersection screened by LASSO and SVM. We identified 11 genes in the intersection. In the Supplementary Table 2, we provide detailed information on diagnosis-related genes selected by the LASSO algorithm, the SVM-RFE algorithm, and their intersection genes. Figure 7E, 7F are ROC curves of the diagnostic model constructed using a set of 12 genes in both the training and test sets. Among them, the AUC in the training set reached 0.995 (CI: 0.986-0.999). The AUC in the test set reached 0.95 (CI: 0.841-1), indicating a high level of diagnostic accuracy. In addition, we also evaluated the AUC for each diagnosis-related gene.

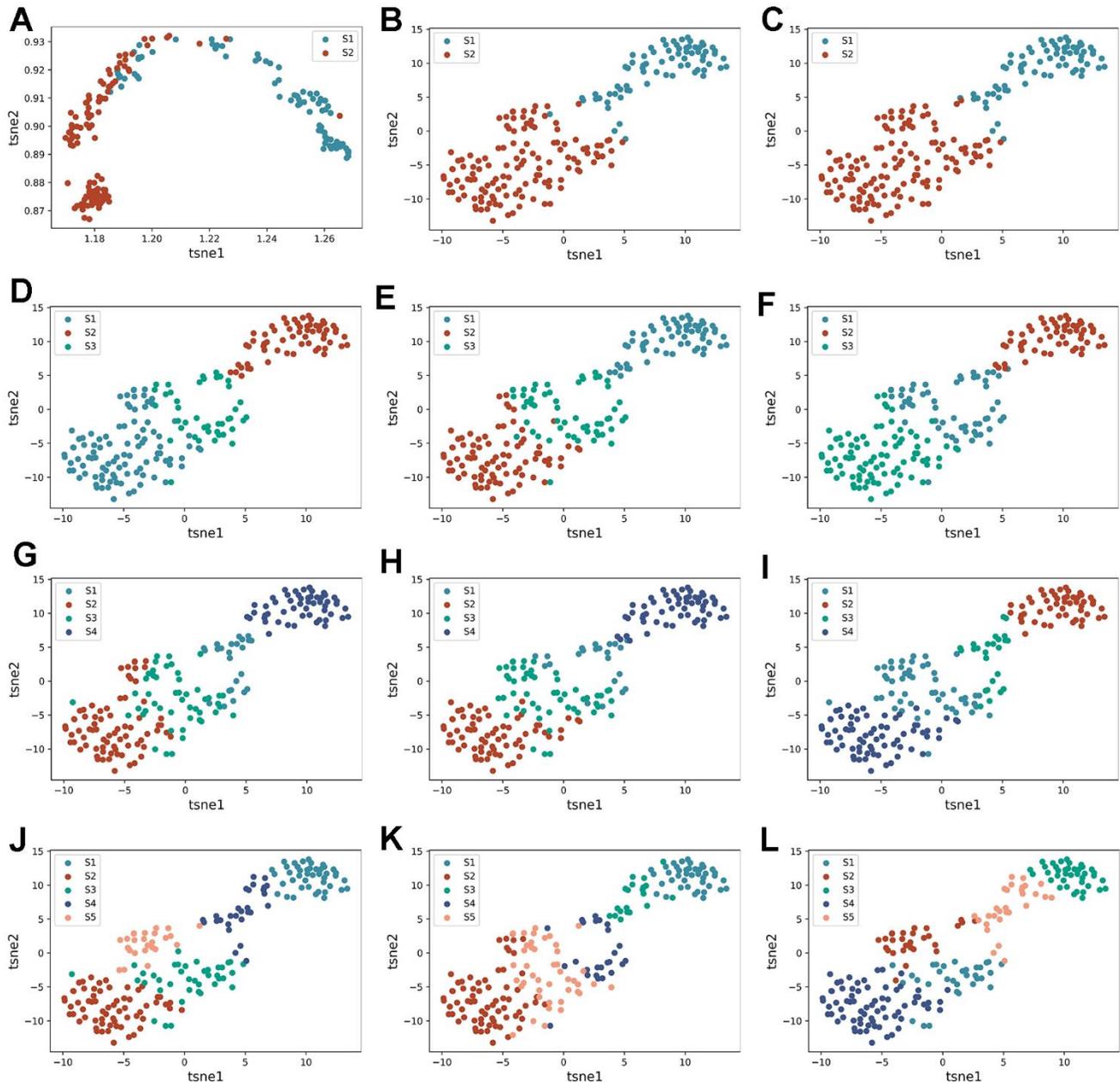


**Figure 3. Identification process of HF subtypes.** (A–C) are the tsne dimensionality reduction scatter plot obtained by AE, DAE, and K-means algorithms when the number of clusters is set to two categories, respectively. (D–F) are the tsne dimension reduction scatter plot obtained by three algorithms when the number of clusters is set to three categories, respectively. (G–I) are the tsne dimension reduction scatter plot obtained by three algorithms when the number of clusters is set to four categories, respectively. (J–L) are the tsne dimension reduction scatter plot obtained by three algorithms when the number of clusters is set to five categories, respectively.



by many cytokines. For a long time, we have been concerned that HF is related to systemic inflammation. Essentially, the progress of HF is attributed to the continuous signal transduction of pro-inflammatory cytokines, and the early stage of HF also shows an inflammatory state [19–21]. The protein encoded by IL-10 gene is a cytokine that plays a pleiotropic role in immune regulation and inflammation and participates in

the regulation of JAK-STAT signaling pathway. TNF encodes a multifunctional pro-inflammatory cytokine belonging to the TNF superfamily that is involved in regulating apoptosis. The relationship between HF and TNF was recognized as early as 1990 [22]. The mortality of patients with HF increases with the increase in TNF- $\alpha$  level [23]. TNF family members may represent a new target for HF treatment [24].



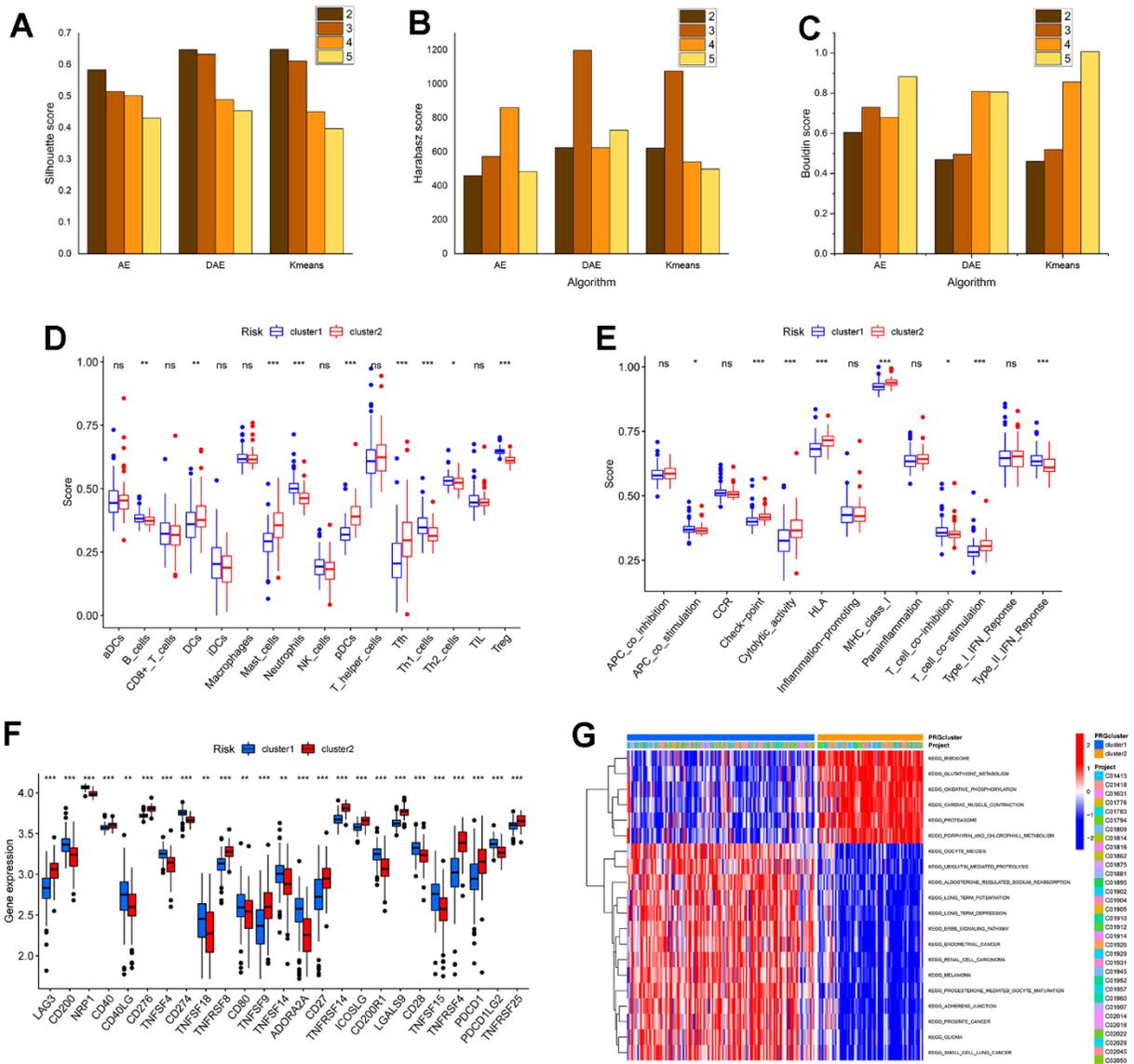
**Figure 5. Re-identification of HF subtypes based on DEGs.** (A–C) are the tsne dimensionality reduction scatter plot obtained by AE, DAE, and K-means algorithms when the number of clusters is set to two categories, respectively. (D–F) are the tsne dimension reduction scatter plot obtained by three algorithms when the number of clusters is set to three categories. (G–I) are the tsne dimension reduction scatter plot obtained by three algorithms when the number of clusters is set to four categories, respectively. (J–L) are tsne dimension reduction scatter plot obtained by three algorithms when the number of clusters is set to five categories, respectively.

