



OPEN

## Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease

Peng-Fei Zheng<sup>1,2</sup>, Lu-Zhu Chen<sup>1</sup>, Yao-Zong Guan<sup>2</sup> & Peng Liu<sup>1</sup>✉

This investigation seeks to dissect coronary artery disease molecular target candidates along with its underlying molecular mechanisms. Data on patients with CAD across three separate array data sets, GSE66360, GSE19339 and GSE97320 were extracted. The gene expression profiles were obtained by normalizing and removing the differences between the three data sets, and important modules linked to coronary heart disease were identified using weighted gene co-expression network analysis (WGCNA). Gene Ontology (GO) functional and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were applied in order to identify statistically significant genetic modules with the Database for Annotation, Visualization and Integrated Discovery (DAVID) online tool (version 6.8; <http://david.abcc.ncifcrf.gov>). The online STRING tool was used to construct a protein–protein interaction (PPI) network, followed by the use of Molecular Complex Detection (MCODE) plug-ins in Cytoscape software to identify hub genes. Two significant modules (green-yellow and magenta) were identified in the CAD samples. Genes in the magenta module were noted to be involved in inflammatory and immune-related pathways, based on GO and KEGG enrichment analyses. After the MCODE analysis, two different MCODE complexes were identified in the magenta module, and four hub genes (*ITGAM*, degree = 39; *CAMP*, degree = 37; *TYROBP*, degree = 28; *ICAM1*, degree = 18) were uncovered to be critical players in mediating CAD. Independent verification data as well as our RT-qPCR results were highly consistent with the above finding. *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* are potential targets in CAD. The underlying mechanism may be related to the transendothelial migration of leukocytes and the immune response.

Inflammation plays a crucial role in the pathophysiology of coronary artery disease (CAD), it is involved in the formation, erosion and final rupture of atherosclerotic plaque, resulting in partial or total occlusion of coronary artery. This might result in myocardial ischemia and hypoxia and thereby an acute myocardial infarction (AMI)<sup>1</sup>. Complete occlusion usually leads to ST-elevation in the electrocardiogram, which is defined as an acute ST-segment elevation myocardial infarction (STEMI). Partial occlusion or occlusion with collateral circulation without ST-segment elevation is classified as unstable coronary syndrome. Unstable coronary syndromes without elevated Troponin T (TnT, a marker of myocardial necrosis) were defined as non-ST-segment elevation acute coronary syndromes, while those with elevated TnT were defined as non-ST-segment elevation myocardial infarction (non-STEMI)<sup>2</sup>. Although with the spread and popularization of emergency percutaneous coronary intervention (PCI) treatment, the prognosis of patients with AMI can be significantly improved by rapidly restoring blood flow in occluded vessels. However, CAD still maintains a high morbidity and mortality, and leading to significant reduction in quality of life of those patients as well as poses a hefty burden on healthcare systems<sup>3,4</sup>. The overall prevalence of CAD, also known as ischemic heart disease, nearly has risen steadily since 1990, reaching 182 million cases and 9.14 million deaths in 2019<sup>5</sup>. Well established risk factors for CAD include high blood pressure, diabetes, a sedentary lifestyle, smoking, family history, obesity, stress and hyperlipidemia. Although lots of efforts have been undertaken in recent years, the prevention and cure of CAD remains a daunting challenge for physicians around the world. There is an urgent need for further exploration of the potential molecular mechanisms correlated with CAD. Existing literature indicates that CAD is primarily mediated by

<sup>1</sup>Department of Cardiology, The Central Hospital of Shao Yang, 36 QianYuan lane, Shaoyang 422000, Hunan, People's Republic of China. <sup>2</sup>Graduate School of Guangxi Medical University, 22 Shuangyong Road, Nanning 530021, Guangxi, People's Republic of China. ✉email: ying\_lpxm@163.com

coronary atherosclerosis<sup>6</sup>. Early intervention in preventing atherosclerosis could significantly decrease CAD, stroke and other ischemic diseases from occurring and developing<sup>7</sup>.

Microarray analysis might serve as a novel and practical approach to identify susceptibility genes correlated with coronary heart disease<sup>8</sup>. However, the reproducibility and sensitivity of microarray analysis based on differentially expressed genes may be limited<sup>9,10</sup>. Exactly, hence, currently, the microarray-based transcriptome analysis has been largely replaced by (or even single cell-based) RNA seq. Furthermore, Gene co-expression network-based methods have been widely used in processing microarray<sup>11,12</sup> and RNA seq data<sup>13</sup> and have especially been used to identify meaningful functional modules. Weighted gene co-expression network analysis (WGCNA) is one of the most effective methods of gene co-expression network analysis. Transcriptome data from different sources within the same species can be grouped together for WGCNA analysis<sup>14</sup>. WGCNA generates a scale-free network of gene–gene interactions, if some genes always have similar expression changes in a physiological process or different tissues, and these genes will be enriched in a common significant module. Furthermore, it can be used to further analyze the correlation between modules and phenotypes or clinical characteristics<sup>15</sup>. Given the capabilities of WGCNA in formulating a co-expression network comprising of significant modules, we are able to glean new information regarding CAD features and may uncover novel insights in CAD-related molecular mechanisms, signaling pathways and genetic biomarkers.

## Results

**Data preprocessing.** Interpatch difference removal and data normalization were carried out to obtain the final gene expression profiles. 113 samples yielded a total of 23,493 gene symbols. Further information regarding gene expression profile and the sample phenotypes are depicted in Supplementary Tables S1A,B and S2.

**Weighted gene co-expression networks.** Weighted gene co-expression networks were constructed based on identified genes after determining the soft threshold ( $\beta = 14$ ) (Fig. 1). In order to create a topological overlap matrix (TOM), the adjacency and correlation matrices of the gene expression profile were calculated. A final gene clustering tree based on the gene–gene non- $\omega$  similarity was produced (Fig. 2). Using the hierarchical average linkage clustering method in combination with the TOM, we proceeded to identify gene modules of each gene network. The dynamic tree cut algorithm highlighted five gene modules (Fig. 3). Genes that did not fit in any modules were discarded from further analyses (presented as gray modules).

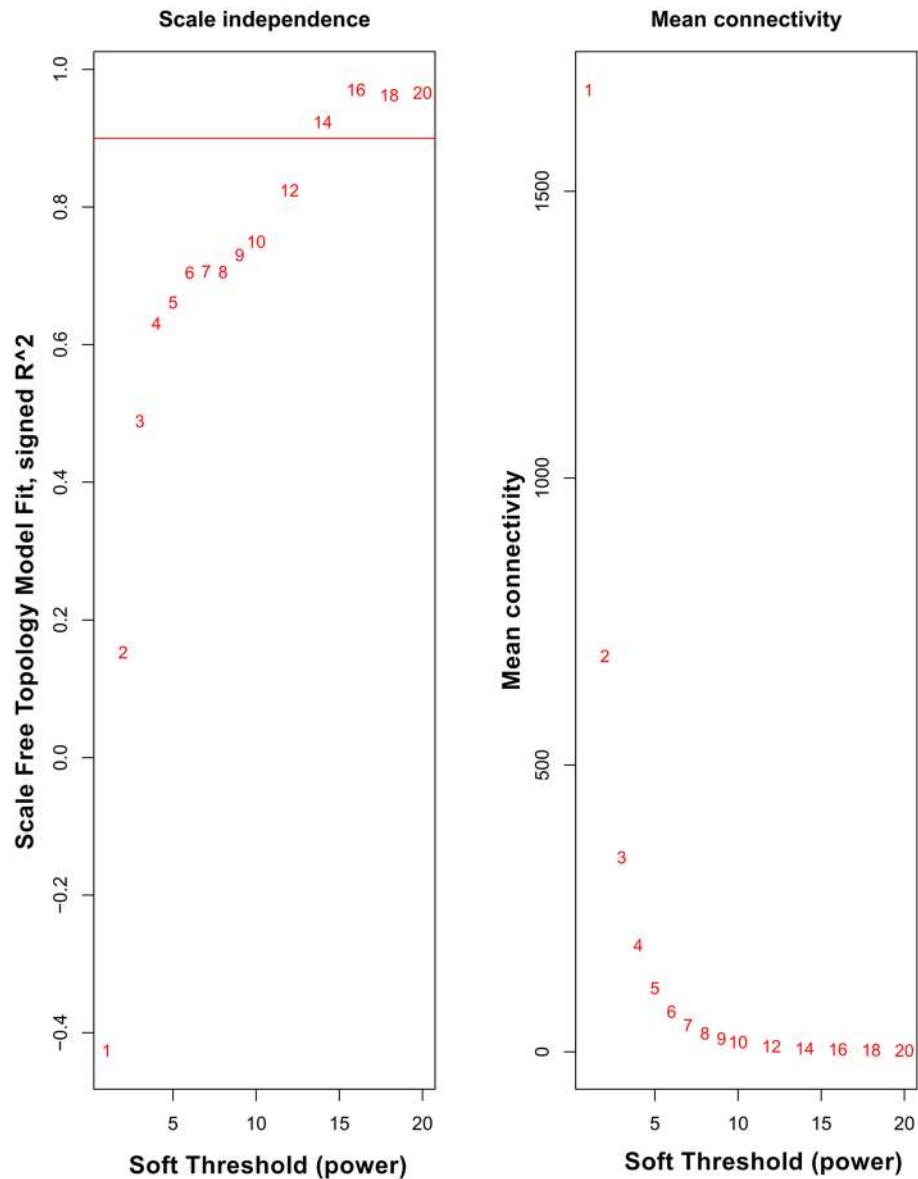
**Identification of the modules of interest.** Biologically significant modules which those that strongly correlated to clinicopathological features. Supplementary Fig. S1 and Fig. 4 show that the green–yellow ( $r^2 = 0.52$ ,  $P = 5E-04$ ) and magenta ( $r^2 = 0.42$ ,  $P = 0.02$ ) modules were highly correlated with CAD. Therefore, subsequent analyses were carried out on genes from both these modules. In addition, Supplementary Fig. S2 shows a highly significant correlation between gene significant (GS) versus module membership (MM) in the green–yellow (A) and magenta (B) modules with CAD.

