BMC Genomics

RESEARCH

Open Access

Identification of cardiomyopathy-related core genes through human metabolic networks and expression data

Zherou Rong, Hongwei Chen, Zihan Zhang, Yue Zhang, Luanfeng Ge, Zhengyu Lv, Yuqing Zou, Junjie Lv, Yuehan He, Wan Li* and Lina Chen*

Abstract

Background: Cardiomyopathy is a complex type of myocardial disease, and its incidence has increased significantly in recent years. Dilated cardiomyopathy (DCM) and ischemic cardiomyopathy (ICM) are two common and indistinguishable types of cardiomyopathy.

Results: Here, a systematic multi-omics integration approach was proposed to identify cardiomyopathy-related core genes that could distinguish normal, DCM and ICM samples using cardiomyopathy expression profile data based on a human metabolic network. First, according to the differentially expressed genes between different states (DCM/ICM and normal, or DCM and ICM) of samples, three sets of initial modules were obtained from the human metabolic network. Two permutation tests were used to evaluate the significance of the Pearson correlation coefficient difference score of the initial modules, and three candidate modules were screened out. Then, a cardiomyopathy risk module that was significantly related to DCM and ICM was determined according to the significance of the module score based on Markov random field. Finally, based on the shortest path between cardiomyopathy known genes, 13 core genes related to cardiomyopathy were identified. These core genes were enriched in pathways and functions significantly related to cardiomyopathy and could distinguish between samples of different states.

Conclusion: The identified core genes might serve as potential biomarkers of cardiomyopathy. This research will contribute to identifying potential biomarkers of cardiomyopathy and to distinguishing different types of cardiomyopathy.

Keywords: Cardiomyopathy, Human metabolic network, Expression data, Module, Core genes

Introduction

Cardiomyopathy is a disease of the heart muscle with major abnormalities in the structure and function of the heart, and cause the myocardium to become weak and ineffective [1]. The World Health Organization separates the various cardiomyopathies into several types based on the main pathophysiology and etiology/pathogenic factors. Dilated cardiomyopathy (DCM) and ischemic cardiomyopathy (ICM) are two major types with essentially different etiology. DCM could be caused by viral infections, autoimmunity, and genetic factors, while ICM was mainly caused by long-term myocardial ischemia due to atherosclerotic lesions. Moreover, DCM and ICM often exhibit similar clinical symptoms [2–4], making them two highly related pathologies that have not been fully characterized. Therefore, effective differentiation between DCM and ICM is of great importance in preventing

*Correspondence: liwan@hrbmu.edu.cn; chenlina@ems.hrbmu.edu.cn
College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang, China

and personalizing the treatment of cardiomyopathy in patients. Giraldo et al. used respiratory sinus arrhythmia (RSA) index of the parasympathetic system quantified by linear and non-linear analysis methods to discriminate between DCM and ICM with high sensitivity and specificity [5]. Sweet et al. used differentially expressed genes (DEGs) and pathway analysis to identify DEG signatures that could correctly classify the phenotypes of ICM and DCM samples [6].

In addition to gene expression, abnormal metabolism can also lead to poor heart function, which can affect the functions of myocardial cell [7]. Zhao et al. evaluated plasma metabolomics of patients with DCM or ICM through comprehensive metabolomic analysis to identify plasma metabolite biomarkers [8]. And some studies have shown that changes in myocardial metabolism are one of the important pathogenic factors of diabetic cardiomyopathy [9]. Metabolic networks can reflect a variety of chemical reactions catalyzed by gene-encoded enzymes and their interaction systems. Wang et al. constructed a lipid metabolism network and identified lipid subnetworks and clusters that involved in the pathogenesis of cardiovascular diseases [10]. Moreover, the identification of genes related to diseases in networks through shortest path analysis has been widely used to study the mechanism of diseases. Yang et al. performed shortest path analysis to explore the key drug targets of LianXia NingXin formulations for the treatment of coronary heart disease-related phenotypes (e.g., co-morbid diseases and symptoms) [11].

Therefore, in this study, a systematic multi-omics integration approach was proposed to identify cardiomyopathy-related core genes based on metabolic networks and expression data. First, the Significant Analysis of Microarray (SAM) method was used to screen DEGs between samples of different states in the expression data, and three sets of initial modules containing DEGs were obtained from the modules mined by Molecular Complexity Detection (MCODE). Two permutation tests were used to evaluate the significance of the Pearson correlation coefficient difference score of the initial modules, and candidate modules were screened out. Then, according to the significance of the module score based on Markov random field (MRF), the cardiomyopathy risk module that was significantly related to DCM and ICM was determined. Finally, based on the shortest path between known genes, 13 core genes closely related to cardiomyopathy were identified (Fig. 1). Our method provided valuable ideas for identifying potential cardiomyopathy genes that could effectively distinguish different types of cardiomyopathy.

Results

Initial and candidate modules

From the reconstructed metabolic network, modules with nodes ≥ 4 were selected. A total of 52 modules were identified from the reconstructed metabolic network using MCODE (see Materials and Methods for details). Of these modules, 8 modules containing DEGs of DCM and ICM samples (DCM_ICM was used to indicate the two states of DCM and ICM.) were screened as initial D_I modules, 21 modules containing DEGs of DCM and normal samples (DCM_NF was used to indicate the two states of DCM and normal.) were screened as initial DCM modules (containing seven initial D_I modules), and 37 modules containing DEGs of ICM and normal samples (ICM_NF was used to indicate the two states of ICM and normal.) were screened as initial ICM modules (containing 8 initial D_I modules and 21 initial DCM modules).

Two permutation tests of Pearson difference scores were performed on all initial modules, and finally 3 modules that were significantly different compared with random modules of the same degree and the same scale were selected as candidate modules (both P values < 0.05), including 2 ICM modules (ICM-module1, ICM-module2), 1 DCM module (DCM-module1) and 1 D_I module (D_I-module1). Among them, DCM-module1 and D_I-module1 were the same module.

The expression values of all genes in the three candidate modules were used as classification features to classify samples of different states (DCM_NF, ICM_NF or DCM_ICM), respectively. The genes of the three modules had good classification performance for normal and disease samples. However, for the classification of ICM and DCM samples, the three modules showed different results. Among them, the DCM-module1/D_I-module1 (containing 4 DCM_ICM DEGs) could effectively distinguish between DCM_ICM samples, while the modules ICM-module1 and ICM-module2 could not (Table 1).

Cardiomyopathy risk modules

The module scores based on MRF (MRFmss) for 3 candidate modules were calculated (see Materials and Methods for details) [12, 13], and compared with that of random modules. One cardiomyopathy risk module, D_I-module1, containing 205 genes, with significantly higher MRFms was identified ($p < 0.05$).

This cardiomyopathy risk module was significantly enriched in pathways and functions (see Materials and Methods for details) related to cardiomyopathy (some are in Fig. 2).

“Purine metabolism” and “Pyrimidine metabolism” are two pathways closely related to nucleotide content. Studies in human and animal models have demonstrated that

