

Research Article

Identification of Biomarkers Associated with Heart Failure Caused by Idiopathic Dilated Cardiomyopathy Using WGCNA and Machine Learning Algorithms

Mengyi Sun ¹ and Linping Li ²

¹Department of Clinical Laboratory, Jining First People's Hospital, Jining, Shandong, China

²Institute of Cardiovascular Diseases of Jining Medical Research Academy, Jining First People's Hospital, Jining, Shandong, China

Correspondence should be addressed to Linping Li; llp0522@163.com

Received 16 November 2022; Revised 3 April 2023; Accepted 12 April 2023; Published 25 April 2023

Academic Editor: Monika Dmitrzak-Weglarz

Copyright © 2023 Mengyi Sun and Linping Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. The genetic factors and pathogenesis of idiopathic dilated cardiomyopathy-induced heart failure (IDCM-HF) have not been understood thoroughly; there is a lack of specific diagnostic markers and treatment methods for the disease. Hence, we aimed to identify the mechanisms of action at the molecular level and potential molecular markers for this disease. **Methods.** Gene expression profiles of IDCM-HF and non-heart failure (NF) specimens were acquired from the database of Gene Expression Omnibus (GEO). We then identified the differentially expressed genes (DEGs) and analyzed their functions and related pathways by using “Metascape”. Weighted gene co-expression network analysis (WGCNA) was utilized to search for key module genes. Candidate genes were identified by intersecting the key module genes identified via WGCNA with DEGs and further screened via the support vector machine-recursive feature elimination (SVM-RFE) method and the least absolute shrinkage and selection operator (LASSO) algorithm. At last, the biomarkers were validated and evaluated the diagnostic efficacy by the area under curve (AUC) value and further confirmed the differential expression in the IDCM-HF and NF groups using an external database. **Results.** We detected 490 genes exhibiting differential expression between IDCM-HF and NF specimens from the GSE57338 dataset, with most of them being concentrated in the extracellular matrix (ECM) of cells related to biological processes and pathways. After screening, 13 candidate genes were identified. Aquaporin 3 (AQP3) and cytochrome P450 2J2 (CYP2J2) showed high diagnostic efficacy in the GSE57338 and GSE6406 datasets, respectively. In comparison to the NF group, AQP3 was significantly down-regulated in the IDCM-HF group, while CYP2J2 was significantly up-regulated. **Conclusion.** As far as we know, this is the first study that combines WGCNA and machine learning algorithms to screen for potential biomarkers of IDCM-HF. Our findings suggest that AQP3 and CYP2J2 could be used as novel diagnostic markers and treatment targets of IDCM-HF.

1. Introduction

As a non-ischemic heart disease, dilated cardiomyopathy (DCM) is characterized by abnormal myocardial structure and function [1, 2]. It is the most common indication of heart transplantation and the third-most prevalent pathogenic factor responsible for heart failure (HF) [3, 4]. Idiopathic and familial diseases are the most common causes of DCM. According to a previous report, around 70% DCM cases are idiopathic [4]. Idiopathic dilated cardiomyopathy (IDCM) is a myocardial disease of unknown etiol-

ogy, which can only be diagnosed if a secondary cause is excluded. Among HF patients, IDCM patients are younger and more at risk than cohorts with other etiologies (ischemic, hypertensive, and valvular) [5, 6]. The pathogenesis and genetic factors involved in the development of IDCM-induced HF (IDCM-HF) are not fully understood, and there is a lack of specific diagnostic markers and hardly any specific treatments for IDCM-HF. Therefore, it is crucial to unveil the specific molecular mechanism of IDCM-HF and to promote the diagnosis and precision treatment of the disease [7].

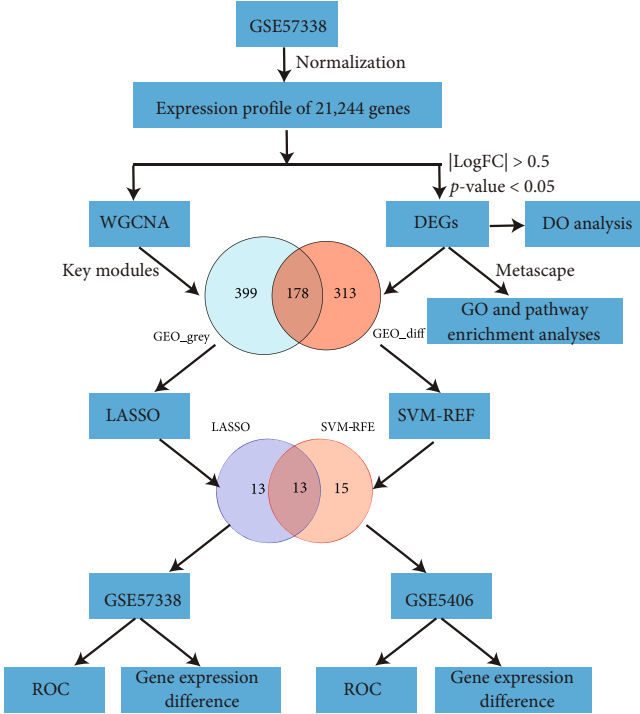


FIGURE 1: Flow chart of the study.

In recent years, gene sequencing and bioinformatics methods have resulted in new ideas for the elucidation of the mechanisms of disease development, disease diagnosis, and personalized precision medicine development. Wang et al. validated the diagnostic value of *MYG1*, *FLOT1*, and *ATG13* and proposed potential immunological mechanisms, biomarkers, and treatment targets in DCM patients with HF [8]. Qiu et al. investigated the mechanism of action of long noncoding RNAs (lncRNAs) including LINC01-AS1, AC061961.2, and RP11-13e1.5, which were implicated with the development of DCM with HF [9]. Zhang et al. found that ASPN, CD163, IL10, and LUM could predict the occurrence of DCM [10]. Despite this, there are few studies on IDCM-HF. Huang et al. [11] explored the key genes associated with the development of IDCM-HF, but their conclusions need to be examined further as the results have not been verified.

In this study, we combined weighted gene co-expression network analysis (WGCNA) with machine learning algorithms, including support vector machine-recursive feature elimination (SVM-RFE) and least absolute shrinkage and selection operator (LASSO), to screen hub genes of IDCM-HF samples for the first time. Results were then verified using the external dataset. Target genes that could be reliably used for exploring the pathogenesis, diagnosing, and managing IDCM-HF.

2. Materials and Methods

2.1. Data Source. Figure 1 illustrates the workflow chart of data preparation, processing, analysis, and validation. Two datasets (GSE57338 [12] and GSE5406 [13]) were acquired

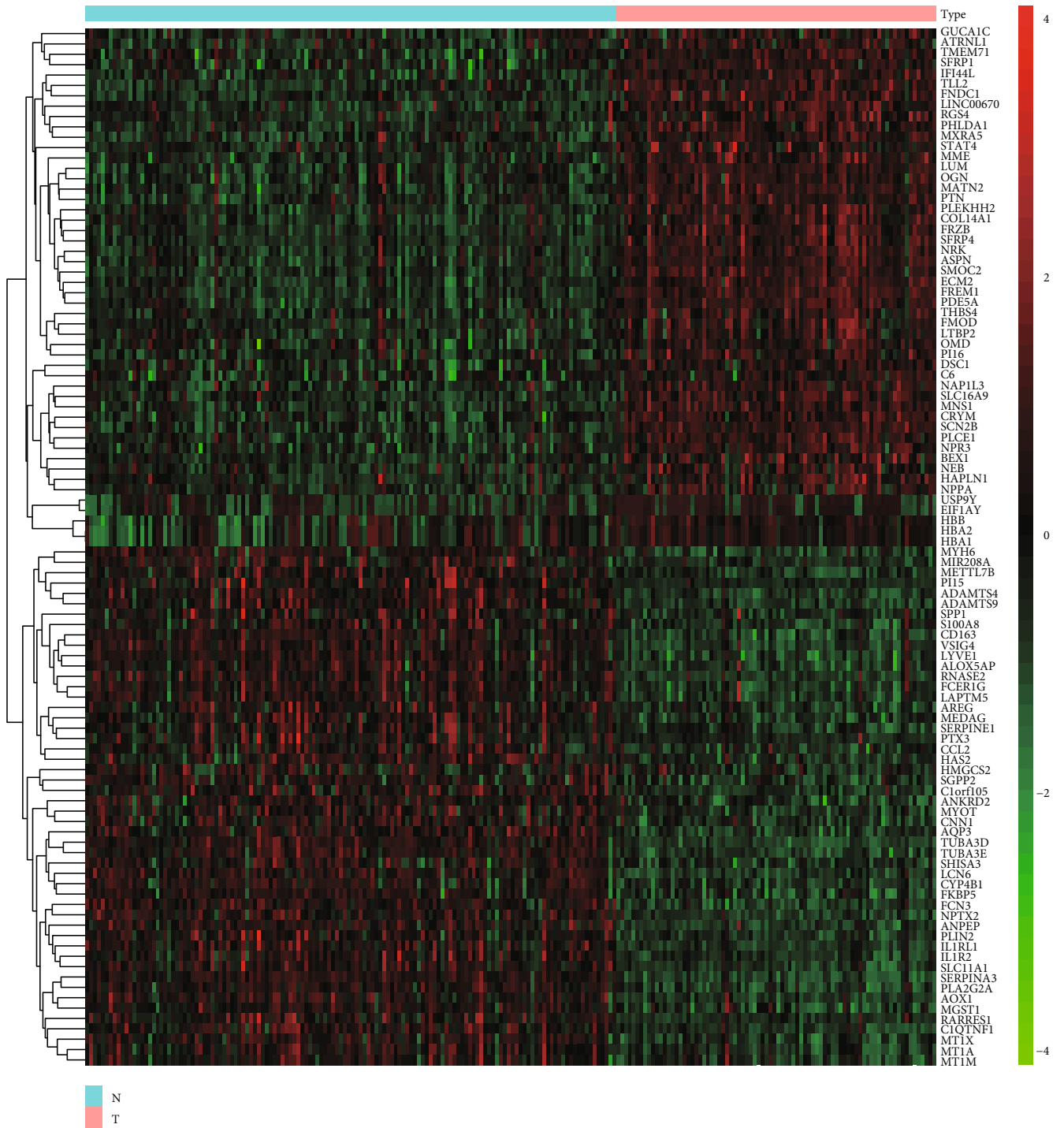
TABLE 1: The basic sample information from GSE 57338 and GSE 5406.

GSE accession	GSE57338	GSE5406
NF		
Mean age	49.36 ± 15.00	—
Male	73	—
Female	63	—
Total patients	136	16
IDCM-HF		
Mean age	51.16 ± 13.98	—
Male	63	—
Female	19	—
Total patients	82	86

from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). Both of the datasets were novel myocardial gene expression signatures of HF caused by IDCM. The GSE57338 dataset (GPL11532 platform), which was updated in 2018, was used for expression profiling analyses of the heart left ventricle of 82 IDCM-HF patients and 136 non-heart failure (NF) controls. The GSE5406 dataset (GPL96 platform) found data regarding expression profiling analyses of the heart left ventricle of 86 IDCM-HF patients and 16 NF controls. The basic sample information from the datasets was shown in Table 1. There were only two datasets (GSE57338 and GSE5406) on HF induced by IDCM found in the GEO database, so we selected the GSE57338 dataset with a larger sample size as the training group and GSE5406 as the verification group, in place of the combined analysis to validate the results.