**Module preservation test.** We performed preservation analysis of the expression profiles of CAD and noticed that there were one weak and two strong preserved modules between CAD and control subjects (Supplementary Fig. S3). The statistical results of medianRank and Zsummary are consistent, which indicates that the module size has little effect on the preservation analysis. We noticed that the green–yellow and magenta modules were highly preserved with CAD; blue and brown modules were weakly preserved with CAD. These findings demonstrated that the gene expression patterns between the CAD and control subjects are different to a large extent.

**Enrichment analysis of interesting modules.** Biological functions of genes in both these modules were then subjected to further GO and KEGG pathway enrichment analyses. A total of 309 genes in the magenta module (Supplementary Table S3) were significantly correlated with the following pathways: leukocyte transendothelial migration signaling pathway (*ITGAM* and *ICAM1*), TNF signaling pathway (*ICAM1*), and the *Staphylococcus aureus* infection (*ITGAM* and *ICAM1*), rheumatoid arthritis (*ICAM1* and *ITGAM*), tuberculosis (*ITGAM* and *CAMP*), natural killer cell-mediated cytotoxicity (*ICAM1* and *TYROBP*), and NF-kappa B signaling pathways (*ICAM1*). The KEGG pathway analysis, molecular functions, biological processes as well as cellular components are depicted Fig. 5, with a more detailed presentation of data included in Supplementary Tables S4 and S5.

**PPI network construction and module analysis of DEGs.** The STRING online tool was used to formulate a PPI network comprising 309 nodes and 3165 edges. Subsequent analysis found that only two MCODEs with scores  $> 6$  were detected. Hub genes *ITGAM* (degree = 39), *CAMP* (degree = 37), and *TYROBP* (degree = 28) were identified in Molecular-1 (A-1), and *ICAM1* (degree = 18) was identified in Molecular-2 (A-2) (Fig. 6). It can be concluded that the identified genes are strongly linked to CAD.

**Results of meta-analysis based on three eligible microarrays.** The meta-analysis included three different datasets from the GEO database and included a total of 80 acute myocardial infarction (AMI) and 67 normal subjects. The integrated gene expression profile was obtained after eliminating the batch effects between three datasets and also shown in Supplementary Table S6A,B. A fixed effect model was used to identify DEGs in INMEX. As a result, a total of 2409 DEGs (1125 upregulated and 1284 downregulated genes) were identified in AMI compared with normal subjects (Supplementary Table S7). Additionally, *ITGAM*, *CAMP*, *TYROBP* and

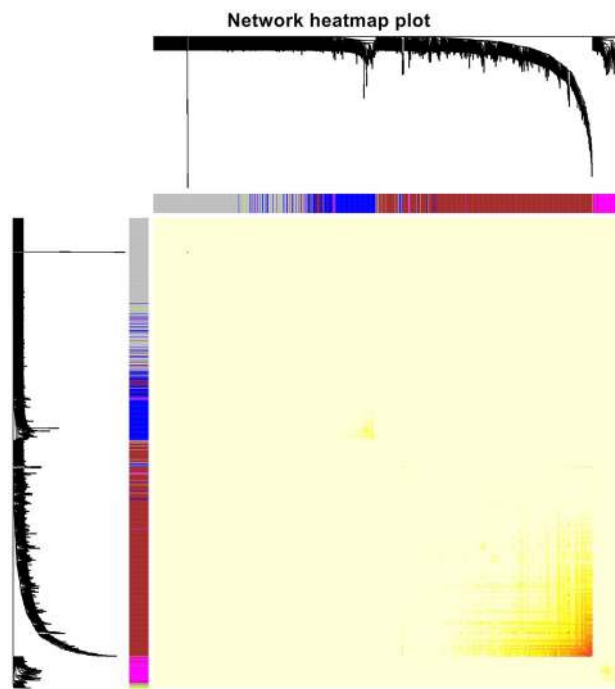


**Figure 1.** Analysis of network topology for various soft-thresholding powers. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis).

*ICAM1* were all included in the upregulated DEGs group (Table 1). The expression pattern of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* across three eligible datasets are also shown in Supplementary Fig. S4. Expressions of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* genes were also markedly raised in individuals with CAD in comparison to healthy subjects.

**Validation analysis by RT-qPCR.** The results of RT-qPCR uncovered that *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* expressions were markedly raised in those with CAD in contrast to individuals without the condition. These findings reflect the results of the microarray analysis (Fig. 7).

**ROC curve for CAD patients.** As shown in Fig. 8, ROC analysis was used to evaluate the predictive values of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* for CAD. The AUC values of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* were 0.714 (95% CI 0.666–0.762;  $P < 0.001$ ) with a sensitivity of 74.3% and a specificity of 76.2%; 0.897 (95% CI 0.865–0.928;  $P < 0.001$ ) with a sensitivity of 84.2% and a specificity of 91.1%; 0.761 (95% CI 0.716–0.807;  $P < 0.001$ ) with a sensitivity of 77.1% and a specificity of 79.7% and 0.848 (95% CI 0.811–0.885;  $P < 0.001$ ) with a sensitivity of 83.3% and a specificity of 86.7% for prediction of CAD risk, respectively. *CAMP* diagnostic performance is depicted to be superior in comparison to other genes.



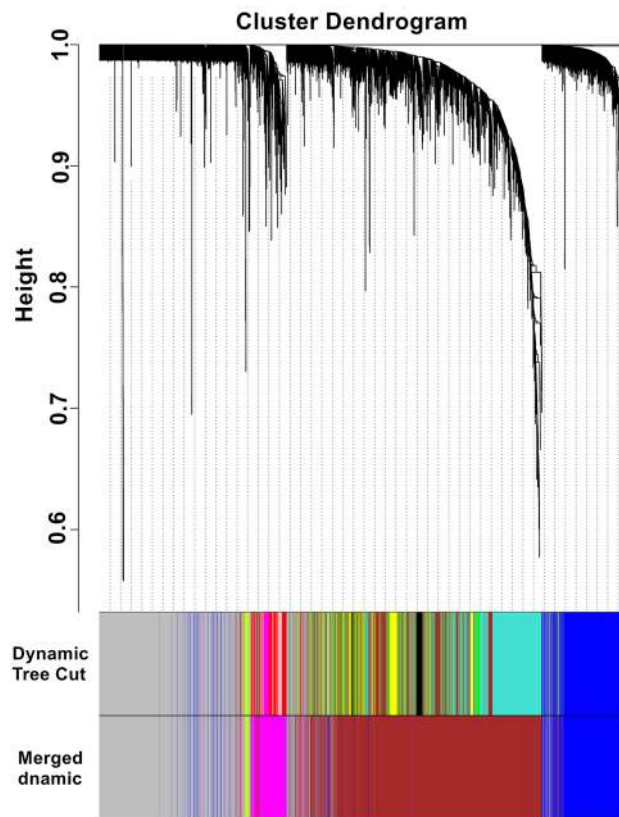
**Figure 2.** Heatmap plot of topological overlap in the gene network. In the heatmap, each row and column correspond to a gene, light color denotes low topological overlap, and progressively darker red denotes higher topological overlap. Darker squares along the diagonal correspond to modules. The gene dendrogram and module assignment are shown along the left and top.

