scientific reports

Check for updates

OPEN

A personalized mRNA signature for predicting hypertrophic cardiomyopathy applying machine learning methods

Jue Gu[1,4], Yamin Zhao[2,4], Yue Ben[1], Siming Zhang[3], Liqi Hua[3], Songnian He[3], Ruizi Liu[3], Xu Chen[3]✉ & Hongzhuan Sheng[1]✉

Hypertrophic cardiomyopathy (HCM) may lead to cardiac dysfunction and sudden death. This study was designed to develop a HCM signature applying bioinformatics and machine learning methods. Data of HCM and normal tissues were obtained from public databases to screen differentially expressed genes (DEGs) using the R software limma package. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were performed for enrichment analysis of HCM-associated DEGs. Hub genes for HCM were determined using weighted gene co-expression network analysis (WGCNA) together with two machine learning algorithms (SVM-RFE and LASSO). Finally, we introduced a zebrafish model to simulate changes in the hub genes in the HCM and to observe their effects on cardiac disease development. The mRNA expression data from a total of 106 HCM tissues and 39 normal samples were collected and we screened 157 DEGs. Enrichment analysis showed that immune pathways played an important role in the pathogenesis of HCM. Three hub genes (FCN3, MYH6 and RASD1) were identified using WGCNA, SVM-RFE, and LASSO analysis. In a zebrafish model, knockdown of MYH6 and RASD1 resulted in cardiac malformations with reduced ventricular capacity and heart rate, which validated the clinical significance of these genes in the diagnosis of HCM. Based on machine learning algorithms, our study created a signature with potential impact on cardiac function and cardiac quality index for HCM. The current findings had important implications for the early diagnosis and treatment of HCM.

Keywords HCM, Machine learning, Zebrafish, Prediction, Bioinformatics

Cardiovascular disease remains the leading cause of death worldwide[1]. In 2019, a prospective urban and rural epidemiological study published by *the Lancet* showed that 40% of all deaths were caused by cardiovascular diseases[2]. Although the etiology of cardiovascular disease is different, heart failure is commonly the final stage. Noticeably, pathological cardiac hypertrophy could easily develop into heart failure and therefore becomes an increasingly important cause of cardiovascular diseases[3,4]. Hypertrophic cardiomyopathy (HCM), which is characterized by myocardial hypertrophy asymmetry, often occurs in ventricular septum, resulting in outflow tract obstruction, left ventricular filling limitation and reduced compliance[5]. In some cases, HCM will lead to heart failure, myocardial ischemia and sudden death. Thus, early detection of HCM becomes highly important. HCM is a genetic heterogeneous disorder associated with mutations in certain genes[6]. Development of gene sequencing technology has raised the significance of using genetic detection in the diagnosis of HCM patients with a family history and in asymptomatic patients to avoid sudden death. The European and American Guidelines also encourage the use of genetic testing for potential HCM patients[7,8].

Machine learning refers to the process of learning and training from data and accurately predicting the system of future events. It is a multidisciplinary specialty and a type of artificial intelligence applied in data processing of bioinformatics[9,10]. Machine learning can identify potential rules through massive amounts of data, outperforming most traditional statistical methods[11,12]. In the medical field, machine learning can be used as a predictive model to guide precision medicine[13]. In protein function research, machine learning can promote prediction

[1]Affiliated Hospital of Nantong University, No.20 Xisi Road, Nantong 226000, Jiangsu Province, China. [2]Nantong Second People's Hospital, Nantong, China. [3]Medical School of Nantong University, Nantong University, Nantong, China. [4]These authors contributed equally: Jue Gu and Yamin Zhao. ✉email: chenxu20211104@163.com; yjshz@ntu.edu.cn

accuracy and enable more comprehensive analysis[14]. In metabolic engineering, the integration of machine learning enriches data analysis techniques and enhances the precision of metabolic outcome prediction[15]. Recently, machine learning method has been applied to the development of prognostic models for various malignant tumors[16–18]. In addition, multiple machine learning algorithms are combined to mine genes for predicting the prevalence of bronchopulmonary dysplasia patients, and results have confirmed that genes discovered by bioinformatics can serve as potential targets for identifying the disease and contribute to its treatment[10]. However, studies about machine learning in HCM is largely limited.

This study screened DEGs from the GSE36961 database for HCM. Applying machine learning method, key genes were mined and the area under the curve (AUC) showed that these genes had a strong predictive performance. Moreover, a zebrafish model was constructed to verify the effectiveness of the models. In this paper, through machine learning, valuable biomarkers for HCM were screened and their clinical diagnostic value and treatment direction were analyzed.

Materials and methods

Data acquisition and screening of DEGs

The microarray dataset GSE36961 of Bos et al.[19], which included mRNA expression data from 106 HCM tissues and 39 normal samples, was downloaded from the Gene Expression Omnibus (GEO) database for analysis. The raw data of GSE36961 was preprocessed using the "limma" package in R software (version 3.6.2). Missing values were filled utilizing the k-nearest neighbor algorithm[20]. Raw data were normalized by running robust multiarray average algorithm[21]. Elimination of batch effect was performed in the "sva" package in R employing the COMBAT method. Identifying the anchor points through principal component analysis (PCA), and assessing the top two principal components using the t-distributed stochastic neighbor embedding (t-SNE) technique to unveil notable groupings. The screening threshold value of DEGs was $|Log2FC| > 1.0$, adjust. p value < 0.05. Subsequently, the heatmaps and volcano maps were plotted by the "pheatmap" and "EnhanceVolcano" packages of the R software, respectively. In addition, we validated the comparison of the expression levels of the screened key genes in HCM and control samples based on the GSE141910 dataset.

Enrichment analysis of the genes

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using the "clusterProfiler" package, a commonly analytical tool in bioinformatics to identify statistically significantly enriched biological terms[22]. We used the list of DEGs as the input for GO analysis using the "enrichGO" function and the "enrichKEGG" function for KEGG pathway analysis. Subsequently, the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/gsea/ index. jsp) was used to obtain Hallmark and C7 immunosignature gene sets. According to the expression level of DEGs as the phenotype annotation, data from HCM patients were divided into the HCM group and normal group under NOM $p < 0.05$ and FDR $q < 0.25$.

Weighted correlation network analysis (WGCNA)

A scale-free weighted gene co-expression network was developed using the "WGCNA" package in R to identify co-expressed genes and modules associated with clinical features[23]. The data were clustered, outliers were detected, and appropriate soft thresholds were set to obtain a network as scale-free as possible. Topological overlap matrix-based hierarchical clustering was used to section gene modules. To determine the correlation of each module with clinical features, Pearson correlation coefficients were calculated to select the module with the greatest clinical correlation and the mRNAs in the module were obtained. The glmnet package was used to construct a LASSO regression model using differential genes, and e1071, caret and kernlab packages were loaded to develop SVM-RFE model. Intersection genes between the two models were defined as the signature genes and Venn diagram was drawn for visualization.

Protein–protein interaction (PPI) network construction

PPI analysis of turquoise module genes screened by WGCNA was performed using the STRING database (https://string-db.org/). The PPI network was visualized using Cytoscape software[24]. In the network, nodes represent proteins and edges represent interactions between proteins. The core genes in the network were identified by calculating the degree of connectivity (degree) of each node.

Support vector machine based recursive feature elimination (SVM-RFE)

Support vector machine (SVM), a robust machine learning algorithm, has been extensively used to predict the functions of biological molecules[25]. Our research employed SVM modeling with the "e1071" R package, with the radial basis function as the kernel function of choice. The SVM learning model was constructed using all original features to calculate the absolute coefficient $|w|$ for each input attribute. Subsequently, the features were sorted based on the square of $|w|$, and those at the lower end of the ranking were removed, while the remaining attributes were subjected to repeated iterative process of SVM model construction and ranking while mirroring the previously taken steps. This cycle continued until all features were eliminated[26]. Features removed at the end of the process were considered more significant as compared to those eliminated earlier. To determine the optimal number of mRNAs for developing a signature for HCM, the dataset was subjected to fivefold cross-validation. By varying the selected sets, SVM models were trained using different number of top mRNAs to calculate the overall prediction error. Finally, receiver operating characteristic (ROC) curves was used to calculate the area under the curve (AUC) value for each chosen mRNA characteristic employing the "pROC" tool in R[27], so as to assess the effectiveness of these features in the diagnosis of HCM.