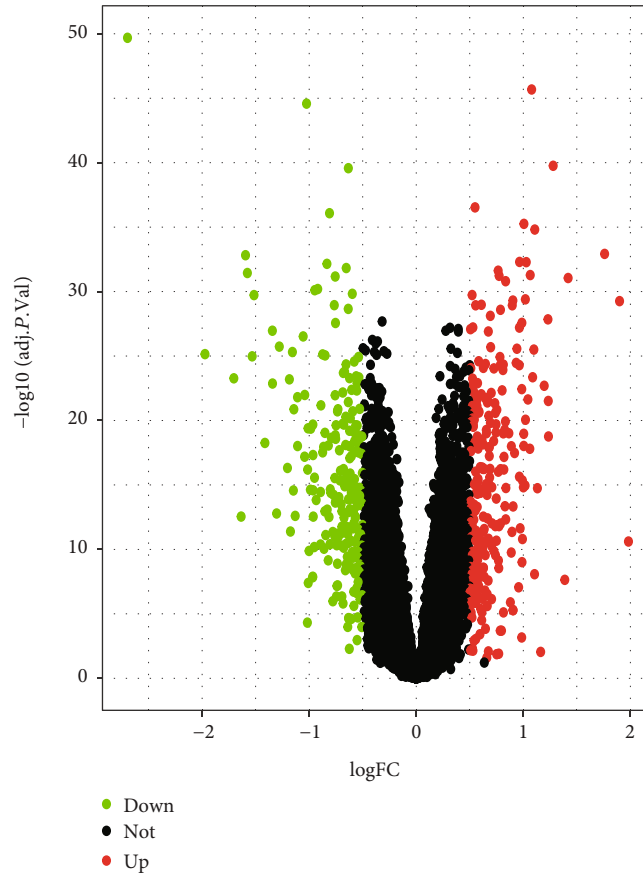
2.2. Identification and Analysis of DEGs. Firstly, we adjusted the threshold with $|\log_2 \text{fold change (FC)}| > 1$ and $|\log_2 \text{FC}| > 0.5$, respectively, to select the cut-off point. Then we utilized the LIMMA package of R to identify the DEGs between IDCM-HF and NF samples in the GSE57338 dataset. Disease Ontology (DO) enrichment analysis was performed using the DOSE package of R. The functions and related pathways of DEGs were analyzed by “Metascape” [14]. During functional analysis, we included gene ontology (GO) analysis; it included an analysis of the molecular functions (MF), cellular components (CC), and biological processes (BP). During pathway analysis, we also performed Kyoto Encyclopedia of Genes and Genomes (KEGG), canonical pathway, Reactome pathway, Wiki pathway, and PANTHER pathway analyses.

2.3. WGCNA Analysis. First, we constructed a clustering tree for samples to identify and remove outliers. Afterwards, the gene expression data of the GSE57338 dataset were analyzed via WGCNA [15]. We performed “pickSoftThreshold” in WGCNA to compute β values (range: 1–20) and select the optimal soft threshold. For network construction, the topological overlap matrix was created via transformation of the adjacency matrix, and the gene tree and color of modules were established based on the dissimilarity degree. Dynamic



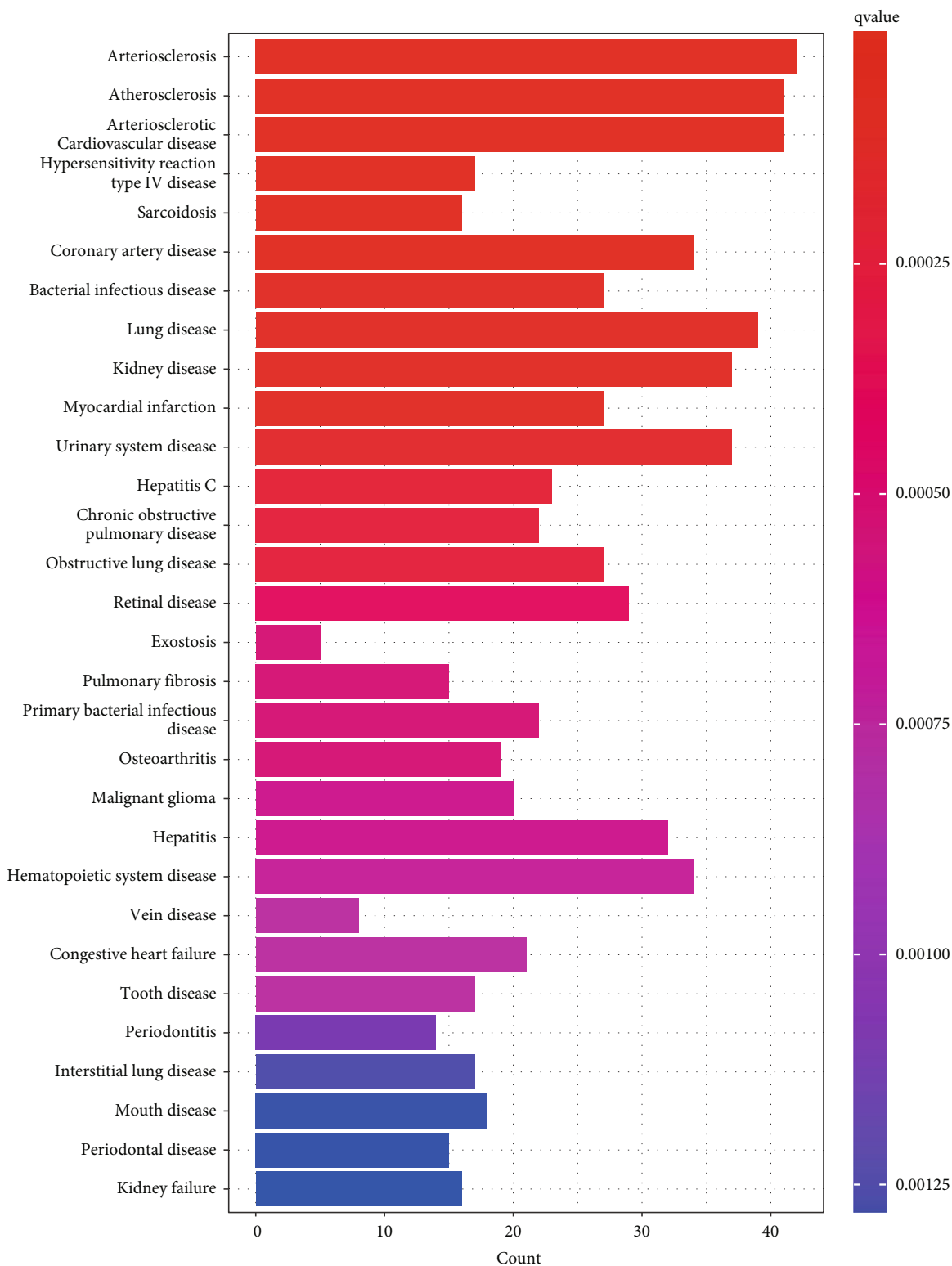
(a)

FIGURE 2: Continued.



(b)

FIGURE 2: Continued.



(c)

FIGURE 2: Continued.

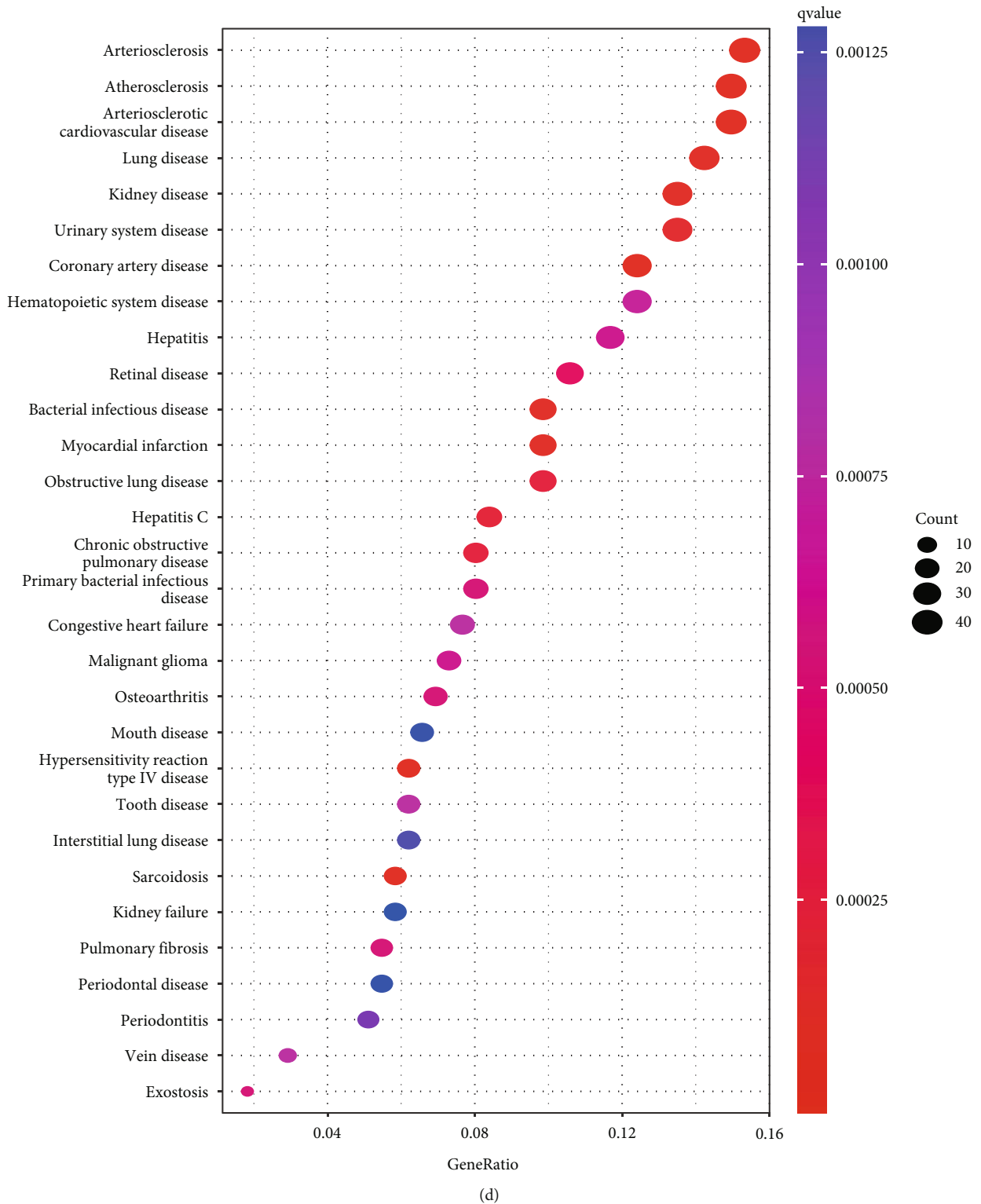


FIGURE 2: Correlation analysis of DEGs in the NF and IDCM-HF groups in the GSE57338 dataset. (a) Heatmap of the 50 most significantly up- and down-regulated DEGs. (b) Volcanic map. (c) Histogram of the top 30 diseases observed during DO analysis. (d) Bubble charts of the top 30 diseases observed during DO analysis. NF: non-heart failure; IDCM-HF: idiopathic dilated cardiomyopathy-induced heart failure; GO: gene ontology.

