

## ORIGINAL ARTICLE

# A module of multifactor-mediated dysfunction guides the molecular typing of coronary heart disease

Yuewei Li<sup>1,2\*</sup> | Maohuan Lin<sup>1,2\*</sup> | Kangjie Wang<sup>3</sup> | YaQing Zhan<sup>4</sup> | Wenli Gu<sup>1,2</sup> |  
Guanghao Gao<sup>1,2</sup> | Yuna Huang<sup>1,2</sup> | Yangxin Chen<sup>1,2</sup> | Tucheng Huang<sup>1,2</sup> |  
Jingfeng Wang<sup>1,2</sup> 

<sup>1</sup>Department of Cardiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China

<sup>2</sup>Laboratory of Cardiac Electrophysiology and Arrhythmia in Guangdong Province, Guangzhou, China

<sup>3</sup>Division of Vascular Surgery, the First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

<sup>4</sup>Department of Anesthesiology, the First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

## Correspondence

Jingfeng Wang and Tucheng Huang, Department of Cardiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510120, China. Email: drwjf\_sysu@126.com (J.-F. W.) and huangtch3@mail.sysu.edu.cn (T.-C. H.)

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 81570329, 81600233 and 81770229

## Abstract

**Background:** Coronary atherosclerotic heart disease (CHD) is the most common cardiovascular disease and has become a leading cause of death globally. Various molecular typing methods are available for the diagnosis and treatment of tumors. However, molecular typing results are not routinely used for CHD.

**Methods and Results:** Aiming to uncover the underlying molecular features of different types of CHD, we screened the differentially expressed genes (DEGs) associated with CHD based on the Gene Expression Omnibus (GEO) data and expanded those with the NCBI-gene and OMIM databases to finally obtain 2021 DEGs. The weighted gene co-expression analysis (WGCNA) was performed on the candidate genes, and six distinctive WGCNA modules were identified, two of which were associated with CHD. Moreover, DEGs were mined as key genes for co-expression based on the module network relationship. Furthermore, the differentially expressed miRNAs in CHD and interactions in the database were mined in the GEO data set to build a multifactor regulatory network of key genes for co-expression. Based on the network, the CHD samples were further classified into five clusters and we defined FTH1, HCAR3, RGS2, S100A9, and TYROBP as the top genes of the five subgroups. Finally, the mRNA levels of FTH1, S100A9, and TYROBP were found to be significantly increased, while the expression of HCAR3 was decreased in the blood of CHD patients. We did not detect measurable levels of RGS2.

**Conclusion:** The screened core clusters of genes may be a target for the diagnosis and treatment of CHD as a molecular typing module.

## KEYWORDS

Coronary heart disease (CHD), difference analysis, disease classification, GEO, network mining, WGCNA

\*Yuewei Li and Maohuan Lin share co-first author.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Molecular Genetics & Genomic Medicine* published by Wiley Periodicals LLC.

## 1 | INTRODUCTION

Coronary atherosclerotic heart disease (CHD), also known as coronary heart disease, is caused by coronary atherosclerosis resulting in stenosis or occlusion of the cardiac coronary arteries. This may lead to myocardial ischemia and hypoxia in patients (Arnett et al., 2019; Thygesen et al., 2018). Over the past century, a dramatic increase in coronary heart disease has been seen in developed and developing countries. Diagnosis of CHD relies on clinical manifestation, electrocardiogram data, as well as levels of cardiac enzymes and troponin. The complexity of this diagnosis in an emergency setting necessitates rapid diagnosis, especially in developing countries. When choosing conservative or invasive treatment for CHD, each patient must weigh the risk of invasive treatment complications, such as the possibility of restenosis or occlusion. Therefore, individualized risk stratification and treatment are urgently needed. Molecular typing facilitates rapid screening and diagnosis and may guide further individualized treatment.

In 1999, the National Cancer Institute (NCI) put forward the concept of tumor molecular typing, based on molecular features through comprehensive molecular analysis techniques, to transform the basis of tumor classification from morphology to a new tumor classification system. Perou et al. proposed the molecular typing concept (Perou et al., 2000), for which various molecular typing methods can be used to diagnose and treat cancers. However, molecular typing results are currently not routinely used for CHD.

Therefore, we used bioinformatics analysis to identify the molecular typing of CHD. We validated multifactor-mediated dysfunction, predicted molecular typing, and confirmed in blood samples of patients with CHD.

## 2 | MATERIALS AND METHODS

### 2.1 | Data sets

All data sets we used were downloaded from the Gene Expression Omnibus (GEO submission: GSE66306, GSE64554, GSE59421, GSE28858, and GSE64563). Two mRNA expression profiles (GSE66360 and GSE64554) of human CHD and healthy samples were downloaded from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) (Barrett et al., 2013). The numbers of each data set are depicted in Table 1. To obtain miRNA profiles of human CHD, the GSE59421, GSE28858, and GSE64563 data sets were fetched from the GEO database. The numbers of each data set are shown in Table 2. CHD-related genes were searched and extracted from the NCBI-gene and OMIM databases.

**TABLE 1** GEO database mRNA expression profile data sets

| Data set ID | Platform | Tissue      | CHD | Normal |
|-------------|----------|-------------|-----|--------|
| GSE66360    | GPL570   | Whole Blood | 49  | 50     |
| GSE64554    | GPL6947  | EAT         | 13  | 10     |
|             |          | SAT         | 13  | 10     |

Abbreviations: CHD, coronary atherosclerotic heart disease; EAT, epicardial adipose tissue; SAT, subcutaneous adipose tissue.

**TABLE 2** GEO database miRNA expression profile data sets

| Data set ID | Platform | Tissue | CHD | Normal |
|-------------|----------|--------|-----|--------|
| GSE59421    | GPL10850 | Blood  | 33  | 63     |
| GSE28858    | GPL8179  | Blood  | 12  | 12     |
| GSE64563    | GPL8179  | EAT    | 13  | 10     |
|             |          | SAT    | 13  | 10     |

Abbreviations: CHD, coronary atherosclerotic heart disease; EAT, epicardial adipose tissue; SAT, subcutaneous adipose tissue.