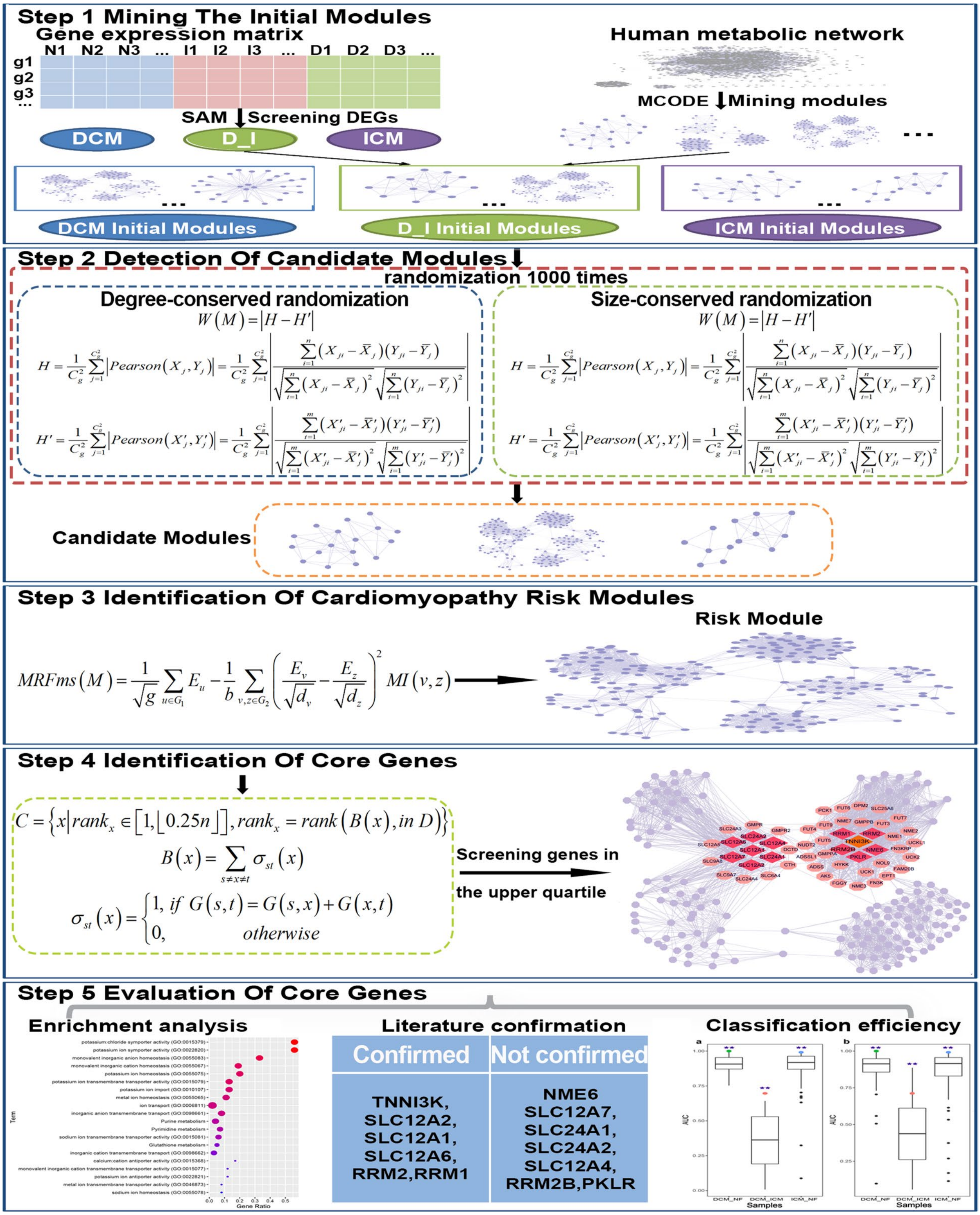
Step 1 Mining The Initial Modules
Gene expression matrix
N1 N2 N3 ... I1 I2 I3 ... D1 D2 D3 ...
g1
g2
g3
...
Human metabolic network
MCODE Mining modules
SAM Screening DEGs
DCM
D_I
ICM
...
DCM Initial Modules
D_I Initial Modules
ICM Initial Modules
Step 2 Detection Of Candidate Modules
randomization 1000 times
Degree-conserved randomization
Size-conserved randomization
$W(M)=|H-H'|$
$H=\frac{1}{C_g^2}\sum_{j=1}^{C_g^2}|Pearson(X_j,Y_j)|=\frac{1}{C_g^2}\sum_{j=1}^{C_g^2}\left|\frac{\sum_{i=1}^{n}(X_{ji}-\bar{X}_j)(Y_{ji}-\bar{Y}_j)}{\sqrt{\sum_{i=1}^{n}(X_{ji}-\bar{X}_j)^2}\sqrt{\sum_{i=1}^{n}(Y_{ji}-\bar{Y}_j)^2}}\right|$
$H'=\frac{1}{C_g^2}\sum_{j=1}^{C_g^2}|Pearson(X'_j,Y'_j)|=\frac{1}{C_g^2}\sum_{j=1}^{C_g^2}\left|\frac{\sum_{i=1}^{m}(X'_{ji}-\bar{X}'_j)(Y'_{ji}-\bar{Y}'_j)}{\sqrt{\sum_{i=1}^{m}(X'_{ji}-\bar{X}'_j)^2}\sqrt{\sum_{i=1}^{m}(Y'_{ji}-\bar{Y}'_j)^2}}\right|$
Candidate Modules
Step 3 Identification Of Cardiomyopathy Risk Modules
Risk Module
$MRFms(M)=\frac{1}{\sqrt{g}}\sum_{u\in G_1}E_u-\frac{1}{b}\sum_{v,z\in G_2}\left(\frac{E_v}{\sqrt{d_v}}-\frac{E_z}{\sqrt{d_z}}\right)^2 MI(v,z)$
Step 4 Identification Of Core Genes
$C=\{x|rank_x\in[1,\lfloor 0.25n\rfloor],rank_x=rank(B(x),in\ D)\}$
$B(x)=\sum_{s\neq x\neq t}\sigma_{st}(x)$
$\sigma_{st}(x)=\begin{cases}1, & if\ G(s,t)=G(s,x)+G(x,t)\\0, & otherwise\end{cases}$
Screening genes in the upper quartile
TNNI3K
RRM1
RRM2
RRM2B
NME6
PKLR
SLC12A1
SLC12A2
SLC12A4
SLC12A6
SLC12A7
SLC24A1
SLC24A2
SLC24A3
SLC24A4
Step 5 Evaluation Of Core Genes
Enrichment analysis
potassium:chloride symporter activity (GO:0015379)
potassium ion symporter activity (GO:0022820)
monovalent inorganic anion homeostasis (GO:0055083)
monovalent inorganic cation homeostasis (GO:0055067)
potassium ion homeostasis (GO:0055075)
potassium ion transmembrane transporter activity (GO:0015079)
potassium ion import (GO:0010107)
metal ion homeostasis (GO:0055065)
ion transport (GO:0006811)
inorganic anion transmembrane transport (GO:0098661)
Purine metabolism
Pyrimidine metabolism
sodium ion transmembrane transporter activity (GO:0015081)
Glutathione metabolism
inorganic cation transmembrane transport (GO:0098662)
calcium:cation antiporter activity (GO:0015368)
monovalent inorganic cation transmembrane transporter activity (GO:0015077)
potassium ion antiporter activity (GO:0022821)
metal ion transmembrane transporter activity (GO:0046873)
sodium ion homeostasis (GO:0055078)
Term
Gene Ratio
0.0 0.1 0.2 0.3 0.4 0.5
Literature confirmation
Confirmed
TNNI3K, SLC12A2, SLC12A1, SLC12A6, RRM2,RRM1
Not confirmed
NME6 SLC12A7, SLC24A1, SLC24A2, SLC12A4, RRM2B,PKLR
Classification efficiency
a
b
AUC
1.00
0.75
0.50
0.25
0.00
DCM_NF DCM_ICM ICM_NF
Samples

Fig. 1 Flow chart of this study. Step 1: mining the initial modules from the network modules mined by MCODE according to DEGs. Step 2: detection of candidate modules based on the significance of the Pearson correlation coefficient difference score of the initial modules. Step 3: identification of the cardiomyopathy risk modules significantly related to DCM and ICM according to the importance of the module score based on Markov random field. Step 4: identification of core genes from the cardiomyopathy risk module based on the shortest path between known disease-causing genes. Step 5: evaluation of core genes from three aspects: enrichment analysis, literature confirmation and classification efficiency

Table 1 Classification performance of candidate modules

Modules	AUC for normal and disease samples	AUC for ICM and DCM samples
ICM-module1	0.93	0.31
ICM-module2	0.79	0.40
DCM-module1/D_I-module1	0.99	0.71

many disorders of purine and pyrimidine nucleotide content in the myocardium played a role in the pathogenesis of muscle dysfunction in diseases such as coronary heart disease and left ventricular hypertrophy [14, 15]. Genes of the cardiomyopathy risk module were widely distributed in both pathways, suggesting the involvement of the cardiomyopathy risk module in muscle dysfunction of the myocardium diseases. For example, in the "Purine metabolism" pathway, the cardiomyopathy risk module genes mainly encode various enzymes involved in energy conversion in the pathway, including various kinases, reductases, hydrolases and synthases (Fig. 3). They are mainly involved in the pathway module Adenine ribonucleotide biosynthesis and Guanine ribonucleotide biosynthesis. These enzymes catalyze the hydrolysis of tetraphosphate to produce ATP, and the mutual conversion of ATP/GTP and ADP/GDP.

Functions including "ion transport", "anion: cation symporter activity", "potassium ion antiporter activity", "sodium ion transmembrane transporter activity" were related to ion transport in cells. Through a series of activities of diverse ion channels, the excitability of cardiac myocytes is caused by ionic fluxes [17, 18]. The "Fatty acid degradation" pathway and "fatty acid biosynthetic process" function were related to fatty acids. Fatty acids were the main energy substrates of the heart, provided energy for myocardial contraction, and were