**Comparison with other methods.** Machine learning, more specifically, deep learning, is one of the most widely used forms of artificial intelligence. As is known that deep learning (DL) is widely used for intelligence medicine to assist disease risk prediction and disease diagnosis based on small sample size, like transcriptomic<sup>16</sup> or genomic<sup>17</sup> data, and imaging data<sup>18</sup>. The application of deep learning in disease detection or diagnosis is particularly important for clinician, as it has the potential to maximize diagnostic performance. Previous research has demonstrated that compared with the traditional Cox model, the risk stratification model based on gene co-expression network and DL would apply deep Convolutional Neural Network (CNN) to high-dimensional gene expression data, and could improve the risk stratification and survival prediction ability of the model<sup>19</sup>. In addition, Liu, et al. also noticed that the multi-model deep learning framework based on CNN has better performance in the diagnosis of Alzheimer's disease (AD) and mild cognitive impairment (MCI) than the single-model method and several other competing methods<sup>18</sup>. However, the construction of deep learning model is tedious and time-consuming, and it is very challenging to build models with deep learning method for some parts with complex structure and difficult to image. Therefore, in the current research, the empirical ROC curves of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* were plotted using non-parametric method<sup>20</sup> by SPSS (Version 22.0) software. Non-parametric method, which do not require any assumptions about data distribution, can be calculated without fitting any ROC curves<sup>21</sup>. Thus, the non-parametric method of area estimation under the ROC curve can be used to evaluate the accuracy of all diagnostic tests because there are no restrictions. However, the diagnostic performance of non-parametric method may be worse than that of a deep learning model.

**Demographic and biochemical characteristics.** Both control and CAD individuals had similar proportions of height, age, gender ratio, and proportion of drinkers (Table 2). Patients with CAD were more likely to be smokers, and possessed higher levels of serum LDL-C, ApoB, TG and TC levels, body mass index (BMI), glucose level, weight, systolic and diastolic blood pressure as well as, glucose level, pulse pressure, in contrast to healthy individuals. Serum HDL-C and ApoA1 levels and the ApoA1/ApoB ratio were significantly higher in the control group.

## Discussion

Despite the plethora of information available regarding CAD, little is known regarding the feasibility of non-invasive diagnostic markers for this debilitating disease<sup>22–24</sup>. To facilitate improved treatment and diagnosis, there needs to be a deeper understanding on the underlying pathophysiology of CAD. Differential gene analysis based on microarray expression data is helpful for us to identify susceptibility genes and elucidate the molecular mechanism of CAD, however, microarray expression data are not always reproducible or are too sensitive to errors<sup>8</sup>. Therefore, the integration of gene expression profile data combined with WGCNA analysis may be an effective method to identify susceptibility genes of CAD. To meet this need, we integrated three different datasets

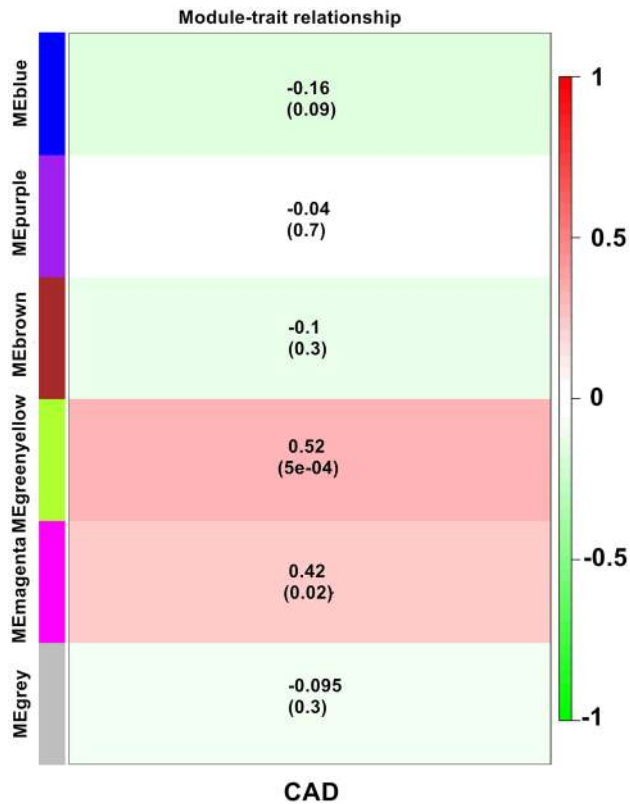


**Figure 3.** Clustering dendrogram of genes. Gene clustering tree (dendrogram) obtained by hierarchical clustering of adjacency-based dissimilarity. The colored row below the dendrogram indicates module membership identified by the dynamic tree cut method, together with assigned merged module colors and the original module colors.

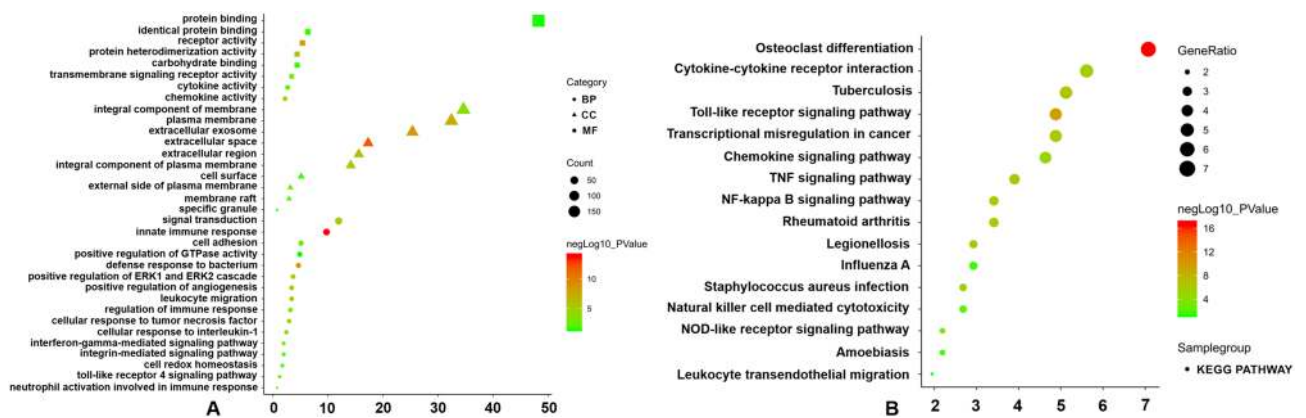
from CAD patients (GSE66360, GSE19339 and GSE97320) in order to carry out WGCNA analysis, which subsequently identified 2 modules (green-yellow and magenta) that were significantly correlated with CAD. Refer to previous researches<sup>25,26</sup>, to determine the reliability of the identified CAD-related modules (green-yellow and magenta), we have conducted a preservation analysis of the expression profiles of CAD and we noticed that the green-yellow and magenta modules were highly preserved. Furthermore, KEGG and GO gene enrichment analyses of these two modules highlighted that those of the magenta module may impart significant biological functions closely related to inflammation, the immune response, and white blood cell activation and migration. Four hub genes (*ITGAM*, *TYROBP*, *ICAM1* and *CAMP*) were identified in two moleculars that were detected by MCODE by analyzing the PPI protein interaction network. Moreover, network-based meta-analysis revealed that the expression levels of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* in CAD patients in GSE60993 and GSE66360 datasets were significantly higher than those in the control group, at the same time, the gene expression levels of *ITGAM*, *CAMP* and *TYROBP* in CAD patients in GSE61144 dataset were also significantly higher than those in the control group. Similarly, our RT-qPCR results strongly correlated with the above results. *ITGAM*, *TYROBP*, *ICAM1* and *CAMP* gene expressions were noted to be raised in individuals with CAD in comparison to those without. Therefore, the identified *ITGAM*, *TYROBP* and *ICAM1* and *CAMP* genes were concluded to be related to CAD onset, but the underlying molecular mechanisms of these genes might be slightly different.