Figure 8A–8K show the ROC curves of the following genes (ATG5 (AUC:0.723,CI:0.67-0.775), CASP1 (AUC:0.883,CI:0.848-0.917), CD8A(AUC:0.703,CI:0.644-0.755), ENTPD1(AUC:0.663,CI:0.610-0.713), IL1R1(AUC:0.748,CI:0.697-0.798), IL10(AUC:0.922,CI:0.893-0.949), IL17RA(AUC:0.842,CI:0.800-0.880), MYD88(AUC:0.690,CI:0.635-0.742), NT5E(AUC:0.913,CI:0.829-0.942), PRF1 AUC:0.869,CI:0.829-0.903, and TNF (AUC:0.677, CI:0.619-0.729)) in the training set. Figure 9A–9K display the ROC curves of the following genes (ATG5(AUC:0.723,CI:0.670-0.775), ASP1(AUC:0.883,CI:0.848-0.917), D8A(AUC:0.703,

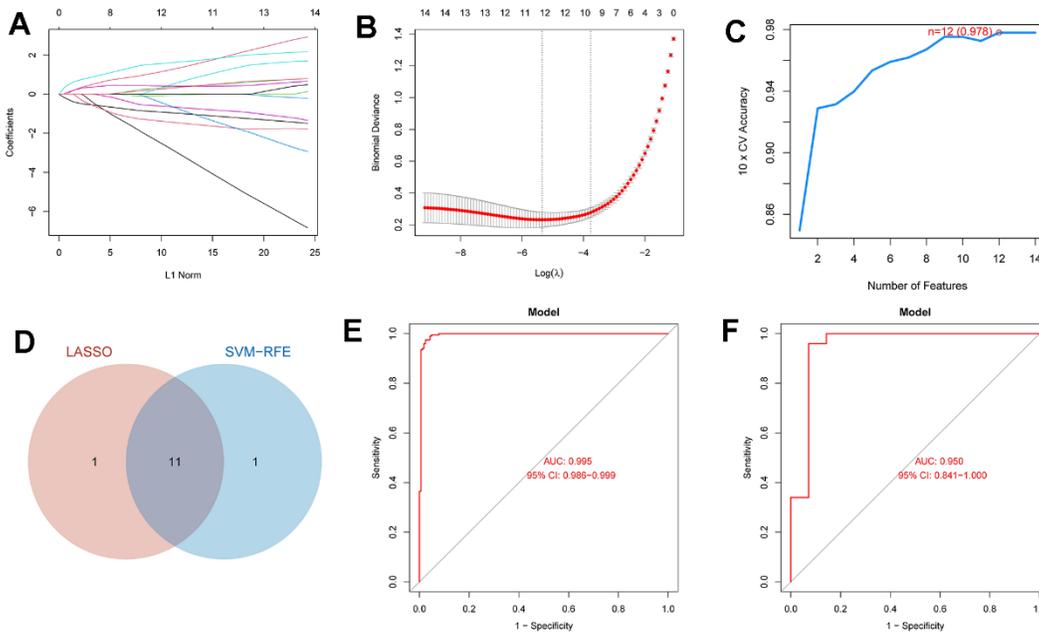
CI:-.644-0.755), NTPD1(AUC:0.663,CI:0.610-0.713), L1R1(AUC:0.748,CI:0.697-0.798), L10(AUC:0.922,CI:0.893-0.949), L17RA(AUC:0.842,CI:0.800-0.880), YD88(AUC:0.690,CI:0.635-0.742), T5E(AUC:0.913,CI:0.879-0.942), RF1(AUC:0.869,CI:0.829-0.903), and TNF(AUC:0.677,CI:0.619-0.729)) in the test set. All the genes had diagnostic significance for HF.

### Construction of nomogram model

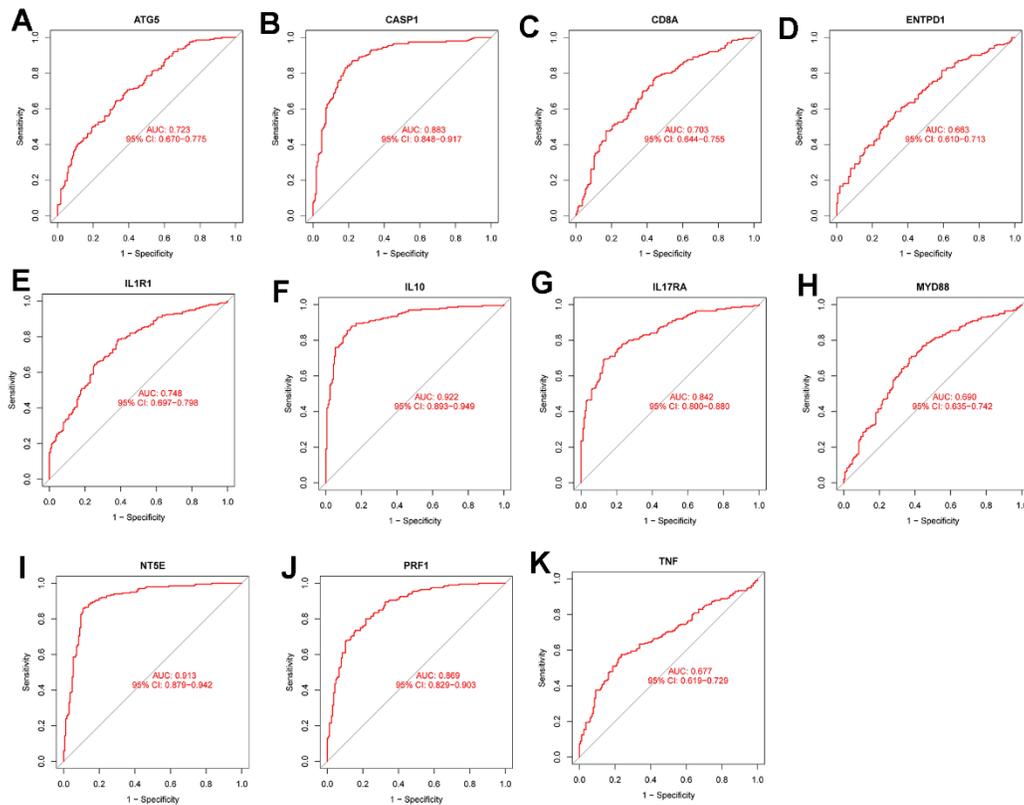
We developed a nomogram model by using diagnosis-associated genes (Figure 10A). The calibration curve in



**Figure 6. The performance of algorithm and the analysis of subtypes based on DEGs identification in the immune microenvironment and biological function characteristics.** (A–C) are the histogram of the sample contour coefficient, Calinski-Harabasz Index, and Davies-Bouldin Index of three algorithms under four clustering numbers, respectively. (D, E) are the difference box diagram of infiltration abundance and immune function of immune cells among different types, respectively. (F) The box diagram of the difference in immune examination sites between different subtypes. (G) The GSVA analysis between subtypes.