## 2.2 | Methods

### 2.2.1 | Ethical compliance

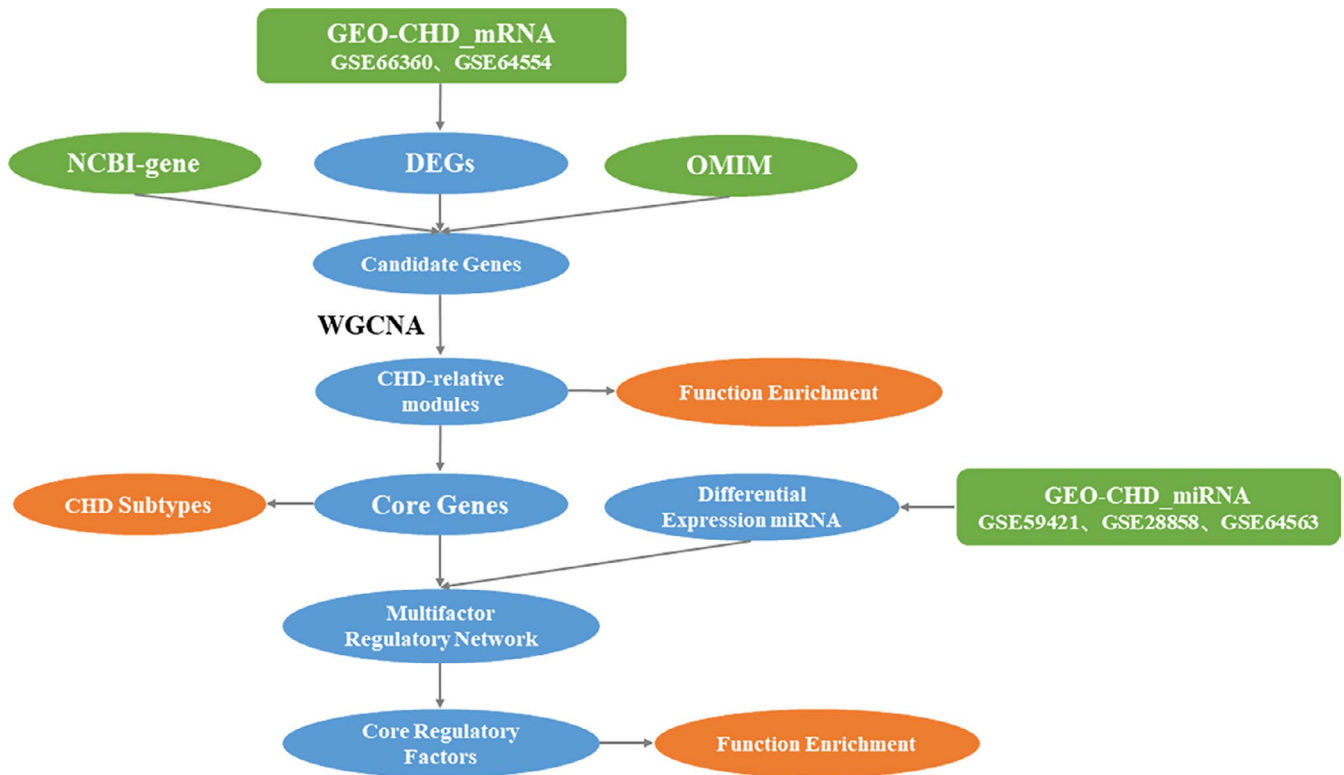
The study protocol was approved by the Ethics Committee of Sun Yat-sen Memorial Hospital (affiliated to Sun Yat-sen University, Guangzhou, China). Written informed consent was obtained from each participant.

### 2.2.2 | Data preprocessing and differential gene analysis

The data processing flowchart is depicted in Figure 1. The downloaded data sets were log<sub>2</sub>-transformed and we quantile-normalized signal intensity. First, after corresponding probes to genes/miRNAs, we removed unloaded probes. Second, we selected the expression median of probes as the expression value of the genes/miRNAs, if more than one probe was mapped to the same gene. Then, we used the R package Limma to perform differential expression analysis of the processed genes/miRNAs. Finally, differentially expressed genes/miRNAs between CHD and control samples were selected by fold changes and significance of differentially expressed analysis.

### 2.2.3 | Candidate gene sets and expression data

Candidate gene sets were mined through the merging of DEGs and CHD-related genes from the GENE and OMIM



**FIGURE 1** Data processing flowchart

databases, and redundancy was removed. Candidate genes also expressed in the GSE66360 data set were selected for further analysis.

## 2.2.4 | Weighted gene co-expression network analysis (WGCNA)

As a systems-biology approach, WGCNA was applied to conduct a scale-free network using genes expression data (Zhang & Horvath, 2005). First, we constructed a similarity matrix based on gene expression data. The absolute value of the Pearson correlation coefficient between two genes was calculated using Formula 1, where  $x_i$  and  $y_j$  represent the gene  $i$  and gene  $j$  expression values, respectively.

Formula 1

$$S_{ij} = \left| \frac{1 + \text{cor}(x_i, y_j)}{2} \right|$$

Next, we converted the similarity matrix of gene expression into an adjacency matrix using Formula 2, with the network type assigned, in which  $\beta$  was the soft threshold. We evaluated the Pearson correlation for each pair of genes to the power of  $\beta$ . As a result, strong correlations were strengthened while weak ones weakened at the exponential level.

Formula 2

$$a_{ij} = \left| \frac{1 + \text{cor}(x_i, y_j)}{2} \right|^\beta$$

Then, we transformed the adjacency matrix into topological overlap measure (TOM) using Formula 3, to depict the degree of correlation between genes.

Formula 3

$$\text{TOM} = \frac{\sum_{u \neq ij} a_{iu} a_{uj} + a_{ij}}{\min(\sum_u a_{iu} + \sum_u a_{ju}) + 1 - a_{ij}}$$

TOM represents the degree of dissimilarity between gene  $i$  and gene  $j$ . With  $1 - \text{TOM}$  used as the distance metric, hierarchical clustering was performed on genes. Then, the dynamic shear tree method was used to identify modules. Module eigenvector genes (ME), the most representative ones, and the first principal component in each module, were calculated by Formula 4, in which  $i$  represents the gene, and  $l$  represents the chip sample in module  $q$ .

Formula 4

$$\text{ME} = \text{princomp}(x_{il}^{(q)})$$

A gene module membership (MM) was evaluated by the Pearson correlation between the gene expression profile in all

samples and the ME expression profile in the module. MM was calculated with Formula 5, where  $x_i$  represents the expression profile of gene  $i$ ,  $ME^q$  represents the ME of module  $q$  and  $ME_i^q$  represents the MM of gene  $i$  in module  $q$ . When  $ME_i^q = 0$ , gene  $i$  was excluded from the module  $q$ . The closer  $ME_i^q$  was to +1 or -1, the more highly correlated gene  $i$  was with module  $q$ . The plus or minus sign indicates whether there was a positive or negative correlation between gene  $i$  and module  $q$ .

Formula 5

$$MM_i^q = \text{cor}(x_i, ME^q)$$

We used gene significance (GS) to assess the association between genes and external information. A higher GS indicates that the gene has a higher biological significance. When  $GS = 0$ , it can be considered that the gene is not involved in the biological process.

The WGCNA was constructed to identify CHD-correlated modules for the selected gene expression data using the WGCNA R package. Gene set enrichment analysis was conducted using the ClusterProfiler R package to evaluate the significant gene sets involved in the CHD-correlated modules using GO and KEGG annotation pathways. A false discovery rate below 0.05 was considered significantly enriched. The interaction subnet was constructed for the co-expression gene sets in the CHD-correlated modules, and the core genes of the module were screened according to the node degree of the network.

### 2.2.5 | Construction of a multi-factor regulatory network

lncRNAs and miRNAs that interacted with core genes were screened out from starBase v2.0 database (<http://starbase.sysu.edu.cn/starbase2/>), which included miRNA-lncRNA and miRNA-mRNA interaction pairs (Li, Liu, Zhou, Qu, & Yang, 2014). Transcription factors (TF) regulating the expression of core genes were screened from the TRRUST v2 database, which included the data of human TF-mRNA

interaction pairs. Based on the identified interaction pairs, a multifactor regulatory network was constructed, including mRNA, TF, lncRNA, and miRNA.

### 2.2.6 | CHD molecular typing mediated by core genes

CHD samples were classified according to the module core gene pairs. The optimal  $K$  value (number of categories) was determined by finding the best sum of squared error (SSE) inflection point. CHD was divided into different subclasses by using the unsupervised clustering method  $K$ -means combined with t-SNE dimension reduction. We examined the expression patterns of these core genes in different subclasses and analyzed the genes with significant differences in different subclasses to find potential marker genes of CHD subclasses.

### 2.2.7 | Blood samples preparation

Blood samples were collected from CHD patients and healthy controls. Five milliliters of peripheral venous blood was collected in an anticoagulant tube with ethylenediaminetetraacetic acid (EDTA). Subsequently, the blood samples were centrifuged in primary blood collection tubes for 10 min at  $1000\times g$  and  $4^\circ\text{C}$  using a swinging bucket rotor. The substrate blood cells were stored at  $-80^\circ\text{C}$  until analysis.