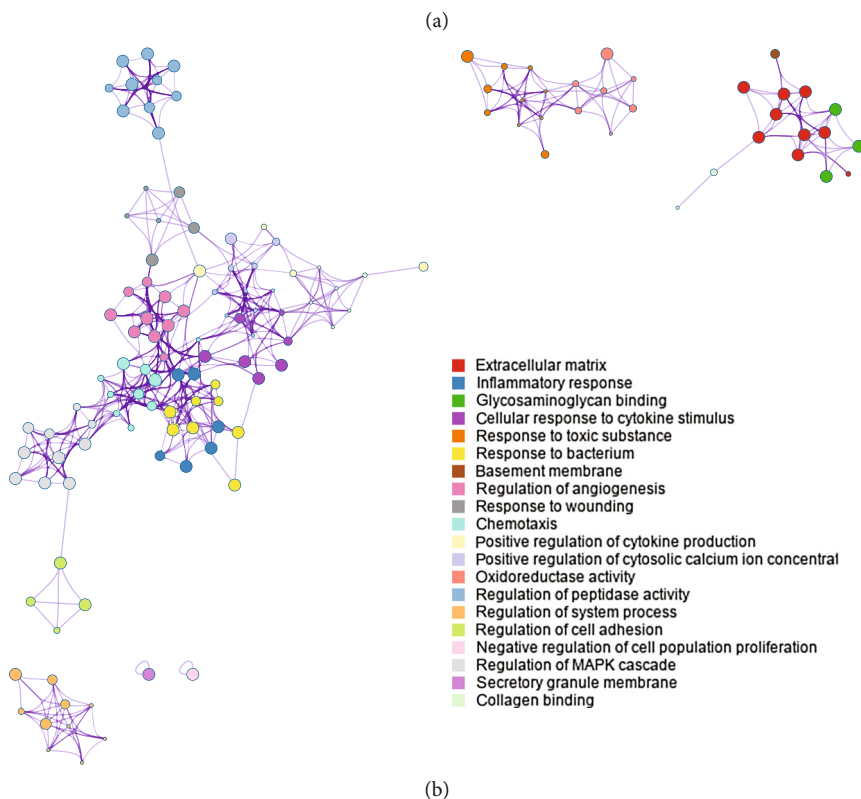
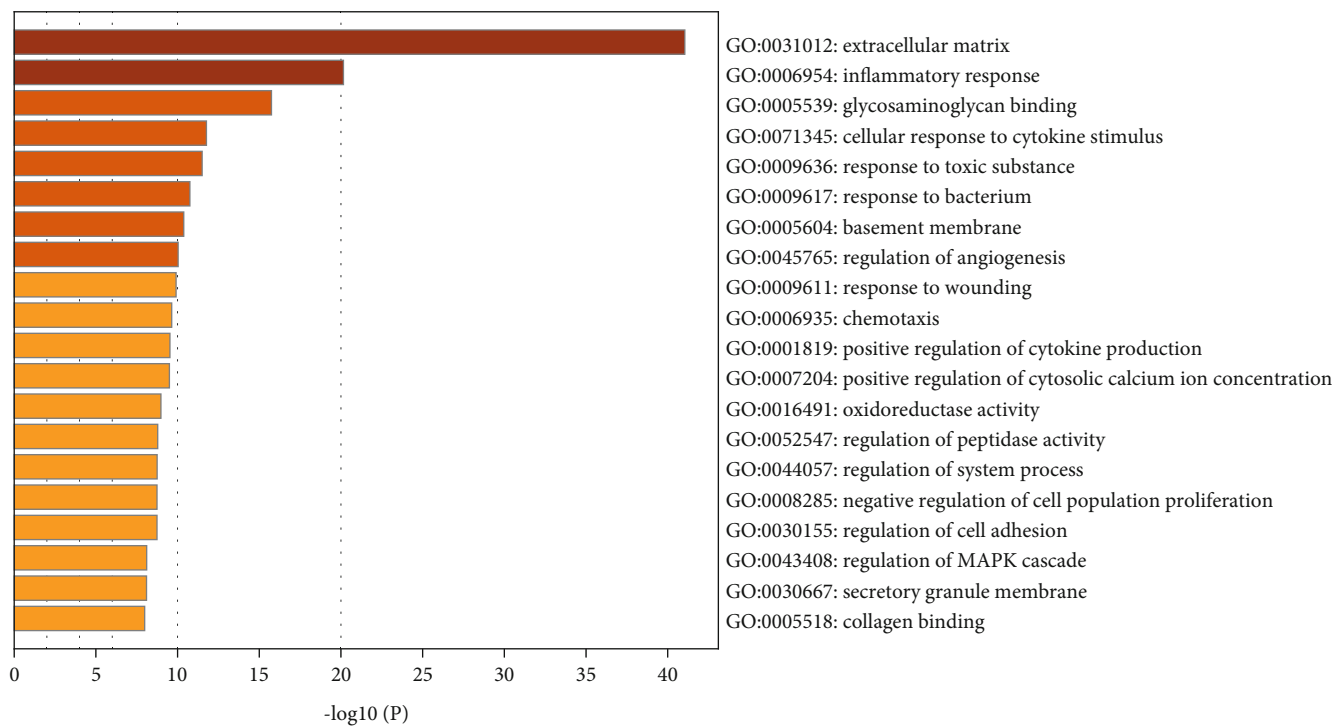
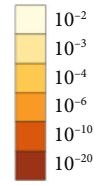
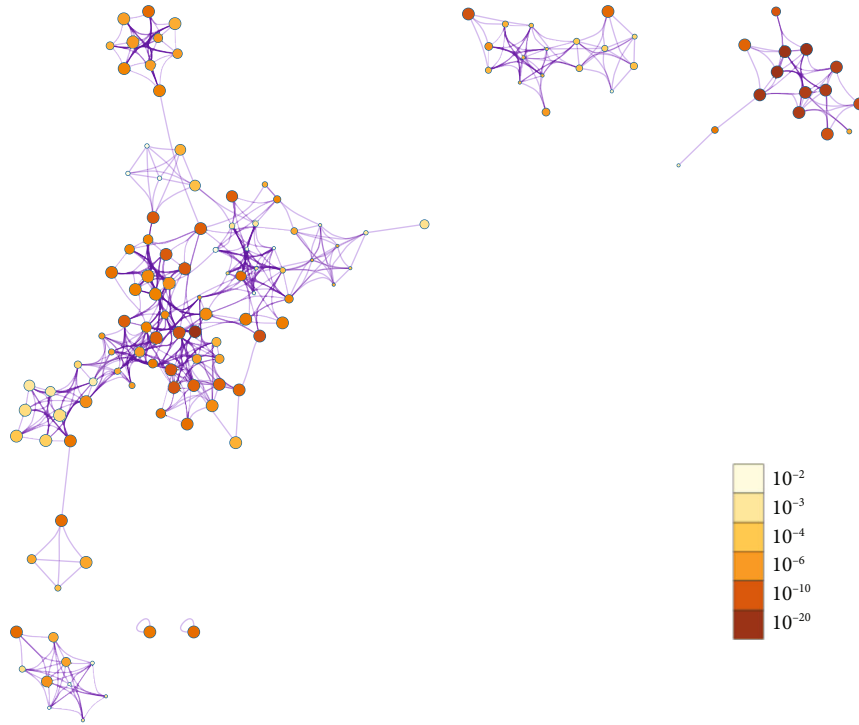
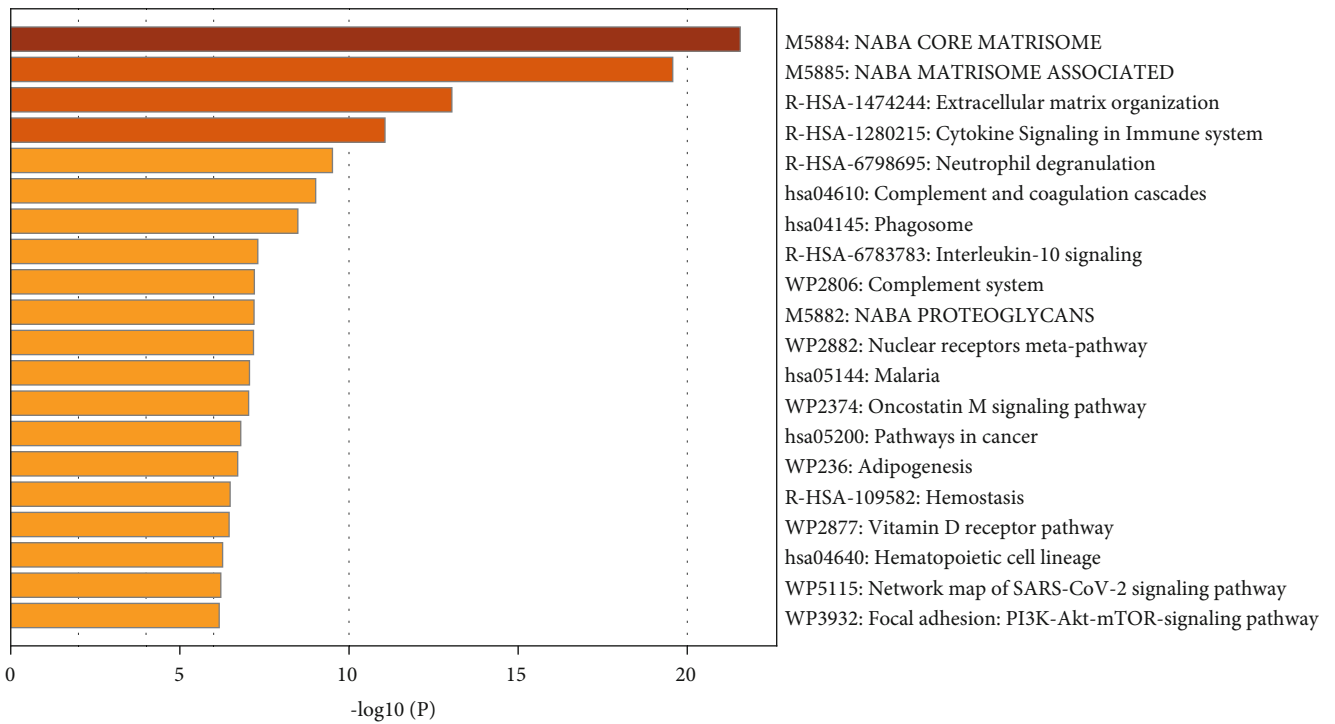


FIGURE 3: Continued.



(c)



(d)

FIGURE 3: Continued.