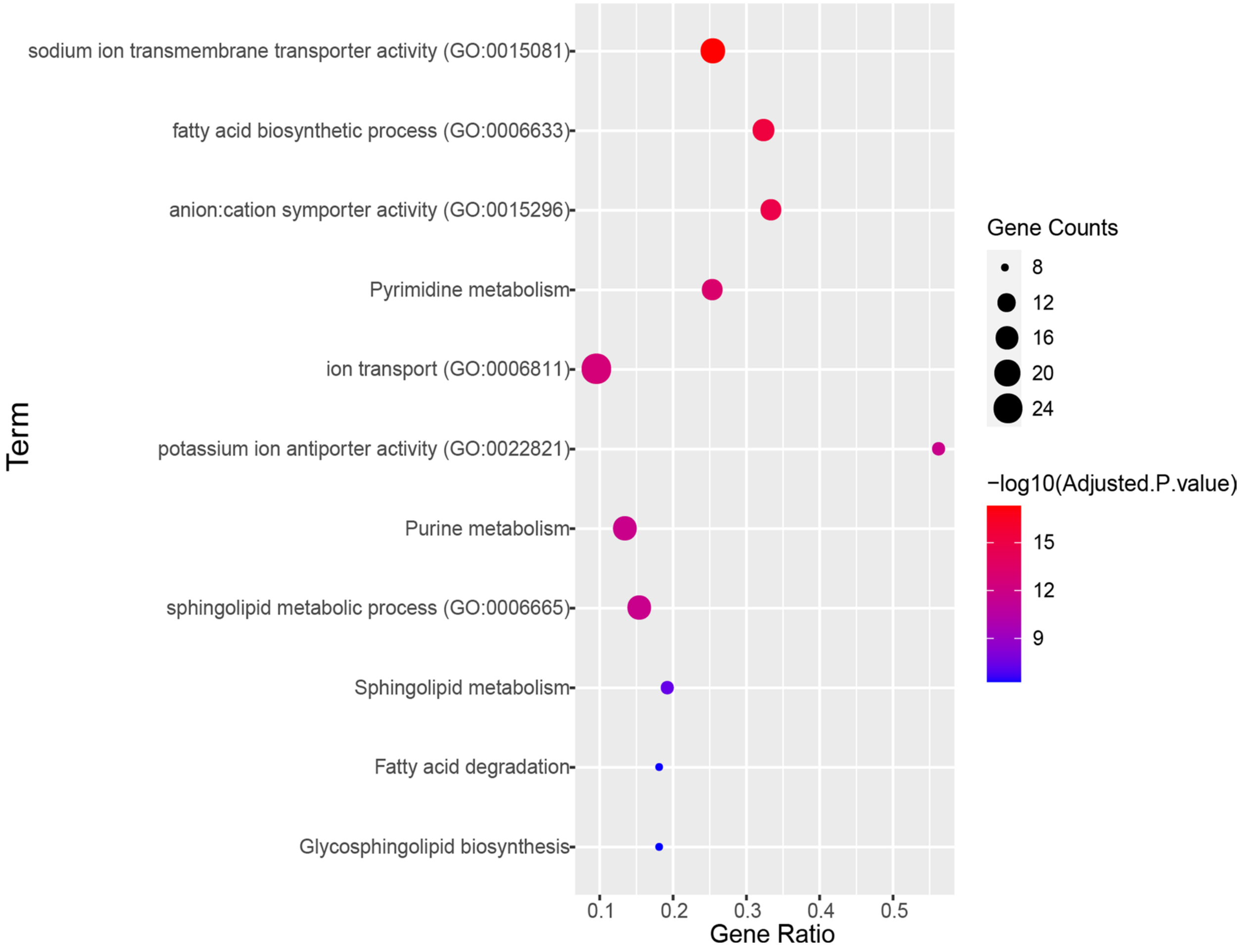

Fig. 2 Some pathways and functions enriched by the cardiomyopathy risk module

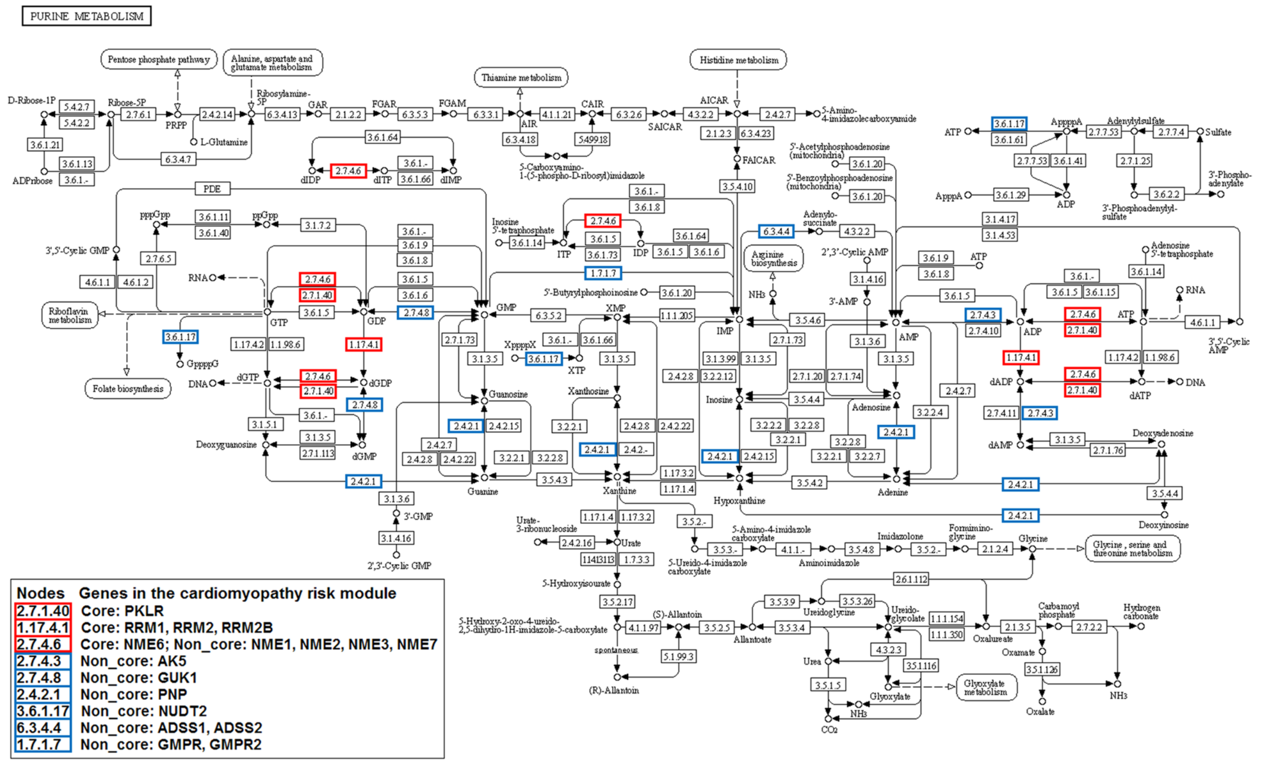
PURINE METABOLISM
Pentose phosphate pathway
Alanine, aspartate and glutamate metabolism
Thiamine metabolism
Histidine metabolism
Riboflavin metabolism
Folate biosynthesis
Arginine biosynthesis
Glyoxylate metabolism
Glycine, serine and threonine metabolism
Nodes Genes in the cardiomyopathy risk module
2.7.1.40 Core: PKLR
1.17.4.1 Core: RRM1, RRM2, RRM2B
2.7.4.6 Core: NME6; Non_core: NME1, NME2, NME3, NME7
2.7.4.3 Non_core: AK5
2.7.4.8 Non_core: GUK1
2.4.2.1 Non_core: PNP
3.6.1.17 Non_core: NUDT2
6.3.4.4 Non_core: ADSS1, ADSS2
1.7.1.7 Non_core: GMPR, GMPR2

Fig. 3 Purine metabolism pathway [16]. The red rectangles indicate the core genes in the cardiomyopathy risk module, and the blue ones indicate the non-core genes

essential substrates for the synthesis of sphingolipids [19, 20]. Sphingolipids regulated many cellular processes that occurred in primary and secondary cardiomyopathy, and were also involved in functional categories and pathways, such as “Glycosphingolipid biosynthesis”, “Sphingolipid metabolism” and “sphingolipid metabolic process “. And a large number of studies have shown that disorders of sphingolipid metabolism can cause changes in the structure and function of cardiomyocytes [21–23].

Cardiomyopathy-related core genes

In the cardiomyopathy risk module, 52 candidate genes were located on the shortest paths between known genes. For each candidate gene, the number of known gene pairs linked by it via the shortest paths was calculated. Finally, 13 genes linked more than 6 (the top quartile) known gene pairs were identified as cardiomyopathy-related core genes of the module (containing 2 ICM_NF DEGs, 1 DCM_NF DEG, and 1 DCM_ICM DEG and 1 known pathogenic gene) (Fig. 4).

The relationship between core genes and cardiomyopathy was analyzed from the following aspects.

Functional enrichment analysis

In order to understand the relationship between core genes and diseases, a functional enrichment analysis of core genes was performed. These core genes were significantly enriched in multiple Gene Ontology (GO) functions and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways associated with cardiomyopathy and related diseases or tissues (FDR adjusted $p < 0.05$) (Fig. 5), including some functions and pathways enriched by the cardiomyopathy risk module, such as “Purine metabolism” pathway (Fig. 3) and other pathways and functional categories. In the “Purine metabolism” pathway, the core genes encode two kinases and one reductase. Two kinases are involved in the hydrolysis and synthesis of the energy substance ATP, while reductase is responsible for the de novo conversion of ribonucleoside diphosphates to deoxyribonucleoside diphosphates, and two kinases and reductase are involved in subsequent DNA synthesis.

The homeostasis of Glutathione (GSH), which could be affected by the “Glutathione metabolism” pathway, was related to the pressure-overloaded heart remodeling and dysfunction [24].