### 2.2.8 | Reverse-transcription polymerase chain reaction

Total RNA of blood cells was extracted using Trizol reagent (Invitrogen). Individual RNA samples were reversely transcribed into cDNA using PrimeScript<sup>®</sup> RT Master Mix (Takara, Japan) according to the manufacturer's instructions. The primers used are listed in Table 3. The relative levels of target gene mRNA transcripts were determined by quantitative real-time polymerase chain reaction (qRT-PCR) using the FastStart Universal SYBR Green Master (Roche Applied

|        | Sense primers (5'–3') | Antisense primers (5'–3') |
|--------|-----------------------|---------------------------|
| FTH1   | GCTTGCGGAATATCTCTT    | AACTGAACAACGGCACTTA       |
| RGS2   | GAATCAGGAAGCCAGTAACT  | TCAACACCATAGCACTCATT      |
| S100A9 | GCTGGAACGCAACATAGA    | CTCCTGATTAGTGGCTGTG       |
| TYROBP | CTGCTGGCTGTAAGTGATT   | GTGTTGAGGTCGCTGTAG        |
| HCAR3  | TGGCGGTAGACAGGTATT    | TGAGCAGAACAGGATGATG       |

**TABLE 3** List of primers used in qRT-PCR

Abbreviations: FTH1, ferritin heavy chain 1; HCAR3: hydroxycarboxylic acid receptor 3; RGS2, regulator of G protein signaling 2; S100A9, S100 calcium binding protein A9; TYROBP, transmembrane immune signaling adaptor TYROBP.

Sciences) and specific primers on a Roche LightCycler 480 Real-Time PCR System.

### 3 | RESULT

#### 3.1 | Differential mRNA gene screening

To obtain the mRNA expression profile of human CHD, we downloaded two data sets, GSE66360 and GSE64554, from the GEO database. The GSE64554 data set was divided into EAT and SAT groups based on tissue locations. Therefore, there were 49\_vs\_50, 13\_vs\_10, and 13\_vs\_10 samples (CHD vs. normal) in the three groups, respectively. The R package Limma was used for differential analysis among the three groups. Differentially-expressed genes (DEGs) were selected based on the fold change ( $|\log\text{FC}| > 0.584962501$ ) and the significance threshold (adj.  $p$ . Val  $< 0.01$ ). We selected 921 DEGs from data set GSE66360, among which 561 DEGs were upregulated while 360 were downregulated, 330 DEGs from GSE64554\_EAT, among which 175 DEGs were upregulated while 173 were downregulated, 407 DEGs from GSE64554\_SAT, among which there were 175 upregulated DEGs and 232 downregulated (Table 4). The volcano diagram of differential genes in each group is shown in Figure 2a. (Partial genes were not concentrated; therefore, only the majority is shown).

#### 3.2 | Candidate gene sets

We downloaded 797 CHD-related genes (CRGs) from the GENE database and 115 CRGs from the OMIM database. By merging DEGs and CRGs from both databases and removing redundancy, 2339 candidate genes were selected. Of those, 2021 genes expressed in GSE66360 were selected for the next WGCNA (Figure 2b).

#### 3.3 | Construction of WGCNA for candidate gene sets

We constructed a weighted co-expression network for candidate gene sets with the R software package WGCNA. The

**TABLE 4** Statistical table of differential expressed genes

| Data set ID  | DEG num | Up  | Down |
|--------------|---------|-----|------|
| GSE66360     | 921     | 561 | 360  |
| GSE64554_EAT | 330     | 175 | 173  |
| GSE64554_SAT | 407     | 175 | 232  |

Note: Differential expressed genes in the Data set of GSE66360, GSE64554\_EAT and GSE64554\_SAT.

Abbreviations: EAT, epicardial adipose tissue; SAT, subcutaneous adipose tissue.

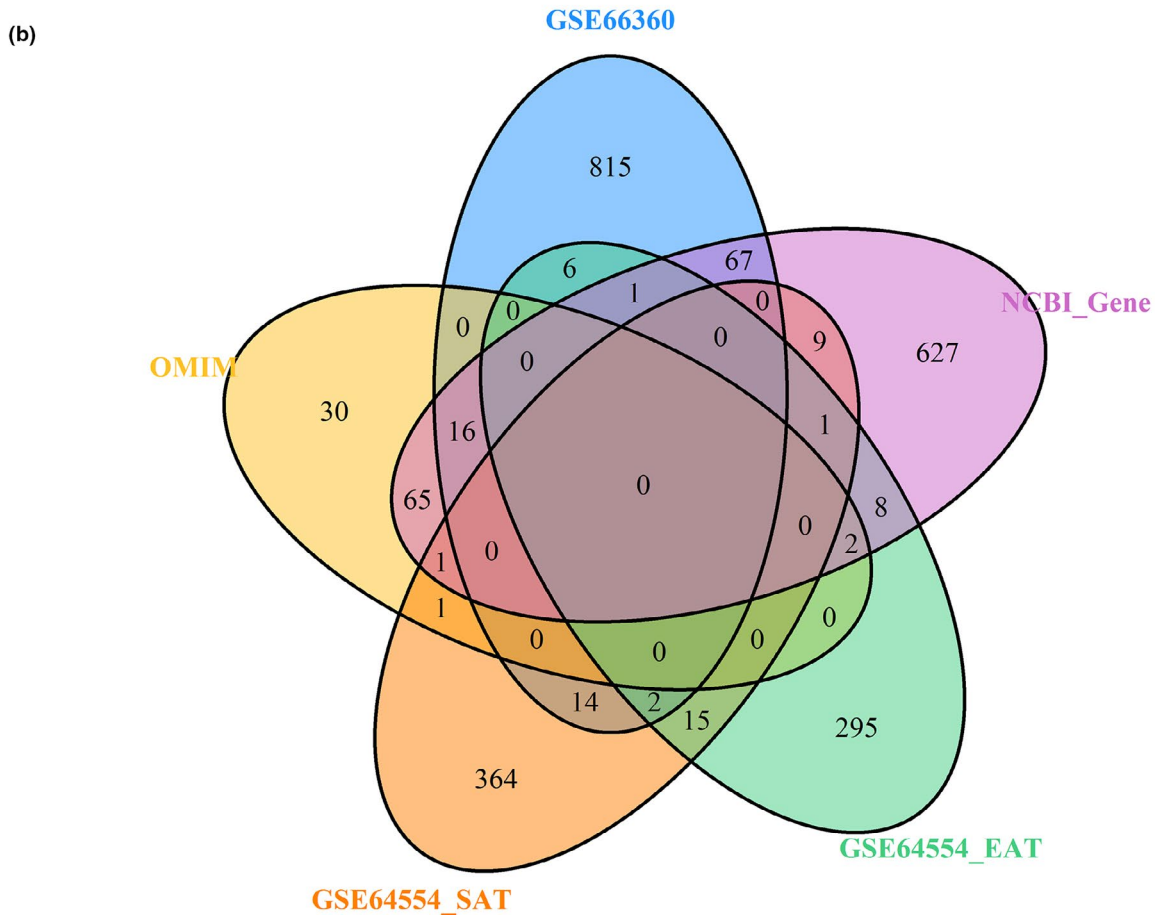
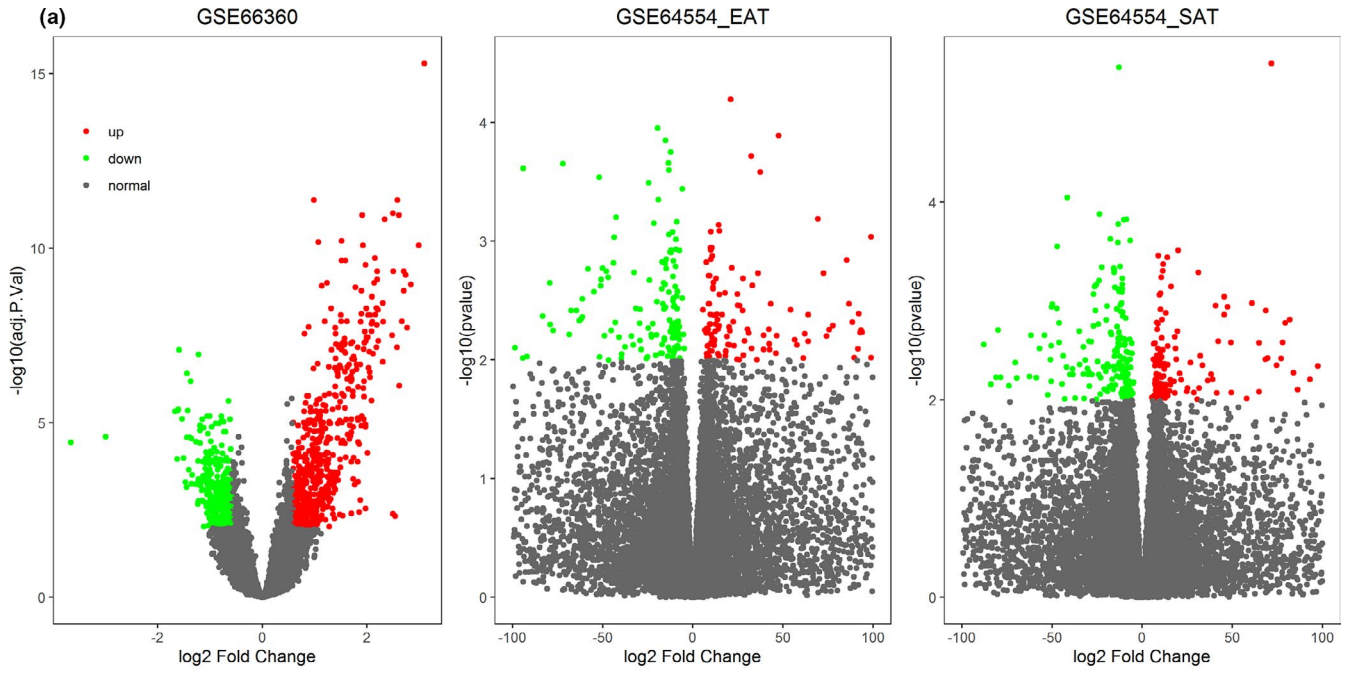
cluster suggested that eight samples were outliers, and the remaining 91 samples were analyzed. This analysis indicated that the network of gene co-expression accorded with the scale-free network. In other words, the logarithm  $\log(k)$  of the node with  $k$  connectivity was negatively correlated with the logarithm  $\log(P(k))$  of the probability of the node, and the correlation coefficient was higher than 0.8. To ensure that the expression network was a scale-free one, we set  $\beta = 6$  (Figure S1). Next, we transformed the expression matrix into an adjacency matrix, which was subsequently transformed into a topological matrix. In the light of TOM, genes were clustered using the average-linkage hierarchical clustering method (Langfelder & Horvath, 2008). The minimum number of genes in each gene network module was set to 30 based on hybrid dynamic shear pruning tree criteria. After using the dynamic shearing method to determine the gene module, the eigengenes value of each module was calculated at once. A clustering analysis was then performed on the modules. Next, we merged modules with close distances. With height set to 0.25, six modules were obtained (Figure 3a). The number of genes in each module is shown in Table 5.