Zebrafish husbandry

Zebrafishes were housed at the Zebrafish Research Center of Nantong University. The animal study protocol was approved by the Nantong University Institutional Animal Care and Use Committee (Item number is IACUC20221008-1001). Zebrafish embryos with Tg (cmlc2: GFP) were obtained by natural mating and maintained at 28.5 c. The embryos 24 h (h) post-fertilization (hpf) was treated with 0.2 mM 1-phenyl-2-thio-urea (PTU).

RNA extraction, reverse transcription, and qRT-PCR

In brief, tissues were homogenized in 1 mL Trizol (Life Technologies). RNA (1 μg) was reverse-transcribed into cDNA using Reverse Transcription Kit (Vazyme, China).

The qRT-PCR with ABI StepOne instrument was carried out in a total volume of 20μL. The reference gene was GAPDH. The MYH6 and RASD1 primers for qRT-PCR were: MYH6 F: 5′-AGAATAAGGATGGAGGGA-3′; R: 5′-CTTTAGATTGAACAGCACC-3′; RASD1 F: 5′- CCTCGGGTCCACCAAAGT-3′; R:5′- GTTCCCTGA AGTATCCAAAA-3′. The qRT-PCR reactions were run in three technical replicates, and the data were collected from at least three independent experiments.

Microinjection

The sgRNA transcription template was prepared by PCR by adding the T7 promoter sequence. The reverse primer was 25 bp sgRNA-R (Table1). The MAXIscript T7 Kit (Invitrogen, USA) for in vitro transcription was used to obtain sgRNAs. mMessage mMachine T7 Kit (Invitrogen, USA) was used to prepare Capped dCas9 mRNA for in vitro transcription. Dcas9 mRNA and sgRNA (Table1) injection concentrations were adjusted to 300:200 (unit: ng/μL).

Imaging and statistics

Heart development in Tg (cmlc2: GFP) zebrafishes was observed by anesthetizing embryos with egg water (1% PTU and 0.6% low agarose). Images were taken using a fluorescence microscope (IX71, Olympus, Japan). The measurement data analysis was performed using ImageJ. Significance of differences were shown using GraphPad Prism 8 and analyzed using Student's t-test.

Ethical approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. The animal protocols used in this investigation were approved by the Nantong University Institutional Animal Care and Use Committee (Item number is IACUC20221008-1001), complying with the ARRIVE guidelines. All authors have participated in the work and have reviewed and agree with the contents of the article for publication.

Results

Identification of DEGs between HCM and control groups

We well distinguished between patients with HCM and normal samples based on PCA, which indicates that the GSE36961 dataset used in this study is of good quality (Supplementary Fig. S1). Subsequently, to screen abnormal expressed genes in HCM, a total of 157 DEGs incorporating 47 up-regulated genes and 110 down-regulated genes were identified between HCM and normal groups (Fig. 1A,B).

Enrichment analysis

Enrichment analysis on the positive regulation of inflammatory response and external stimuli was performed using GO analysis (Fig. 2A). According to the KEGG analysis, the genes were enriched in the complement and coagulation cascades (Fig. 2B). Combining the GO and KEGG analysis, we found the DEGs were associated with the immune pathways. The immunologic gene sets and genes in the HCM group were enriched in the C7

Primer	sgRNA Sequence (5′–3′)
MYH6-1	ATGCTTTAATGGCAGAGTTCGGG
MYH6-2	GCTCTTTGTCAGATTTCCTCAGG
MYH6-3	CAAACGCAGAGCCAAATAAGGG
MYH6-4	GGTTTATCTGCAGTTAAATGTGG
MYH6-5	AGTTGAGTTTCAAAGAGGGGGGG
RASD1-1	GGTGGCTTCGGAGCCTCTGAAGG
RASD1-2	TCCCGGCGAAAAACTGCTACCGG
RASD1-3	GCAGTACACGCCGACCATTGAGG
RASD1-4	GTCCTCTCAGTACACTTTCCTGG
RASD1-5	AAAAATAGATAGACTAATGTAGG
sgRNA-R	AAAAAAAGCACCGACTCGGTGCCAC

Table 1. The sgRNAs in this study.

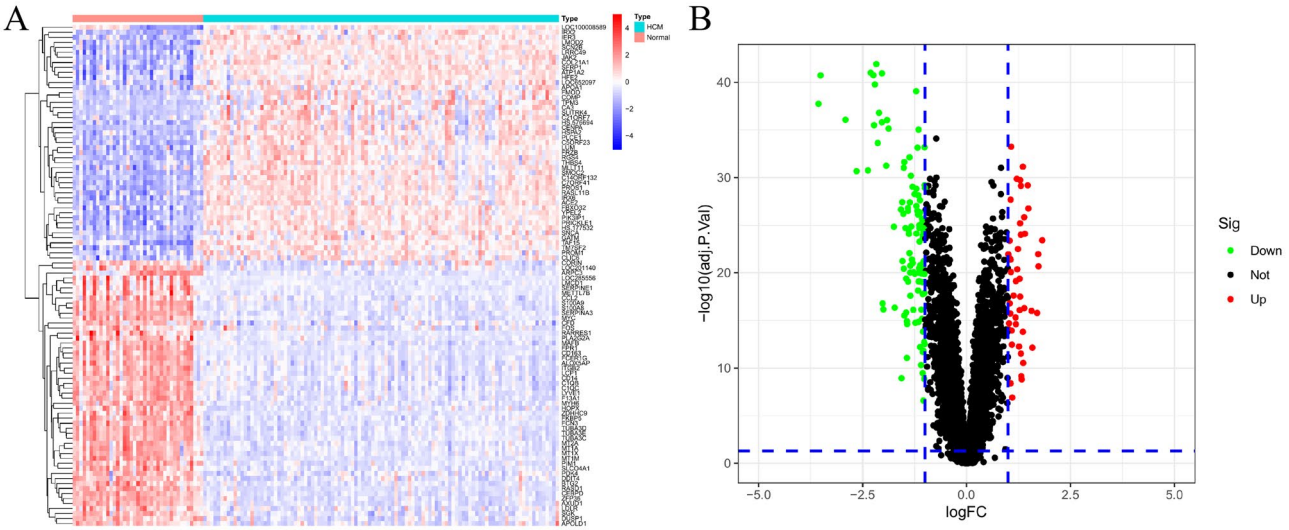

Figure 1. Based on the GSE36961 dataset to screen for differentially expressed genes (DEGs) in HCM and normal controls. Heat maps (**A**) and volcano maps (**B**) showing DEGs, respectively.