Five function terms were related to ion homeostasis in cells, such as “metal ion homeostasis”, “monovalent inorganic cation homeostasis” and “monovalent inorganic

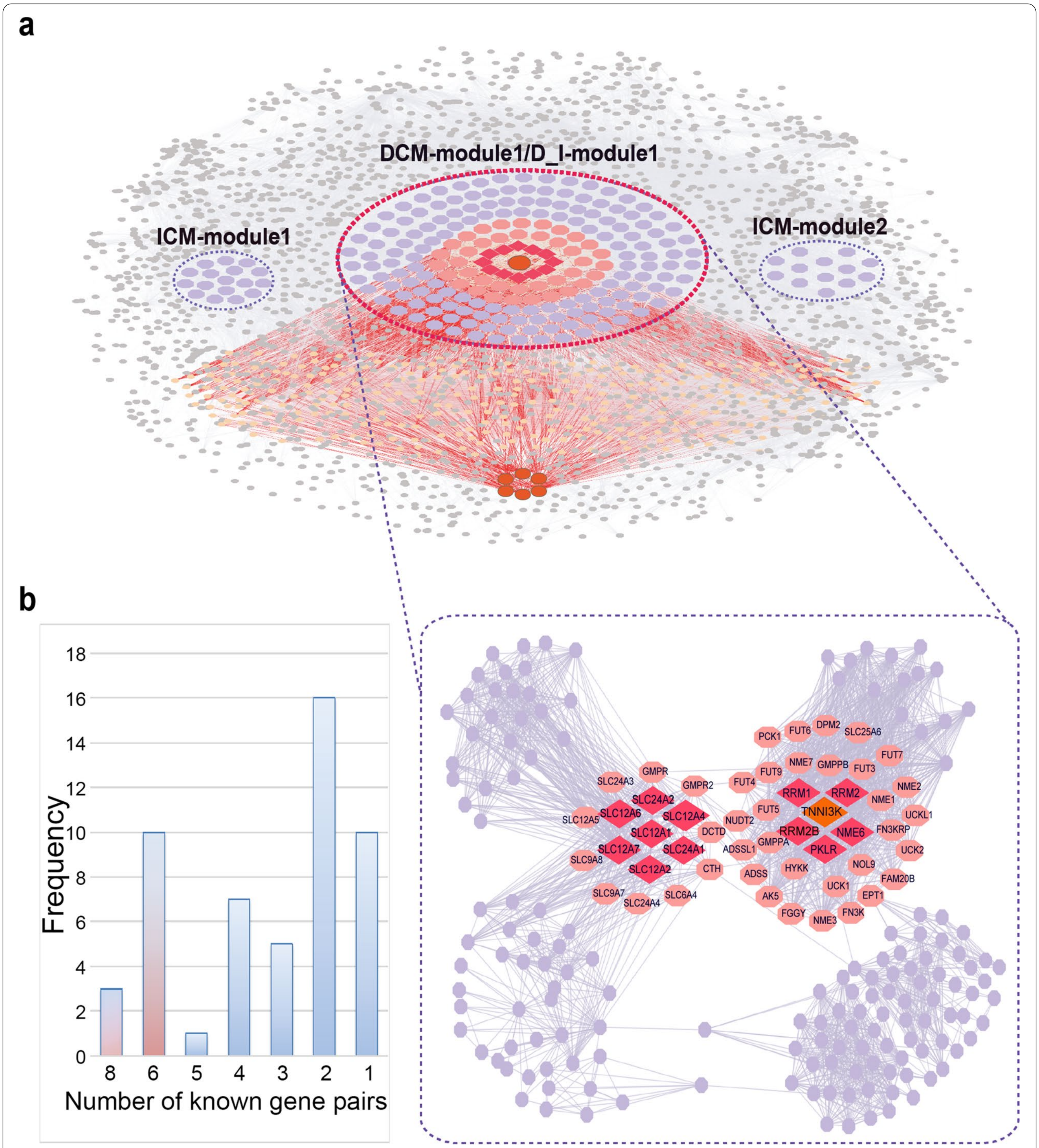

Fig. 4 Identification of core genes. **a** The overall reconstructed metabolic network. Genes in the red dashed circle comprised the cardiomyopathy risk module. Nodes in light orange are the genes on the shortest paths between known genes. The red lines are the shortest paths between known genes. **b** The picture on the left is the distribution of the number of known gene pairs linked by candidate genes via the shortest paths, and the red bars are the top 25%. The picture on the right is the enlarged picture of the cardiomyopathy risk module. Nodes in purple are the genes in the module, in light red are the candidate genes, in dark red and orange diamonds are the core genes. The dark orange ones are known genes of cardiomyopathy

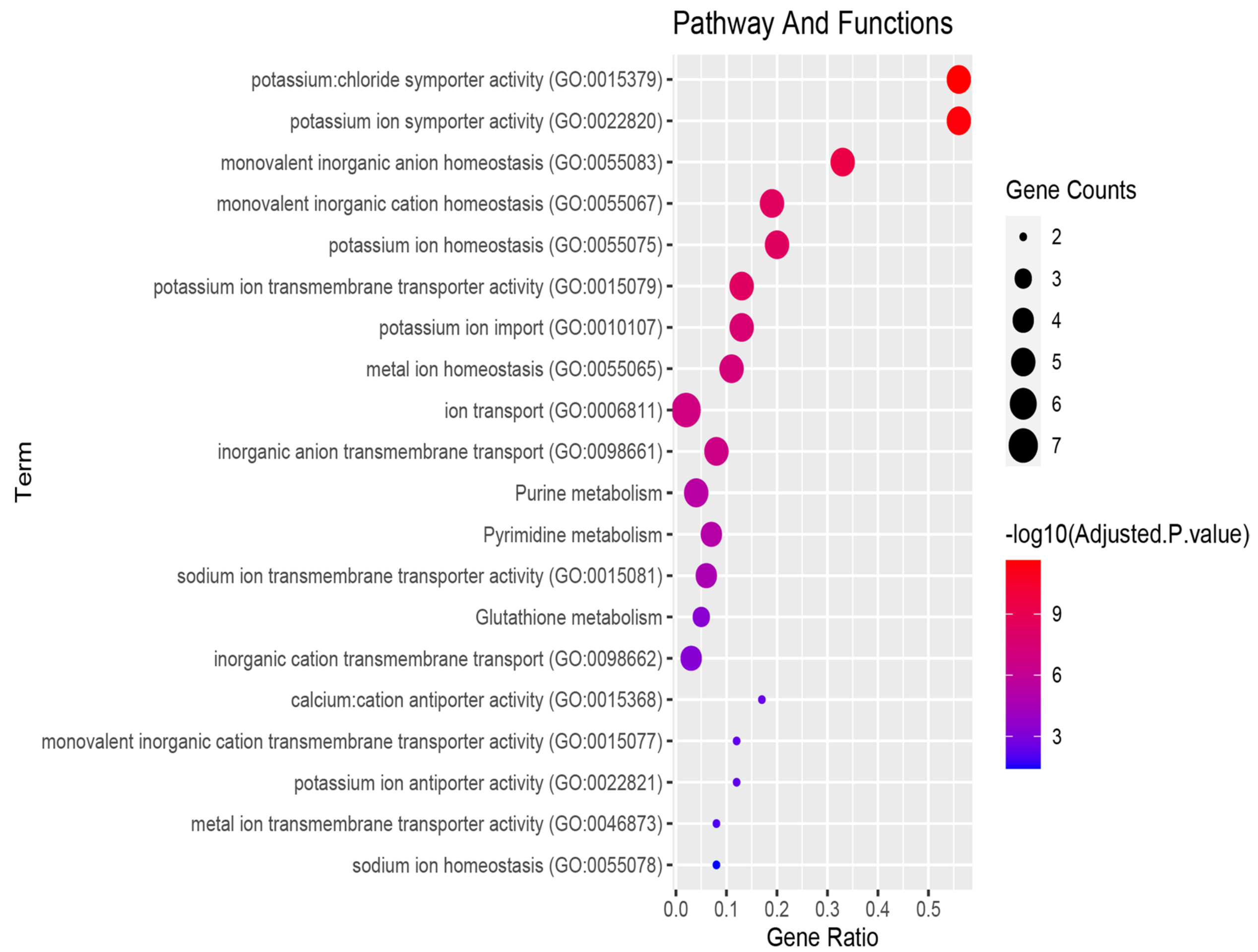

Fig. 5 Enriched functions and pathways of core genes

anion homeostasis". Abnormal ion homeostasis is related to the process of cardiomyopathy and reperfusion injury after myocardial ischemia, and often occurs in patients with various heart diseases [25, 26]. Among ions, K^+, Na^+ and Ca^{2+} were related to core genes, since core genes could be enriched in K^+, Na^+, Ca^{2+}-related functions. For example, "potassium ion homeostasis", "potassium ion transmembrane transporter activity" and other functions were closely related to the steady state and transport of potassium ions. Adenosine triphosphate-sensitive potassium channels (KATP) exist on the cell surface and mitochondrial membrane of cardiomyocytes, and can adapt electrical activity to metabolic challenges, thereby maintaining the normal biological functions of myocytes [27]. "calcium: cation antiporter activity " and "sodium ion transmembrane transporter activity" were closely related to the transportation and homeostasis of Na^+ and Ca^{2+}. Studies have found that changes in intracellular Ca^{2+} homeostasis and late Na^+ current increased the possibility of early depolarization and delayed depolarization, which caused arrhythmia in diseased cardiomyocytes [28].

Literature confirmation

Among the 13 core genes, 7 genes (containing one known disease gene) have been confirmed to be related to cardiomyopathy or other heart diseases by literature.

TNNI3K was a confirmed pathogenic gene of DCM that has been implicated in various cardiac phenotypes and diseases [29, 30]. Na^+-K^+-$2Cl^-$ cotransporter 1 (NKCC1), encoded by gene SLC12A1, and Na^+-K^+-$2Cl^-$ cotransporter 2 (NKCC2), encoded by gene SLC12A2, were two variants or isoforms of Na-K-2Cl-cotransporter (NKCC), which was one of the most important sodium transport mechanisms that could cause the intracellular sodium concentration to increase. The increase in intracellular Na^+ and Ca^{2+} concentration transduced nuclear signals, and triggered cardiac remodeling and hypertrophy [31]. Studies have shown that SLC12A6 is more specifically present in cardiomyocytes, vascular smooth muscle cells and various neurons [32, 33]. The genes RRM1 and RRM2 encoded two subunit proteins of ribonucleotide reductase (RNR), and RRM2B encoded the small subunit of p53-inducible RNR. The increase of RNR and/or dATP pools in heart cells could significantly alter the cycle of actin-myosin bridges, thereby enhancing the contractile function of patients with heart failure [34, 35].