The Pearson correlation coefficient was calculated for the ME of each module and CHD sample. The higher the coefficient, the more important the module was considered. The rows in Figure 3b represent the eigenvector genes of each module, and the correlation coefficients are presented from high to low (red to blue). The numbers in each cell represent the correlation coefficient between the gene module and the CHD sample, while the numbers in brackets represent the  $p$  value of significance. The GS value of each gene module was then calculated. The larger the GS value, the more relevant the module was for the CHD sample. From the results of the above two methods of analyzing the correlation between modules and phenotypes, the modules most relevant to CHD were colored yellow and blue (Figure 3c).

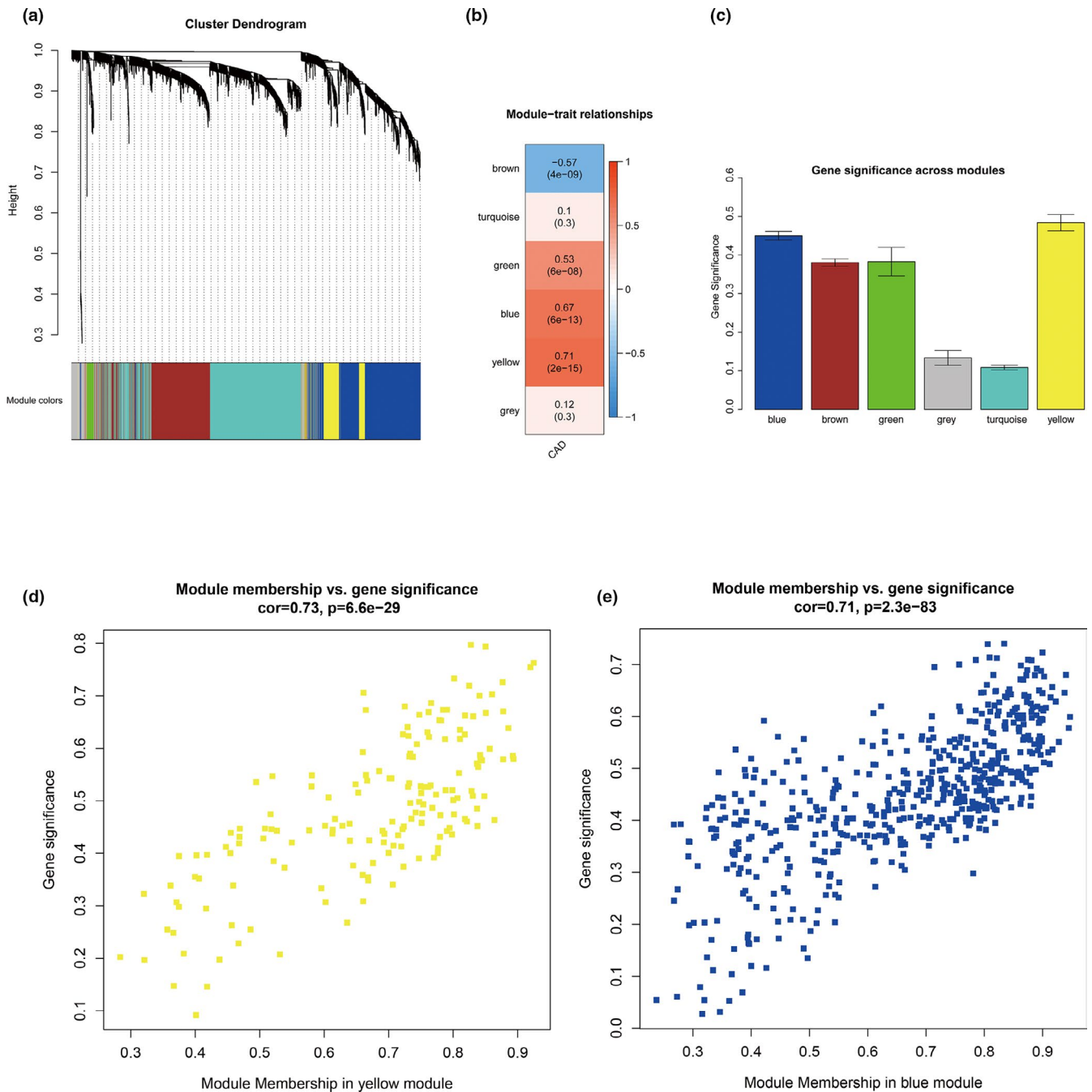
Pathway enrichment analysis was conducted on the yellow gene and blue gene modules (only the top 10 are shown) (Figure S2). Genes in the yellow module were associated with biological processes such as positive regulation of cell activation, muscle cell proliferation, cell activation, and regulation of leukocyte-cell adhesion. In contrast, genes in the blue module were associated with signaling pathways such as fluid shear stress and atherosclerosis, and the NF-kappa B signaling pathway. Moreover, genes in both modules correlated with CHD (Figure 3d,e).