A recent study has proven that the occurrence of CAD is caused by a variety of factors, which is a result of interaction between alterations in plasma lipid levels, lifestyle, environmental factors and genomic background<sup>27</sup>. Atherosclerosis is generally regarded as the pathological foundation of CAD<sup>6</sup>. Atherosclerosis is a combination of abnormal lipid metabolism and a chronic inflammatory process<sup>28</sup>. Transendothelial migration and subintimal aggregation of monocytes are some of the most important features of early human atherosclerotic lesions. After the differentiation into macrophages and the ingestion of lipids, which causes the formation of foam cells, the arterial wall may develop atherosclerotic plaques composed of foam cells, calcium, lipids, and other components<sup>29</sup>. Activated leukocytes can promote vascular endothelial injury and inflammatory response, and secrete a series of inflammatory factors, such as interleukin-1 (*IL-1*), *TNF- $\alpha$*  and *IL-6*, resulting in cellular adhesion, infiltration of inflammatory cells, matrix degradation, all of which culminates in plaque rupture, thereby accelerating the progression of atherosclerosis<sup>30</sup>. Through a comprehensive search of the NCBI GENE database, we discovered that *ITGAM* (also known as *CR3A*; *MO1A*; *CD11B*; *MAC-1*; *MAC1A*; and *SLEB6*; gene ID: 3684, HGNC: 6149, OMIM: 120980) is located on chromosome 16p11.2 (exon count: 31) and encodes the integrin  $\alpha$ M chain, which plays a crucial role in several inflammatory reactions, including the monocyte and neutrophil



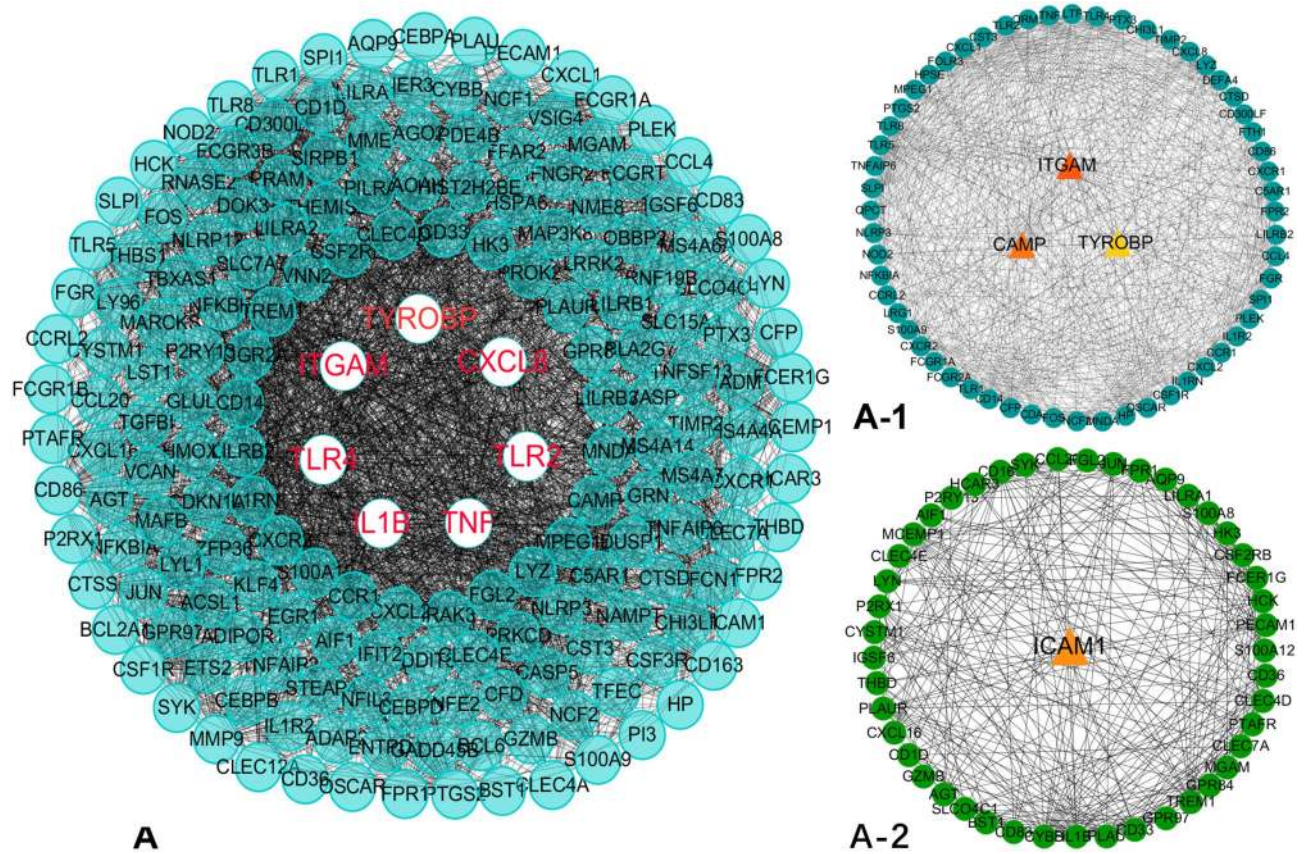


**Figure 4.** Module-feature associations. Each row corresponds to a moduleEigengene and the column to the clinical phenotype. Each cell contains the corresponding correlation in the first line and the P-value in the second line. The table is color-coded by correlation according to the color legend.



**Figure 5.** GO functional and KEGG pathway enrichment analyses for genes in the object module. The x-axis shows the number of genes and the y-axis shows the GO and KEGG pathway terms. The  $-\log_{10}$  (P-value) of each term is colored according to the legend. (A) GO functional enrichment analysis. (B) KEGG pathway enrichment analysis<sup>64</sup>.

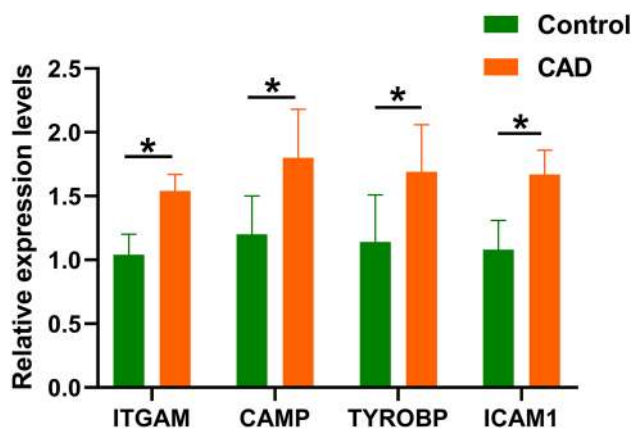
adhesion to damaged endothelial cells and transendothelial migration, and integrin  $\alpha M$  is also involved in CD40L-mediated inflammation during atherosclerosis<sup>31</sup>. Several novel studies also proved that *ITGAM* could act as a complement component 3 receptor that is involved in the inflammatory response<sup>32,33</sup>. Additionally, Ayari et al. suggested that *ITGAM* and *TYROBP* expression levels were raised in human carotid artery plaques<sup>34</sup>. Yongming Pan et al. also found similar expression trends for *ITGAM* and *TYROBP* in a novel Tibetan minipig atherosclerosis model<sup>35</sup>. *TYROBP* (also known as *DAPI2*; *KARAP*; *PLOSL*; *PLOSL1*; gene ID: 7305, HGNC: 12449, OMIM: 221770) encodes a transmembrane signaling polypeptide and is a type of transmembrane receptor that is ubiquitously found in macrophages/monocytes, natural killer (NK) cells and neutrophils. In recent years, NK cells, especially NKT cells, have been considered to be important participants in inflammatory cells chemotaxis, adhesion between inflammatory cells and endothelial cells, and other processes that are active in