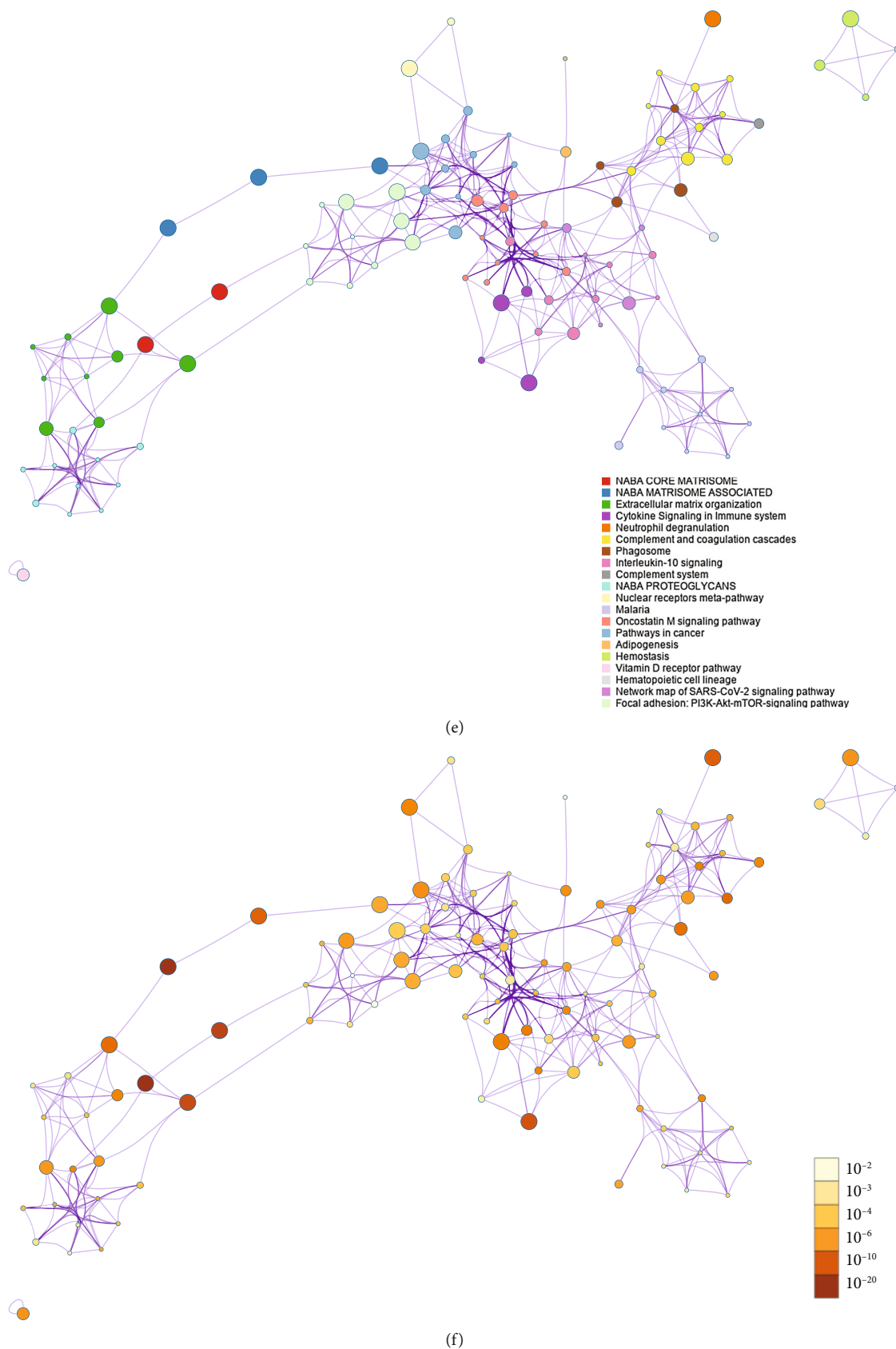
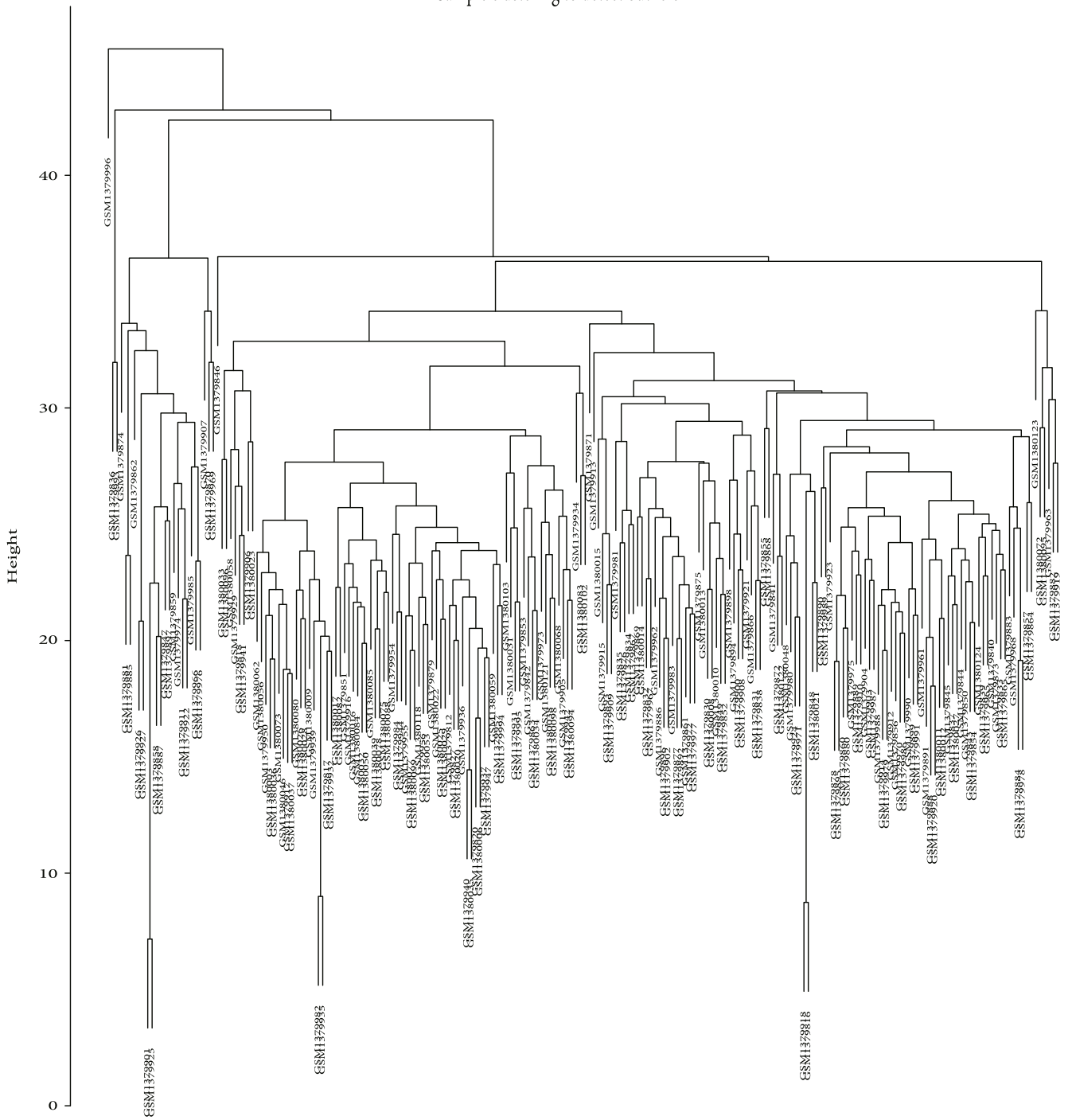


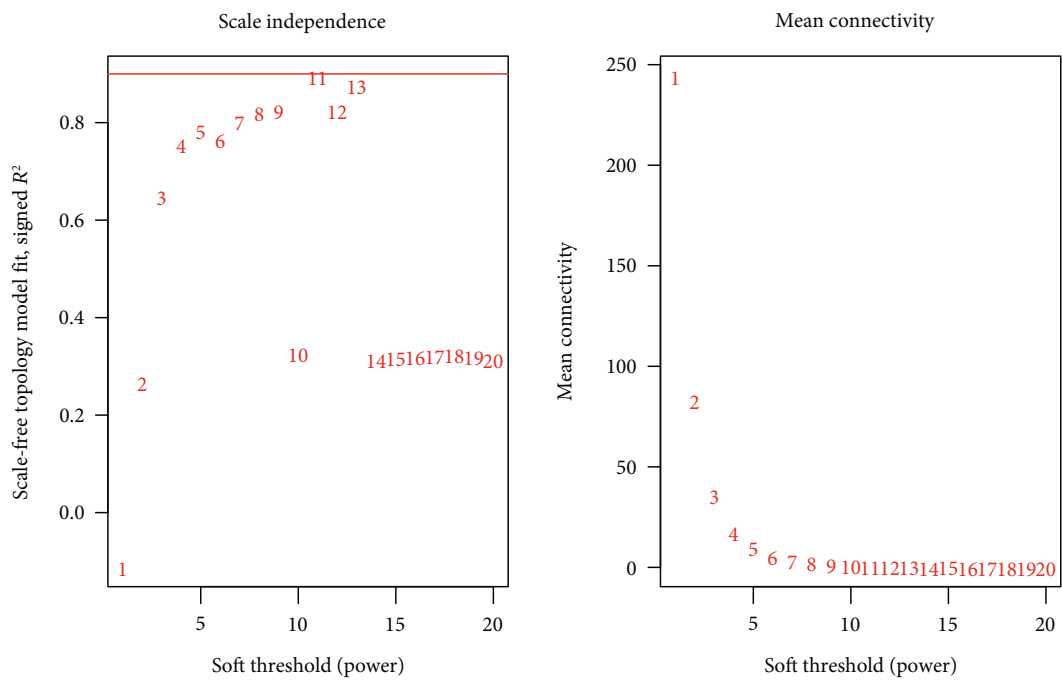
FIGURE 3: Functional enrichment analyses. (a) GO annotation. Construction of the GO functional annotation network based on the cluster ID (b) and *P*-value (c). (d) Related pathway analysis. Construction of the enrichment analysis network in accordance with the cluster ID (e) and *P*-value (f). GO: gene ontology.

Sample clustering to detect outliers

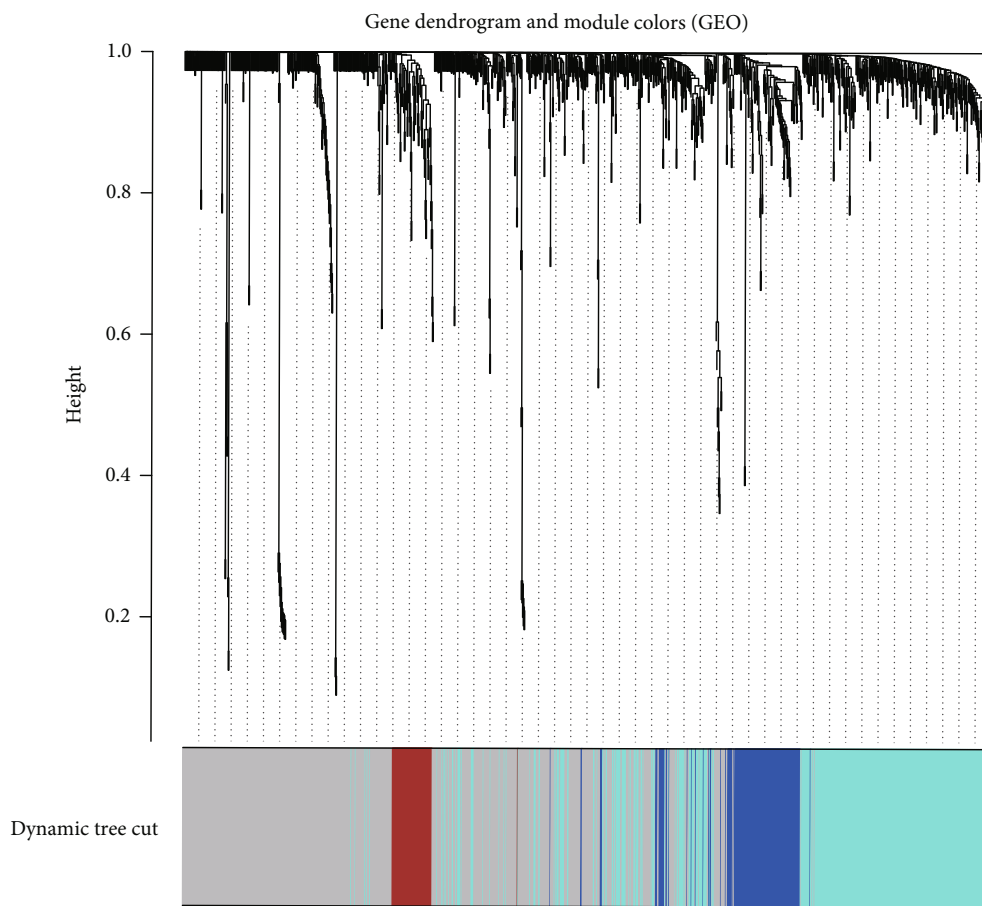


(a)

FIGURE 4: Continued.