Although 6 core genes have not been reported to be related to cardiomyopathy, they were enriched in the pathways and functions related to cardiomyopathy. The role of these 6 core genes in cardiomyopathy is worthy of further study.

Classification efficiency

To further reveal the relationship between the core genes and cardiomyopathy, the expression values of the core genes were used as the classification features to classify samples of different states (DCM_NF, ICM_NF, or DCM_ICM) in the cardiomyopathy expression data (GSE116250). The results showed that the core genes had good classification performance for normal and DCM samples (AUC= 0.996), ICM and normal samples (AUC=0.989), and for ICM and DCM samples (AUC=0.708).

In addition, two sets of random genes were selected for comparison to verify the classification efficiency of core genes. The first set of random genes was comprised of genes with the same number as the core genes randomly selected from 39 candidate genes after removing the core genes. The second set contained 13 randomly selected genes with the same number of differentially and non-differentially expressed genes as the core genes. The expression values of genes in the random gene sets were used as classification features to classify samples of different states (DCM_NF, DCM_ICM or ICM_NF). The randomization was performed 100 times. The difference between the AUC value of the core genes and the AUC values of random gene sets were tested by Wilcoxon Signed Rank Test. The results demonstrated significant differences between the core genes and the two random gene sets (p <0.01, Fig. 6). And the classification efficiency of core genes was significantly better than that of random genes.

To further verify the performance of the core genes in classifying samples of different states, two other publicly published independent microarray datasets GSE21610 (8 normal samples, 42 DCM samples, and 18 ICM samples) and GSE1145 (15 DCM samples, 11 ICM samples and 11 NF samples) from the Gene Expression Omnibus (GEO) database were used. The Support vector machine (SVM) algorithm was applied to these datasets, respectively. It was demonstrated that the core genes could accurately classify samples of different states in GSE21610 (AUC > 0.70) and GSE1145 (AUC > 0.80).

The above results showed that core genes could not only efficiently distinguish between different samples, but also correctly classify samples of other expression profiles. These core genes were expected to become markers for DCM and ICM.

Discussion

Cardiomyopathy is a type of myocardial disease with abnormal heart structure and myocardial function caused by different causes. DCM and ICM are two common types of cardiomyopathy with similar clinical manifestations, and difficult to distinguish [36, 37]. Here, an integrated method was proposed to identify cardiomyopathy-related core genes in a human metabolic network using cardiomyopathy-related expression data. Three groups of initial modules were determined from the reconstructed human metabolic network. Furthermore, three candidate modules with significant differences by permutation tests were selected.

One of them was identified as a cardiomyopathy risk module, which was able to distinguish between DCM_NF samples as well as DCM_ICM samples. And 13 core genes closely related to cardiomyopathy were identified. They could effectively distinguish between samples of different states (DCM_NF, ICM_NF or DCM_ICM) and were enriched in pathways and functions related to cardiomyopathy.

Generally speaking, network modules are selected with a relatively large number of nodes, such as module with at least 5 nodes [38]. To obtain more and comprehensive initial modules, we reduced the threshold to 4 nodes, which also appeared in previous studies [39]. We further selected the initial module with the number of nodes ≥3 for analysis, and the cardiomyopathy risk module and core genes finally identified remained unchanged.

The risk module showed significant differences for both DCM_NF and DCM_ICM samples. Although the risk module was not significantly different between normal and ICM samples, the core genes identified from it could distinguish ICM samples from normal ones.

Genes of the other two candidate modules (ICM-module1 and ICM-module2) were effective to separate ICM samples from normal samples, illustrating its significant differences between the normal state and the ICM state (ICM-module1 (AUC = 0.93), ICM-module2 (AUC = 0.79)).

The classification efficiency of the core genes was compared with the classification efficiency of initial modules to prove its effectiveness. Specifically, when

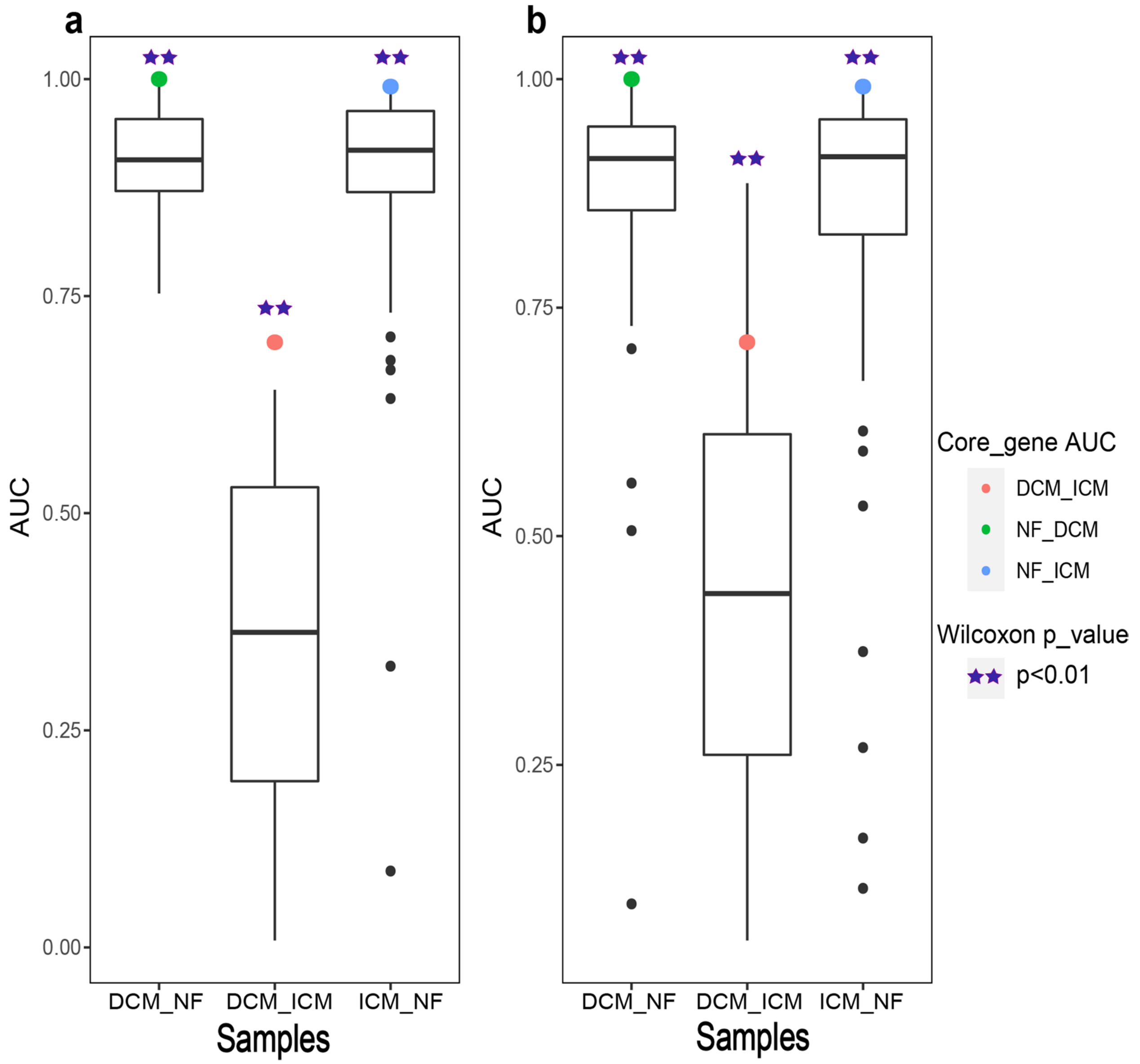

Fig. 6 AUC values of core genes and two random sets. **a** AUC value distribution of the first set of random genes. **b** AUC value distribution of the second set of random genes. The box plots represent the AUC values of random sets classifying DCM_NF, DCM_ICM and ICM_NF samples, and the green, red and blue marks are the AUC values of core genes, respectively. The purple asterisk indicates the degree of difference between the AUC value of the core genes and the AUC values of the random sets

distinguishing ICM_NF, DCM_NF and DCM_ICM samples, the classification efficiency of core genes was compared with that of 37 ICM initial modules, 21 DCM initial modules and 8 D_I initial modules, respectively. The initial modules were compared at two levels. First, the expression values of all genes in each initial module (the cardiomyopathy risk module included) were used as classification features to classify samples in different states. Second, since the classification of more genes might obtain higher classification accuracy, for modules with more than 13 genes, 100 sets of genes with the same number of differentially and non-differentially expressed genes as the core genes were randomly selected. The expression values were used as the classification features to classify samples in different states. The average of 100 random AUC values was used for each module. The classification efficiency for ICM_NF, DCM_NF and DCM_ICM samples of core genes were significantly better than that of the initial modules of the two levels (Wilcoxon signed rank test, Fig. 7). Our identified core genes could classify samples with high accuracy.