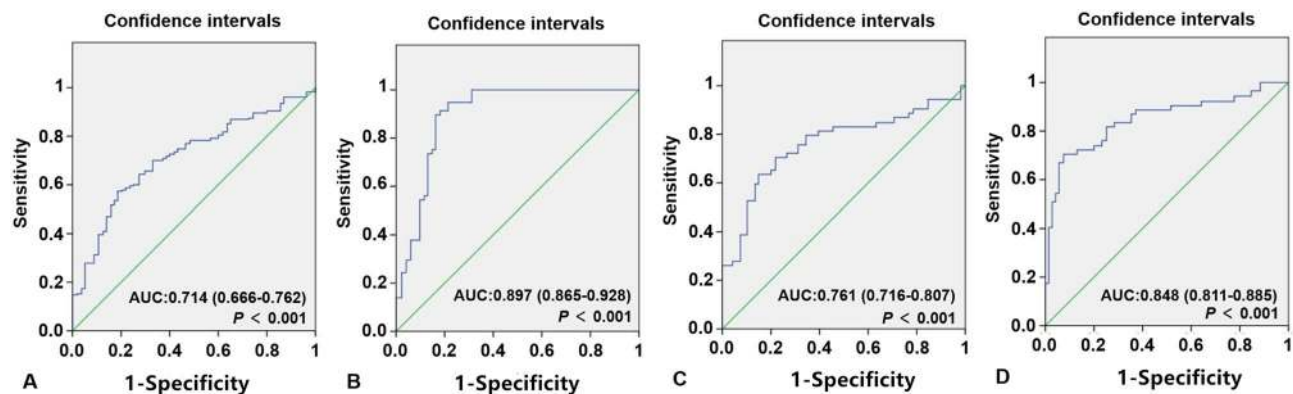
**Figure 6.** PPI network construction and identification of hub genes. (A) PPI network of genes in magenta module. The edge shows the interaction between two genes. Significant modules identified from the PPI network using the MCODE with a score >6.0. (A-1) Molecular-1 with MCODE score = 26.2. (A-2) Molecular-2 with MCODE score = 13.4.

Entrez ID	Gene symbol	Gene name	Combined ES	P value
3684	ITGAM	Integrin subunit alpha M	1.0900	1.83E-05
820	CAMP	Cathelicidin antimicrobial peptide	1.0255	7.24E-07
7305	TYROBP	Transmembrane immune signaling adaptor TYROBP	1.1046	7.40E-08
3383	ICAM1	Intercellular adhesion molecule 1	1.0241	1.16E-06

**Table 1.** Four upregulated DEGs (*ITGAM*, *CAMP*, *TYROBP* and *ICAM1*) in CAD relative to normal subjects.



**Figure 7.** Four identified hub genes were verified by RT-qPCR. \**P* < 0.001.



**Figure 8.** The ROC curves for the predictive values of ITGAM, CAMP, TYROBP and ICAM1 to identify CAD patients from healthy controls. (A) The AUC of ITGAM in CAD was 0.714 with a sensitivity of 74.3% and a specificity of 76.2%. (B) The AUC of CAMP in CAD was 0.897 with a sensitivity of 81.2% and a specificity of 91.1%. (C) The AUC of TYROBP in CAD was 0.761 with a sensitivity of 77.1% and a specificity of 79.7%. (D) The AUC of ICAM1 in CAD was 0.848 with a sensitivity of 83.3% and a specificity of 86.7%.

Characteristic	Control (n = 216)	CAD (n = 230)	Test-statistic	P
Male/female	150/66	169/61	0.890	0.345
Age (years)	54.02 ± 11.66	53.17 ± 10.19	0.931	0.352
Height (cm)	164.40 ± 7.5	165.22 ± 6.97	- 1.227	0.221
Weight (kg)	58.70 ± 9.08	66.08 ± 10.73	- 7.820	3.89E-14
BMI (kg/m <sup>2</sup> )	20.00 ± 3.77	24.14 ± 3.23	- 12.460	9.09E-31
Smoking, n %	71 (32.9)	98 (42.6)	4.489	0.034
Alcohol, n %	57 (26.4)	62 (27.0)	0.018	0.892
SBP (mmHg)	130.88 ± 18.75	136.74 ± 22.56	- 2.974	0.003
DBP (mmHg)	79.95 ± 11.04	82.54 ± 12.05	- 2.221	0.027
PP (mmHg)	50.93 ± 13.40	54.19 ± 16.66	- 1.974	0.049
Glu (mmol/L)	6.12 ± 1.64	6.47 ± 1.86	- 2.104	0.036
TC (mmol/L)	4.49 ± 1.02	4.74 ± 1.36	- 2.164	0.031
TG (mmol/L)	1.36 ± 1.17	1.61 ± 1.07	- 2.864	0.008
HDL-C (mmol/L)	1.65 ± 0.47	1.13 ± 0.30	13.861	1.47E-36
LDL-C (mmol/L)	2.79 ± 0.97	3.06 ± 1.11	- 2.713	0.007
ApoA1 (g/L)	1.42 ± 0.34	0.99 ± 0.31	14.057	2.18E-37
ApoB (g/L)	0.88 ± 0.21	0.95 ± 0.27	- 3.146	0.002
ApoA1/ApoB	1.69 ± 0.52	1.12 ± 0.49	11.732	7.13E-28

**Table 2.** Comparison of demographic, lifestyle characteristics and serum lipid levels of the participants. SBP Systolic blood pressure; DBP Diastolic blood pressure; PP Pulse pressure; Glu Glucose; HDL-C high-density lipoprotein cholesterol; LDL-C low-density lipoprotein cholesterol; Apo Apolipoprotein; TC Total cholesterol; TG Triglyceride. <sup>a</sup>Continuous data were presented as means ± SD and determined by two side *t*-test. <sup>b</sup>A Chi-square analysis was used to evaluate the difference of the rate between the groups.

the early stages of atherosclerosis<sup>36,37</sup>. Previous research has demonstrated that *TYROBP* acts as one of the key drivers of a variety of inflammatory pathways<sup>38</sup>. Wang et al. revealed that APOE mice demonstrated plaques which richly expressed *TYROBP*, a feature thought to result in *TREM-1/DAP12* pathway-mediated accelerated atherosclerosis progression<sup>39</sup>.

A central tenet of the inflammatory process involves endothelial cell binding by leukocytes through integrins. Intercellular adhesion molecule 1 (*ICAM1*; *CD54*) is a representative ligand of integrin that is key to mediating leukocyte adhesion to the endothelial cell surface<sup>40</sup>. *ICAM1*-mediated endothelial chemokines attract and activate leukocytes, leading to a severe inflammatory response<sup>41</sup>. Silvia Dragoni et al. proved that *ICAM1*-mediated intra-endothelial signaling plays a critical role in regulating lymphocyte transendothelial migration and modulating vascular permeability, thereby propagating chronic endothelial inflammation<sup>42</sup>. In addition, activation of *ICAM1* also increased the expression of inflammatory genes correlated with coronary heart disease, such as *IL-1B*<sup>40</sup>, *CXCL8*, *CCL5*<sup>41</sup>, and *VCAM-1*<sup>43</sup>. Similar results were also confirmed in KEGG pathways and GO enrichment analysis, we noticed that *ITGAM*, *TYROBP* and *ICAM1* were mainly involved in the following inflammation-related signaling pathways and biological processes: leukocyte transendothelial migration (*ITGAM* and *ICAM1*), *Staphylococcus aureus* infection (*ITGAM* and *ICAM1*), rheumatoid arthritis (*ICAM1*),



integrin-mediated signaling pathway (*ITGAM*), Toll-like receptor 4 signaling pathway (*ITGAM*), tuberculosis (*ITGAM*), natural killer cell-mediated cytotoxicity (*ICAM1* and *TYROBP*), NF-kappa B signaling pathway (*ICAM1*), and TNF signaling pathway (*ICAM1*). Therefore, we speculated that *ITGAM*, *TYROBP* and *ICAM1* may be involved in atherosclerosis by mediating the inflammatory pathways described above.