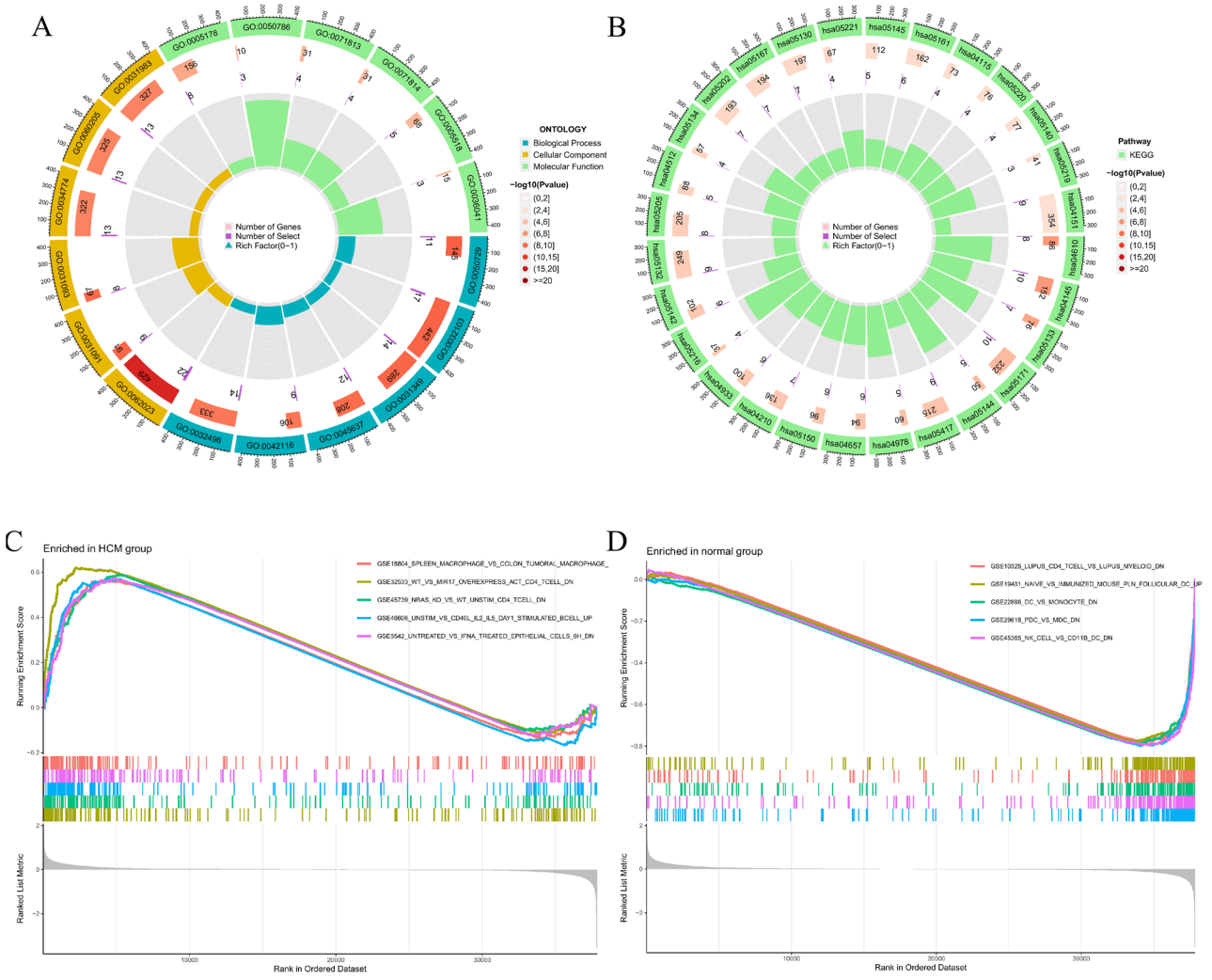

Figure 2. Enrichment analysis. (**A**) GO analysis (**B**) KEGG analysis (**C**) Enriched immunologic gene sets in C7 collection by the HCM group. (**D**) Enriched immunologic gene sets in C7 collection by the normal group.

collection (Fig. 2C). As compared to the normal group, the enriched genes were shown in Fig. 2D. These results confirmed that DEGs were correlated with immune-related signals.

Identification of hub genes by WGCNA and machine learning methods

To screen hub genes, we performed WGCNA co-expression network on 157 DEGs. As shown in Fig. 3A,B, the scale-free network was generated under the soft threshold of 10 (R2 = 0.88), as supported by the adjacency matrix and topological overlap matrix. Then, the modules were sectioned according to mean hierarchical clustering and dynamic tree cutting (Fig. 4A). The turquoise module was significantly correlated with HCM patients, which was therefore selected for further analysis (Fig. 4B). As shown in Supplementary Fig. S2, we performed PPI on the genes of the module to observe the interactions of these genes. This indicates that these genes may be involved in the occurrence and development of HCM through multiple interactions.

In addition, we further identified signature genes with diagnostic significance for HCM by LASSO regression analysis. As shown in Fig. 5A,B, the coefficients of the model gradually approached 0 as the penalty parameter λ increased and finally reached 14 to characterize genes for subsequent studies. Meanwhile, we used the SVM-RFE method to recursively remove features and obtained 28 key marker genes (Fig. 5C). Integrating the LASSO

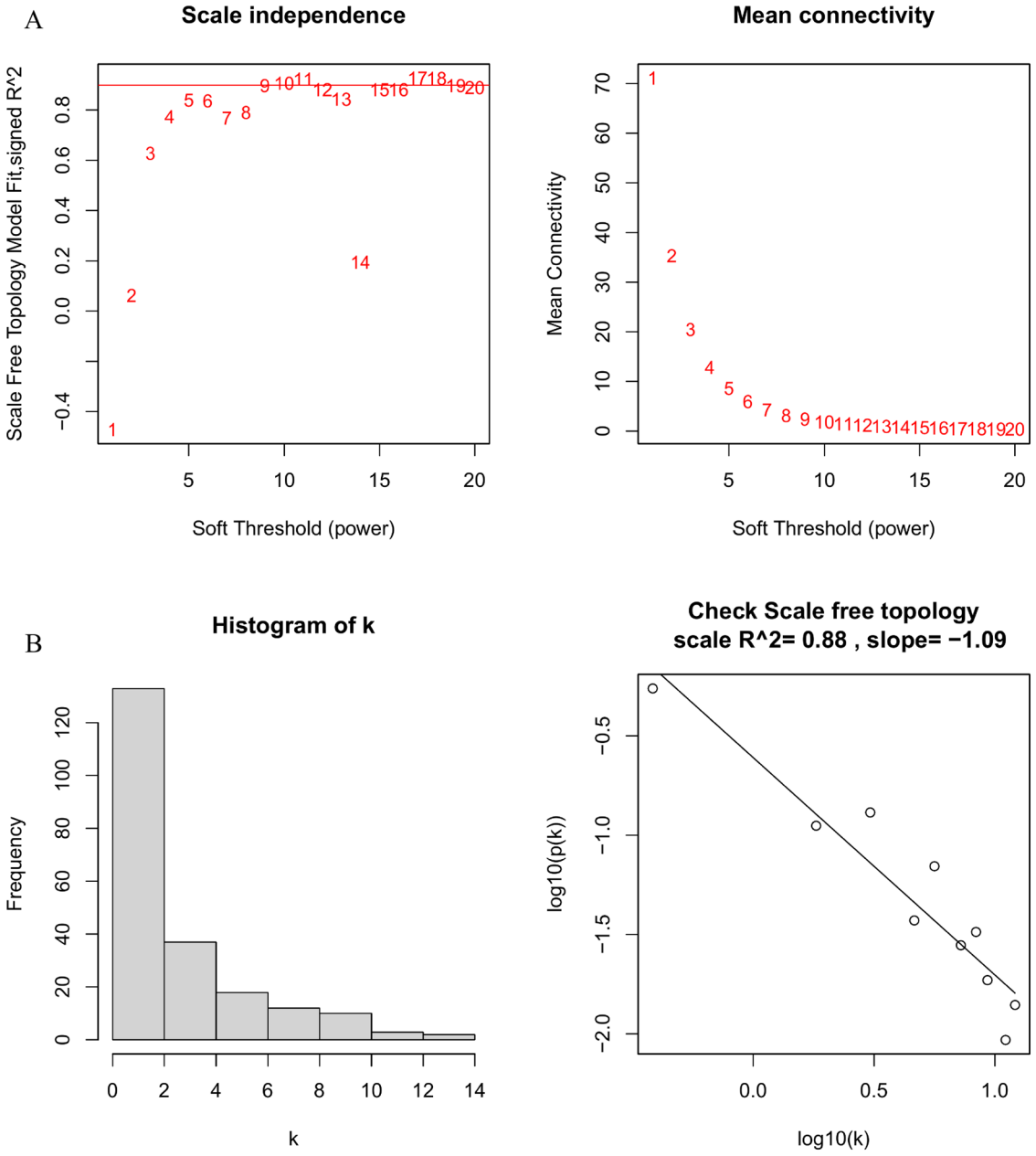

Figure 3. The soft-threshold power in the WGCNA. (**A**) Calculating the mean connectivity and fit of the scale-free topology mode to determine the optimal soft-thresholding power β. (**B**) Checking the scale free topology.