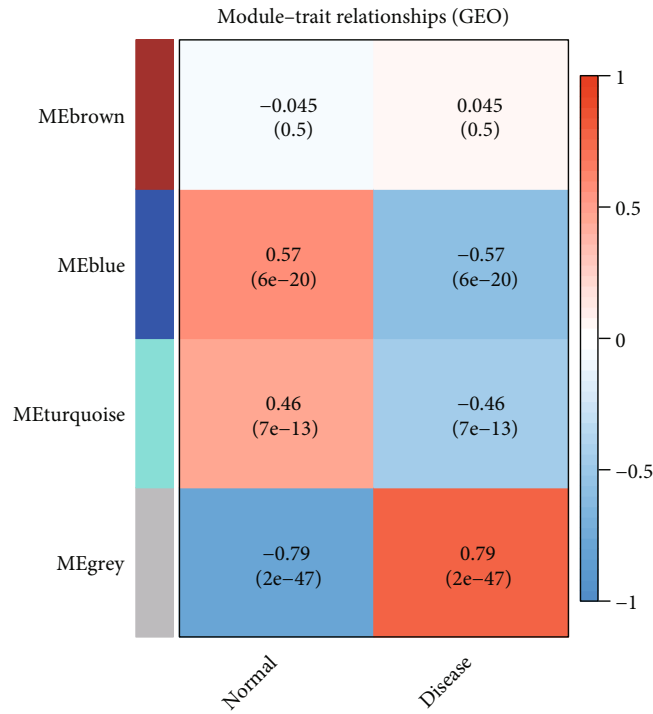


(b)



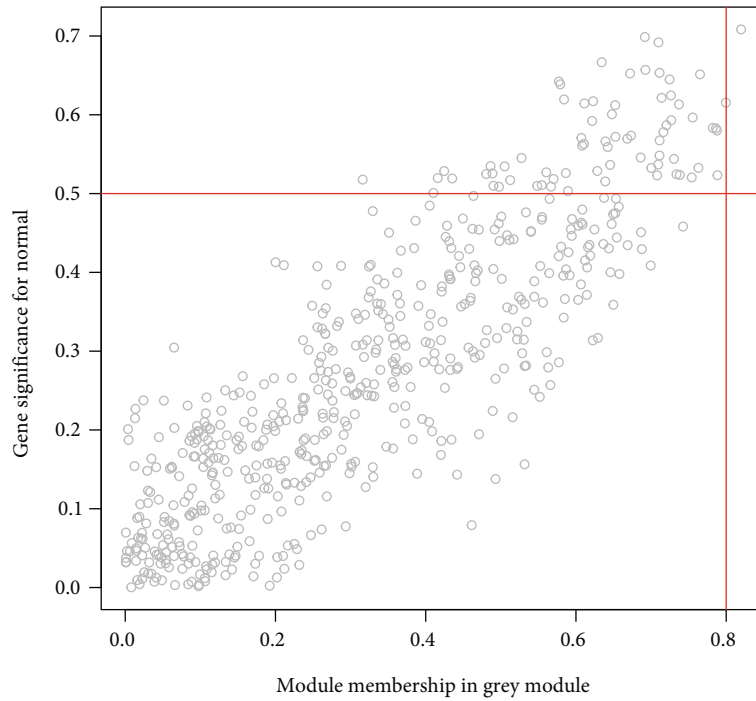
(c)

FIGURE 4: Continued.



(d)

Module membership vs. gene significance
cor = 0.86, p = 3.8e-170



(e)

FIGURE 4: Continued.

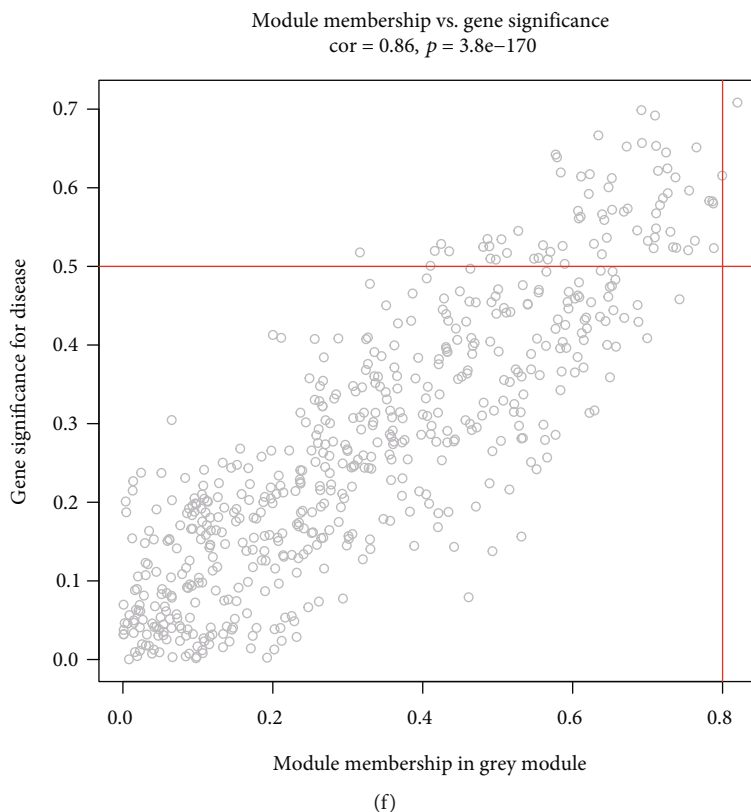


FIGURE 4: WGCNA of the GSE57338 dataset. (a) Clustering of samples for the detection of an outlier. (b) The left and right panels represent scale-free index and mean connectivity analyses, respectively, at different threshold powers. (c) Clustering dendrogram of genes and colors corresponding to each module. (d) Module–trait associations. Module membership vs. gene significance of the MEgrey module in the (e) NF and (f) IDCM-HF groups. NF: non-heart failure; IDCM-HF: idiopathic dilated cardiomyopathy-induced heart failure.

tree cutting was implemented for division of the generated modules, after which comparable modules were combined. Key modules were picked according to the correlation between the module members and the significance of genes. Genes from key modules were used for subsequent analysis.

2.4. Identification of Hub Genes in IDCM-HF. The intersecting genes between DEGs and genes in the key modules via WGCNA were identified to obtain candidate genes. Then, we implemented LASSO and SVM-RFE by adopting the GLMnet [16] and e1071 [17] packages in R, respectively, to further identify hub genes, which were the genes identified by both LASSO and SVM-RFE.

2.5. Hub Genes Verification. Hub genes were verified by evaluating the diagnostic efficacy using receiver operating characteristic (ROC) curves and appraising the expression profile in the training group (GSE57338) and the verification group (GSE5406). Firstly, we plotted ROC curves using pROC packages [18] and determined the area under ROC curves (AUC). An AUC value more than 0.8 signified that the data was a good fit for the gene, and a P -value less than 0.05 indicated that the value was of statistical significance. Secondly, we confirmed the difference in hub gene expression in the IDCM-HF and NF groups. The validated

genes were identified as robust diagnostic biomarkers for IDCM-HF.

3. Results

3.1. Identification of DEGs. The GSE57338 dataset downloaded from GEO contains 20,134 genes in 218 left ventricular myocardia from 82 IDCM-HF patients and 136 NF controls. The DEGs were identified between IDCM-HF and NF samples in the GSE57338 dataset with selected thresholds of $|\log_2 FC| > 0.5$ and adjusted p -value < 0.05 . There were totally 490 DEGs identified based on the criteria mentioned in the Materials and Methods section. The heat map shows the expression profile of the top 50 differential genes (Figure 2(a)). 269 and 221 genes of DEGs exhibited significantly higher and lower expression, respectively, in the IDCM-HF patients compared with the controls, as displayed by the volcano map (Figure 2(b)).

3.2. DO Enrichment Analysis. We performed DO enrichment analysis of DEGs and found that their expression was implicated with the occurrence of heart diseases such as atherosclerosis, atherosclerotic heart disease, myocardial infarction, coronary heart disease, and congestive HF (Figures 2(c) and 2(d)).

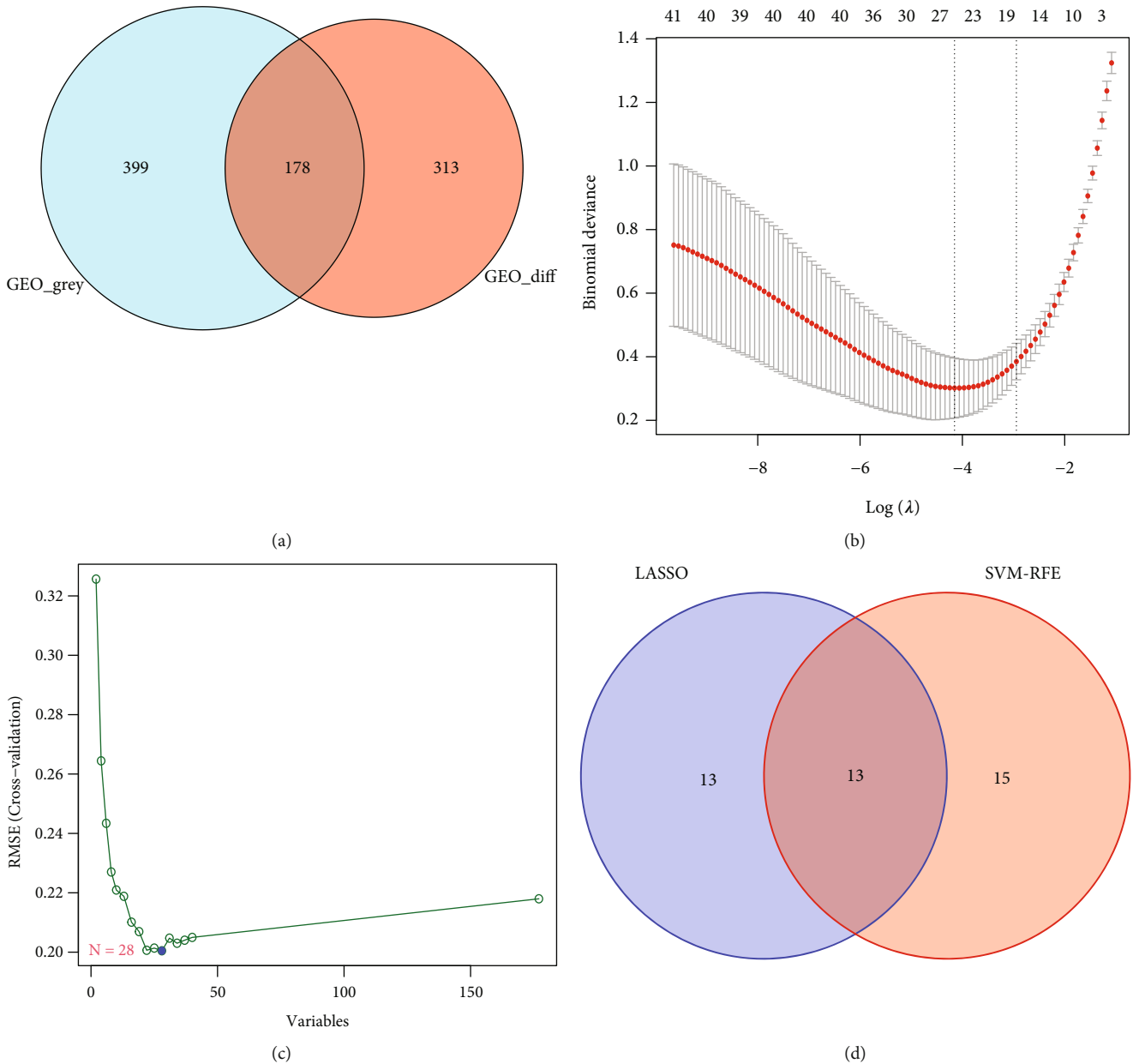
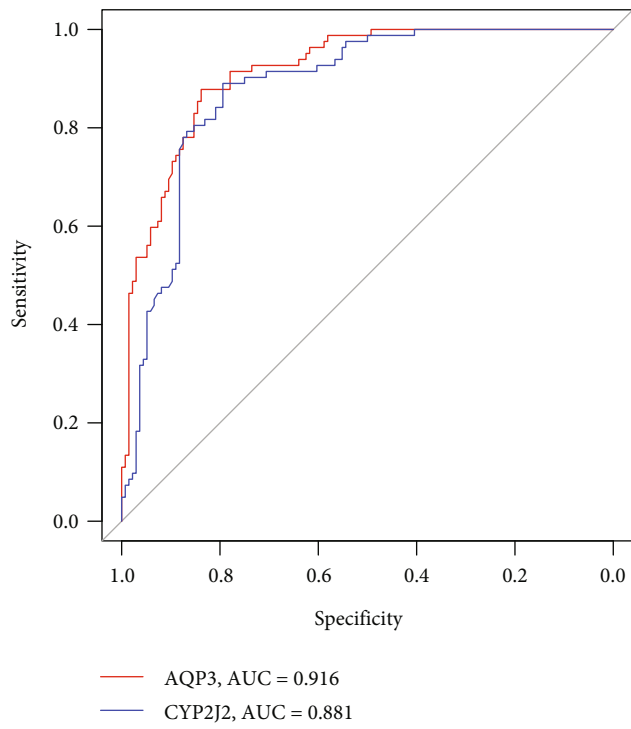


FIGURE 5: Identification of candidate genes. (a) Venn diagram showing the 178 intersecting genes between DEGs and genes in MEBrown. (b) LASSO algorithm for the screening of candidate feature genes. (c) SVM-RFE algorithm for the screening of candidate genes. (d) Venn diagram of 13 overlapping candidate genes among the genes identified by the SVM-RFE and LASSO algorithms. DEG: differentially expressed gene; SVM-RFE: support vector machine-recursive feature elimination; LASSO: least absolute shrinkage and selection operator.