**Figure 7. The diagnosis model and ROC analysis of the model based on LASSO and SVM.** (A) The relationship curve between L1 norm and coefficient in the Lasso regression of DEICDRGs. (B) The cross-validation result of Lasso regression. (C) The result of using SVM regression to filter features. (D) The Wayne diagram of the intersection of characteristic genes screened by Lasso and SVM. (E, F) are ROC analysis of the diagnosis model in the training set and test set, respectively.



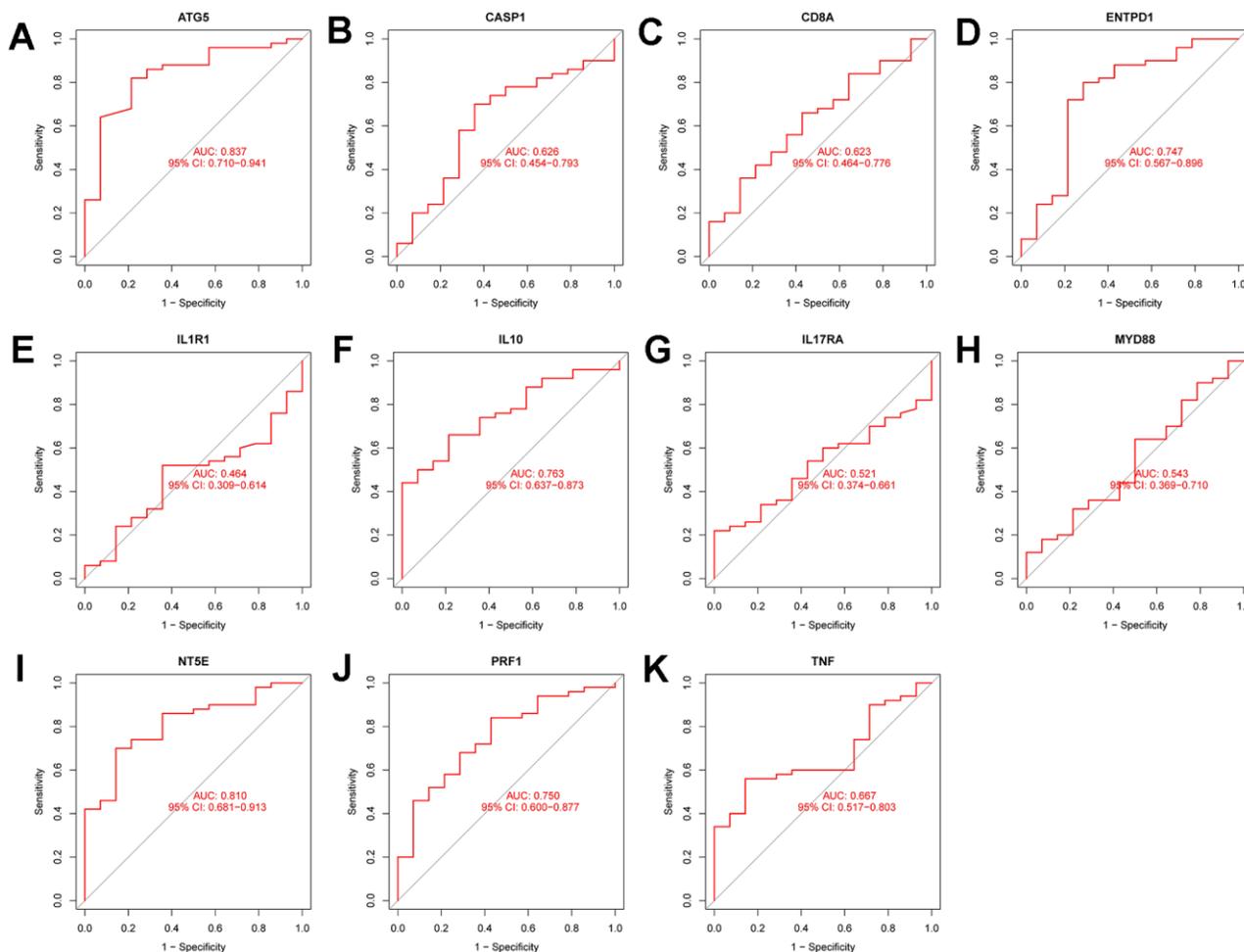
**Figure 8. Diagnostic performance verification of diagnostic genes and diagnostic models in the training set.** (A-K) are ROC curves of ATG5, CASP1, CD8A, ENTPD1, IL1R1, IL10, IL17RA, MYD88, NT5e, PRF1 and TNF in the training set, respectively.

Figure 10B illustrated that the nomogram model had excellent diagnostic ability. The DCA of Figure 10C proved that the nomogram model has more excellent clinical utility than a single diagnosis-related gene. The clinical influence curve of Figure 10D demonstrated that the nomogram model had outstanding diagnostic ability.

### Correlation analysis of immune infiltration and construction of the regulatory network

We developed a nomogram model utilizing diagnosis-related genes (Figure 10A). The calibration curve in Figure 10B demonstrated that the nomogram model had excellent diagnostic ability. The DCA in Figure 10C indicated that the Nomogram model exhibited higher clinical utility than a single diagnosis-related gene. The clinical influence curve in Figure 10D showed that the nomogram model had remarkable diagnostic ability. We present a scatter plot of the correlation between immune cells/functions and diagnostic genes in Figure 11A–11V.

For the miRNA-mRNA interaction network (Figure 12), we confirmed that some miRNAs are related to the pathogenesis of HF by consulting the literature. MiR-423-5p was initially identified as a circulating biomarker of heart disease. Tijssen et al. proved that the circulating level of miR-423-5p in patients with clinical HF increased [25]. Deng et al. also studied and determined that miR-423-5p is a potential target for the diagnosis and treatment of HF [26]. The protein coding gene regulated by miR-107 and the gene regulated by miR-139-5p were identified as genes that play a role in HF [27]. For the miRNA-TF interaction network, we know that CREB3 can promote the expression of inflammatory genes. RELA, also known as NF-kappa-B, is a pleiotropic transcription factor, exists in almost all cell types and is the endpoint of a series of signal transduction events, which are triggered by a large number of stimuli related to many biological processes, such as inflammation, immunity, differentiation, cell growth, tumorigenesis,



**Figure 9. Diagnostic performance verification of diagnostic genes and diagnostic models in the test set. (A–K) are ROC curves of ATG5, CASP1, CD8A, ENTPD1, IL1R1, IL10, IL17RA, MYD88, NT5e, PRF1, and TNF in the test set, respectively.**

and apoptosis. ATF1 can regulate the expression of downstream target genes to affect cell physiological processes, which is related to soft tissue melanoma, and the relationship between melanoma and HF has been confirmed in previous literature [9–11].