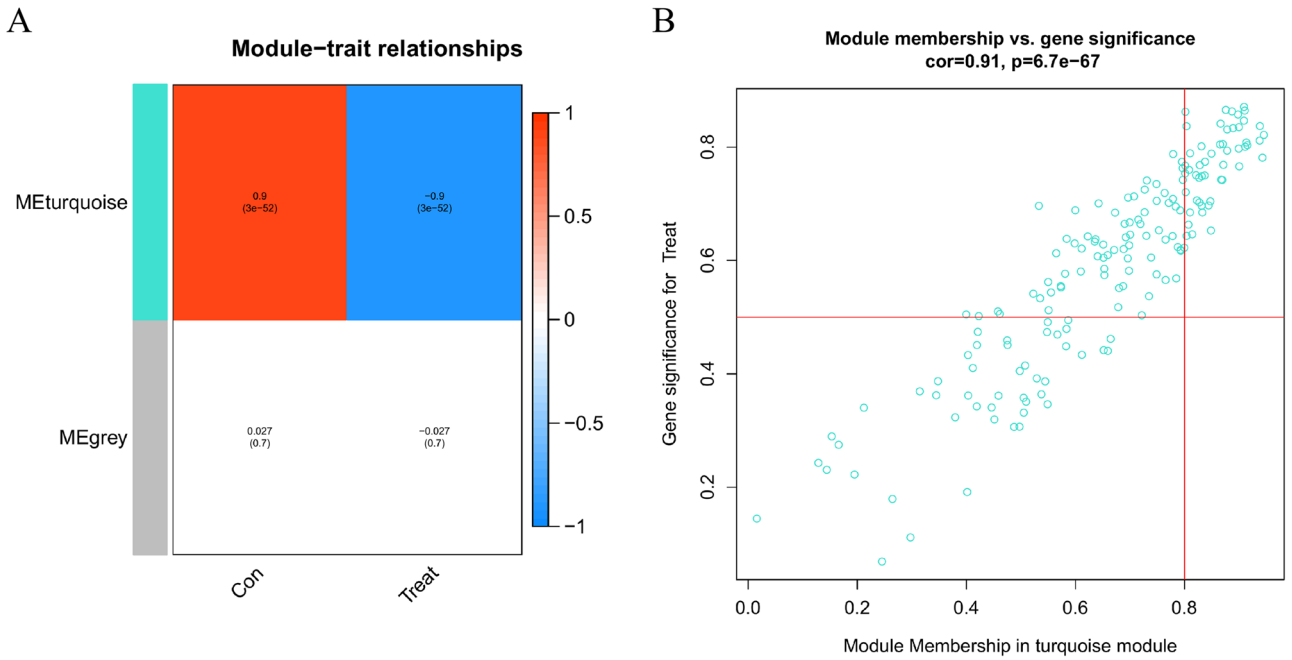

Figure 4. Identification of modules associated with HCM patients. (**A**) Heatmap of the correlation between the module and HCM patients. (**B**) The MEturquoise module associated with the HCM patients.

regression model genes, SVM model genes and WGCNA module genes (Fig. 5D), FCN3, MYH6 and RASD1 can be used as hub genes in HCM group.

Validation of the hub genes

We found that the expression of these key signature genes (FCN3, MYH6, and RASD1) was significantly down-regulated in HCM patients relative to normal patients (p value < 0.05) (Fig. 6A–C). Similarly, we also performed another validation based on the GSE141910 dataset and showed that the expression of these three genes was significantly down-regulated in HCM samples relative to normal control samples (Supplementary Fig. S3). In addition, the ROC curves of FCN3 (AUC = 0.968, CI = 0.917–0.998), MYH6 (AUC = 0.954, CI = 0.899–0.995), and RASD1 (AUC = 0.978, CI = 0.952–0.997) also showed relatively higher sensitivity for the diagnosis of HCM patients (Fig. 6D–F).

Importantly, our in vivo zebrafish-based model also validated the effects of MYH6 and RASD1 on the animal hearts. Here, we used CRISPRi technology and Tg (cmlc2: GFP) as the research background to construct the zebrafish model with MYH6 and RASD1 knockdown. It was found that knockdown of MYH6 and RASD1 resulted in malformation to the zebrafish hearts (Figs. 7A, 8A). The bar graphs showed a significant downregulation in the expression of the MYH6 and RASD1 knockdown groups (Figs. 7B, 8B). In addition, after knockdown of MYH6 and RASD1, we also observed a reduced heart rate frequency, prolonged bulbus arteriosus- sinus venosus (BA-SV) distance, and a decrease in fractional shortening percentage (FS%), which suggested the significant effects of these two genes on regulating cardiac rhythms and cardiac pumping capacity (Figs. 7C–E, 8C–E). Significant differences in ventricular diastole (VD) and systole (VS) phases were also observed after knocking down these two genes. By analyzing the ventricular volume of the zebrafishes with MYH6 (48hpf) and RASD1 (72hpf) knockdown. The results showed that atrial and ventricular embryos could form after MYH6 and RASD1 knockdown, but the ventricles were smaller than those in wild-type group (Figs. 7F–H and 8F–H). Knockout of the MYH6 and RSD1 genes slowed heartbeat and reduced ventricular function of the zebrafishes, which was similar symptoms to patients with HCM. These results revealed that MYH6 and RSD1 were the potential driver genes in HCM development.

Discussion

Cardiovascular myocardial hypertrophy is an important cause of a variety of major cardiovascular diseases. The process of cardiomyocyte enlargement is accompanied by increased myocardial interstitial, which will lead to myocardial hypoxia and cardiac remodeling due to excessive capillary supply. If without timely intervention, the heart will further develop into an irreversible decompensation period, accompanied by cardiac dysfunction, ultimately leading to heart failure and death. HCM is caused by myocardial hypertrophy, manifesting as fatigue dyspnea, chest pain, sudden death. It has been demonstrated that HCM is genetically heterogeneous[28], therefore, this provides a basis for exploring the potential biological information of HCM gene expression and helps to efficiently mine HCM-related target genes through a large amount of data to be able to help reduce adverse cardiovascular events. In this study, we obtained 106 HCM patients and 39 normal controls from the GEO database, from which a total of 157 DEGs were screened. Subsequent enrichment analysis showed that