#### 3.4 | Module core gene analysis

To determine the CHD-related core gene, functional analysis of the yellow and blue modules was performed. Based



**FIGURE 2** Candidate gene sets were selected from DEGs. (a) DEGs volcano map mined from the GSE66360 and GSE64554 datasets. (b) Venn diagram of CHD-related genes from different sources



**FIGURE 3** The results of WGCNA. (a) Gene dendrogram and module colors; (b) Module-trait relationships; (c) Gene significance across modules. (d–e) Relationships between module membership and gene significance in CHD.

on the gene expression relationship in the yellow module, a co-expression network diagram was constructed with co-expression pairs with the connection threshold  $\geq 0.13$  as its edge, and the genes with node degree  $\geq 10$  ( $N = 10$ ) as its core genes. Similarly, based on the blue module's gene expression relationships, we created a co-expression network diagram with co-expression pairs with the connection threshold  $\geq 0.25$  as its edge, and the genes with node degree  $\geq 10$  ( $N = 9$ ) as its core genes (Figure 4a). Finally, 19 genes were determined as co-expression key genes.

### 3.5 | Differential miRNA screening

miRNA data of the GSE59421, GSE28858, and GSE64563 CHD and control samples were fetched from the GEO database. The GSE64563 data set was divided into EAT and SAT groups based on tissue locations. There were 33\_vs\_63, 12\_vs\_12, 13\_vs\_10, and 13\_vs\_10 samples (CHD vs. normal) in the four groups. The R package Limma was used for analysis between the four groups, among which DEGs were selected based on the difference multiple ( $\log_2\text{FCI} > 0.2$ ) and

**TABLE 5** Statistical results of corresponding genes for each module

| Module    | No. of genes |
|-----------|--------------|
| Blue      | 536          |
| Brown     | 448          |
| Green     | 41           |
| Grey      | 119          |
| Turquoise | 711          |
| Yellow    | 166          |

Note: Six modules were defined by using weighted gene co-expression network analysis. There were 536 genes in blue modules, 448 genes in brown modules, 41 genes in green modules, 119 genes in grey modules, 711 genes in turquoise modules, 166 genes in yellow modules, respectively.

significance threshold (adj.  $p$ . Val <0.05). We selected 38 differentially-expressed miRNAs from dataset GSE59421, among which 9 miRNAs were upregulated and 19 were downregulated, 138 differentially expressed miRNAs from data set GSE28858, among which 74 miRNAs were upregulated and 64 were downregulated, 47 differentially expressed miRNAs from GSE64563\_EAT, among which 17 miRNAs were upregulated and 30 were downregulated, 28 differentially expressed miRNAs from GSE64563\_SAT, among which 12 were upregulated and 16 were downregulated (Table 6). Differential genes from each group were displayed in a volcano diagram (Figure S3a). (Partial miRNAs were not concentrated, so only the majority is shown). We acquired 227 candidate miRNAs by merging the differentially expressed miRNAs of the four groups (Figure S3b).

### 3.6 | Construction of multifactor regulatory network

From the starBase v2.0 database, the number of supporting experiments  $\geq$ high stringency ( $\geq 3$ ), number of cancer types (pan-cancer)  $\geq 1$  cancer type, miRNA-lncRNA interactions were screened, and 376 miRNA-lncRNAs were obtained. In addition to the above two conditions, the screening of miRNA-mRNA interactions from starBase v2.0 also met the condition that the interaction pairs could be identified by at least one of these software packages: targetScan, picTar, RNA22, PITA, or miRanda/mirSVR. Finally, 160,774 miRNA-mRNA interactions were obtained. Human TF-mRNA regulatory relationships were obtained from the TRRUST v2 database, resulting in 9396 TF-mRNA interactions.

Firstly, TF interacting with the 19 co-expressed key genes were selected. Next, five miRNAs, which interacted with co-expressed key genes and belonged to candidate miRNA, were selected as the final miRNA. Then, we screened for lncRNA interacting with these five miRNAs. Finally, a multifactor

regulatory network containing 63 interaction pairs was obtained (Figure 4b). The degree of each regulator (miRNA, lncRNA, TF) was calculated, and the top 10 regulators were defined as core regulators (degree >1) (Table 7). To explore the functions and signaling pathways of these core TFs, 5,706 target genes interplaying with core TFs were selected from miRNA-lncRNA, miRNA-mRNA, and TF-mRNA interactions. Pathway enrichment analysis on target genes revealed that these genes were related to the TNF signaling pathway and protein processing in the endoplasmic reticulum (Figure 4c).

### 3.7 | CHD clustering based on dysfunction module genes

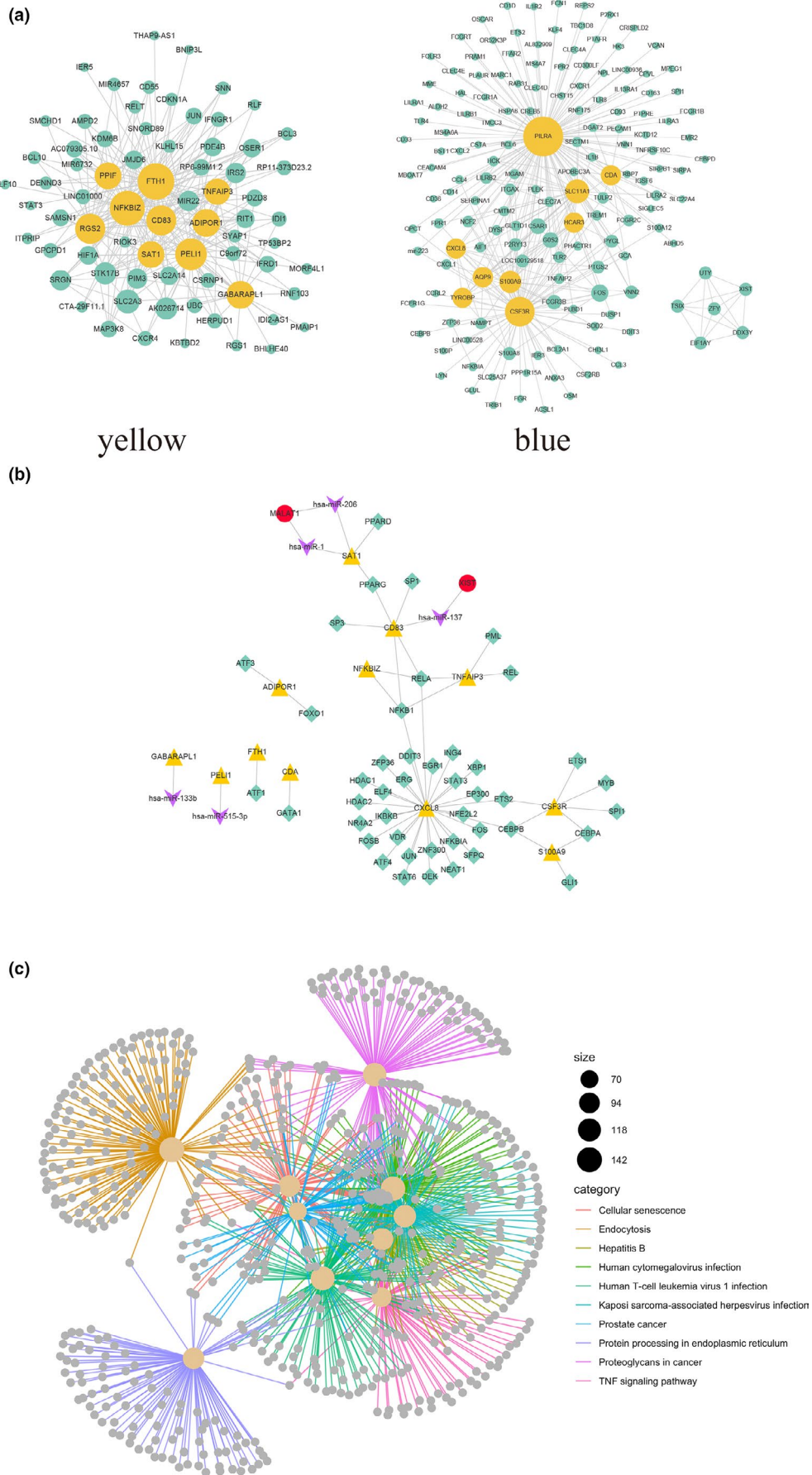
The 19 co-expressed key genes were mapped to GSE66360 for unsupervised clustering, selecting the CHD samples to be classified by  $K$ -means unsupervised clustering. First, the optimal value of  $K$  was determined by the SSE node, which represents the sum of the squares of the distances from all points to the center of the cluster to which they belong. As depicted in Figure S4a, SSE decline slowed down when  $K > 5$ , therefore, we defined  $K = 5$ . We used the R software package Rtsne for dimension reduction of the gene expression data. As displayed in Figure S4b, all CHD samples could be classified into five categories. The expression of co-expressed key genes in all CHD samples is shown in Figure 5a. Although there was no taxonomic difference in the expression of a single gene, the results were highly consistent with those combining 19 co-expressed key genes in Figure S4b. Therefore, we inferred that these 19 co-expressed key genes were of significance for CHD typing.