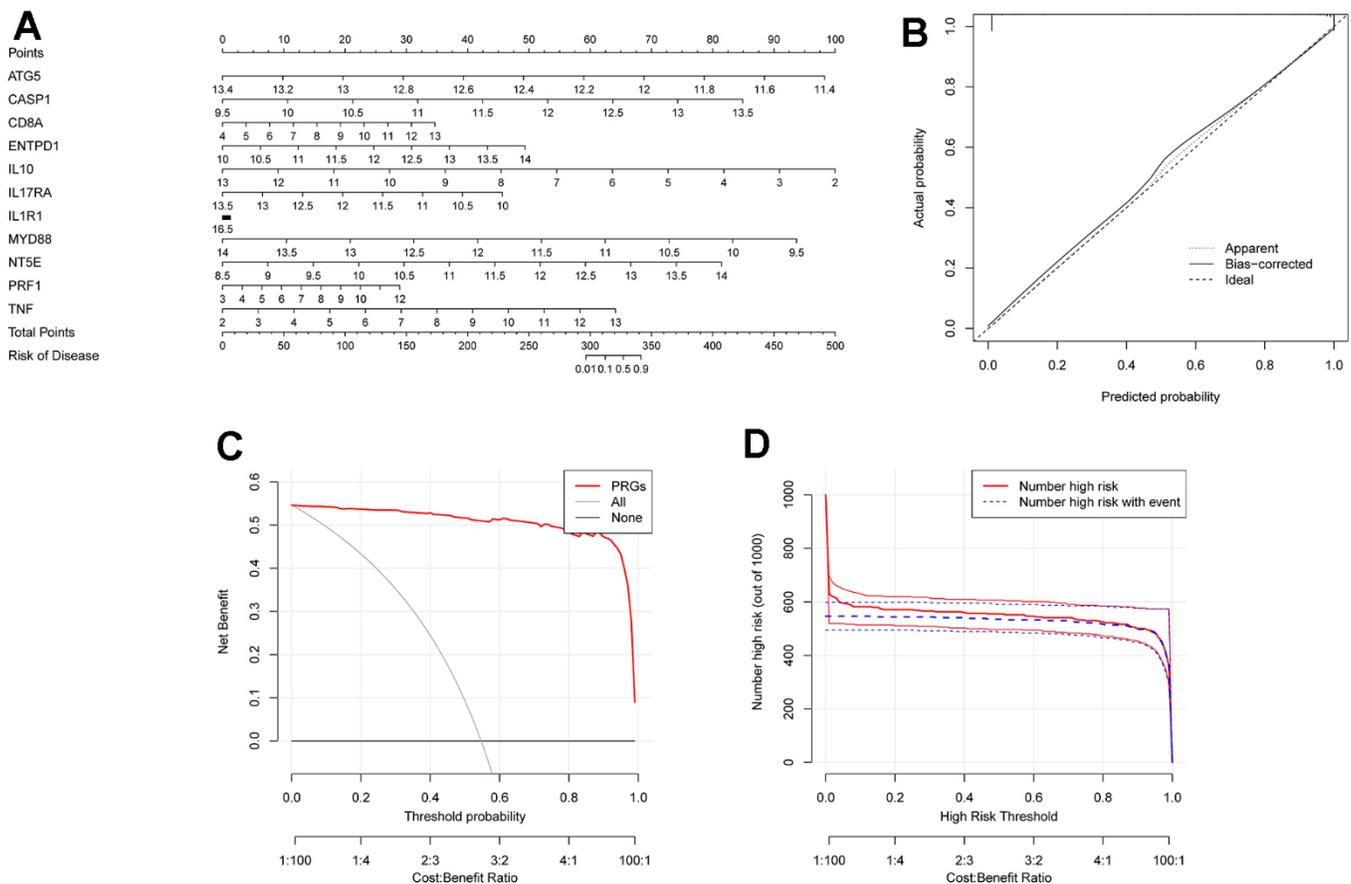
### Results qRT-PCR experimental verification

As shown in Figure 13, the expression levels of diagnosis-related genes MYD8, TNF, ATG5, CD8A, ENTPD1, IL17RA, NT5E, IL1R1, PRF1, IL-10 and CASP1 in the HF-cell model and control cell model. Among them, the p-value of MYD8, ATG5, ENTPD1, and IL1R1 expression in both groups was less than 0.05. MCAD, CD8A, IL17RA, NT5E, and CASP1 had p-value less than 0.01 in both groups. PRF1 and IL-10 had p-value less than 0.001 in both groups. The expression trends of these genes through qRT-PCR experiments were consistent with the results of differential expression analysis.

### DISCUSSION

Heart failure is a deterioration in heart function resulting from various heart diseases, and patients will show symptoms such as shortness of breath, fatigue, and palpitation. Complex HF often occurs in the elderly and is challenging to diagnose. ICD can elicit various immune reactions, but the function of ICDRGs in HF is still unclear. Therefore, this paper uses three clustering algorithms to identify the subtypes of HF based on ICDRGs. The DAE model is divided into two clusters by three clustering indexes for the optimal result. Significant differences exist between the two subtypes in the infiltration levels of immune cells, immune function, and the expression of genes related to immune surveillance sites.

We also cluster again according to DEGs between subtypes to verify the reliability of the above clustering results. We found that there were significant differences



**Figure 10. Construction of column diagram model.** (A) is a nomogram model constructed on selected diagnostically-relevant genes (ATG5, CASP1, CD8A, ENTPD1, IL1R1, IL10, IL17RA, MYD88, NT5E, PRF1, and TNF). (B) is the calibration curve used to evaluate the diagnostic capability of the nomogram model. (C) is based on DCA, and the Nomogram model had higher clinical utility than a single diagnostically-relevant gene. (D) is the clinical impact curve.

in the immune microenvironment and functional enrichment among different subtypes of samples obtained by re-clustering. Furthermore, we used the LASSO algorithm and the SVM algorithm to select genes related to the diagnosis of HF (ATG5, CASP1, CD8A, ENTPD1, IL1R1, IL10, IL17RA, MYD88, NT5E, PRF1, and TNF) and constructed the diagnosis model of HF.

Finally, based on the diagnosis-related genes, we constructed the miRNA-mRNA interaction network and the mRNA-TF interaction network, respectively. MiRNA is a small noncoding RNA molecule with a

length of about 22 nucleotides that regulates gene translation by silencing or degrading the target mRNA. They are involved in many biological processes, including differentiation and proliferation, metabolism, hemostasis, apoptosis or inflammation, and the pathophysiology of many diseases.

## CONCLUSIONS

This is a study to identify HF subtypes based on immunogenic cell death related genes and through multiple advanced deep-learning techniques. The two subtypes have significant differences in immunological