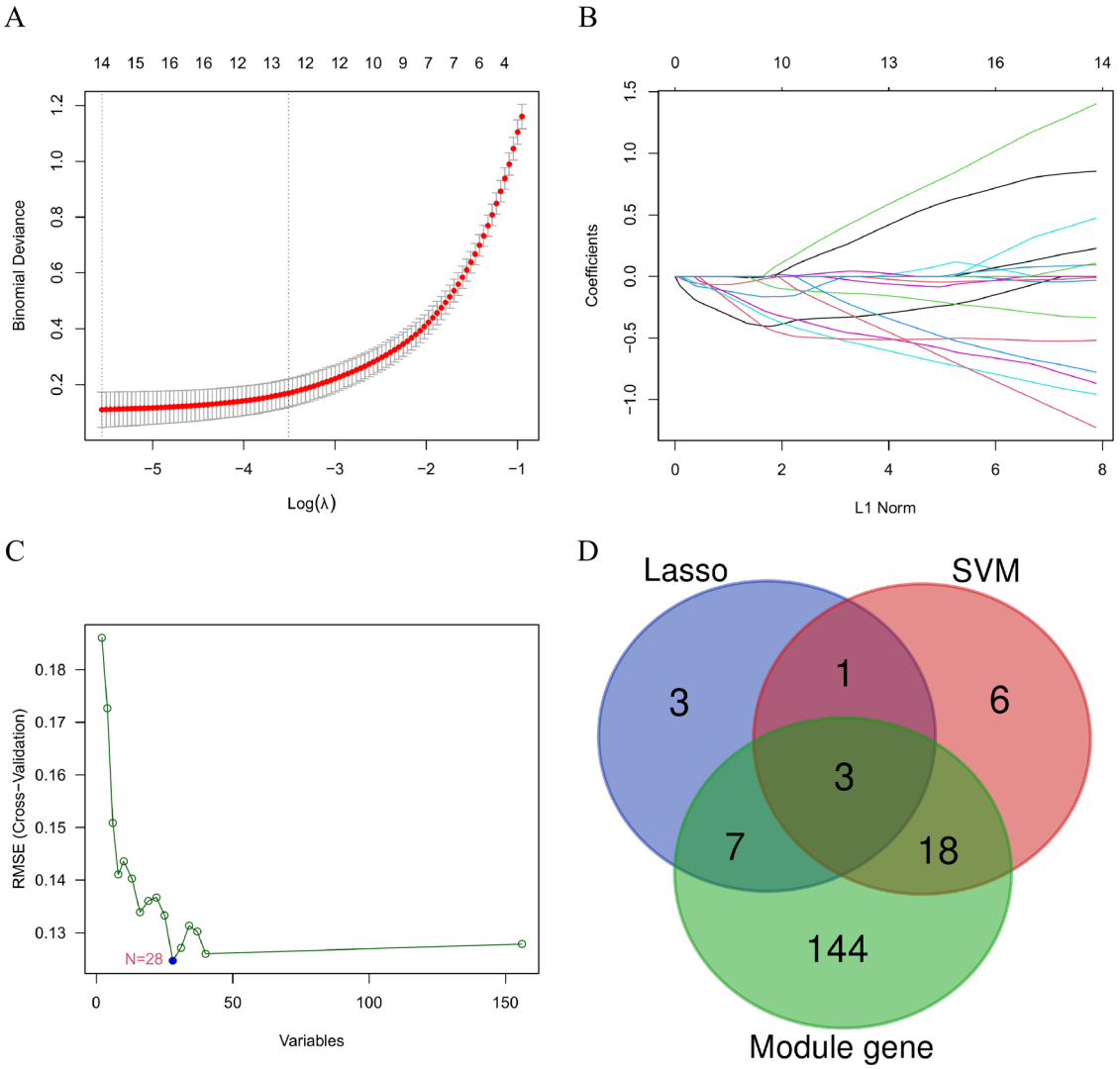

Figure 5. Screening the hub genes by machine learning. (**A**,**B**). LASSO algorithm to screen genes. (**C**) SVM-RFE algorithm to screen genes. (**D**) Venn diagram showing key genes identified based on WGCNA, LASSO regression analysis, and SVM-RFE have been screened for 3 signatures used to diagnose HCM.

these DEGs were mainly enriched in immune and inflammatory response pathways. Past findings have found the presence of mild systemic and local inflammation in individuals with HCM. Patients with HCM have mild chronic inflammatory cell infiltration in their myocardium[29–31]. Specifically, levels of circulating inflammatory markers were high in HCM, including tumor necrosis factor (TNF)-α and interleukin (IL)-6[32–34]. In addition, the immune system maintains the normal physiological function of the heart in HCM, and damage to its immune system leads to the development of abnormal inflammatory responses and myocardial remodeling[35,36]. These findings could support the crucial role of immune and inflammatory pathway mechanisms in HCM.

The use of high-throughput sequencing technologies and bioinformatics offers the possibility to further identify and detect HCM predictors. Yu et al. combined multiple machine algorithms to screen three key signature genes related to hypoxia and immunity from by HCM-related datasets[37]. Li et al. used SVM-RFE and random forest algorithms to select m6A regulators associated with HCM and used LASSO to establish a gene signature that can distinguish HCM patients from normal controls[38]. Specifically, in our study, three key hub genes (MYH6, RASD1 and FCN3) were screened by LASSO, SVM-RFE and WGCNA. Study confirmed that MYH6 mutations present phenotypes such as cardiac sudden death and heart failure associated with HCM in Japanese families[39]. Previous studies have found that MYH6$^{-/-}$ mutation causes dysfunction of the sinus node to slow down the heartbeat[40,41]. We also found the same phenotype by knocking down MYH6 in zebrafishes. Sinus node dysfunction is a clinical disease presented with bradycardia symptoms. In our study, we observed that zebrafishes with MYH6 knocked down also had bradycardia, further confirming the relationship between MYH6 and sinus node dysfunction. In addition, we also observed that after MYH6 knockdown, the FS, systolic and diastolic ventricular volume of zebrafishes were significantly reduced. Low FS and ventricle volume represent heart malformation and cardiac dysfunction. Low FS in zebrafishes with GTP3 mutation has also been previously defined[42]. Studies found low ventricle volume and malformation in zebrafishes with mutations of the genes, which was consistent

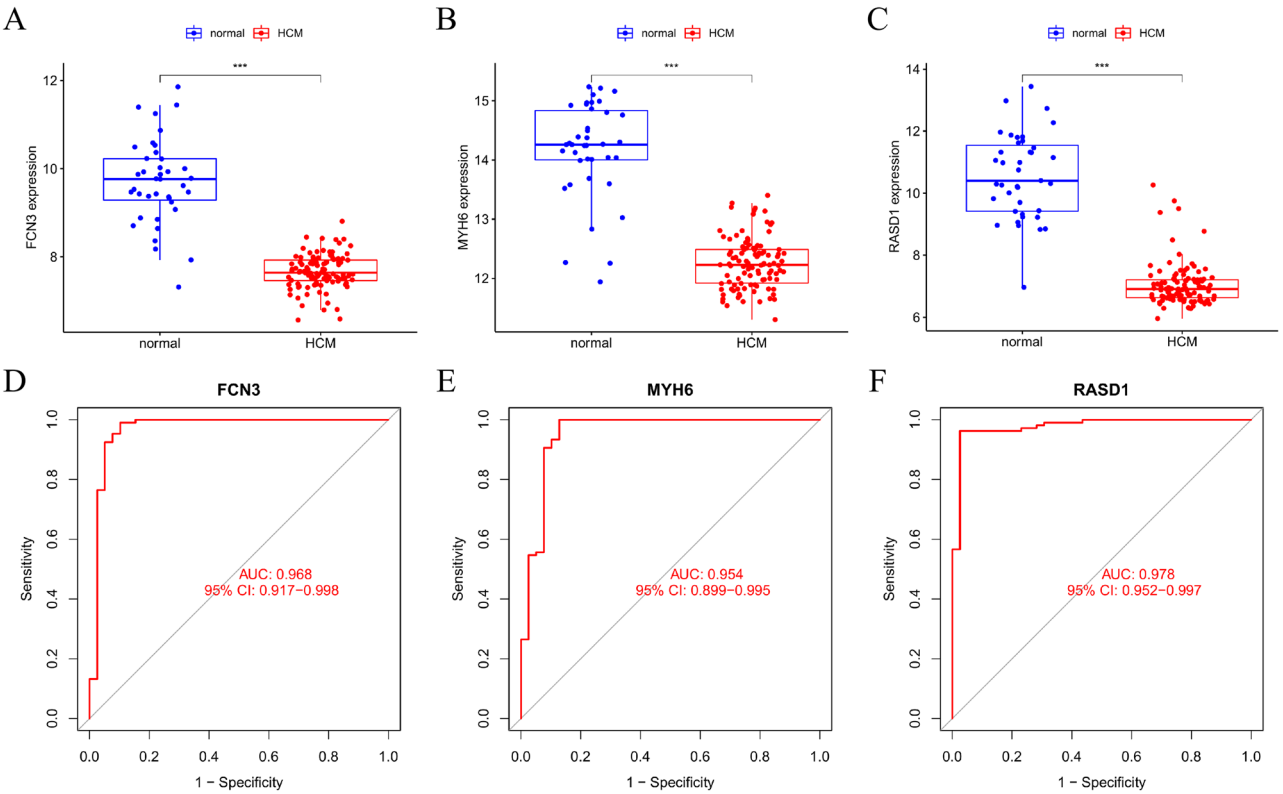

Figure 6. Based on the GSE36961 dataset to validate the expression and diagnostic value of the 3 key signatures. (**A**,**B**) Differential expression analysis of FCN3 (**A**), MYH6 (**B**), and RASD1 (**C**) in normal control and HCM groups. (**D**,**E**) ROC curves to validate the diagnostic value of FCN3 (**D**), MYH6 (**E**), and RASD1 (**F**) in patients with HCM.