To further explain the differential expression of the 19 co-expressed key genes in different clusters, the mean value of each gene in each category was viewed as the expression value of genes in this category. *FTH1*, *HCAR3*, *RGS2*, *S100A9*, and *TYROBP* were differentially expressed in five clusters (Figure 5b). To explore the expression of these five genes in different CHD subtypes, we looked separately at the expression of these five genes in all CHD samples. We found significant differences among the five genes in different categories ( $p < 0.01$ ), suggesting these genes could be used as marker genes for different CHD subcategories (Figure 5c–g).

### 3.8 | Verification of the DEGs in patient samples

To estimate the levels of the five DEGs likely implicated in CHD molecular typing, RT-qPCR was performed on blood samples from CHD patients and healthy volunteers. As shown in Figure 5h, the mRNA expression levels of S100A9, TYROBP, and FTH1 in the blood of CHD patients were





**FIGURE 4** Module core gene analysis and construction of multifactor regulation network. (a) module gene network visualization. (b) Multifactor regulation network: yellow represents mRNA, green represents TF, purple represents miRNA, and red represents lncRNA. (c) Enrichment analysis of target genes of core regulators

**TABLE 6** Differential Expressed miRNAs

| Data set ID  | DEG num | Up | Down |
|--------------|---------|----|------|
| GSE59421     | 38      | 9  | 29   |
| GSE28858     | 138     | 74 | 64   |
| GSE64563_EAT | 47      | 17 | 30   |
| GSE64563_SAT | 28      | 12 | 16   |

Note: Differential expressed miRNAs in the Data set of GSE59421, GSE28858, GSE64563\_EAT and GSE64563\_SAT. EAT: Epicardial Adipose Tissue; SAT: Subcutaneous Adipose Tissue.

**TABLE 7** Core regulatory factors

| Factor      | Spe    | Degree |
|-------------|--------|--------|
| NFKB1       | TF     | 4      |
| RELA        | TF     | 4      |
| CEBPB       | TF     | 3      |
| MALAT1      | lncRNA | 2      |
| hsa-miR-1   | miRNA  | 2      |
| hsa-miR-137 | miRNA  | 2      |
| hsa-miR-206 | miRNA  | 2      |
| CEBPA       | TF     | 2      |
| ETS2        | TF     | 2      |
| PPARG       | TF     | 2      |

Note: 10 core regulators were screened out by constructing a multifactor regulatory network.

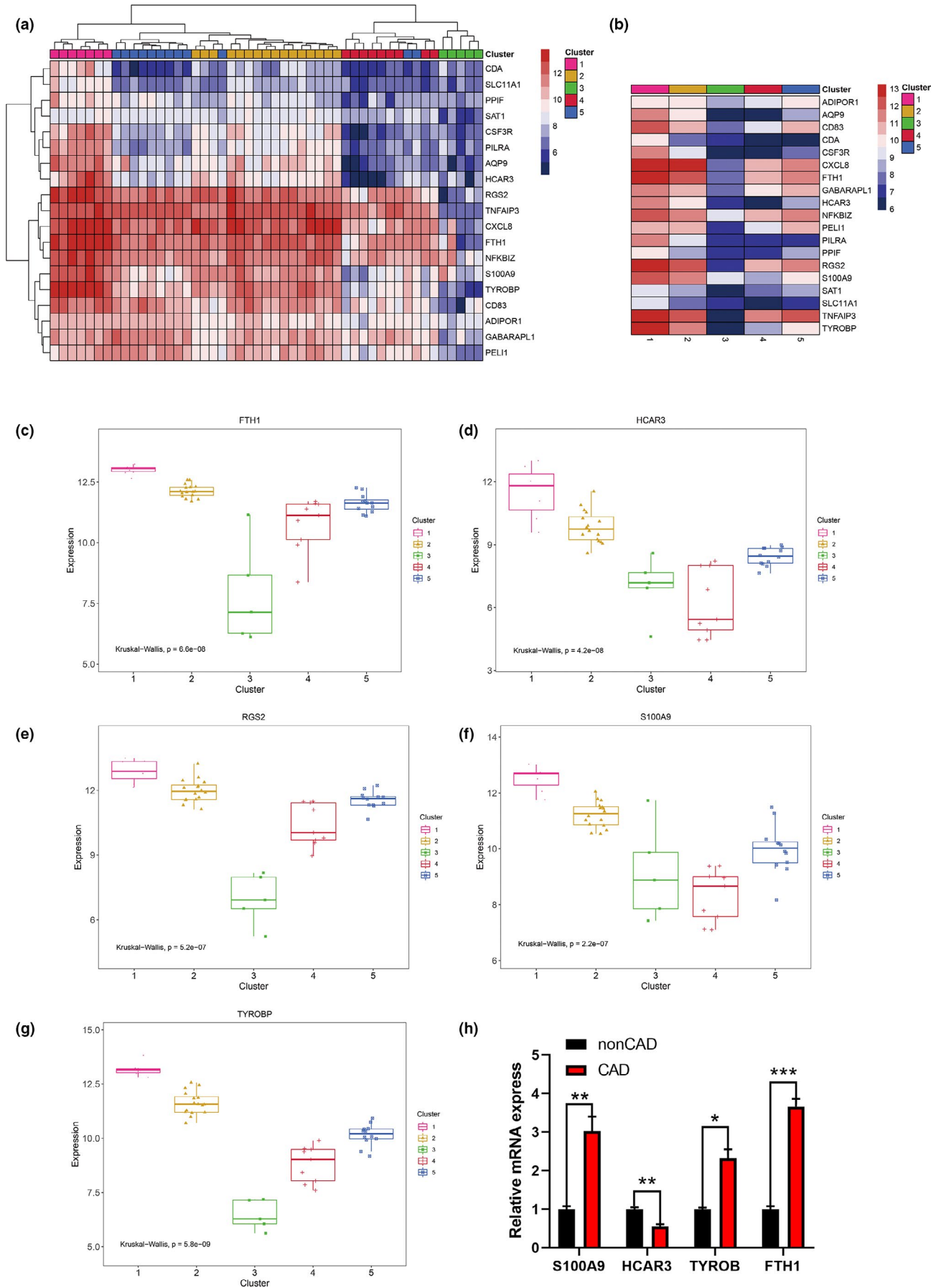
significantly higher than that in healthy subjects, while a decreased expression level of HCAR3 was observed in patients with CHD. However, RGS2 could hardly be detected at the transcriptional level. These results indicate that S100A9, TYROBP, FTH1, and HCAR3 could be used for classifying the molecular typing of CHD, while RGS2 should be excluded as it could not be detected.