WGCNA is a popular tool for network analysis and mining modules and hub genes based on gene co-expression [40]. To further evaluate our approach, WGCNA was performed using genes from the reconstructed metabolic network based on the GSE116250 dataset. Sixteen co-expression modules were identified using the one-step network construction function of the WGCNA R package. Genes with |gene significance| > 0.2 and |module membership| > 0.8 were selected as hub genes in each significant module. However, the core genes that we identified were not in these genes. The hub genes of these modules were respectively used as characteristics to classify samples of different states in GSE116250. The results showed

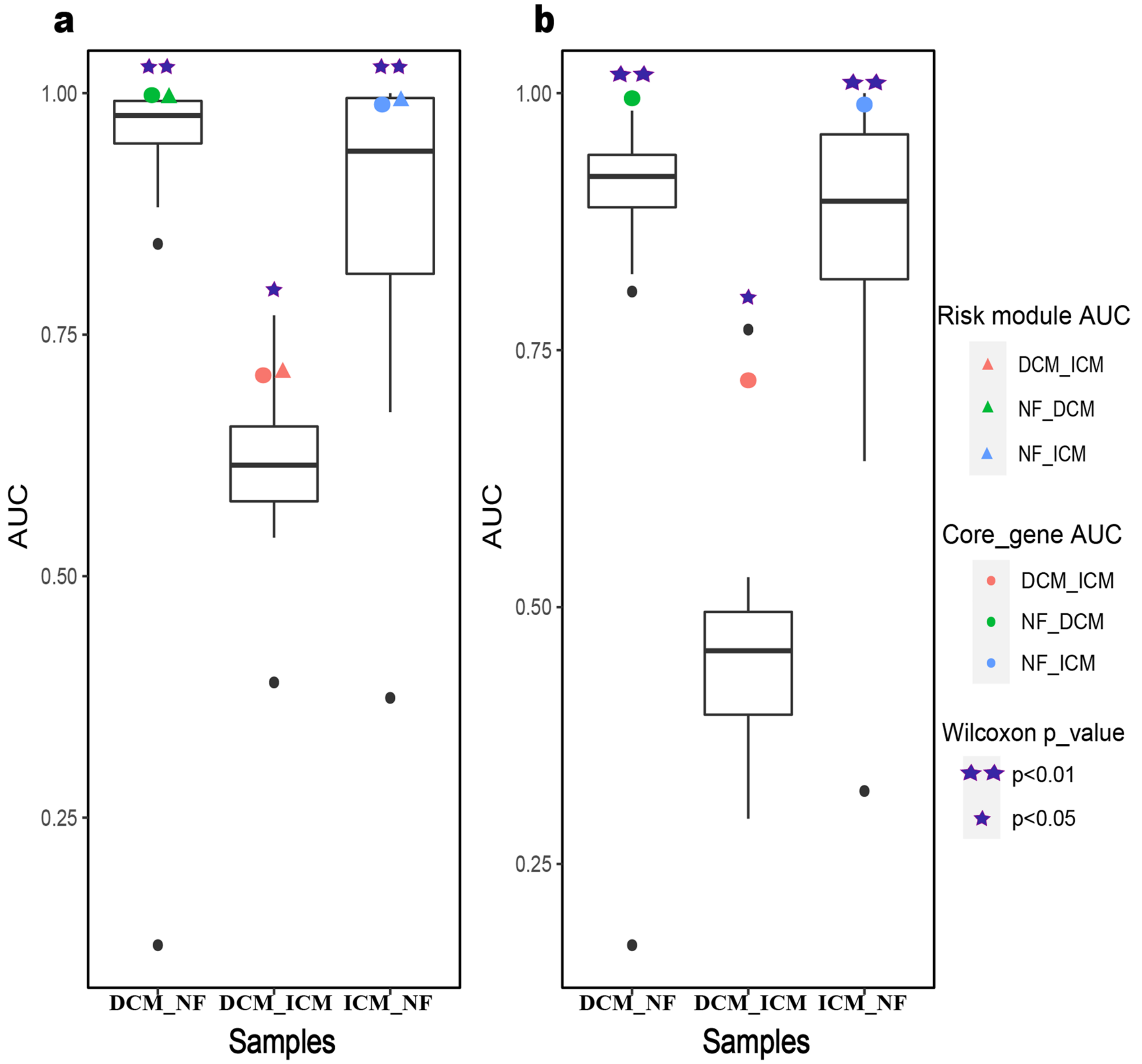

Fig. 7 AUC values of core genes and initial modules. **a** AUC value distribution of the initial modules by the first level. **b** AUC value distribution of the initial modules by the second level. The box plots represent the AUC values of initial modules classifying DCM_NF, DCM_ICM and ICM_NF samples. The green, red and blue circles are the AUC values of core genes, while the green, red and blue triangles are the AUC values of the cardiomyopathy risk module, respectively. The purple asterisk indicates the degree of difference between the AUC value of the core genes and the AUC values of the initial modules

that the hub genes of most modules could effectively distinguish normal and disease samples, and the hub genes of two modules could distinguish DCM and ICM samples. Our approach and WGCNA identified cardiomyopathy-related genes from different perspectives. Our approach was a systematic multi-omics integrated approach based on a metabolic network, while WGCNA conducted network analysis for co-expression between genes.

The limitation of our research was that although Recon 3 contains relatively complete metabolic reaction information, the reconstructed metabolic network was not large enough. Therefore, some known genes and DEGs were not in the network. A more comprehensive network might help to improve results, obtain more candidate modules, and identify more cardiomyopathy risk modules and core genes related to diseases.

Conclusions

In summary, a comprehensive method based on a human metabolic network using cardiomyopathy expression data was proposed to identify cardiomyopathy-related core genes. A total of 13 core genes were identified from the cardiomyopathy risk module based on the shortest paths between known genes. These core genes could distinguish both between normal and disease samples and between DCM and ICM samples. This research will contribute to identifying potential biomarkers of cardiomyopathy and to distinguishing different types of cardiomyopathy.

Materials and methods

Data

Screening of DEGs

The expression profile by high throughput sequencing GSE116250 was downloaded from the GEO (https://www.ncbi.nlm.nih.gov/geo/), which contained 14 normal samples, 13 ICM samples, and 37 DCM samples [41]. All expressed data used in this study were processed using the following process. (1) The probes or genes with more than 50% missing values were deleted, and the remaining missing values were filled with the k-Nearest Neighbor method using the knnImputation function in R package "DMwR". Specifically, for each missing value, its k nearest expression values were searched based on the Euclidean Distance, and the weighted average of these values was used to fill in the missing value. (2) Probes corresponding to multiple genes were deleted. (3) For multiple probes corresponding to the same gene, the average expression value of these probes was used as the expression value of the gene.

The SAM algorithm was used to find the DEGs between ICM and normal samples, between DCM and normal samples, and between ICM and DCM samples, through the R package "samr". Finally, 1802 DEGs between DCM and normal (indicated as DCM_NF in this paper) samples (DCM_NF DEGs), 3253 DEGs between ICM and normal (indicated as ICM_NF) samples (ICM_NF DEGs), and 358 DEGs between DCM and ICM (indicated as DCM_ICM) samples (DCM_ICM DEGs) with $|\log2(FC)| > 1$ and FDR adjusted p-value < 0.05 were obtained.

Metabolic network reconstruction

On the basis of metabolic responses extracted from Recon 3 of the Virtual Metabolic Human Database (https://www.vmh.life) [42], a metabolic network composed of protein-coding genes (nodes) and their interactions (edges) was reconstructed. Recon 3 was created by expanding Recon 2 through the addition of new publicly available metabolomics data. The metabolic network was reconstructed by the following process. The two enzymes were thought to interact if the product of the reaction catalyzed by one enzyme was the substrate of the reaction catalyzed by the other enzyme. The genes encoding the proteins that make up the two enzymes were connected in the reconstructed network. Ubiquitous metabolites such as H_2O, CO_2 and ADP were excluded to avoid bias due to their extreme connections. The reconstructed metabolic network contained 3105 nodes (containing 257 ICM_NF DEGs, 141 DCM_NF DEGs, and 34 DCM_ICM DEGs) and 85,880 edges.

Mining the initial modules

Subsequently, the metabolic network was visualized with the help of Cytoscape software (version 3.7.0). In addition, the MCODE (version 1.6.1) plug-in in Cytoscape software was used to explore important modules in the metabolic network [43]. The advanced options were set to degree cutoff = 2, K-Core = 3, and node score cutoff = 0.2. The modules containing DEGs were screened as initial modules.

Detection of candidate modules

Candidate modules with significant differences were detected from initial modules using two steps.

W was evaluated by the difference between the average of absolute values of Pearson correlation coefficients H and $H^{'}$ for samples of different states (DCM_NF, ICM_NF, or DCM_ICM).

$$W(M) = |H - H'|$$

where H and H' were calculated according to expression values for all gene pairs.

$$H = \frac{1}{C_g^2}\sum_{j=1}^{C_g^2}\left|Pearson\left(X_j, Y_j\right)\right| = \frac{1}{C_g^2}\sum_{j=1}^{C_g^2}\left|\frac{\sum_{i=1}^{n}\left(X_{ji}-\overline{X}_j\right)\left(Y_{ji}-\overline{Y}_j\right)}{\sqrt{\sum_{i=1}^{n}\left(X_{ji}-\overline{X}_j\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_{ji}-\overline{Y}_j\right)^2}}\right|$$

$$H' = \frac{1}{C_g^2}\sum_{j=1}^{C_g^2}\left|Pearson\left(X_j', Y_j'\right)\right| = \frac{1}{C_g^2}\sum_{j=1}^{C_g^2}\left|\frac{\sum_{i=1}^{m}\left(X_{ji}'-\overline{X}_j'\right)\left(Y_{ji}'-\overline{Y}_j'\right)}{\sqrt{\sum_{i=1}^{m}\left(X_{ji}'-\overline{X}_j'\right)^2}\sqrt{\sum_{i=1}^{m}\left(Y_{ji}'-\overline{Y}_j'\right)^2}}\right|$$

C_g^2 is the number of all gene pairs. n and m are the number of samples of different states (DCM_NF, ICM_NF, or DCM_ICM), and X_j, Y_j and X_j', Y_j' are the expression values of the j-th gene pair in two different states, X_{ji}, X_{ji}' and Y_{ji}, Y_{ji}' are the expression values of the j-th gene pair in the i-th sample, and $\overline{X}_j, \overline{X}_j'$ and $\overline{Y}_j, \overline{Y}_j'$ are their average expression value, respectively.