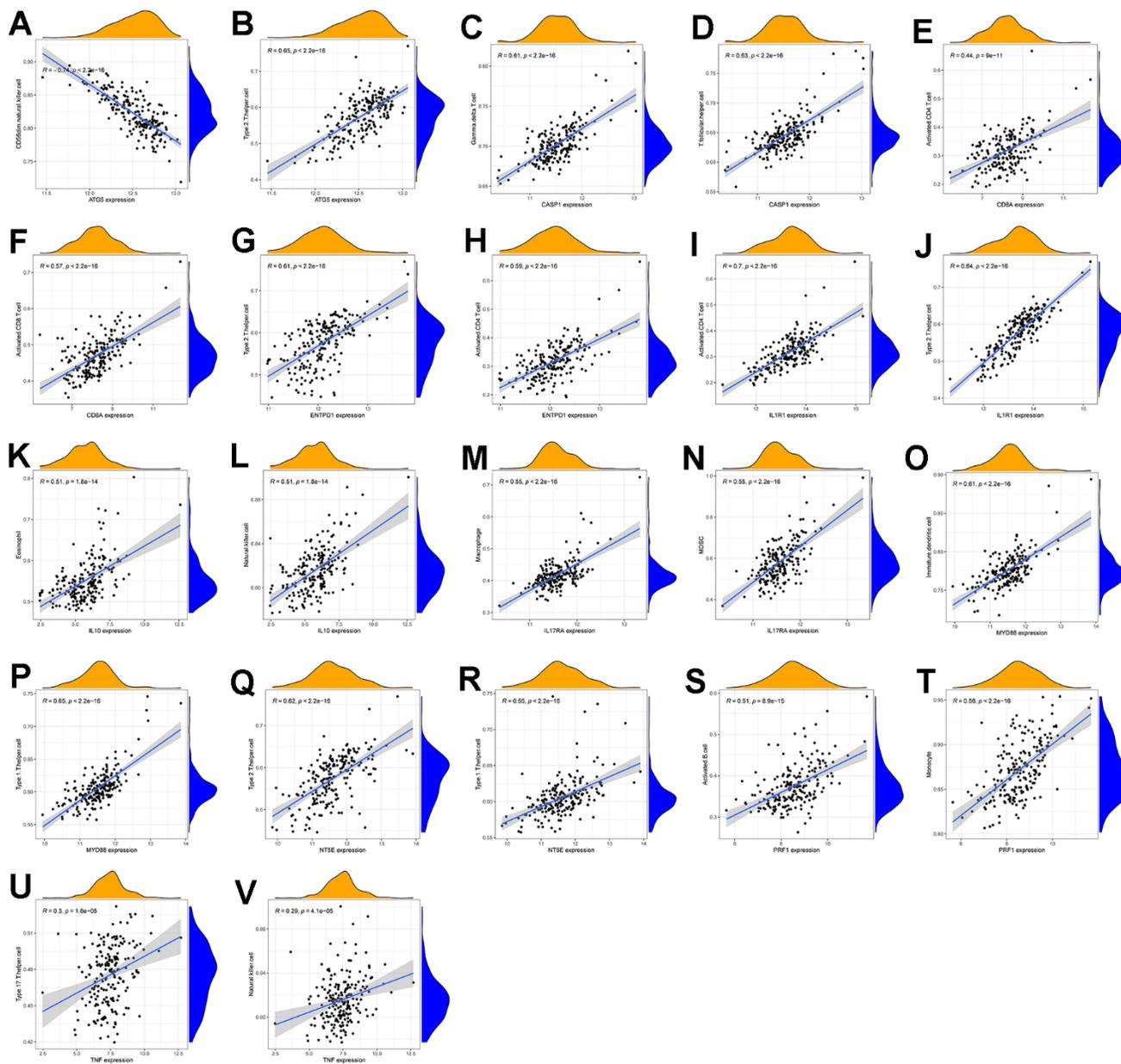
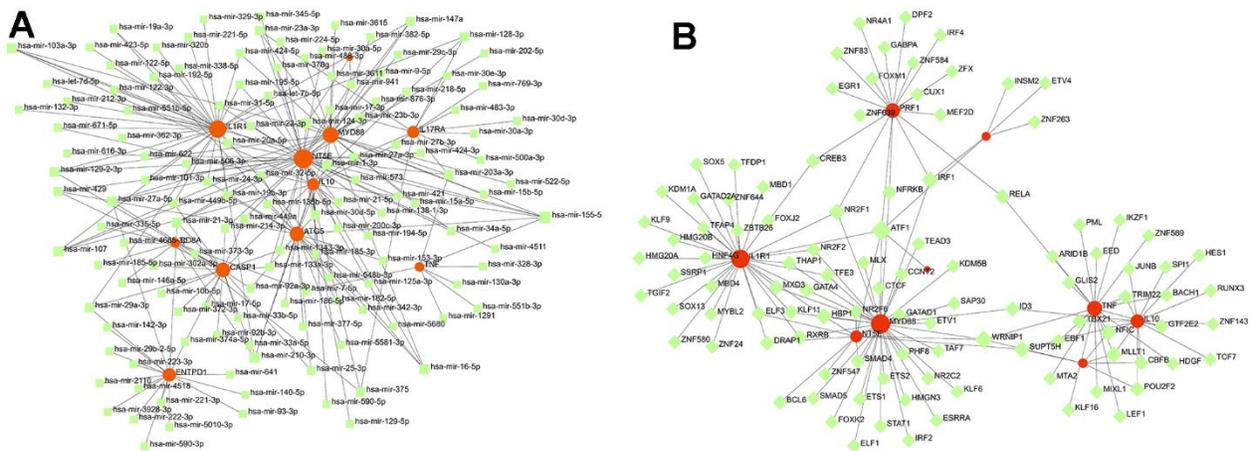
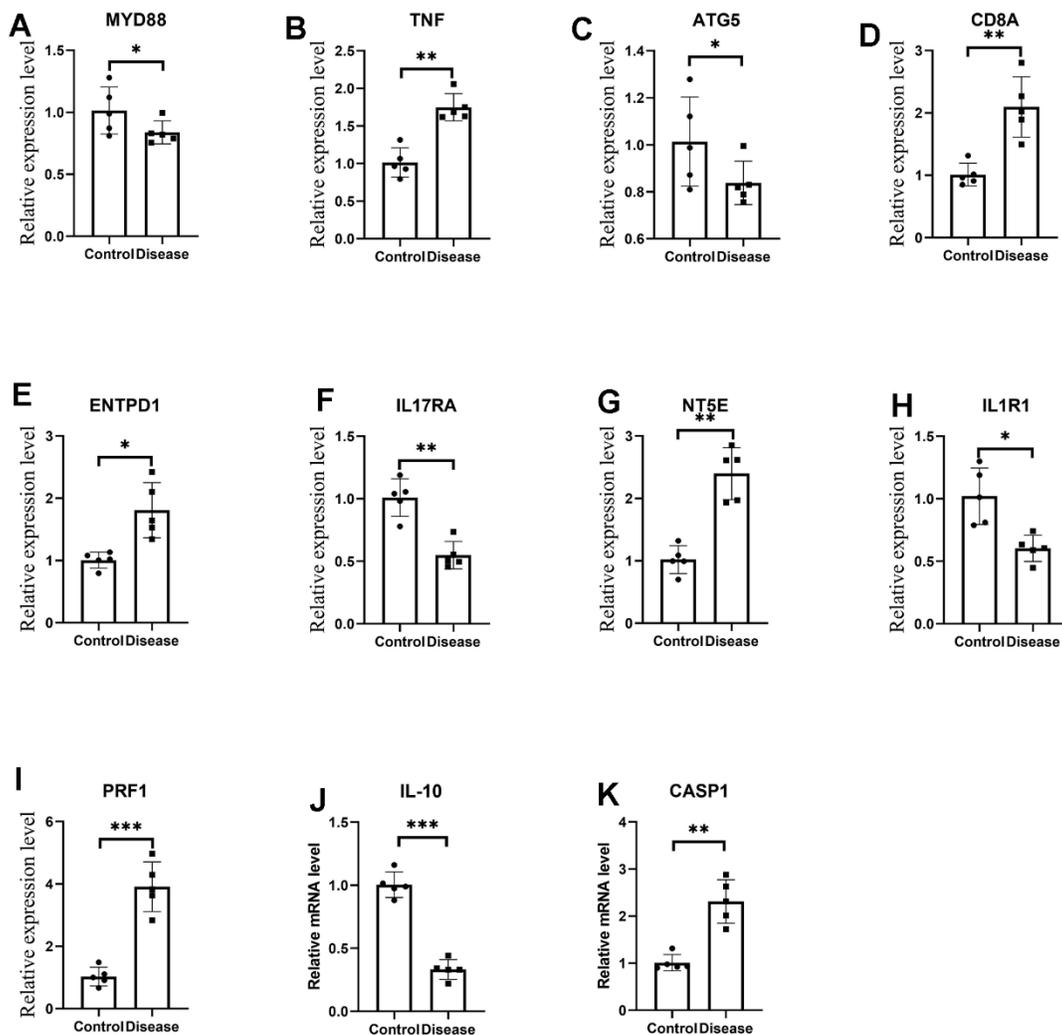


Figure 11. (A–V) are scatter plots of immune cell/function and diagnostic gene correlations, respectively.



**Figure 12.** Exploration of the interaction between diagnosis-related genes and miRNA and TF. (A) The miRNA-mRNA interaction network of diagnosis-related genes. (B) The mRNA-TF interaction network of diagnosis-related genes.