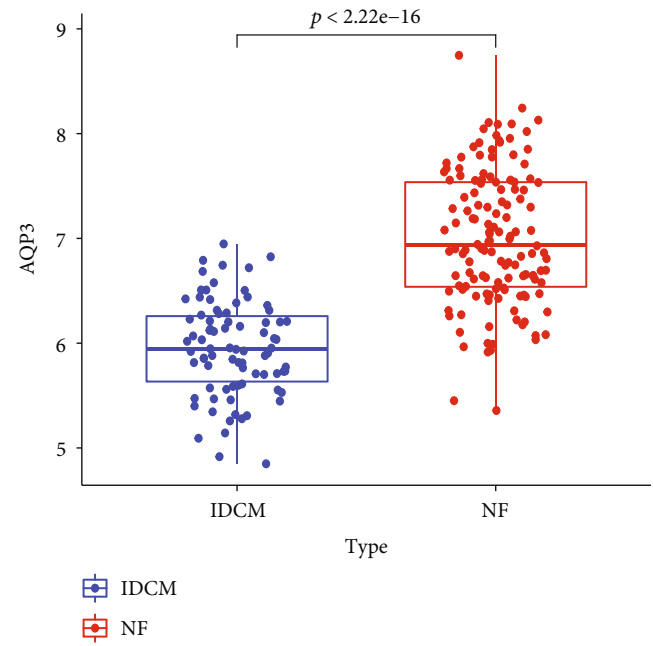
3.3. *Functional Enrichment Analyses.* To further explore the DEGs' biofunctions, GO annotation and pathway analysis were performed. As shown in Figures 3(a), 3(b), and 3(c), according to GO annotation, most of the DEGs were associated with the extracellular matrix (ECM); they were also associated with processes such as inflammatory response generation, glycosaminoglycan binding, response generation to cytokine stimulation, response generation to toxic substances and bacteria, and regulation of certain processes. Consistent with GO enrichment analysis, the pathways with which DEGs were associated were mainly related to the ECM, NABA_MATRISOME_ASSOCIATED,

NABA_CORE_MATRISOME, and ECM organization-related pathways (Figures 3(d), 3(e), and 3(f)).

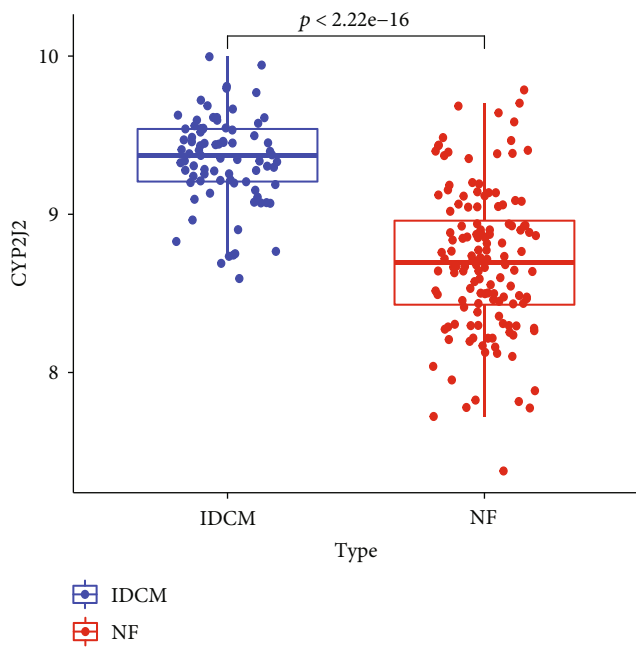
3.4. *Establishment of a Network Based on WGCNA and Identification of Hub Modules.* The detailed process of WGCNA analysis was shown in Additional file 3. After verifying the missing values, we constructed a sample clustering tree (Figure 4(a)) and selected 11 as the optimal soft threshold power. The R^2 value was 0.896. Then, the network topology was analyzed with 1–20 threshold weights (Figure 4(b)). After merging similar modules, a total of four modules were identified (Figure 4(c)). Based on the



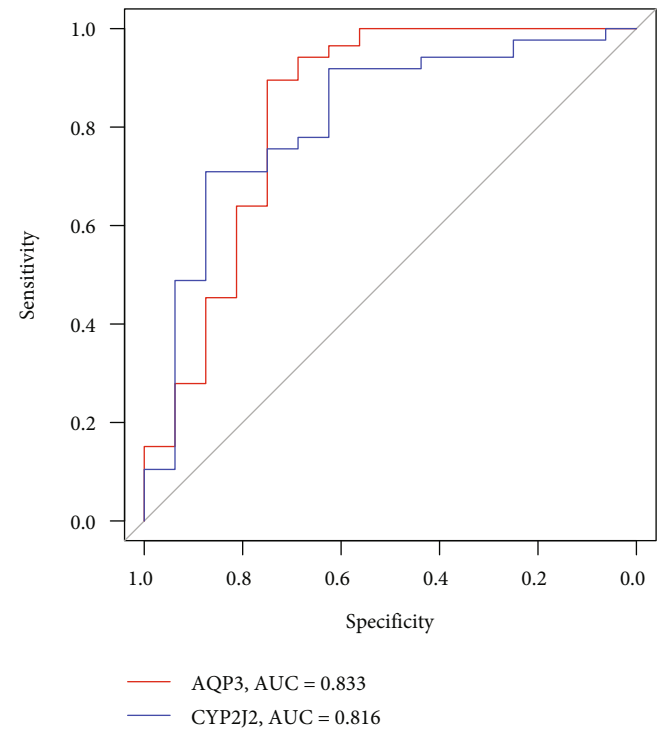
(a)



(b)



(c)



(d)

FIGURE 6: Continued.

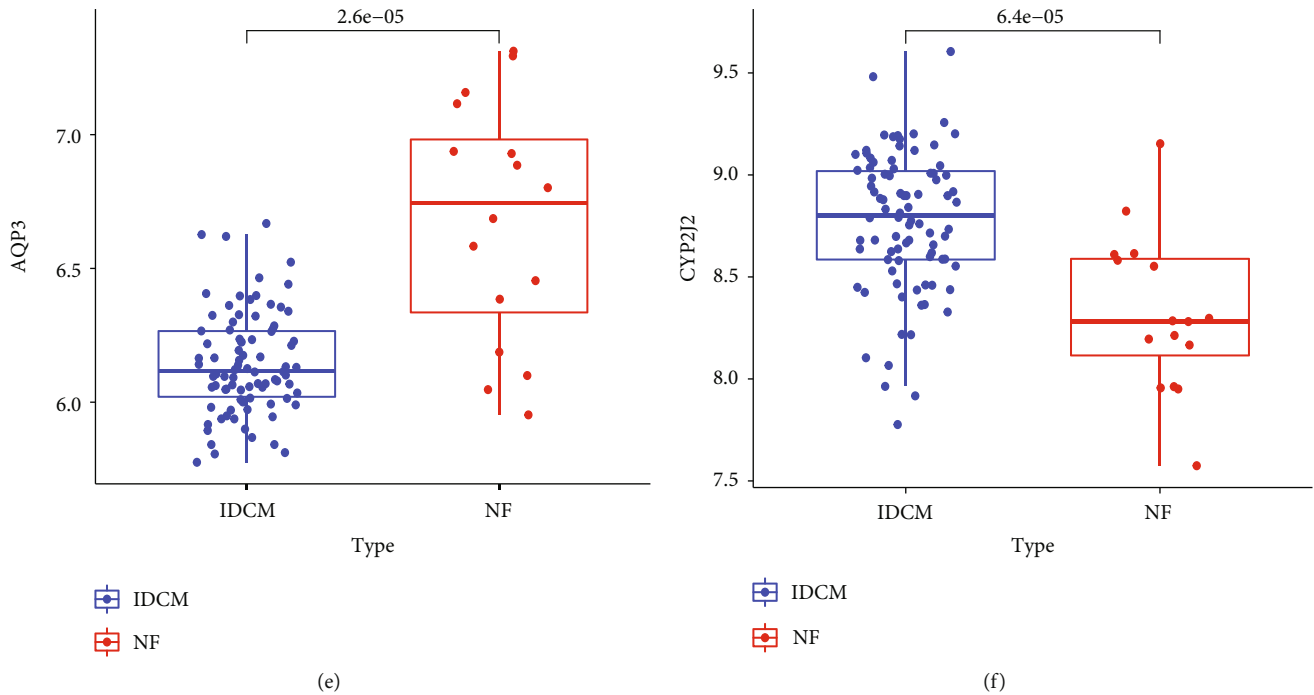


FIGURE 6: Verification of hub genes. (a) ROC curves of the hub genes in the GSE57338 dataset. Validation of expression levels of AQP3 (b) and CYP2J2 (c) in GSE53778. (d) ROC curves of the hub genes in GSE5406. Validation of expression levels of AQP3 (e) and CYP2J2 (f) in GSE5406. ROC: receiver operating characteristics.

module–trait associations (Figure 4(d)), the MEgrey module was picked as a key module. The scatter diagram (Figures 4(e) and 4(f)) shows the significance of the module membership versus gene in this module. In the NF and IDCM groups, the correlation coefficient was 0.86. The genes in MEgrey were subjected to further investigations.

3.5. Identification of Hub Genes. We intersected the DEGs and genes from the MEgrey module identified using WGCNA and obtained 178 intersected genes (Figure 5(a)). We conducted LASSO and SVM-REF screening in order to further identify the hub genes. Through LASSO screening (Figure 5(b)), we obtained 26 genes (Table 1). Through SVM-REF screening (Figure 5(c)), we obtained 28 genes (Table 1). After the intersection of the genes screened using LASSO and SVM-REF, 13 hub genes were obtained (Figure 5(d), Table 1); these included *LCN6*, *SLC16A9*, *AQP3*, *C16orf89*, *USP13*, *EDNRB*, *CYP2J2*, *NT5E*, *CREB5*, *MPP3*, *MID1IP1*, *SLCO2A1*, and *DHCR24*.