Second, permutation tests were performed on the DCM, ICM and D_I initial modules to screen the modules with significant differences. The null hypothesis was that the initial modules had no difference between different states (DCM_NF, ICM_NF, or DCM_ICM). From the reconstructed metabolic network, 1000 degree-conserved random modules (the same degree of nodes as the initial module) and 1000 size-conserved random modules (the same number of nodes as the initial module) were constructed for each initial module. The Pearson difference scores of every random module (degree-conserved and size-conserved) were calculated and compared with the score of the corresponding initial module, respectively. For each way of randomization, the p values of initial modules were defined as follows.

$$p = 1 - \frac{t}{1000}$$

where t is the number of random modules whose Pearson difference score is less than that of the initial module.

For initial modules with p value < 0.05, the null hypothesis should be rejected, so they were significantly different between different states. The initial modules that were significant in both cases (degree-conserved and size-conserved) were retained as candidate modules (both p values < 0.05).

The performance of candidate modules in classifying samples of different states can further reveal their relationship with cardiomyopathy. A SVM classifier was constructed to classify samples of different states (DCM_NF, ICM_NF, DCM_ICM) with the expression values of all genes in each module as features. The kernel function of SVM was set to “radial”. The performance was evaluated using leave-one-out cross validation (LOOCV). LOOCV draws one sample at a time as the test set, and the rest as the training set. Then the receiver operating characteristic (ROC) curve was drawn, and the AUC was calculated to measure the classification performance according to the classification results of the test sets.

Identification of cardiomyopathy risk modules

Markov random field (MRF) refers to the random field with Markov characteristics, which is often used to build mathematical models to identify protein interaction subnetworks [12, 13]. In our study, an MRF model was used to evaluate the expression difference considering both DEGs and non-DEGs in a candidate module by module score based on MRF (MRFms).

For a candidate module M with g genes, it was assumed that the expression difference $E=(E_1,\dots,E_g)$ between samples of different states formed an MRF. According to the properties of Markov random fields, the expression difference of gene g depends on the difference value of its one-step neighbor genes. Gibbs distribution was employed to specify the joint probability of E:

$$P(E) = \frac{1}{K}e^{-\frac{1}{T}F(E)}$$

where K is a constant that guarantees the probability sum to be 1, T is a temperature parameter controlling the distribution sharpness, and

$$F(E) = -\frac{1}{\sqrt{g}}\sum_{i\in G_1}E_i + \frac{1}{b}\sum_{i,j\in G_2}\left(\frac{E_i}{\sqrt{d_i}} - \frac{E_j}{\sqrt{d_j}}\right)^2 MI\left(i,j\right)$$

Based on our previous study [44] and the calculation process in [45], the MRFms for module M incorporating Mutual Information (MI) was defined as

$$MRFms(M) = \frac{1}{\sqrt{g}}\sum_{u\in G_1} E_u - \frac{1}{b}\sum_{v,z\in G_2}\left(\frac{E_v}{\sqrt{d_v}} - \frac{E_z}{\sqrt{d_z}}\right)^2 MI(v,z)$$

where b is the number of edges, G_1 and G_2 are the set of DEGs and non-DEGs in the module; E_u, E_v and E_z are the expression differences of genes u, v and z between different states (DCM_NF, ICM_NF, or DCM_ICM), and d_v and d_z are the degrees of genes v and z in the network, respectively. $MI(v,z)$ is the mutual information of genes v and z.

Then the same permutation test as in the previous step was used to screen out cardiomyopathy risk modules. Finally, the modules that were significant in both cases (degree-conserved and size-conserved random modules) were identified as cardiomyopathy risk modules (both p values < 0.05).

Identification of core genes

According to the connection between genes and known genes in the network, core genes were further screened in the cardiomyopathy risk modules. From the Online Mendelian Inheritance in Man database [46], 43 known genes of cardiomyopathy were extracted, and 7 of them (PPCS, RAF1, TNNI3K, ABCC9, EYA4, SDHA, and TTN) were in the metabolic network. Then, the shortest paths between known gene pairs were searched, and genes in cardiomyopathy risk modules that appeared on these shortest paths were selected as candidate genes. The number of known gene pairs linked by gene x via these shortest paths $B(x)$ was counted as follows.

$$B(x) = \sum_{s\neq x\neq t}\sigma_{st}(x)$$

$$\sigma_{st}(x) = \begin{cases} 1 & if\ G(s,t) = G(s,x) + G(x,t) \\ 0 & \quad otherwise \end{cases}$$

where $G\ (s,\ t)$ is the length of the shortest path between two nodes s and t. s and t are known genes for cardiomyopathy in the metabolic network. $\sigma_{st}(x)$ is a variable that indicates whether any shortest path between nodes s and t passes through node x. If so, it is 1, otherwise it is 0.

Genes linked more known gene pairs (top upper quartile) via shortest paths were identified as core genes.

$$C = \left\{x \middle| \text{rank}_x \in [1, \lfloor 0.25n \rfloor], \text{rank}_x = \text{rank}(B(x), in\ D)\right\}$$

where D is the set of $B(x)$ for all x, $rank_x$ is the rank of $B(x)$ when ranking all $B(x)$s in set D in descending order, n is the number of candidate genes.

Evaluation of core genes

In order to reflect the association of core genes with cardiomyopathy, they were analyzed from three aspects: literature verification, enrichment analysis and classification performance. Literature verification was conducted by searching literature showing the relationship between core genes and cardiomyopathy in the PubMed database (https://www.ncbi.nlm.nih.gov/pubmed). Enrichr was used for GO functional annotation and KEGG pathway enrichment of core genes [47]. The PubMed database was also used to validate the association of significantly enriched functional classes and pathways (FDR adjusted $p < 0.05$) with the disease. The expression values of the core genes were further used as features to classify samples of different states (DCM_NF, ICM_NF, or DCM_ICM) in the expression profile and independent microarray datasets. The classification performance of the core genes and of random gene sets was compared to further evaluate the classification performance of the core genes. Random gene sets were composed of randomly selected differentially and non-differentially expressed genes from the cardiomyopathy risk modules with the same number as the core genes.

Abbreviations

DCM: Dilated cardiomyopathy; ICM: Ischemic cardiomyopathy; DEGs: Differentially expressed genes; SAM: Significant Analysis of Microarray; DCM_NF: The two states of DCM and normal; ICM_NF: The two states of ICM and normal; DCM_ICM: The two states of DCM and ICM; D_I modules: The modules containing DEGs of DCM_ICM; DCM modules: The modules containing DEGs of DCM_NF; ICM modules: The modules containing DEGs of ICM_NF; MRF: Markov random field; MRFms: Module score based on MRF; RNR: Ribonucleotide reductase; GEO: Gene Expression Omnibus; SVM: Support vector machine; MCODE: Molecular Complexity Detection; LOOCV: Leave-one-out cross validation; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; AUC: Area under the curve; MI: Mutual Information.

Acknowledgements

Not applicable.

Authors' contributions

W.L. and Z.R. provided ideas for this article. Z.Z., H.C., Y.Z., J.L. and Y.H. collected and processed data. Z.R., L.G., Z.L. and Y.Z. analyzed and interpreted the results. Z.R., Z.Z. and H.C. wrote and revised the article. W.L. and L.C. reviewed and helped writing the manuscript. All authors have read and approved the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China [grant numbers 61702141 and 61272388]; the Natural Science Foundation of Heilongjiang Province [grant number LH2021F043]; the Fundamental Research Funds for the Provincial Universities in Heilongjiang Province [grant number 2017-KYYWF-0303]; the Innovative Scientific Research Funding Project of Harbin Medical University [grant number 2017JCZX46]; the Heilongjiang Postdoctoral Funds for Scientific Research Initiation [grant number LBH-Q17132]; the National College Student Innovation and Entrepreneurship

Training Program [grant number 201710226011]; the University Student Innovation and Entrepreneurship Training Program in Harbin Medical University [grant number 201910226372].