**Figure 13.** (A-K) are the mRNA expression level of MYD88, TNF, ATG5, CD8A, ENTPD1, IL17RA, NT5E, IL1R1, PRF1, IL-10 and CASP1 by qRT-PCR. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001.

characteristics and physiological functions. In addition, a robust heart failure diagnosis model was constructed based on machine learning models. Biomarker genes including ATG5, CASP1, CD8A, ENTPD1, IL1R1, IL10, IL17RA, MYD88, NT5E, PRF1, and TNF were identified. Finally, the interplay between biomarker genes, miRNAs, and transcription factors was explored by constructing a nomogram model. In conclusion, this article demonstrates the potential diagnostic utility of genes associated with immunogenic cell death in HF and hopes to help improve the risk stratification of HF and provide potential therapeutic targets.

## MATERIALS AND METHODS

### Clustering algorithm

#### Autoencoders

Autoencoders (AE) is a deep neural network that consists of an encoder and a decoder. Both the encoder and decoder are composed of multilayer feedforward neural networks. They are connected by the bottleneck layer. The encoder and decoder are represented by Formula (2) and Formula (3), respectively.

$$z = f_{\text{encoder}}(x) \quad (1)$$

$$x' = f_{\text{decoder}}(z) \quad (2)$$

Where  $z$  is the output of the encoder, which can be regarded as a reduced-dimension representation of data.  $x'$  is the output of the decoder.  $f_{\text{encoder}}$  and  $f_{\text{decoder}}$  are multilayer neural networks. In this paper, all encoder-based models are implemented using PyTorch. Among them, the loss function used by AE is MSE loss functions. For all encoder parts, the number of network layers was set to [10, 5, cluster\_num]. Cluster\_num represents the number of clusters. For all decoder parts, the network layer number was set to [10, 5, cluster\_num]. For all models based on the self-encoder, epoch was set to 100 during training.

#### Denoising autoencoders

Different from AE, denoising autoencoders (DAE) construct partially damaged data by adding noise to the input data, and then restores it to the original input data by encoding and decoding. The newly generated  $\tilde{x}$  can be expressed by the following formula.

$$\tilde{x} = q_D(\tilde{x} | x) \quad (3)$$

Where  $q_D$  represents random mapping and obeys the unit normal distribution  $N(0, 1)$ . Its encoder and decoder can be represented by Formula (4) and Formula (5), respectively.

$$z = f_{\text{encoder}}(\tilde{x}) \quad (4)$$

$$x' = f_{\text{decoder}}(z) \quad (5)$$

#### K-means clustering algorithm

K-means clustering is a classical clustering algorithm, and its implementation steps are as follows: Firstly,  $k$  clustering centers are randomly selected. Then the distance from each sample point is calculated, and the cluster center will divide it into the nearest cluster, forming  $k$  clusters. Next, the centroid (mean) of each cluster is recalculated. Repeat the above process until the position of the center of mass no longer changes or reaches the set number of iterations. In this paper, the algorithm was implemented with default parameters based on the scikit-learn package of Python.

#### Algorithm evaluation index

In this paper, three evaluation indexes of clustering performance were introduced, including the sample contour coefficient (the value was between -1 and 1, the closer to 1, the better the clustering effect), the Calinski-Harabasz Index (the value was greater than 0, the better the clustering effect), and the Davies-Bouldin Index (the value was greater than 0, the closer to 0, the better the clustering effect). All of them were implemented by Python's scikit-learn package.

#### Data acquisition

All the data in this paper came from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>). Specifically, we used GSE141910 data set (126 diseased samples and 240 control samples) as the training set, and the GSE116250 data set (50 diseased samples and 14 control samples) as the test set. The GSE141910 data set is derived from left ventricular free wall tissue harvested during cardiac surgery from HF subjects undergoing transplantation and from unused donor hearts with apparently normal function. Cold myocardial paralysis was perfused before cardiac resection to block contraction and prevent ischemic damage, and tissue specimens were frozen in liquid nitrogen. The GSE141910 data set comes from 64 samples of human left ventricular tissue.

#### The expression of ICDRGs before and after renal ischemia-reperfusion

The differential expression analysis of the GSE141910 data set was carried out using the "limma" package, and 8885 differentially expressed genes were obtained. We set Adj.P.Val<0.05 as the threshold for screening differentially expressed genes. Then, this paper collected 20 ICDRGs from the previous work and

**Table 1. The primer sequences.**

Primers		Sequence (5'→3')
MYD88	Forward	AGTGGGATGGGGAGAACAGA
	Reverse	TGTAGTCCAGCAACAGCCAG
TNF	Forward	CCGTGAACTGCTACAGCGTG
	Reverse	GACACATCACCTTCCCGAT
ATG5	Forward	GGACAGTTGCACACACTAGGA
	Reverse	CCGGGTAGCTCAGATGTTCA
CD8A	Forward	AAATCGGGAGACAAGCCCAG
	Reverse	ACACAGGGAGGAAGACTGGA
ENTPD1	Forward	AGTTCTGTGCTCAGCCTTGG
	Reverse	TAGCCTTGCAGAAGGAGGGA
IL17RA	Forward	GCCCAGACCAGAAGAGTTCC
	Reverse	AAGAAGGGCTGGATCTGCAC
PRF1	Forward	GACAACGAGGTGGAGGACTG
	Reverse	AAGGAGGCCGTCATCTTGTG
IL-10	Forward	CCGTGGAGCAGGTGAAGAAT
	Reverse	GCCACCCTGATGTCTCAGTT
CASP1	Forward	ATCCACAATGGGCTCTGTTT
	Reverse	CTCTTTCAGTGGTGGGCATCT

intersected them with differentially expressed genes to obtain intersection genes (DEICDRGs). The expression of ICDRGs in the diseased group and the control group in the GSE141910 data set was displayed as a box graph. In order to evaluate the correlation between ICDRGs, the Pearson correlation coefficient of DEICDRGs in the sample is calculated and visualized by “corrplot” in R software.

### Enrichment analysis of different clusters

In this paper, the enrichment analysis of gene ontology (GO) and genome encyclopedia (KEGG) was realized by using the R package “clusterProfile.” And the analysis of gene set variation (GSVA) was realized by using R-packet “GSVA.” Among them, c2.cp.kegg.v7.4.symbols.gmt was downloaded from the MSigDB database and used to study the changes in biological signal pathways. The R package “ggplot2” was used to visualize the enrichment results.

### Immunoassay

The ssGSEA algorithm was used to estimate the infiltration abundance of immune cells and the score of immune function in the sick group and the control group. Then, we also collected the genes related to immune inspection sites and explored the difference in expression levels of genes related to immune inspection sites between the two groups by box chart.

### Construction and verification of HF-related diagnosis model

For the purpose of screening the diagnosis-related genes of HF, we adopt a small absolute contraction and selection operator (LASSO) and support vector machine recursive feature elimination (SVM-RFE). The R-packet “glmnet” was used to implement the LASSO algorithm in the case of ten times cross-validation. SVM-RFE was also implemented by R script under the condition of ten times cross-validation. Finally, we draw ROC curves of diagnostic genes and diagnostic models by using the R-package “pROC.”

### Construction of nomogram

The study constructed the nomogram model based on diagnosis-related genes using the R package “rms”. Then the validity of the nomogram model was evaluated by the calibration curve. The clinical practicability of the nomogram model was evaluated by a decision curve. Finally, the high-risk probability stratification was predicted by the clinical impact curve when the population size was 1000.

### Construction of interaction network

We conducted online analysis of hub genes using the NetworkAnalyst database (<https://www.networkanalyst.ca/NetworkAnalyst/>) to construct a transcription factor

(TF) hub gene network and a miRNA hub gene interaction network. In this research, miRNA and TF interacting with diagnosis-related genes were queried by using the miRTarBase database (<http://mirtarbase.cuhk.edu.cn/php/index.php>) and TargetScan database ([http://www.targetscan.org/vert\\_72/](http://www.targetscan.org/vert_72/)) respectively. Then, an interaction network between diagnostic-related genes and miRNA was constructed, the same between diagnostic-related genes and TF.

### Experimental validation of diagnosis related genes

In this paper, the mRNA expression of diagnosis related genes was detected by qRT-PCR. Total RNA was isolated from the cell culture of H9C2 cells for HF and HL-1 cells for normal using the TRI Reagent. Five samples were in the HF group and five in the normal group. The mRNA expression level was calculated with the  $2^{-\Delta\Delta Ct}$ . All data were expressed by means  $\pm$  SD, and the statistical differences between groups were tested by T-Test, and  $p < 0.05$  indicated a significant difference. The primer sequences are shown in Table 1.