Previous studies have shown that CAD and several autoimmune diseases, such as psoriasis<sup>44</sup>, systemic lupus erythematosus<sup>45</sup>, and rheumatoid arthritis<sup>46</sup>, share a common pathogenesis, which suggests that the development of atherosclerosis is highly dependent on autoimmunity. The protein encoded by *CAMP* belongs to the antimicrobial peptide group that was previously established to be an autoantigen in psoriasis which is involved in cell chemotaxis, inflammatory response regulation and immune mediator induction<sup>47,48</sup>. Several compelling studies also proved that there was abnormal expression of *CAMP* in atherosclerotic plaques and suggested that the autoimmune response mediated by *CAMP* may be related to the development of atherosclerosis<sup>49,50</sup>, and these processes may partially account for the significantly raised CAD risk in patients with psoriasis<sup>51</sup>. Additionally, Peter M et al. further confirmed that *CAMP* was a potential autoantigen implicated in the atherosclerotic immune response<sup>52</sup>. Furthermore, previous studies have proven that *CAMP* has been defined as a pro-atherosclerotic molecule<sup>53,54</sup> and that *CAMP*-deficient transgenic mice have a reduced risk of atherosclerosis<sup>55</sup>. In addition, we noticed that *CAMP* was mainly enriched in the following signaling pathways and biological processes: tuberculosis, innate immune response, cell redox homeostasis, cellular response to interleukin-1, and cellular response to tumor necrosis factor. These findings indicated that *CAMP* may be involved in atherosclerosis by mediating autoimmune or inflammatory responses. Furthermore, the independent verification data as well as our RT-qPCR results also revealed that CAD patients had significantly raised *ITGAM*, *TYROBP*, *ICAM1* and *CAMP* expression levels in contrast to healthy subjects. In addition, based on ROC curve analysis, we propose *CAMP* to function as a potential diagnostic biomarker for CAD.

This research had several limitations. Firstly, we have only identified and verified the hub genes that were associated with CAD, but have not constructed the transcriptional regulatory network, which may lead to more meaningful discoveries. Secondly, the sample sources of the three datasets selected in the current research are different and the biological differences will inevitably produce an impact on our findings. Thirdly, there was only one disease phenotype of clinical features in this study, thus, more clinical features are needed in order to further define the phenotype-genotype relationship. Fourthly, this is a single-center study comprising of a small patient number, and large multi-center studies are necessary to validate our findings. Lastly, the molecular mechanisms of *ITGAM*, *TYROBP*, *ICAM1* and *CAMP* involved in CAD are still not fully defined and require further cytology and animal experiments to further outline their respective roles in vivo and in vitro.

In summary, we determined that *ITGAM*, *TYROBP*, *ICAM1* and *CAMP* may possess significant roles in mediating the chronic inflammatory process that eventually culminates in atherosclerosis and CAD. The underlying mechanism may be related to transendothelial migration of leukocytes and the immune response. Independent verification data, combined with our RT-qPCR results were similar to those derived from the microarray analysis, which further increased the credibility of the conclusion.

## Materials and methods

**CAD microarray data sets were used to identify hub genes.** Three microarray data sets originating from individuals with CAD (GSE66360, GSE19339 and GSE97320) were extracted from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) database. This data was based on the GPL570 Affymetrix Human Genome U133 Plus 2.0 array and was used to construct a co-expression network. The expression profile data from 56 CAD samples and 57 normal samples across three data sets were analyzed using integrated analysis. Before we analyzed the data, some powerful and accurate R packages, such as *affy*, *affyPLM*, and *RColorBrewer* were used to perform quality testing of microarray data. Some functions in *affyPLM* package can be used to fit the original data of microarray and generate the weights and residuals diagram, relative the relative log expression (RLE) and the relative standard deviation (NUSE, Normalized unscaled standard errors) box diagram<sup>25</sup>. After confirming that there are no outliers, RMA methods were used to normalize gene expression value matrices which were extracted from the original files in CEL format. Furthermore, *KNN* function of the *impute* package is used to calculate and supplement missing values. *SVA* methods were then used to remove and batch differences via the R software (version 4.0.0)<sup>12</sup>. After this, gene symbols were designated based on probe identification numbers (IDs) using the *Bioconductor* package<sup>13</sup>. Average expression values were used in cases where multiple probe IDs corresponded to the same gene.

**Construction of the weighted gene co-expression network.** WGCNA is a widely used systems biology method that is able to transform gene expression data profiles into a scale-free network<sup>9</sup>. Outlier samples were excluded to maintain the reliability of network construction results. The appropriate soft threshold power (soft power = 14) was chosen with reference standard scale-free networks, with the power function used to calculate adjacency values between all differentially expressed genes. A topological overlap matrix (TOM) was then formulated based on the adjacency values in order to calculate the corresponding dissimilarity (1-TOM) values. Module identification was accomplished with the dynamic tree cut method by hierarchically clustering genes using 1-TOM as the distance measure with a minimum size cutoff of 30 and a deep split value of 2 for the resulting dendrogram. A module preservation function was used to verify the stability of the identified modules by calculating module preservation and quality statistics in the WGCNA package<sup>14</sup>.

**Preservation analysis of five network modules.** Refer to the methods described in a previous study<sup>26</sup>, a composite preservation statistics method based on the module Preservation function in the WGCNA R package was used to verify the conservativeness of the five modules. The Z-summary statistic was used to measure

module density and intramodular connectivity metrics in each module. In the corresponding network, Zdensity (function 1) was used to conduct the 4 density preservation statistics, Zconnectivity (function 2) was used to conduct the 3 connectivity-based statistics, the combines module density and intramodular connectivity metrics was measured by the Zsummary (function 3) and defined as follows: Z density = median (Z meanCor, Z meanAdj, Z propVarExpl, Z meanKME) (function 1); Z connectivity = median (Z cor.kIM, Z cor.kME, Z cor.cor) (function 2); Z summary = (Z density + Z connectivity)/2 (function 3). In addition, if Z summary < 2 indicated no evidence that the module preserved; if  $2 < Z \text{ summary} < 10$  indicated weak to moderate preservation; if Z summary > 10 indicated high preservation among modules. The module size has a strong influence on Z statistics. Therefore, the medianRank for preservation analysis was conducted to compare the preservation statistics of different sized modules. It indicates that modules with lower median rank tend to show better preservation statistics than those with higher median rank.

**Identification of the module of interest and functional annotation.** The relationship between clinicopathological characteristics and modules were discerned using the Pearson correlation analysis in order to determine CAD-related biological modules. Significant gene modules were subsequently processed with the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) online tool (version 6.8; <http://david.abcc.ncifcrf.gov>).  $P < 0.05$  was set as the cutoff criterion.

**Hub gene analysis.** The degree of module membership (MM) was defined as the correlation between module eigengenes (Mes) and gene expression profiles. The correlation between the gene and external traits was determined to be the degree of gene significance (GS). Generally, identified modules with increased MS and GS values were then subjected to further scrutinization for their biological function<sup>15</sup>. The Search Tool for the Retrieval of Interacting Genes database (version 11.0; <http://www.string-db.org>) was used to construct protein–protein interaction (PPI) gene networks based on the chosen module<sup>16</sup>. This network was then visualized with the Cytoscape software<sup>17</sup>, with the most valuable clustering module identified using molecular complex detection (MCODE)<sup>18</sup>. Modules with an MCODE score > 6 were selected for further analysis.

**Sample verification and diagnostic criteria.** Two hundred thirty unrelated patients with CAD were recruited from the Shao Yang Central Hospital. CAD was defined as significant coronary artery stenosis ( $\geq 50\%$ ) in at least one of the three main coronary vessels or their main branches (branch diameter  $\geq 2$  mm)<sup>19</sup>. The diagnostic criteria for the three types of CAD patients are as follows: (i) stable exertional angina (n = 122), defined as episodes with reversible ischemic chest pain. (ii) Non-ST-elevation acute coronary syndromes (NSTE-ACS) that included non-ST-elevated myocardial infarction (NSTEMI) patients and unstable angina (n = 66), defined as angina at crescendo or rest angina. (iii) ST-elevated myocardial infarction (STEMI) (n = 42) defined as elevated plasma levels of Troponin T (TnT; at least one value above the 99th percentile) together with ST-segment elevation or new left bundle branch block in the electrocardiogram and ischemic symptoms. All subjects had no history of autoimmune, thyroid, renal, neoplastic, hematologic, liver disease or type 1 diabetes. The control group consisted of two hundred sixteen healthy controls matched by age, ethnicity (Han Chinese) and gender and were randomly recruited from the Physical Examination Center of the Shao Yang Central Hospital, in the same period. All control individuals were assessed with questionnaires, clinical history and examination to ensure the absence of type 2 diabetes mellitus (T2DM), previous myocardial infarction or CAD as well as ischemic stroke (IS). Written, informed consent was gained from all individuals prior to participation and all experiments were performed in accordance with relevant named guidelines and regulations. The research proposal was approved by the Ethics Committee of the Shao Yang Central Hospital (No: KY 2020-023-08).