with our findings[43,44]. Ras-related protein 1 (RASD1) is a dexamethasone induced monomeric Ras-like G protein that oscillates in the suprachiasmatic nucleus (SCN). As a novel signaling protein involved in a variety of cellular processes, RASD1 plays an important role in regulating uterine remodeling dynamics during the estrous cycle in utero[45]. In the cardiovascular system, knockdown of RASD1 significantly increases the secretion of atrial Natriuretic Factor (ANF) and negatively affects cardiovascular homeostasis[46]. In this study, we knocked down RASD1 in zebrafishes, and found that the decrease of RASD1 affected the heartbeat and ventricle volume, causing bradycardia and small ventricle volume, which may be explained by the interaction between RASD1 and renin as RASD1 is involved in renin transcriptional regulation[47]. It has been demonstrated that the renin-angiotensin system is an important candidate for susceptibility for left ventricle hypertrophic (LVH). Gene polymorphisms of the renin-angiotensin system plays a key role to HCM[48,49]. The renin-angiotensin system regulates human blood pressure and the expression of cardiac hypertrophy, affecting myocardial remodeling. The knockdown of RASD1 influences the renin-angiotensin system and ANF regulation, leading to the failure of myocardial remodeling and heart malformation. FCN3, also known as thermo- labile β-2 macroglycoprotein, is a protein of lectin-pathway that mainly express in the lung and liver[50]. In cardiovascular system, studies found that the FCN3 is involved in heart failure[51,52] and hypertension[53]. Moreover, bioinformatics analysis showed that FCN3 may be the potential key dysfunctional gene of HCM[54], which was similar to our study. However, we could not use zebrafish model with FCN3 knockdown as we did not have the expression of homologous gene in zebrafishes. Thus, more subsequent experiments are needed to improve and verify the relationship between FCN3 and HCM.

However, due to the complex etiology and pathogenesis of myocardial hypertrophy, this study still had some limitations, for instance, the current sample size was relatively small and we did not perform specific experiments for verification. In the future, we will expand the sample size to verify the present findings. In addition, as we did not find a homologous gene for FCN3 expression in zebrafishes, we were unable to construct a knockdown model for FCN3. Therefore, we will utilize other organism models to investigate the role of FCN3 in future studies.

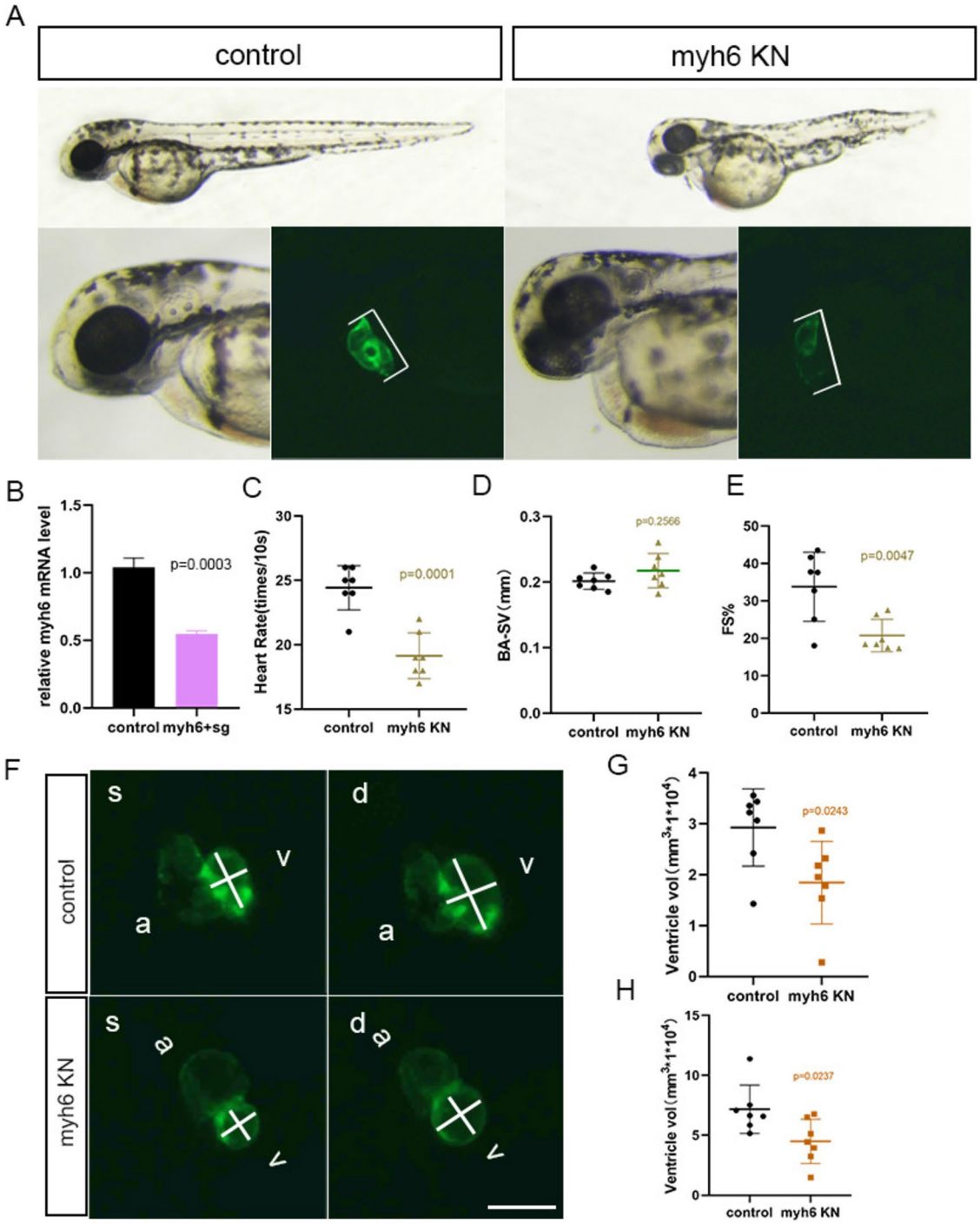

Figure 7. Phenotype of MYH6 cardiac malformation in a knocked down zebrafish. (**A**) the lateral body and enlarged heart of MYH6 zebrafish, control, n = 7, MYH6 KN, n = 7, white line, venous sinus-arterial distance. (**B**) Down-regulation efficiency of MYH6 gene. (**C**) heart rate (times/10 s), control, n = 10, MYH6 KN, n = 7. (**D**) BV-SA distance, control, n = 7, MYH6 KN, n = 7. (**E**) ventricular shortening fraction analysis, control, n = 7, MYH6 KN, n = 7. (**F**–**H**) the ventricle volume of zebrafish in systole and diastole, control, n = 7, MYH6 KN, n = 7, *a* atrium, *v* ventricle, *s* systole, *d* diastole; white line, the major and minor axes of the ventricles; bar = 200um.

Conclusion

In conclusion, we found that the expression of FCN3, MYH6 and RASD1 was downregulated in patients with HCM. The effects of MYH6 and RASD1 on cardiac function and cardiac mass index have been proved using a zebrafish model. A combined use of these genes may be useful for HCM diagnosis.