### Data availability

The data used in the paper was downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>).

### AUTHOR CONTRIBUTIONS

Zhihui Ma: Conceptualization, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing-review and editing. Shixin Ma: Conceptualization, Methodology, Software, Visualization, Writing-review and editing. Bin Chen: Conceptualization, Methodology, Software, Writing-review and editing. Yongjun Zhang, and Jinmei Zeng: Data curation, Software, Writing-review and editing. Jianping Tao: Visualization, Writing-review and editing. Yu Hu: Data curation, Writing-review and editing. All authors read and approved the manuscript.

### CONFLICTS OF INTEREST

The authors have no relevant financial or non-financial interest to disclose.

### FUNDING

The biological experiments involved in this paper (qRT-PCR) were performed by the first author at his own expense.

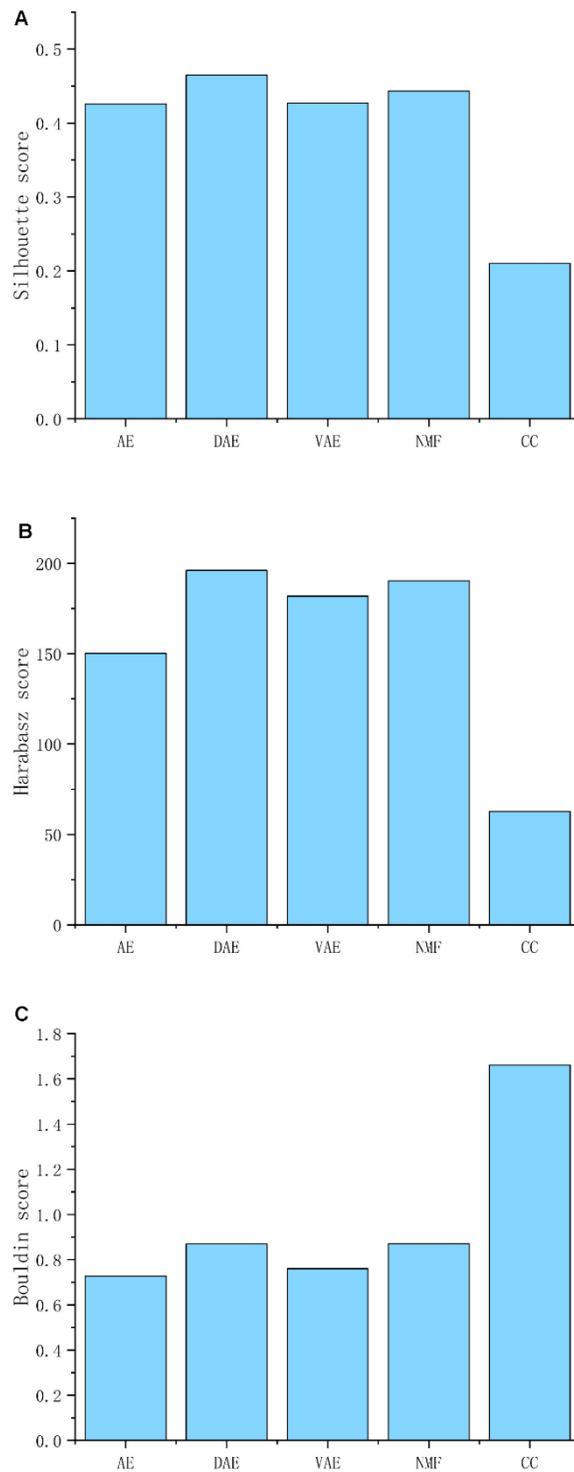
### REFERENCES

1. Rich MW. Heart failure. *Cardiol Clin*. 1999; 17:123–35. [https://doi.org/10.1016/s0733-8651\(05\)70060-6](https://doi.org/10.1016/s0733-8651(05)70060-6) PMID:[10093769](https://pubmed.ncbi.nlm.nih.gov/10093769/)
2. Tanai E, Frantz S. Pathophysiology of Heart Failure. *Compr Physiol*. 2015; 6:187–214. <https://doi.org/10.1002/cphy.c140055> PMID:[26756631](https://pubmed.ncbi.nlm.nih.gov/26756631/)
3. Goldstein D, Frishman WH. Diastolic Heart Failure: A Review of Current and Future Treatment Options. *Cardiol Rev*. 2021; 29:82–8. <https://doi.org/10.1097/CRD.0000000000000303> PMID:[32101908](https://pubmed.ncbi.nlm.nih.gov/32101908/)
4. Pourafkari L, Tajlil A, Nader ND. Biomarkers in diagnosing and treatment of acute heart failure. *Biomark Med*. 2019; 13:1235–49. <https://doi.org/10.2217/bmm-2019-0134> PMID:[31580155](https://pubmed.ncbi.nlm.nih.gov/31580155/)
5. Kolur V, Vastrad B, Vastrad C, Kotturshetti S, Tengli A. Identification of candidate biomarkers and therapeutic agents for heart failure by bioinformatics analysis. *BMC Cardiovasc Disord*. 2021; 21:329. <https://doi.org/10.1186/s12872-021-02146-8> PMID:[34218797](https://pubmed.ncbi.nlm.nih.gov/34218797/)
6. Li Y, Liu X, Zhang X, Pan W, Li N, Tang B. Immunogenic cell death inducers for enhanced cancer immunotherapy. *Chem Commun (Camb)*. 2021; 57:12087–97. <https://doi.org/10.1039/d1cc04604g> PMID:[34714302](https://pubmed.ncbi.nlm.nih.gov/34714302/)
7. Kroemer G, Galluzzi L, Kepp O, Zitvogel L. Immunogenic cell death in cancer therapy. *Annu Rev Immunol*. 2013; 31:51–72. <https://doi.org/10.1146/annurev-immunol-032712-100008> PMID:[23157435](https://pubmed.ncbi.nlm.nih.gov/23157435/)
8. Markham A. Lurbinedin: First Approval. *Drugs*. 2020; 80:1345–53. <https://doi.org/10.1007/s40265-020-01374-0> PMID:[32816202](https://pubmed.ncbi.nlm.nih.gov/32816202/)
9. Rusconi C, Faggiano P, Ghizzoni G, Sorgato A, Minzioni G, Sabatini T. Congestive heart failure due to rapid right ventricular obliteration by metastatic malignant melanoma. *Minerva Cardioangiol*. 1996; 44:123–5. PMID:[8767611](https://pubmed.ncbi.nlm.nih.gov/8767611/)
10. Tesolin M, Lapierre C, Oligny L, Bigras JL, Champagne M. Cardiac metastases from melanoma. *Radiographics*. 2005; 25:249–53. <https://doi.org/10.1148/rg.251045059> PMID:[15653599](https://pubmed.ncbi.nlm.nih.gov/15653599/)
11. Wang CY, Zoungas S, Voskoboynik M, Mar V. Cardiovascular disease and malignant melanoma. *Melanoma Res*. 2022; 32:135–41.