3.6. Verification of Hub Genes. We further evaluated the diagnostic efficacy of these 13 genes using ROC curves in the GSE57338 dataset. As exhibited by Figure 6(a) and Additional file 1, these genes were associated with high levels of accuracy. The AUC was greater than 0.8 for all genes except *DHCR24*, for which the AUC was 0.75. Meanwhile, we appraised the expression profile of the 13 genes. As shown in Figures 6(b) and 6(c), and Additional file 2, these genes displayed significantly differential expression between the NF and IDCM groups. In order to further appraise the diagnostic significance of candidate genes and obtain reliable diagnostic markers, we conducted validation with the external dataset

GSE5406. In the GSE5406 dataset, two genes (*AQP3* and *CYP2J2*) intersected with 13 target genes, and the results obtained for *AQP3* and *CYP2J2* were consistent with those obtained with GSE57338, and the corresponding AUC was 0.833 and 0.816, respectively (Figure 6(c)). In comparison with the NF group, *AQP3* expression was down-regulated and *CYP2J2* expression was up-regulated in the IDCM-HF group, and the differences were statistically significant (Figures 6(d), 6(e), and 6(f)).

4. Discussion

IDCM is the most prevalent cause of non-ischemic HF. In most cases, the mechanisms of action are unknown, and the main pathologic disorders involve patchy interstitial fibrosis, myocardial cell degeneration, and ventricular dilatation [19]. Myocardial injury, as well as genetic and environmental factors, might be the pathogenic mechanism of IDCM [20]. The heterogeneity of etiology of the disease makes it difficult to achieve a rapid diagnosis, but the use of high-throughput molecular biology techniques might represent a new solution. We used a combination of WGCNA, LASSO, and SVM-RFF for the first time, to screen the target genes associated with human IDCM-HF and performed external dataset validation. Our findings could shed new lights for diagnosing and managing IDCM-HF patients.

The authors of the GSE 57338 identified genes with distinct expression patterns between failing and non-failing hearts. The authors of the GSE5406 determined HF genes differentially expressed and found that a discrete set of cardiac transcription factors (TFs) was associated with human HF. In this bioinformatics analysis, based on the results of

TABLE 2: The list of genes obtained using LASSO, SVM-REF, and the intersection of LASSO and SVM-REF.

No.	LASSO genes	SVM-REF genes	Inter genes
1	AQP3	AQP3	AQP3
2	C16orf89	C16orf89	C16orf89
3	CREB5	CREB5	CREB5
4	CYP2J2	CYP2J2	CYP2J2
5	DHCR24	DHCR24	DHCR24
6	DUSP13	DUSP13	DUSP13
7	EDNRB	EDNRB	EDNRB
8	MID1IP1	MID1IP1	MID1IP1
9	MPP3	MPP3	MPP3
10	NT5E	NT5E	NT5E
11	SLC16A9	SLC16A9	SLC16A9
12	SLCO2A1	SLCO2A1	SLCO2A1
13	LCN6	LCN6	LCN6
14	FGF7	LINC00670	
15	FLJ30064	LRRC17	
16	GNMT	MME	
17	GPAT3	ABCG2	
18	LIPH	ADAMTS15	
19	NQO1	ARRDC3	
20	SLC7A8	C1orf105	
21	TNNT1	MYH6	
22	CPNE5	NAP1L3	
23	BEX1	NPTX2	
24	CCL5	PHLDA1	
25	DIO2	PI16	
26	DPT	SCUBE2	
27		FAM46B	
28		FCGBP	

SVM-RFE: support vector machine-recursive feature elimination; LASSO: least absolute shrinkage and selection operator.

gene microarray of GSE57338, we used WGCNA and machine learning algorithms to identify biomarkers associated with HF caused by IDCM and further verified the diagnostic efficacy and the expression profile of the hub genes in the dataset GSE 5406. The validated hub genes could shed new lights for diagnosing and managing IDCM-HF patients.

There were 490 DEGs in the IDCM-HF and NF control groups that were associated with atherosclerosis and atherosclerotic and coronary heart diseases. This is consistent with the findings of the following published studies. Recent studies suggest that the coronary atherosclerotic load was a significant factor predictive of major adverse cardiovascular events in non-ischemic DCM patients [21]. Studies show that the heart function in patients exhibiting a DCM-based deteriorating mechanism was likely to be associated with myocardial perfusion inadequacy and microvascular level of coronary artery disease [22], and the coronary microcirculation conditions for IDCM were consistent with the extent of progression of HF [23]; around 50% IDCM patients exhibited endothelial dysfunction of the coronary artery [24].

In accordance with the findings reported by Huang et al. [11], we found that most of the DEGs were associated with the extracellular matrix, which is a complicated network of fibrin (mainly types I and III collagen), elastic fiber, glycosaminoglycan, glycoprotein, and adhesive proteoglycan [25]. In case of a cardiac or extracardiac damage, ECM regulation might be crucial for the remodeling and fibrosis of ventricle. Gunja-Smith et al. were the first to demonstrate that an increase in the level of matrix deposition was caused by an increase in the collagen content in DCM. The newly produced collagen cannot form stable crosslinks, resulting in the dilatation of the ventricular wall [26]. In comparison to the healthy heart, the expression of metalloproteinases (MMPs) was up-regulated and that of its inhibitor was down-regulated in the heart of DCM patients. The up-regulation of MMPs could promote ECM degradation [27]. In addition, some studies have found that the increased level of serum tenascin C (TNC) in patients with DCM might aggravate cardiac fibrosis [27]. ECM functions both as a structural scaffold and as a center for the transmission of signals resulting in cascade reactions that are essential to the functioning of cells. In addition, as a repository of growth factors, it could release growth factors that regulate cellular behavior and activate the process of repair after injury, thereby playing an essential role in HF pathogenesis [28]. However, little is known about the role of specific ECM-dependent molecular pathways in the regulation of the ECM, and future studies need to examine this aspect.

AQPs constitute a highly conserved family of membrane channel proteins responsible for transmembrane water and uncharged small molecule transport. There have been reports regarding 13 human AQPs, from AQP0–AQP12, till date [29–33]. AQP3 is an aqua glycerol channel protein that can permeate through solutes such as water, glycerin, urea [34], transport H_2O_2 [35], and ammonia [36]. AQP3 is distributed extensively in the eye, inner ear, heart, lung, kidney, skin, gastrointestinal tract, reproductive tract, adipose tissues, and cartilage tissues [29, 37]. AQP3 is extensively involved in the development of many diseases. It was found to control water balance and urine concentration within the cells in the collecting duct. It also promotes cell migration, proliferation, and re-epithelialization during the healing of skin wounds [38–40]. The reduced expression of AQP3 during sepsis might play a role in immune cell migration [41]. It was also involved in the toll-like receptor 4-induced activation of macrophages during inflammation [30]. In addition, increased levels of AQP3 expression have been found in many cancers, and increased AQP3 expression enhances the proliferation, invasion, and migration of cancer cells and exacerbates the epithelial–mesenchymal transition-mediated cancer progression [42]. At present, there are few studies on the association between AQPs and cardiovascular disease. AQPs 1 and 3 are the most common endothelial AQPs. A loss of AQP1 is associated with endothelial dysfunction and atherosclerotic progression [43]. Increased AQP1 levels could lead to myocardial hypertrophy [44]. In human stomach cancer cells, up-regulation of AQP3 was associated with the increased expression of MMP2 and MMP9 [45]. AQP3 promotes ECM degradation [37]. The

excessive degradation of MMPs could lead to an aneurysm or rupture in the left ventricle (LV), while excessive accumulation could lead to left ventricular stiffness, increased non-compliance, and HF [25]. In this work, we observed that AQP3 displayed lower expression in IDCM-HF patients. It is hypothesized that AQP3 might alleviate myocardial fibrosis and HF via the reduction in the MMP expression level and degradation of the ECM. This would provide a new direction to the research conducted in this field. It is crucial to better understand the activation, gating, and trafficking mechanisms of AQPs to facilitate the development of novel targeted therapies.

We also observed that CYP2J2 displayed a higher expression level in the IDCH-HF group than in the controls. Cytochrome P450 (CYP) is a monooxygenase located on the cell membrane, which can oxidize exogenous and endogenous substances in a pH-dependent manner. CYP2J2 is the only cyclooxygenase found in humans and is mainly expressed in the heart [46–49]. It converts arachidonic acid into epoxyeicosatrienoic acids (EETs), such as 5,6-, 8,9-, 11,12-, and 14,15-EETs. CYP2J2 and EETs have been repeatedly demonstrated to possess protective effects against various cardiovascular diseases. Previous reports have described the function of EETs during cardiac remodeling inhibition in cardiac hypertrophy, fibrosis, angiogenesis, inflammation, and apoptosis [49, 50]. The findings of another study indicated that overexpression of CYP2J2 promoted the production of EET, alleviated cardiac function impairment, and enhanced the fibrotic response in CYP2J2 transgenic mice [51]. Ma et al. found that the overexpression of CYP2J2 alleviated diabetic cardiomyopathy and tumor necrosis factor (TNF)- α -induced heart tissue damage [52]. CYP2J2 expression was up-regulated in the IDCH-HF group. This was thought to inhibit cardiac fibrosis and reduce cardiac remodeling. The mechanism of action of CYP2J2 still needs to be explored further.

Several limitations are associated with this study. First, because specific cardiac tissues had to be used, the sample size was small. Second, only one microarray dataset was analyzed, which inevitably resulted in a deviation. Third, further in vitro and animal experiments need to be conducted to elaborate the mechanisms underlying the regulatory effects of AQP3 and CYP2J2 on IDCM-HF. Fourth, IDCM is a primary disease with hereditary, infectious, and autoimmune etiology, and its pathogenesis needs to be further explored according to the etiology. Fifth, the lack of clinical information in the datasets may lead to population bias.

5. Conclusion

In summary, we used WGCNA and the machine learning algorithm for the first time to identify biomarkers of IDCM-HF and obtained two hub genes, AQP3 and CYP2J2, which have the potential to serve as targets for the diagnosis and management of IDCM-HF.

Data Availability

All the data generated or analyzed during this study are included in this published article.

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors.

Consent

All authors agree to publish the final version of this manuscript.

Disclosure

A preprint has previously been published [53].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

LLP conceptualized and designed the study and helped with manuscript revision. SMY and LLP managed the data and carried out bioinformatics analyses. SMY drafted the manuscript. The final version of this manuscript has been approved by all the authors, who have agreed to be responsible for all aspects of this study, including data accuracy and integrity.

Acknowledgments

The authors would like to thank TopEdit (<http://www.topeditsci.com>) for language polish of the manuscript.

Supplementary Materials

Additional file 1. ROC curves of the 13 hub genes in the GSE57338 dataset. Additional file 2. The differential expression of 13 hub genes in the GSE57338 dataset. Additional file 3. The detailed process of WGCNA analysis. (*Supplementary Materials*)

References

- [1] H.-P. Schultheiss, D. Fairweather, A. L. P. Caforio et al., "Dilated cardiomyopathy," *Nature Reviews. Disease Primers*, vol. 5, no. 1, p. 32, 2019.
- [2] M. Merlo, A. Cannatà, M. Gobbo, D. Stolfo, P. M. Elliott, and G. Sinagra, "Evolving concepts in dilated cardiomyopathy," *European Journal of Heart Failure*, vol. 20, no. 2, pp. 228–239, 2018.
- [3] N. Orphanou, E. Papatheodorou, and A. Anastasakis, "Dilated cardiomyopathy in the era of precision medicine: latest concepts and developments," *Heart Failure Reviews*, vol. 27, no. 4, pp. 1173–1191, 2022.
- [4] A. R. Ednie, A. R. Parrish, M. J. Sonner, and E. S. Bennett, "Reduced hybrid/complex N-glycosylation disrupts cardiac electrical signaling and calcium handling in a model of dilated cardiomyopathy," *Journal of Molecular and Cellular Cardiology*, vol. 132, pp. 13–23, 2019.