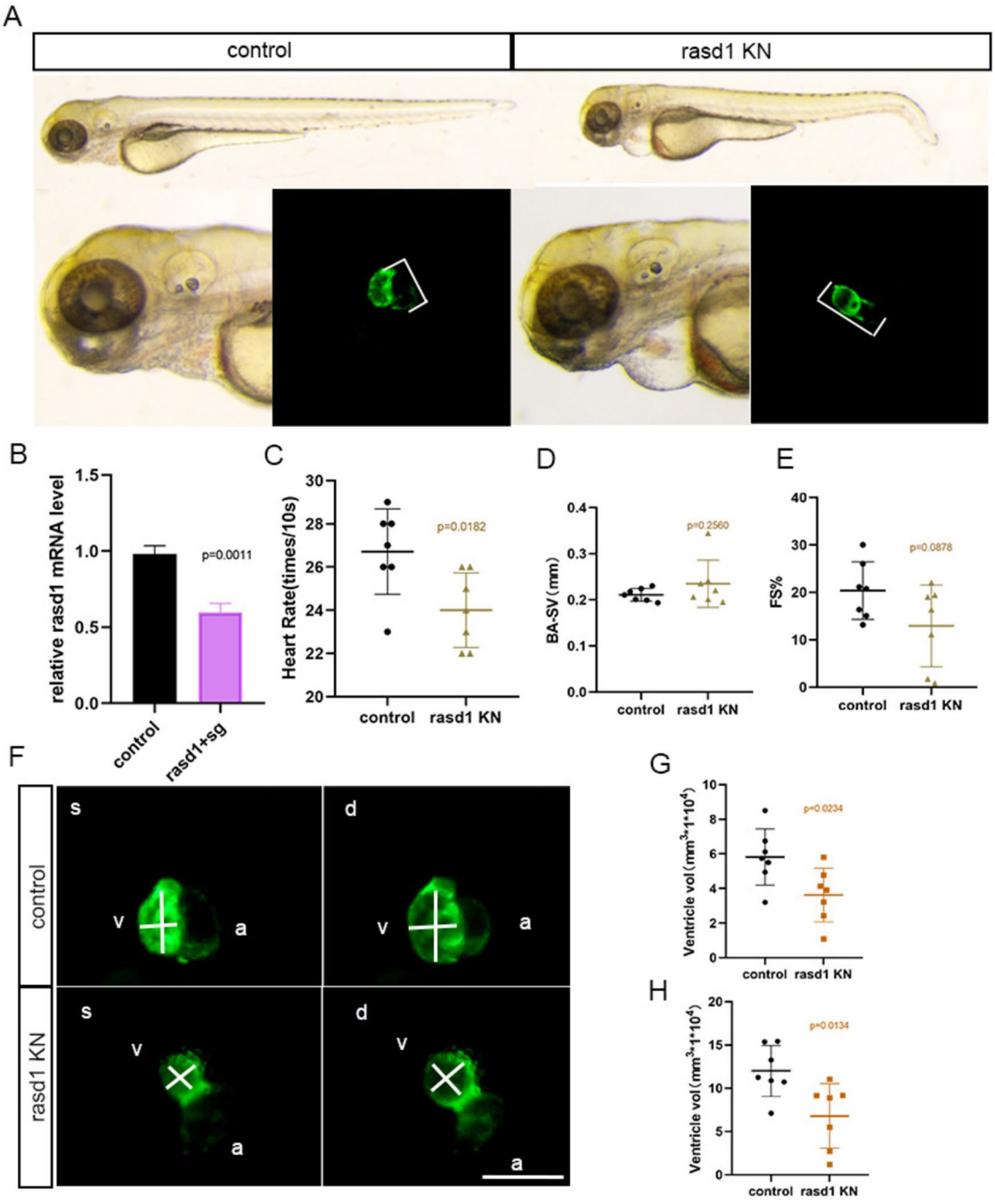

Figure 8. Phenotype of RASD1 cardiac malformation in a knocked down zebrafish. (**A**) The lateral body and enlarged heart of RASD1 zebrafish, control, n = 7, RASD1 KN, n = 7, white line, venous sinus-arterial distance. (**B**) Down-regulation efficiency of RASD1 gene. (**C**) Heart rate (times/10 s), control, n = 7, RASD1 KN, n = 7. (**D**) BV-SA distance, control, n = 7, RASD1 KN, n = 7. (**E**) Ventricular shortening fraction analysis, control, n = 7, RASD1 KN, n = 7. (**F**–**H**) the ventricle volume of zebrafish in systole and diastole, control, n = 7, RASD1 KN, n = 7, *a* atrium, *v* ventricle, *s* systole, *d* diastole; white line, the major and minor axes of the ventricles; bar = 200 um.

Data availability

Publicly available datasets were analyzed in this study. These data can be found here: all relevant raw data used in the study can be accessed from GSE repository [GSE36961] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36961) and GSE repository [GSE141910] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141910).