**Network-based meta-analysis to verify the identified hub genes.** Acute myocardial infarction (AMI) samples from three datasets (GSE60993, GSE61144 and GSE66360) were selected to verify the identified hub genes, which are based on the platform of GPL6884 Illumina HumanWG-6 v3.0 expression beadchip, GPL6106 Sentrix Human-6 v2 Expression BeadChip and GPL570 Affymetrix Human Genome U133 Plus 2.0 array and used to construct the co-expression network. Raw data (.CEL) were processed using software package in R (version 4.0.0) for microarray quality assessment such as RNA degradation plots, normalized unscaled standard errors (NUSE) and relative log expression (RLE). Several samples with abnormal distribution in each dataset were removed. A robust multi-array average (RMA) algorithm with background adjustment, log transformation and normalization was used to pre-process all data<sup>56</sup>. The Bioconductor package was used to transform the probe identification numbers (IDs) into gene symbols<sup>13</sup>. Average expression values were used in cases where multiple probe IDs corresponded to the same gene. INMEX was used to carry out microarray-based meta-analysis in order to incorporate multiple gene expression datasets<sup>57</sup>, in compliance to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for meta-analysis<sup>58</sup>. Each eligible gene expression profile was uploaded to INMEX. After each gene expression profile integrity check, the ComBat option in INMEX was used to eliminate batch effects between three different gene expression profiles using empirical Bayes methods<sup>59</sup>. To eliminate the differences in platform usage and study design, heterogeneity among microarray datasets, a fixed effect model (FEM) was selected for the meta-analysis in compliance to the between-study heterogeneity based on Cochran's Q test<sup>60</sup>. Differentially expressed genes (DEGs) from the integrated dataset were obtained by the GeneVenn web tool from INMEX<sup>61</sup>.

**Quantitative real-time PCR.** Total RNA was extracted from isolated peripheral blood monocytes (PBMCs) using the TRIzol reagent. cDNA was then reverse-transcribed with the PrimeScript RT reagent kit

Gene	Forward primer	Reverse primer
ITGAM	TCCCGGAAAACTCAGAGGTC	TGAGGCCGTGAAGTTGAGAT
CAMP	AGGTAAGTGTGGAAAGCCTGC	GACCCATTGGATGGTCCACA
TYROBP	CCTCAACTCACCCTCTGCC	TTCAAGGTTGGGGGTGCTT
ICAM1	TCCTCACCGCCTGTTGTATC	ACTTCCCCTCTCATCAGGCT
GAPDH	AGAGAGAGGCCCTCAGTTGCT	TTGTGAGGGAGATGCTCAGTGT

**Table 3.** PCR primers for quantitative real-time PCR.

(Takara Bio, Japan). RT-qPCR was then carried out with the resultant cDNA as a template. Hub gene-specific primers used in these experiments were designed by Sangon Biotech (Shanghai, China) and are detailed in Table 3. An ABI Prism 7500 sequence-detection system (Applied Biosystems, USA) using a Taq PCR Master Mix Kit (Takara) was used to perform quantitative RT-PCR was performed using RT Reaction Mix in a total volume of 20  $\mu$ L with the following reaction conditions: predenaturation at 95 °C for 30 s, then 40 cycles of 95 °C for 30 s and 60 °C for 30 s.

**Diagnostic criteria.** Serum triglyceride (TG; 0.56–1.70 mmol/L), ApoA1 (1.20–1.60 g/L), total cholesterol (TC; 3.10–5.17 mmol/L), apolipoprotein (Apo) B (0.80–1.05 g/L), low-density lipoprotein cholesterol (LDL-C; 2.70–3.10 mmol/L) and high-density lipoprotein cholesterol (HDL-C; 1.16–1.42 mmol/L) and the ApoA1/ApoB ratio (1.00–2.50) levels were defined as their respective normal values at our Clinical Science Experiment Center. The diagnostic criteria of diabetes<sup>20</sup>, hypertension<sup>21</sup>, obesity, normal weight, overweight<sup>22</sup> and hyperlipidemia<sup>23</sup> were based on previous studies.

**Statistical analyses.** All data was analysed with the SPSS (Version 22.0). Continuous data is depicted in terms of mean  $\pm$  SD. Independent-samples t tests were used to assess the general characteristics differences between controls and individuals with CAD patients and controls. The Chi-square test was utilized to evaluate the differences in the amount of alcohol consumers, age distribution and proportion of smokers between controls and individuals with CAD. Referring to previous studies<sup>62,63</sup>, receiver operating characteristic (ROC) curves were built based on plasma levels of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* to evaluate the specificity, sensitivity, and respective areas under the curves (AUCs) with 95% CI. The optimal cut-off value for diagnosis was investigated by maximising the sum of sensitivity and specificity and minimising the overall error (square root of the sum  $[1 - \text{sensitivity}]^2 + [1 - \text{specificity}]^2$ ), and by minimising the distance of the cutoff value to the top-left corner of the ROC curve, and the corresponding empirical ROC curve of *ITGAM*, *CAMP*, *TYROBP* and *ICAM1* were drawn by a nonparametric method using SPSS software (Version 22.0). R software (version 4.0.0) was used to carry out bioinformatic analysis and heat mapping of the correlation models.

Received: 18 August 2020; Accepted: 12 March 2021

Published online: 23 March 2021

## References

- Orrem, H. L. *et al.* IL-6 receptor inhibition by tocilizumab attenuated expression of C5a receptor 1 and 2 in non-ST-elevation myocardial infarction. *Front. Immunol.* **9**, 2035–2035 (2018).
- Roffi, M. *et al.* 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur. Heart J.* **37**, 267–315 (2016).
- Houston, M. The role of noninvasive cardiovascular testing, applied clinical nutrition and nutritional supplements in the prevention and treatment of coronary heart disease. *Ther. Adv. Cardiovasc. Dis.* **12**, 85–108 (2018).
- Zhang, S. *et al.* Critical appraisal of guidelines for coronary artery disease on dual antiplatelet therapy: More consensus than controversies. *Clin. Cardiol.* **42**, 1170–1180 (2019).
- Roth, G. A. *et al.* Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **76**, 2982–3021 (2020).
- Libby, P. & Theroux, P. Pathophysiology of coronary artery disease. *Circulation* **111**, 3481–3488 (2005).
- Herrington, W. *et al.* Epidemiology of atherosclerosis and the potential to reduce the global burden of atherothrombotic disease. *Circ. Res.* **118**, 535–546 (2016).
- Sinnaeve, P. R. *et al.* Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. *PLoS ONE* **4**, e7037 (2009).
- Ntzani, E. E. & Ioannidis, J. P. Predictive ability of DNA microarrays for cancer outcomes and correlates: An empirical assessment. *Lancet* **362**, 1439–1444 (2003).
- Ein-Dor, L. *et al.* Outcome signature genes in breast cancer: Is there a unique set?. *Bioinformatics* **21**, 171–178 (2005).
- Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput. Biol.* **4**, e1000117 (2008).
- Ruan, J., Dean, A. K. & Zhang, W. A general co-expression network-based approach to gene expression analysis: Comparison and applications. *BMC Syst. Biol.* **4**, 8 (2010).
- Le, T. T. *et al.* Identification and replication of RNA-Seq gene network modules associated with depression severity. *Transl. Psychiatry* **8**, 180–180 (2018).
- Wu, S. *et al.* De novo transcriptome assembly of Chinese Kale and global expression analysis of genes involved in glucosinolate metabolism in multiple tissues. *Front. Plant Sci.* **8**, 92–92 (2017).