Availability of data and materials

Our metabolic network data were extracted from the Virtual Metabolic Human Database (https://www.vmh.life). The Expression data were collected from the Gene Expression Omnibus (GEO). Among them, GEO's accession IDs are GSE116250 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116250), GSE1145 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1145) and GSE21610 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21610. All of the above data is publicly available. The code files for main steps of this study are available at https://github.com/wendyliwan/Identification-of-cardiomyopathy-related-core-genes.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 14 May 2021 Accepted: 15 December 2021
Published online: 12 January 2022

References

1. Burke MA, Cook SA, Seidman JG, Seidman CE. Clinical and mechanistic insights into the genetics of cardiomyopathy. J Am Coll Cardiol. 2016;68(25):2871–86.
2. McKenna WJ, Maron BJ, Thiene G. Classification, epidemiology, and global burden of cardiomyopathies. Circ Res. 2017;121(7):722–30.
3. Tompkins BA, Rieger AC, Florea V, Banerjee MN, Natsumeda M, Nigh ED, et al. Comparison of Mesenchymal stem cell efficacy in ischemic versus nonischemic dilated cardiomyopathy. J Am Heart Assoc. 2018;7(14):e008460.
4. Hare JM, Walford GD, Hruban RH, Hutchins GM, Deckers JW, Baughman KL. Ischemic cardiomyopathy: endomyocardial biopsy and ventriculographic evaluation of patients with congestive heart failure, dilated cardiomyopathy and coronary artery disease. J Am Coll Cardiol. 1992;20(6):1318–25.
5. Giraldo BF, Pericas MF, Schroeder R, Voss A. Respiratory sinus arrhythmia quantified with linear and non-linear techniques to classify dilated and ischemic cardiomyopathy. Annu Int Conf IEEE Eng Med Biol Soc. 2018;2018:4860–3.
6. Sweet ME, Cocciolo A, Slavov D, Jones KL, Sweet JR, Graw SL, et al. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. BMC Genomics. 2018;19(1):812.
7. Ritterhoff J, Tian R. Metabolism in cardiomyopathy: every substrate matters. Cardiovasc Res. 2017;113(4):411–21.
8. Zhao J, Yang S, Jing R, Jin H, Hu Y, Wang J, et al. Plasma Metabolomic profiles differentiate patients with dilated cardiomyopathy and ischemic cardiomyopathy. Front Cardiovasc Med. 2020;7:597546.
9. Alonso N, Moliner P, Mauricio D. Pathogenesis, clinical features and treatment of diabetic cardiomyopathy. Adv Exp Med Biol. 2018;1067:197–217.
10. Wang DD, Zheng Y, Toledo E, Razquin C, Ruiz-Canela M, Guasch-Ferre M, et al. Lipid metabolic networks, Mediterranean diet and cardiovascular disease in the PREDIMED trial. Int J Epidemiol. 2018;47(6):1830–45.
11. Yang Y, Yang K, Hao T, Zhu G, Ling R, Zhou X, et al. Prediction of molecular mechanisms for LianXia NingXin formula: a network pharmacology study. Front Physiol. 2018;9:489.
12. Chen L, Xuan J, Riggins RB, Wang Y, Clarke R. Identifying protein interaction subnetworks by a bagging Markov random field-based method. Nucleic Acids Res. 2013;41(2):e42.
13. Wei Z, Li H. A Markov random field model for network-based analysis of genomic data. Bioinformatics. 2007;23(12):1537–44.
14. Swain JL, Sabina RL, Peyton RB, Jones RN, Wechsler AS, Holmes EW. Derangements in myocardial purine and pyrimidine nucleotide metabolism in patients with coronary artery disease and left ventricular hypertrophy. Proc Natl Acad Sci U S A. 1982;79(2):655–9.
15. DeBoer LW, Ingwall JS, Kloner RA, Braunwald E. Prolonged derangements of canine myocardial purine metabolism after a brief coronary artery occlusion not associated with anatomic evidence of necrosis. Proc Natl Acad Sci U S A. 1980;77(9):5471–5.
16. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–51.
17. Soltysinska E, Olesen SP, Christ T, Wettwer E, Varro A, Grunnet M, et al. Transmural expression of ion channels and transporters in human nondiseased and end-stage failing hearts. Pflugers Arch. 2009;459(1):11–23.
18. Gronich N, Kumar A, Zhang Y, Efimov IR, Soldatov NM. Molecular remodeling of ion channels, exchangers and pumps in atrial and ventricular myocytes in ischemic cardiomyopathy. Channels (Austin). 2010;4(2):101–7.
19. Park TS, Goldberg IJ. Sphingolipids, lipotoxic cardiomyopathy, and cardiac failure. Heart Fail Clin. 2012;8(4):633–41.
20. Carpentier AC. Abnormal myocardial dietary fatty acid metabolism and diabetic cardiomyopathy. Can J Cardiol. 2018;34(5):605–14.
21. Kovilakath A, Cowart LA. Sphingolipid mediators of myocardial pathology. J Lipid Atheroscler. 2020;9(1):23–49.
22. Sasset L, Zhang Y, Dunn TM, Di Lorenzo A. Sphingolipid De novo biosynthesis: a rheostat of cardiovascular homeostasis. Trends Endocrinol Metab. 2016;27(11):807–19.
23. Park TS, Yamashita H, Blaner WS, Goldberg IJ. Lipids in the heart: a source of fuel and a source of toxins. Curr Opin Lipidol. 2007;18(3):277–82.
24. Watanabe Y, Watanabe K, Kobayashi T, Saito Y, Fujioka D, Nakamura T, et al. Chronic depletion of glutathione exacerbates ventricular remodelling and dysfunction in the pressure-overloaded heart. Cardiovasc Res. 2013;97(2):282–92.
25. Doohan MM, Rasmussen HH. Myocardial cation transport. J Hypertens. 1993;11(7):683–91.
26. Pierce GN, Czubryt MP. The contribution of ionic imbalance to ischemia/reperfusion-induced injury. J Mol Cell Cardiol. 1995;27(1):53–63.
27. Liu Z, Cai H, Dang Y, Qiu C, Wang J. Adenosine triphosphate-sensitive potassium channels and cardiomyopathies (review). Mol Med Rep. 2016;13(2):1447–54.
28. Coppini R, Santini L, Olivotto I, Ackerman MJ, Cerbai E. Abnormalities in sodium current and calcium homoeostasis as drivers of arrhythmogenesis in hypertrophic cardiomyopathy. Cardiovasc Res. 2020;116(9):1585–99.
29. Fan LL, Huang H, Jin JY, Li JJ, Chen YQ, Zhao SP, et al. Whole exome sequencing identifies a novel mutation (c.333+2T>C) of TNNI3K in a Chinese family with dilated cardiomyopathy and cardiac conduction disease. Gene. 2018;648:63–7.
30. Milano A, Lodder EM, Bezzina CR. TNNI3K in cardiovascular disease and prospects for therapy. J Mol Cell Cardiol. 2015;82:167–73.
31. Baartscheer A, van Borren MM. Sodium ion transporters as new therapeutic targets in heart failure. Cardiovasc Hematol Agents Med Chem. 2008;6(4):229–36.
32. Garneau AP, Marcoux AA, Noel M, Frenette-Cotton R, Drolet MC, Couet J, et al. Ablation of potassium-chloride Cotransporter type 3 (Kcc3) in mouse causes multiple cardiovascular defects and isosmotic polyuria. PLoS One. 2016;11(5):e0154398.
33. Race JE, Makhlouf FN, Logue PJ, Wilson FH, Dunham PB, Holtzman EJ. Molecular cloning and functional characterization of KCC3, a new K-cl cotransporter. Am J Phys. 1999;277(6):C1210–9.
34. Kolwicz SC Jr, Odom GL, Nowakowski SG, Moussavi-Harami F, Chen X, Reinecke H, et al. AAV6-mediated cardiac-specific overexpression of Ribonucleotide Reductase enhances myocardial contractility. Mol Ther. 2016;24(2):240–50.
35. Korte FS, Dai J, Buckley K, Feest ER, Adamek N, Geeves MA, et al. Upregulation of cardiomyocyte ribonucleotide reductase increases intracellular 2 deoxy-ATP, contractility, and relaxation. J Mol Cell Cardiol. 2011;51(6):894–901.
36. Lin F, Gong X, Yu P, Yue A, Meng Q, Zheng L, et al. Distinct circulating expression profiles of long noncoding RNAs in heart failure patients

with ischemic and nonischemic dilated cardiomyopathy. Front Genet. 2019;10:1116.

37. Schaufelberger M. Cardiomyopathy and pregnancy. Heart. 2019;105(20):1543–51.
38. Luo S, Gan L, Luo Y, Zhang Z, Li L, Wang H, et al. Application of molecular Nanoprobes in the analysis of differentially expressed genes and prognostic models of primary hepatocellular carcinoma. J Biomed Nanotechnol. 2021;17(6):1020–33.
39. Patra S, Saxena S, Sahu N, Pradhan B, Roychowdhury A. Systematic network and Meta-analysis on the antiviral mechanisms of probiotics: a preventive and treatment strategy to mitigate SARS-CoV-2 infection. Probiotics Antimicrob Proteins. 2021;13(4):1138–56.
40. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
41. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991–5.
42. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, et al. The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res. 2019;47(D1):D614–24.
43. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
44. Li W, Deng G, Zhang J, Hu E, He Y, Lv J, et al. Identification of breast cancer risk modules via an integrated strategy. Aging (Albany NY). 2019;11(24):12131–46.
45. Shi X, Wang X, Shajahan A, Hilakivi-Clarke L, Clarke R, Xuan J. BMRF-MI: integrative identification of protein interaction network by modeling the gene dependency. BMC Genomics. 2015;16(Suppl 7):S10.
46. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online Mendelian inheritance in man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015;43(Database issue):D789–98.
47. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90–7.

Publisher's Note

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