- [5] M. Luciani and F. Del Monte, "Insights from second-line treatments for idiopathic dilated cardiomyopathy," *Journal of Cardiovascular Development and Disease*, vol. 4, no. 3, p. 12, 2017.
- [6] R. Pecini, D. V. Møller, C. Torp-Pedersen, C. Hassager, and L. Køber, "Heart failure etiology impacts survival of patients with heart failure," *International Journal of Cardiology*, vol. 149, no. 2, pp. 211–215, 2011.
- [7] M. Hazebroek, R. Dennert, and S. Heymans, "Idiopathic dilated cardiomyopathy: possible triggers and treatment strategies," *Netherlands Heart Journal*, vol. 20, no. 7–8, pp. 332–335, 2012.
- [8] X. Wang, H. Guan, W. Liu et al., "Identification of immune markers in dilated cardiomyopathies with heart failure by integrated weighted gene coexpression network analysis," *Genes*, vol. 13, no. 3, p. 393, 2022.
- [9] Z. Qiu, B. Ye, L. Yin, W. Chen, Y. Xu, and X. Chen, "Downregulation of AC061961.2, LING01-AS1, and RP11-13E1.5 is associated with dilated cardiomyopathy progression," *Journal of Cellular Physiology*, vol. 234, no. 4, pp. 4460–4471, 2019.
- [10] Q. Zhang, M. Fan, X. Cao et al., "Integrated bioinformatics algorithms and experimental validation to explore robust biomarkers and landscape of immune cell infiltration in dilated cardiomyopathy," *Frontiers in Cardiovascular Medicine*, vol. 9, article 809470, 2022.
- [11] H. Huang, B. Luo, B. Wang, Q. Wu, Y. Liang, and Y. He, "Identification of potential gene interactions in heart failure caused by idiopathic dilated cardiomyopathy," *Medical Science Monitor*, vol. 24, pp. 7697–7709, 2018.
- [12] Y. Liu, M. Morley, J. Brandimarto et al., "RNA-Seq identifies novel myocardial gene expression signatures of heart failure," *Genomics*, vol. 105, no. 2, pp. 83–89, 2015.
- [13] S. Hannenhalli, M. E. Putt, J. M. Gilmore et al., "Transcriptional genomics associates FOX transcription factors with human heart failure," *Circulation*, vol. 114, no. 12, pp. 1269–1276, 2006.
- [14] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, p. 1523, 2019.
- [15] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, 2008.
- [16] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.
- [17] W. S. Noble, "What is a support vector machine?," *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [18] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.
- [19] S. Roura, C. Gálvez-Montón, and A. Bayes-Genis, "Umbilical cord blood-derived mesenchymal stem cells: new therapeutic weapons for idiopathic dilated cardiomyopathy?," *International Journal of Cardiology*, vol. 177, no. 3, pp. 809–818, 2014.
- [20] S. Roura and A. Bayes-Genis, "Vascular dysfunction in idiopathic dilated cardiomyopathy," *Nature Reviews. Cardiology*, vol. 6, no. 9, pp. 590–598, 2009.
- [21] M. Canu, L. Margerit, I. Mekhdoul et al., "Prognosis of coronary atherosclerotic burden in non-ischemic dilated cardiomyopathies," *Journal of Clinical Medicine*, vol. 10, no. 10, p. 2183, 2021.
- [22] D. Neglia and A. L'Abbate, "Coronary microvascular dysfunction and idiopathic dilated cardiomyopathy," *Pharmacological Reports*, vol. 57, no. Suppl, pp. 151–155, 2005.
- [23] K. E. Gil, A. Pawlak, M. Frontczak-Baniewicz, R. J. Gil, and A. Nasierowska-Guttmejer, "The proposed new classification of coronary microcirculation as the predictor of the heart failure progression in idiopathic dilated cardiomyopathy," *Cardiovascular Pathology*, vol. 24, no. 6, pp. 351–358, 2015.
- [24] F. Bitar, A. Lerman, M. W. Akhter et al., "Variable response of conductance and resistance coronary arteries to endothelial stimulation in patients with heart failure due to nonischemic dilated cardiomyopathy," *Journal of Cardiovascular Pharmacology and Therapeutics*, vol. 11, no. 3, pp. 197–202, 2006.
- [25] M. L. Lindsey, M. Jung, M. E. Hall, and K. Y. DeLeon-Pennell, "Proteomic analysis of the cardiac extracellular matrix: clinical research applications," *Expert Review of Proteomics*, vol. 15, no. 2, pp. 105–112, 2018.
- [26] Z. Gunja-Smith, A. R. Morales, R. Romanelli, and J. F. Woessner Jr., "Remodeling of human myocardial collagen in idiopathic dilated cardiomyopathy. Role of metalloproteinases and pyridinoline cross-links," *The American Journal of Pathology*, vol. 148, no. 5, pp. 1639–1648, 1996.
- [27] L. Louzao-Martinez, A. Vink, M. Harakalova, F. W. Asselbergs, M. C. Verhaar, and C. Cheng, "Characteristic adaptations of the extracellular matrix in dilated cardiomyopathy," *International Journal of Cardiology*, vol. 220, pp. 634–646, 2016.
- [28] N. G. Frangogiannis, "The extracellular matrix in ischemic and nonischemic heart failure," *Circulation Research*, vol. 125, no. 1, pp. 117–146, 2019.
- [29] C. Li and W. Wang, "Molecular biology of aquaporins," *Advances in Experimental Medicine and Biology*, vol. 969, pp. 1–34, 2017.
- [30] I. V. da Silva, C. Cardoso, H. Martínez-Banaclocha, A. Casini, P. Pelegrín, and G. Soveral, "Aquaporin-3 is involved in NLRP3-inflammasome activation contributing to the setting of inflammatory response," *Cellular and Molecular Life Sciences*, vol. 78, no. 6, pp. 3073–3085, 2021.
- [31] M. M. Salman, P. Kitchen, A. J. Yool, and R. M. Bill, "Recent breakthroughs and future directions in drugging aquaporins," *Trends in Pharmacological Sciences*, vol. 43, no. 1, pp. 30–42, 2022.
- [32] J. M. Carbrey and P. Agre, "Discovery of the aquaporins and development of the field," *Handbook of Experimental Pharmacology*, vol. 190, pp. 3–28, 2009.
- [33] L. S. King, D. Kozono, and P. Agre, "From structure to disease: the evolving tale of aquaporin biology," *Nature Reviews. Molecular Cell Biology*, vol. 5, no. 9, pp. 687–698, 2004.
- [34] K. Ishibashi, S. Sasaki, K. Fushimi et al., "Molecular cloning and expression of a member of the aquaporin family with permeability to glycerol and urea in addition to water expressed at the basolateral membrane of kidney collecting duct cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 14, pp. 6269–6273, 1994.
- [35] E. W. Miller, B. C. Dickinson, and C. J. Chang, "Aquaporin-3 mediates hydrogen peroxide uptake to regulate downstream intracellular signaling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 36, pp. 15681–15686, 2010.
- [36] A. Tesse, E. Grossini, G. Tamma et al., "Aquaporins as targets of dietary bioactive phytochemicals," *Frontiers in Molecular Biosciences*, vol. 5, p. 30, 2018.

- [37] K. Wagner, L. Unger, M. M. Salman, P. Kitchen, R. M. Bill, and A. J. Yool, "Signaling mechanisms and pharmacological modulators governing diverse aquaporin functions in human health and disease," *International Journal of Molecular Sciences*, vol. 23, no. 3, p. 1388, 2022.
- [38] M. Hara-Chikuma and A. S. Verkman, "Aquaporin-3 facilitates epidermal cell migration and proliferation during wound healing," *Journal of Molecular Medicine (Berlin, Germany)*, vol. 86, no. 2, pp. 221–231, 2008.
- [39] R. Sebastian, E. Chau, P. Fillmore et al., "Epidermal aquaporin-3 is increased in the cutaneous burn wound," *Burns*, vol. 41, no. 4, pp. 843–847, 2015.
- [40] S. Hamed, Y. Ullmann, D. Egozi et al., "Topical erythropoietin treatment accelerates the healing of cutaneous burn wounds in diabetic pigs through an aquaporin-3-dependent mechanism," *Diabetes*, vol. 66, no. 8, pp. 2254–2265, 2017.
- [41] K. Rump and M. Adamzik, "Function of aquaporins in sepsis: a systematic review," *Cell and Bioscience*, vol. 8, no. 1, p. 10, 2018.
- [42] S. Marlar, H. H. Jensen, F. H. Login, and L. N. Nejsum, "Aquaporin-3 in cancer," *International Journal of Molecular Sciences*, vol. 18, no. 10, p. 2106, 2017.
- [43] P. Wintmo, S. H. Johansen, P. B. L. Hansen et al., "The water channel AQP1 is expressed in human atherosclerotic vascular lesions and AQP1 deficiency augments angiotensin II-induced atherosclerosis in mice," *Acta Physiologica*, vol. 220, no. 4, pp. 446–460, 2017.
- [44] V. Montiel, R. Bella, L. Y. M. Michel, H. Esfahani, D. De Mulder, and E. L. Robinson, "Inhibition of aquaporin-1 prevents myocardial remodeling by blocking the transmembrane transport of hydrogen peroxide," *Science Translational Medicine*, vol. 12, no. 564, 2020.
- [45] H. Xu, Y. Xu, W. Zhang, L. Shen, L. Yang, and Z. Xu, "Aquaporin-3 positively regulates matrix metalloproteinases via PI3K/AKT signal pathway in human gastric carcinoma SGC7901 cells," *Journal of Experimental and Clinical Cancer Research*, vol. 30, no. 1, p. 86, 2011.
- [46] S. Wu, C. R. Moomaw, K. B. Tomer, J. R. Falck, and D. C. Zeldin, "Molecular cloning and expression of CYP2J2, a human cytochrome P450 arachidonic acid epoxygenase highly expressed in heart," *The Journal of Biological Chemistry*, vol. 271, no. 7, pp. 3460–3468, 1996.
- [47] E. A. Evangelista, T. Aliwarga, N. Sotoodehnia et al., "CYP2J2 modulates diverse transcriptional programs in adult human cardiomyocytes," *Scientific Reports*, vol. 10, no. 1, p. 5329, 2020.
- [48] A. Das, A. T. Weigle, W. R. Arnold, J. S. Kim, L. N. Carnevale, and H. C. Huff, "CYP2J2 molecular recognition: a new axis for therapeutic design," *Pharmacology and Therapeutics*, vol. 215, p. 107601, 2020.
- [49] J. Lai and C. Chen, "The role of epoxyeicosatrienoic acids in cardiac remodeling," *Frontiers in Physiology*, vol. 12, p. 642470, 2021.
- [50] X. X. Guan, D. N. Rao, Y. Z. Liu, Y. Zhou, and H. H. Yang, "Epoxyeicosatrienoic acids and fibrosis: recent insights for the novel therapeutic strategies," *International Journal of Molecular Sciences*, vol. 22, no. 19, p. 10714, 2021.
- [51] Z. He, Y. Yang, Z. Wen et al., "CYP2J2 metabolites, epoxyeicosatrienoic acids, attenuate Ang II-induced cardiac fibrotic response by targeting Gα(12/13)," *Journal of Lipid Research*, vol. 58, no. 7, pp. 1338–1353, 2017.
- [52] B. Ma, X. Xiong, C. Chen et al., "Cardiac-specific overexpression of CYP2J2 attenuates diabetic cardiomyopathy in male streptozotocin-induced diabetic mice," *Endocrinology*, vol. 154, no. 8, pp. 2843–2856, 2013.
- [53] M. Y. Sun and L. P. Li, Identification of aquaporin 3 and cytochrome P450 2J2 as potential biomarkers associated with heart failure caused by idiopathic dilated cardiomyopathy, 10 August 2022, PREPRINT (Version 1) available at Research Square.