Received: 5 October 2023; Accepted: 9 July 2024
Published online: 24 July 2024

References

1. Wang, L. *et al.* Assessment of reversibility in pulmonary hypertension related to congenital heart disease by using biomarkers and clinical features. *Congenit. Heart Dis.* **17**(1), 87–97 (2022).
2. Park, J.-S. *et al.* Factors affecting the genetic diagnostic rate in congenital heart disease. *Congenit. Heart Dis.* **17**(6), 653–673 (2022).
3. Yusuf, S. *et al.* Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): A prospective cohort study. *Lancet* **395**(10226), 795–808 (2020).
4. McCarty, M. F. Nutraceutical, dietary, and lifestyle options for prevention and treatment of ventricular hypertrophy and heart failure. *Int. J. Mol. Sci.* **22**(7), 3321 (2021).
5. Xin, L. *et al.* Integrative expression analyses revealed potential biomarkers in hypertrophic cardiomyopathy. *J. Biol. Regul. Homeost. Agents* **37**(6), 3141–3150 (2023).
6. Maron, B. J. & Maron, M. S. Hypertrophic cardiomyopathy. *Lancet* **381**(9862), 242–255 (2013).
7. Elliott, P. M. *et al.* 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: The Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur. Heart J.* **35**(39), 2733–2779 (2014).
8. Ommen, S. R. *et al.* 2020 AHA/ACC guideline for the diagnosis and treatment of patients with hypertrophic cardiomyopathy: Executive summary: A report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. *Circulation* **142**(25), e533–e557 (2020).
9. Verma, A. A. *et al.* Implementing machine learning in medicine. *CMAJ* **193**(34), E1351-e1357 (2021).
10. Chunguang, L. *et al.* Identification of prognostic biomarkers for gastric cancer using a machine learning method. *J. Biol. Regul. Homeostat. Agents* **37**(1), 259–270 (2023).
11. Zhao, E., Xie, H. & Zhang, Y. Predicting diagnostic gene biomarkers associated with immune infiltration in patients with acute myocardial infarction. *Front. Cardiovasc. Med.* **7**, 586871 (2020).
12. Jia, S. *et al.* Integrative machine learning algorithms for developing a consensus RNA modification-based signature for guiding clinical decision-making in bladder cancer. *Oncologie* **26**(2), 269–285 (2024).
13. Deo, R. C. Machine learning in medicine. *Circulation* **132**(20), 1920–1930 (2015).
14. Avery, C., Patterson, J., Grear, T., Frater, T. & Jacobs, D. J. Protein function analysis through machine learning. *Biomolecules* **12**(9), 1246 (2022).
15. Patra, P., Disha, B. R., Kundu, P., Das, M. & Ghosh, A. Recent advances in machine learning applications in metabolic engineering. *Biotechnol. Adv.* **62**, 108069 (2023).
16. Lynch, C. M. *et al.* Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int. J. Med. Inform.* **108**, 1–8 (2017).
17. Zhou, C. M. *et al.* Machine learning to predict the cancer-specific mortality of patients with primary non-metastatic invasive breast cancer. *Surg. Today* **51**(5), 756–763 (2021).
18. Yang, W. *et al.* Machine learning to improve prognosis prediction of metastatic clear-cell renal cell carcinoma treated with cytoreductive nephrectomy and systemic therapy. *Biomol. Biomed.* **23**(3), 471–482 (2023).
19. Bos, J. M. *et al.* Marked up-regulation of ACE2 in hearts of patients with obstructive hypertrophic cardiomyopathy: Implications for SARS-CoV-2-mediated COVID-19. *Mayo Clin. Proc.* **95**(7), 1354–1368 (2020).
20. Lakshminarayan, K., Harp, S. A. & Samad, T. Imputation of missing data in industrial databases. *Appl. Intell.* **11**(3), 259–275 (1999).
21. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**(3), 307–315 (2004).
22. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics* **16**(5), 284–287 (2012).
23. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
24. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11), 2498–2504 (2003).
25. Yang, Y., Zheng, H., Wang, C., Xiao, W. & Liu, T. Predicting apoptosis protein subcellular locations based on the protein overlapping property matrix and tri-gram encoding. *Int. J. Mol. Sci.* **20**(9), 2344 (2019).
26. Huang, S. *et al.* Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genom. Proteomics* **15**(1), 41–51 (2018).
27. Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77 (2011).
28. Maron, B. J. *et al.* Management of hypertrophic cardiomyopathy: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **79**(4), 390–414 (2022).
29. Allen, R. D., Edwards, W. D., Tazelaar, H. D. & Danielson, G. K. Surgical pathology of subaortic septal myectomy not associated with hypertrophic cardiomyopathy: A study of 98 cases (1996–2000). *Cardiovasc. Pathol.* **12**(4), 207–215 (2003).
30. Phadke, R. S., Vaideeswar, P., Mittal, B. & Deshpande, J. Hypertrophic cardiomyopathy: An autopsy analysis of 14 cases. *J. Postgrad. Med.* **47**(3), 165–170 (2001).
31. Baandrup, U. & Olsen, E. G. Critical analysis of endomyocardial biopsies from patients suspected of having cardiomyopathy. I: Morphological and morphometric aspects. *Br. Heart J.* **45**(5), 475–86 (1981).
32. Zen, K. *et al.* Analysis of circulating apoptosis mediators and proinflammatory cytokines in patients with idiopathic hypertrophic cardiomyopathy: Comparison between nonobstructive and dilated-phase hypertrophic cardiomyopathy. *Int. Heart J.* **46**(2), 231–244 (2005).
33. Högye, M., Mándi, Y., Csanády, M., Sepp, R. & Buzás, K. Comparison of circulating levels of interleukin-6 and tumor necrosis factor-alpha in hypertrophic cardiomyopathy and in idiopathic dilated cardiomyopathy. *Am. J. Cardiol.* **94**(2), 249–251 (2004).
34. Fang, L. *et al.* Systemic inflammation is associated with myocardial fibrosis, diastolic dysfunction, and cardiac hypertrophy in patients with hypertrophic cardiomyopathy. *Am. J. Transl. Res.* **9**(11), 5063–5073 (2017).
35. Maron, B. J. *et al.* Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. *Circulation* **92**(4), 785–9 (1995).
36. Prabhu, S. D. & Frangogiannis, N. G. The Biological basis for cardiac repair after myocardial infarction: From inflammation to fibrosis. *Circ. Res.* **119**(1), 91–112 (2016).
37. Yu, H. *et al.* Identification and analysis of key hypoxia- and immune-related genes in hypertrophic cardiomyopathy. *Biol. Res.* **56**(1), 45 (2023).
38. Li, Y., Zhang, W., Dai, Y. & Chen, K. Identification and verification of IGFBP3 and YTHDC1 as biomarkers associated with immune infiltration and mitophagy in hypertrophic cardiomyopathy. *Front. Genet.* **13**, 986995 (2022).
39. Suzuki, T. *et al.* A double heterozygous variant in MYH6 and MYH7 associated with hypertrophic cardiomyopathy in a Japanese Family. *J. Cardiol. Cases* **25**(4), 213–217 (2022).
40. Ishikawa, T. *et al.* Novel mutation in the α-myosin heavy chain gene is associated with sick sinus syndrome. *Circulation* **8**(2), 400–408 (2015).

41. Shih, Y. H. *et al.* Cardiac transcriptome and dilated cardiomyopathy genes in zebrafish. *Circulation* **8**(2), 261–269 (2015).
42. Chen, D. *et al.* Deletion of Gtpbp3 in zebrafish revealed the hypertrophic cardiomyopathy manifested by aberrant mitochondrial tRNA metabolism. *Nucleic Acids Res.* **47**(10), 5341–5355 (2019).
43. Zhang, Q. *et al.* Ablation of Mto1 in zebrafish exhibited hypertrophic cardiomyopathy manifested by mitochondrion RNA maturation deficiency. *Nucleic Acids Res.* **49**(8), 4689–4704 (2021).
44. Da'as, S. I. *et al.* Hypertrophic cardiomyopathy-linked variants of cardiac myosin-binding protein C3 display altered molecular properties and actin interaction. *Biochem. J.* **475**(24), 3933–3948 (2018).
45. Kim, H. R. *et al.* Rapid expression of RASD1 is regulated by estrogen receptor-dependent intracellular signaling pathway in the mouse uterus. *Mol. Cell. Endocrinol.* **446**, 32–39 (2017).
46. McGrath, M. F., Ogawa, T. & de Bold, A. J. Ras dexamethasone-induced protein 1 is a modulator of hormone secretion in the volume overloaded heart. *Am. J. Physiol.* **302**(9), H1826-37 (2012).
47. Tan, J. J., Ong, S. A. & Chen, K. S. Rasd1 interacts with Ear2 (Nr2f6) to regulate renin transcription. *BMC Mol. Biol.* **12**, 4 (2011).
48. Ishanov, A. *et al.* Angiotensin II type 1 receptor gene polymorphisms in patients with cardiac hypertrophy. *Jpn. Heart J.* **39**(1), 87–96 (1998).
49. Orenes-Piñero, E. *et al.* Impact of polymorphisms in the renin-angiotensin-aldosterone system on hypertrophic cardiomyopathy. *J. Renin-Angiotensin-Aldosterone Syst.* **12**(4), 521–530 (2011).
50. Munthe-Fog, L. *et al.* Immunodeficiency associated with FCN3 mutation and ficolin-3 deficiency. *N. Engl. J. Med.* **360**(25), 2637–2644 (2009).
51. Lidani, K. C. F. *et al.* Ficolin-3 in chronic Chagas disease: Low serum levels associated with the risk of cardiac insufficiency. *Parasite Immunol.* **43**(6), e12829 (2021).
52. Jiang, Y., Zhang, Y. & Zhao, C. Integrated gene expression profiling analysis reveals SERPINA3, FCN3, FREM1, MNS1 as candidate biomarkers in heart failure and their correlation with immune infiltration. *J. Thorac. Dis.* **14**(4), 1106–1119 (2022).
53. Lu, J. *et al.* A common genetic variant of FCN3/CD164L2 is associated with essential hypertension in a Chinese population. *Clin. Exp.* **34**(5), 377–82 (2012).
54. Cui, Y., Liu, C., Luo, J. & Liang, J. Dysfunctional network and mutation genes of hypertrophic cardiomyopathy. *J. Healthc. Eng.* **2022**, 8680178 (2022).

Acknowledgements

We would like to thank Zhangji Dong from the Key Laboratory of Neural Regeneration, Nantong University for professional suggestions to the zebrafish experiments.

Author contributions

Conception and design: Jue Gu, Yamin Zhao and Chen Xu; Administrative support: Hongzhuan Sheng; Provision of study materials or patients: Hongzhuan Sheng; Collection and assembly of data: Yue Ben and Ruizi Liu; Data analysis and interpretation: Liqi Hua and Songnian He; Manuscript writing: All authors; Final approval of manuscript: All authors.

Funding

This study supported by Nantong Municipal Health Commission project: QN2022030 and QNZ2023055.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-67201-8.

Correspondence and requests for materials should be addressed to X.C. or H.S.