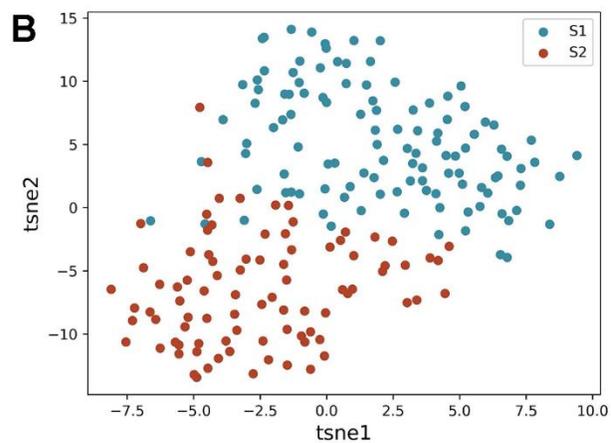
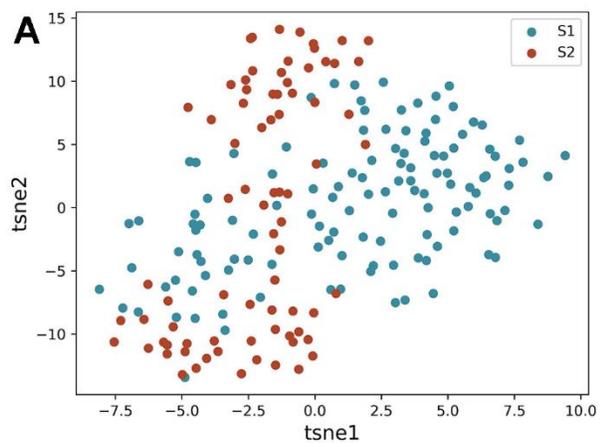
- <https://doi.org/10.1097/CMR.0000000000000817>  
PMID:[35377865](https://pubmed.ncbi.nlm.nih.gov/35377865/)
12. Sabbah HN, Sharov VG. Apoptosis in heart failure. *Prog Cardiovasc Dis*. 1998; 40:549–62.  
[https://doi.org/10.1016/s0033-0620\(98\)80003-0](https://doi.org/10.1016/s0033-0620(98)80003-0)  
PMID:[9647609](https://pubmed.ncbi.nlm.nih.gov/9647609/)
  13. Chen QM, Tu VC. Apoptosis and heart failure: mechanisms and therapeutic implications. *Am J Cardiovasc Drugs*. 2002; 2:43–57.  
<https://doi.org/10.2165/00129784-200202010-00006>  
PMID:[14727998](https://pubmed.ncbi.nlm.nih.gov/14727998/)
  14. Liao M, Xie Q, Zhao Y, Yang C, Lin C, Wang G, Liu B, Zhu L. Main active components of Si-Miao-Yong-An decoction (SMYAD) attenuate autophagy and apoptosis via the PDE5A-AKT and TLR4-NOX4 pathways in isoproterenol (ISO)-induced heart failure models. *Pharmacol Res*. 2022; 176:106077.  
<https://doi.org/10.1016/j.phrs.2022.106077>  
PMID:[35026404](https://pubmed.ncbi.nlm.nih.gov/35026404/)
  15. Booz GW, Day JN, Baker KM. Interplay between the cardiac renin angiotensin system and JAK-STAT signaling: role in cardiac hypertrophy, ischemia/reperfusion dysfunction, and heart failure. *J Mol Cell Cardiol*. 2002; 34:1443–53.  
<https://doi.org/10.1006/jmcc.2002.2076>  
PMID:[12431443](https://pubmed.ncbi.nlm.nih.gov/12431443/)
  16. Okonko DO, Marley SB, Anker SD, Poole-Wilson PA, Gordon MY. Erythropoietin resistance contributes to anaemia in chronic heart failure and relates to aberrant JAK-STAT signal transduction. *Int J Cardiol*. 2013; 164:359–64.  
<https://doi.org/10.1016/j.ijcard.2011.07.045>  
PMID:[21821297](https://pubmed.ncbi.nlm.nih.gov/21821297/)
  17. Du J, Liu Y, Fu J. Autophagy and Heart Failure. *Adv Exp Med Biol*. 2020; 1207:223–7.  
[https://doi.org/10.1007/978-981-15-4272-5\\_16](https://doi.org/10.1007/978-981-15-4272-5_16)  
PMID:[32671751](https://pubmed.ncbi.nlm.nih.gov/32671751/)
  18. Gao G, Chen W, Yan M, Liu J, Luo H, Wang C, Yang P. Rapamycin regulates the balance between cardiomyocyte apoptosis and autophagy in chronic heart failure by inhibiting mTOR signaling. *Int J Mol Med*. 2020; 45:195–209.  
<https://doi.org/10.3892/ijmm.2019.4407>  
PMID:[31746373](https://pubmed.ncbi.nlm.nih.gov/31746373/)
  19. Shirazi LF, Bissett J, Romeo F, Mehta JL. Role of Inflammation in Heart Failure. *Curr Atheroscler Rep*. 2017; 19:27.  
<https://doi.org/10.1007/s11883-017-0660-3>  
PMID:[28432635](https://pubmed.ncbi.nlm.nih.gov/28432635/)
  20. Schiattarella GG, Rodolico D, Hill JA. Metabolic inflammation in heart failure with preserved ejection fraction. *Cardiovasc Res*. 2021; 117:423–34.  
<https://doi.org/10.1093/cvr/cvaa217>  
PMID:[32666082](https://pubmed.ncbi.nlm.nih.gov/32666082/)
  21. Triposkiadis F, Xanthopoulos A, Starling RC, Iliodromitis E. Obesity, inflammation, and heart failure: links and misconceptions. *Heart Fail Rev*. 2022; 27:407–18.  
<https://doi.org/10.1007/s10741-021-10103-y>  
PMID:[33829388](https://pubmed.ncbi.nlm.nih.gov/33829388/)
  22. Levine B, Kalman J, Mayer L, Fillit HM, Packer M. Elevated circulating levels of tumor necrosis factor in severe chronic heart failure. *N Engl J Med*. 1990; 323:236–41.  
<https://doi.org/10.1056/NEJM199007263230405>  
PMID:[2195340](https://pubmed.ncbi.nlm.nih.gov/2195340/)
  23. Müller-Ehmsen J, Schwinger RH. TNF and congestive heart failure: therapeutic possibilities. *Expert Opin Ther Targets*. 2004; 8:203–9.  
<https://doi.org/10.1517/14728222.8.3.203>  
PMID:[15161427](https://pubmed.ncbi.nlm.nih.gov/15161427/)
  24. Ueland T, Yndestad A, Dahl CP, Gullestad L, Aukrust P. TNF revisited: osteoprotegerin and TNF-related molecules in heart failure. *Curr Heart Fail Rep*. 2012; 9:92–100.  
<https://doi.org/10.1007/s11897-012-0088-6>  
PMID:[22453763](https://pubmed.ncbi.nlm.nih.gov/22453763/)
  25. Tijssen AJ, Creemers EE, Moerland PD, de Windt LJ, van der Wal AC, Kok WE, Pinto YM. MiR423-5p as a circulating biomarker for heart failure. *Circ Res*. 2010; 106:1035–9.  
<https://doi.org/10.1161/CIRCRESAHA.110.218297>  
PMID:[20185794](https://pubmed.ncbi.nlm.nih.gov/20185794/)
  26. Vilella-Figuerola A, Gallinat A, Escate R, Mirabet S, Padró T, Badimon L. Systems Biology in Chronic Heart Failure-Identification of Potential miRNA Regulators. *Int J Mol Sci*. 2022; 23:15226.  
<https://doi.org/10.3390/ijms232315226>  
PMID:[36499552](https://pubmed.ncbi.nlm.nih.gov/36499552/)
  27. Deng J, Zhong Q. Advanced research on the microRNA mechanism in heart failure. *Int J Cardiol*. 2016; 220:61–4.  
<https://doi.org/10.1016/j.ijcard.2016.06.185>  
PMID:[27372044](https://pubmed.ncbi.nlm.nih.gov/27372044/)

## SUPPLEMENTARY MATERIALS

### Supplementary Figures



**Supplementary Figure 1.** t-sne results for AE, DAE, VAE, NMF, and CC when the number of clusters is 2. (A–C) are silhouette, harabasz, and bouldin scores of these algorithms, respectively.



**Supplementary Figure 2. t-sne results of NMF and CC when the number of clusters is 2. (A, B) are t-sne scatter plots of NMF and CC, respectively.**

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 1.

### Supplementary Table 1. Intersection\_case.

### Supplementary Table 2. Intersection\_of\_related\_diagnosis\_genes.

Only LASSO	Only SVM-RFE	LASSO AND SVM-RFE
IFNGR1	TLR4	ATG5
		CASP1
		CD8A
		ENTPD1
		IL10
		IL17RA
		IL1R1
		MYD88
		NT5E
		PRF1
		TNF