## 4 | DISCUSSION

This study was motivated by the molecular typing for cancers, which could identify the molecular typing based on their gene expression profiles or network data. We screened the differential expression of genes associated with CHD based on the GEO data and expanded those data with the NCBI-gene and OMIM databases to obtain the candidate genes associated with CHD. Next, a co-expression network analysis was performed on the candidate genes for CHD using WGCNA, and the two most relevant modules of CHD were mined, including 75 CHD and 70 healthy mRNA samples. Both modules correlated with CHD, indicating that the mining results of WGCNA were ideal. Core genes were mined as key genes for co-expression based on their module network

relationship, and the differential miRNAs of CHD and interactions in the database were mined in the GEO data set to build a multi-factor regulatory network of key genes for co-expression. The functions of target genes of core regulators and CHD correlated, indicating that the co-expressed key genes obtained by mining are valuable. The classification of CHD samples based on the co-expression of key genes suggested significant differences in five genes between the analyzed subgroups, and these five genes could be used as biomarker genes to differentiate CHD subtypes.

*FTH1*, *HCAR3*, *RGS2*, *S100A9*, and *TYROBP* genes were differentially expressed in five clusters. Several studies have shown that S100A9 (acute coronary syndrome biomarker) is related to CHD (Buyukterzi et al., 2017; Li et al., 2019). The FTH1 gene encodes the heavy subunit of ferritin, the major intracellular iron storage protein in prokaryotes and eukaryotes. It is a substrate of ferroptosis, which emerges to play vital roles in CHD (Fisher et al., 2007; Kobayashi et al., 2018, Sukseree et al., 2020). Mitochondrial ferritin is a protein that regulates iron storage and H-ferritin-like protein, which exists only in mitochondria and possesses ferroxidase activity, plays a critical role in cell survival and reactive oxygen species regulation in acute myocardial injury (Lee et al., 2014), including percutaneous coronary intervention. However, the physiological roles of FTH1 in CHD are poorly understood. In this study, we verified by RT-qPCR in blood samples that FTH1 was significantly increased in CHD patients. We speculate that patients with coronary heart disease with elevated FTH1 may have plaque rupture or myocardial injury, but whether it is associated with coronary no-reflow injury or myocardial fibrosis remains to be confirmed. HCAR3 is a hydroxycarboxylic acid receptor involved in fatty acid metabolism (Offermanns et al., 2011). Human HCAR3 is activated by MAPK cascades in the PLC/PKC and MMP/EGFR pathways, which substantially increases ATP production by enhancing free fatty acid oxidation under conditions such as diabetic ketoacidosis and fasting (Offermanns et al., 2011; Zhou et al., 2012). Free fatty acid oxidation is vital in myocardial metabolism. The myocardial free fatty acid metabolism of CHD patients with elevated HCAR3 might be better than those with low levels, affecting CHD symptoms. Symptoms of coronary heart disease depend on the balance between oxygen and energy metabolism. RGS2 has GTPase-activating protein activity and acts as an essential controller of the magnitude and kinetics of cell responses after GPCR activation (Biesemann et al., 2014). Though we could not detect RGS2 mRNA in human blood, previous research indicated that the upregulation of RGS2 protected against myocardial injury via inhibition of cAMP levels and maintenance of cardiomyocyte contractility in a failing heart (Sjogren, Parra, Atkins, Karaj, & Neubig, 2016). Moreover, RGS2 facilitates uterine artery adaptation during pregnancy in mice (Koch, Dahlen, Owens, & Osei-Owusu, 2019), but it remains unclear whether it



**FIGURE 5** The expression of co-expressed key genes in CHD samples or in each cluster and validation in blood samples. (a,b) the expression of 19 co-expressed key genes in CHD samples or in each cluster. (c–g) Gene expression in different clusters (each dot represents a sample). (h) mRNA expression of S100A9, HCAR3, TYROBP, and FTH1 in blood samples

increases coronary artery perfusion in CHD patients, in particular in acute myocardial infarction patients. As a biomarker of immunity and inflammation, *S100A9* is related to CHD, especially acute coronary syndrome (Buyukterzi et al., 2017; Li et al., 2019). One study showed that *S100A9* blockage protects against myocardial infarction by reducing postischemic myocardial damage and improving cardiac function (Marinkovic, 2019). *TYROBP* is a key regulator of the immune system and is significantly upregulated in the brains of patients with Alzheimer's disease (Haure-Mirande et al., 2017). Two co-expression analyses showed that *TYROBP* is a promising target for preventing unstable carotid plaque formation in the carotid artery (Chen et al., 2019; Liu, Huan, Wu, Zou, & Qu, 2020). Whether *TYROBP* plays a role in unstable plaque in CHD remains unclear, and further research is required to confirm such effects. *FTH1*, *HCAR3*, *S100A9*, and *TYROBP* genes were differentially expressed in the blood of CHD patients, but the significance and function of these genes involved in the development of CHD are different. Perhaps the molecular typing of CHD can be guided by the changes and dysfunction of these five genes to guide the individual diagnosis and treatment of CHD patients.

All in all, *HCAR3*, *RGS2*, and *TYROBP* may play a beneficial role for CHD patients, but *FTH1* and *S100A9* may adversely affect CHD patients. In this study, we used co-expression analysis to explore the core regulators (miRNA, lncRNA, TF) and performed an enrichment analysis on target genes to reveal the mechanisms behind CHD. We found that these clusters of genes were related to the TNF signaling pathway, the hepatitis B pathway, and protein processing in the endoplasmic reticulum. TNF- $\alpha$  interacts with apoptosis and autophagy in CHD (Dong et al., 2019). Several studies have shown that endoplasmic reticulum stress affects coronary artery endothelial cell function in patients with CHD and participates in CHD development. Abnormal endoplasmic reticulum function affects protein synthesis and disturbs cellular metabolism (Bi et al., 2018; Chang et al., 2019; Li, Hu, Gong, & Yan, 2011). These core factors show the different mechanisms involved in CHD, which may lead to distinctive, individualized treatment of CHD. Therefore, these co-expression clusters of genes contribute to the classification of CHD molecular typing and could guide the clinical management of the disease.

However, we only verified our results in a limited number of patients. The molecular typing guided by the dysfunction of different factors described in this study needs further research to confirm and verify its impact on clinical management and prognosis.

## 5 | CONCLUSION

Taken together, the hub-genes *FTH1*, *HCAR3*, *RGS2*, *S100A9*, and *TYROBP* are differentially expressed in five

clusters associated with CHD, and the top microRNAs hsa-miR-1, hsa-miR-206, and hsa-miR-137, the top TF NFKB1, RELA, CEBPB, CEBPA, ETS2, and PPARG, the lncRNA MALAT1, may be potential targets for diagnosis and treatment of CHD. The core genes may contribute to the classification of the molecular typing of CHD. Further research to confirm and verify its impact on clinical management and prognosis is required.

## 5.1 | Limitations

There were several limitations to this study. First, this study was based on bioinformatics analysis of gene expression and did not include assessment of changes in proteins. Therefore, some translational modifications or posttranslational modifications were not considered. Second, the small sample size for validation could not fully explain the scientific issues. Third, the study lacked mechanism verification, and the roles of the dysfunction of the core genes in molecular typing needs to be confirmed. Large and prospective clinical studies are necessary to validate our results.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NSFC No. 81600233; 81570329; 81770229).

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

## AUTHOR CONTRIBUTIONS

The contributions of the authors are as follows: data analysis, Yuewei Li, Maohuan Lin, Kangjie Wang; sample collection, YaQing Zhan, Wenli Gu, Guanghao Gao, Yuna Huang, Yangxin Chen; experiment, Yuewei Li, Maohuan Lin; writing, Yuewei Li, Tucheng Huang and Jingfeng Wang; overall concept and coordination: Yuewei Li, Tucheng Huang and Jingfeng Wang.

## ORCID

Jingfeng Wang  <https://orcid.org/0000-0003-4598-5256>

## REFERENCES

- Arnett, D. K., Blumenthal, R. S., Albert, M. A., Buroker, A. B., Goldberger, Z. D., Hahn, E. J., ... Ziaeian, B. (2019). 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: Executive summary: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Journal of the American College of Cardiology*, 74, 1376–1414.
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., ... Soboleva, A. (2013). NCBI GEO: Archive for

- functional genomics data sets—update. *Nucleic Acids Research*, *41*, D991–D995.
- Bi, X., Zhang, G., Wang, X., Nguyen, C., May, H. I., Li, X., ... Hill, J. A. (2018). Endoplasmic reticulum chaperone GRP78 protects heart from ischemia/reperfusion injury through Akt activation. *Circulation Research*, *122*, 1545–1554.
- Biesemann, N., Mendler, L., Wietelmann, A., Hermann, S., Schafers, M., Kruger, M., ... Braun, T. (2014). Myostatin regulates energy homeostasis in the heart and prevents heart failure. *Circulation Research*, *115*, 296–310.
- Buyukterzi, Z., Can, U., Alpaydin, S., Guzelant, A., Karaarslan, S., Kocyigit, D., & Gurses, K. M. (2017). Enhanced S100A9 and S100A12 expression in acute coronary syndrome. *Biomarkers in Medicine*, *11*, 229–237.
- Chang, L., Wang, Z., Ma, F., Tran, B., Zhong, R., Xiong, Y., ... Zhu, Y. Z. (2019). ZYZ-803 mitigates endoplasmic reticulum stress-related necroptosis after acute myocardial infarction through downregulating the RIP3-CaMKII signaling pathway. *Oxidative Medicine and Cellular Longevity*, *2019*, 6173685.
- Chen, R., Ge, T., Jiang, W., Huo, J., Chang, Q., Geng, J., & Shan, Q. (2019). Identification of biomarkers correlated with hypertrophic cardiomyopathy with co-expression analysis. *Journal of Cellular Physiology*, *234*, 21999–22008.
- Dong, Y., Chen, H., Gao, J., Liu, Y., Li, J., & Wang, J. (2019). Molecular machinery and interplay of apoptosis and autophagy in coronary heart disease. *Journal of Molecular and Cellular Cardiology*, *136*, 27–41.
- Fisher, J., Devraj, K., Ingram, J., Slagle-Webb, B., Madhankumar, A. B., Liu, X., ... Connor, J. R. (2007). Ferritin: A novel mechanism for delivery of iron to the brain and other organs. *American Journal of Physiology. Cell Physiology*, *293*, C641–C649.
- Haure-Mirande, J. V., Audrain, M., Fanutza, T., Kim, S. H., Klein, W. L., Glabe, C., ... Ehrlich, M. E. (2017). Deficiency of TYROBP, an adapter protein for TREM2 and CR3 receptors, is neuroprotective in a mouse model of early Alzheimer's pathology. *Acta Neuropathologica*, *134*, 769–788.
- Kobayashi, M., Suhara, T., Baba, Y., Kawasaki, N. K., Higa, J. K., & Matsui, T. (2018). Pathological roles of iron in cardiovascular disease. *Current Drug Targets*, *19*, 1068–1076.
- Koch, J. N., Dahlen, S. A., Owens, E. A., & Osei-Owusu, P. (2019). Regulator of G protein signaling 2 facilitates uterine artery adaptation during pregnancy in mice. *Journal of the American Heart Association*, *8*, e010917.
- Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559.
- Lee, Y. K., Ho, P. W., Schick, R., Lau, Y. M., Lai, W. H., Zhou, T., ... Siu, C. W. (2014). Modeling of Friedreich ataxia-related iron overloading cardiomyopathy using patient-specific-induced pluripotent stem cells. *Pflugers Archiv. European Journal of Physiology*, *466*, 1831–1844.
- Li, J. H., Liu, S., Zhou, H., Qu, L. H., & Yang, J. H. (2014). starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, *42*, D92–D97.
- Li, L., Hu, B. C., Gong, S. J., & Yan, J. (2011). Homocysteine-induced caspase-3 activation by endoplasmic reticulum stress in endothelial progenitor cells from patients with coronary heart disease and healthy donors. *Bioscience, Biotechnology, and Biochemistry*, *75*, 1300–1305.
- Li, Y., Chen, B., Yang, X., Zhang, C., Jiao, Y., Li, P., ... Du, J. (2019). S100a8/a9 signaling causes mitochondrial dysfunction and cardiomyocyte death in response to ischemic/reperfusion injury. *Circulation*, *140*, 751–764.
- Liu, Y., Huan, W., Wu, J., Zou, S., & Qu, L. (2020). IGFBP6 is down-regulated in unstable carotid atherosclerotic plaques according to an integrated bioinformatics analysis and experimental verification. *Journal of Atherosclerosis and Thrombosis*. <https://doi.org/10.5551/jat.52993>
- Marinkovic, G., Grauen Larsen, H., Yndigeegn, T., Szabo, I. A., Mares, R. G., De Camp, L., ... Schiopu, A. (2019). Inhibition of pro-inflammatory myeloid cell responses by short-term S100A9 blockade improves cardiac function after myocardial infarction. *European Heart Journal*, *40*, 2713–2723.
- Offermanns, S., Colletti, S. L., Lovenberg, T. W., Semple, G., Wise, A., & Ap, I. J. (2011). International Union of basic and clinical pharmacology. LXXXII: Nomenclature and classification of hydroxy-carboxylic acid receptors (GPR81, GPR109A, and GPR109B). *Pharmacological Reviews*, *63*, 269–290.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., ... Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*, 747–752.
- Sjogren, B., Parra, S., Atkins, K. B., Karaj, B., & Neubig, R. R. (2016). Digoxin-mediated upregulation of RGS2 protein protects against cardiac injury. *Journal of Pharmacology and Experimental Therapeutics*, *357*, 311–319.
- Sukseree, S., Schwarze, U. Y., Gruber, R., Gruber, F., Quiles Del Rey, M., Mancias, J. D., ... Eckhart, L. (2020). ATG7 is essential for secretion of iron from ameloblasts and normal growth of murine incisors during aging. *Autophagy*, 1–7. <https://doi.org/10.1080/15548627.2019.1709764>
- Thygesen, K., Alpert, J. S., Jaffe, A. S., Chaitman, B. R., Bax, J. J., Morrow, D. A., ... Executive Group on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction. (2018). Fourth universal definition of myocardial infarction (2018). *Journal of the American College of Cardiology*, *72*, 2231–2264.
- Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, Article17.
- Zhou, Q., Li, G., Deng, X. Y., He, X. B., Chen, L. J., Wu, C., ... Zhou, N. M. (2012). Activated human hydroxy-carboxylic acid receptor-3 signals to MAP kinase cascades via the PLC-dependent PKC and MMP-mediated EGFR pathways. *British Journal of Pharmacology*, *166*, 1756–1773.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Li Y, Lin M, Wang K, et al. A module of multifactor-mediated dysfunction guides the molecular typing of coronary heart disease. *Mol Genet Genomic Med*. 2020;8:e1415. <https://doi.org/10.1002/mgg3.1415>