15. Miao, L. *et al.* Weighted gene co-expression network analysis identifies specific modules and hub genes related to hyperlipidemia. *Cell. Physiol. Biochem.* **48**, 1151–1163 (2018).
16. Xie, L. *et al.* Deep learning-based transcriptome data classification for drug-target interaction prediction. *BMC Genom.* **19**, 667–667 (2018).
17. Kalinin, A. A. *et al.* Deep learning in pharmacogenomics: From gene regulation to patient stratification. *Pharmacogenomics* **19**, 629–650 (2018).
18. Liu, M. *et al.* A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *Neuroimage* **208**, 116459 (2020).
19. Choi, H. & Na, K. J. A risk stratification model for lung cancer based on gene coexpression network and deep learning. *Biomed. Res. Int.* **2018**, 2914280–2914280 (2018).
20. Obuchowski, N. A. Nonparametric analysis of clustered ROC curve data. *Biometrics* **53**, 567–578 (1997).
21. Mallett, S., Halligan, S., Collins, G. S. & Altman, D. G. Exploration of analysis methods for diagnostic imaging tests: Problems with ROC AUC and confidence scores in CT colonography. *PLoS ONE* **9**, e107633–e107633 (2014).
22. Cai, Y. *et al.* Circulating “lncRNA OTTHUMT00000387022” from monocytes as a novel biomarker for coronary artery disease. *Cardiovasc. Res.* **112**, 714–724 (2016).
23. Cai, Y. *et al.* Circulating “lncPPARδ” from monocytes as a novel biomarker for coronary artery diseases. *Medicine* **95**, e2360–e2360 (2016).
24. Yang, Y. *et al.* Plasma long non-coding RNA, CoroMarker, a novel biomarker for diagnosis of coronary artery disease. *Clin. Sci. (Lond.)* **129**, 675–685 (2015).
25. Li, H. *et al.* Co-expression network analysis identified hub genes critical to triglyceride and free fatty acid metabolism as key regulators of age-related vascular dysfunction in mice. *Aging* **11**, 7620–7638 (2019).
26. Chen, J. *et al.* Genetic regulatory subnetworks and key regulating genes in rat hippocampus perturbed by prenatal malnutrition: Implications for major brain disorders. *Aging (Albany N. Y.)* **12**, 8434–8458 (2020).
27. Zhang, Q.-H. *et al.* TRIB1 and TRPS1 variants, G × G and G × E interactions on serum lipid levels, the risk of coronary heart disease and ischemic stroke. *Sci. Rep.* **9**, 2376–2376 (2019).
28. Li, B., Li, W., Li, X. & Zhou, H. Inflammation: A novel therapeutic target/direction in atherosclerosis. *Curr. Pharm. Des.* **23**, 1216–1227 (2017).
29. Bhui, R. & Hayenga, H. N. An agent-based model of leukocyte transendothelial migration during atherogenesis. *PLoS Comput. Biol.* **13**, e1005523 (2017).
30. Glass, C. K. & Witztum, J. L. Atherosclerosis. The road ahead. *Cell* **104**, 503–516 (2001).
31. Zirlik, A. *et al.* CD40 ligand mediates inflammation independently of CD40 by interaction with Mac-1. *Circulation* **115**, 1571–1580 (2007).
32. Young, J. J. Platelet glycoprotein IIb/IIIa inhibition as adjunctive therapy during primary percutaneous coronary intervention for acute ST-segment elevation myocardial infarction. *J. Invasive Cardiol.* **17**, 300–301 (2005).
33. Salam, A. M. & Al Suwaidi, J. Platelet glycoprotein IIb/IIIa antagonists in clinical trials for the treatment of coronary artery disease. *Expert Opin. Investig. Drugs* **11**, 1645–1658 (2002).
34. Ayari, H. & Bricca, G. Identification of two genes potentially associated in iron-heme homeostasis in human carotid plaque using microarray analysis. *J. Biosci.* **38**, 311–315 (2013).
35. Pan, Y. *et al.* Bioinformatics analysis of vascular RNA-seq data revealed hub genes and pathways in a novel Tibetan minipig atherosclerosis model induced by a high fat/cholesterol diet. *Lipids Health Dis.* **19**, 54–54 (2020).
36. Clerc, G. & Rouz, P. M. Lymphocyte subsets in severe atherosclerosis before revascularization. *Ann. Intern. Med.* **126**, 1004–1005 (1997).
37. Mao, Z., Wu, F. & Shan, Y. Identification of key genes and miRNAs associated with carotid atherosclerosis based on mRNA-seq data. *Medicine (Baltimore)* **97**, e9832 (2018).
38. Mäkinen, V.-P. *et al.* Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS Genet.* **10**, e1004502–e1004502 (2014).
39. Wang, H. M., Gao, J. H. & Lu, J. L. Pravastatin improves atherosclerosis in mice with hyperlipidemia by inhibiting TREM-1/DAP12. *Eur. Rev. Med. Pharmacol. Sci.* **22**, 4995–5003 (2018).
40. Koyama, Y. *et al.* Cross-linking of intercellular adhesion molecule 1 (CD54) induces AP-1 activation and IL-1β transcription. *J. Immunol.* **157**, 5097–5103 (1996).
41. Sano, H. *et al.* Cross-linking of intercellular adhesion molecule-1 induces interleukin-8 and RANTES production through the activation of MAP kinases in human vascular endothelial cells. *Biochem. Biophys. Res. Commun.* **250**, 694–698 (1998).
42. Dragoni, S. *et al.* Endothelial MAPKs direct ICAM-1 signaling to divergent inflammatory functions. *J. Immunol. (Baltimore, Md.: 1950)* **198**, 4074–4085 (2017).
43. Lawson, C., Ainsworth, M., Yacoub, M. & Rose, M. Ligation of ICAM-1 on endothelial cells leads to expression of VCAM-1 via a nuclear factor-κB-independent mechanism. *J. Immunol.* **162**, 2990–2996 (1999).
44. Picard, D. *et al.* Increased prevalence of psoriasis in patients with coronary artery disease: results from a case-control study. *Br. J. Dermatol.* **171**, 580–587 (2014).
45. Amaya-Amaya, J. *et al.* Cardiovascular disease in Latin American patients with systemic lupus erythematosus: A cross-sectional study and a systematic review. *Autoimmune Dis.* **2013**, 794383–794383 (2013).
46. Urman, A., Taklalsingh, N., Sorrento, C. & McFarlane, I. M. Inflammation beyond the joints: Rheumatoid arthritis and cardiovascular disease. *Scifed J. Cardiol.* **2**, 1000019 (2018).
47. Sorensen, O. E. *et al.* Human cathelicidin, hCAP-18, is processed to the antimicrobial peptide LL-37 by extracellular cleavage with proteinase 3. *Blood* **97**, 3951–3959 (2001).
48. Lande, R. *et al.* The antimicrobial peptide LL37 is a T-cell autoantigen in psoriasis. *Nat. Commun.* **5**, 5621 (2014).
49. Edfeldt, K. *et al.* Involvement of the antimicrobial peptide LL-37 in human atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1551–1557 (2006).
50. Ciornei, C. D. *et al.* Human antimicrobial peptide LL-37 is present in atherosclerotic plaques and induces death of vascular smooth muscle cells: A laboratory study. *BMC Cardiovasc. Disord.* **6**, 49 (2006).
51. Jung, K. J. *et al.* Increased risk of atherosclerotic cardiovascular disease among patients with psoriasis in Korea: A 15-year nationwide population-based cohort study. *J. Dermatol.* **46**, 859–866 (2019).
52. Mihailovic, P. M. *et al.* The cathelicidin protein CRAMP is a potential atherosclerosis self-antigen in ApoE(−/−) mice. *PLoS ONE* **12**, e0187432–e0187432 (2017).
53. Doring, Y. *et al.* Auto-antigenic protein-DNA complexes stimulate plasmacytoid dendritic cells to promote atherosclerosis. *Circulation* **125**, 1673–1683 (2012).
54. Zhang, Z. *et al.* Mitochondrial DNA-LL-37 complex promotes atherosclerosis by escaping from autophagic recognition. *Immunity* **43**, 1137–1147 (2015).
55. Doring, Y. *et al.* Lack of neutrophil-derived CRAMP reduces atherosclerosis in mice. *Circ. Res.* **110**, 1052–1056 (2012).
56. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
57. Xia, J. *et al.* INMEX—A web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res.* **41**, W63–70 (2013).



58. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* **8**, 336–341 (2010).
59. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
60. Xu, M. *et al.* Re-exploring the core genes and modules in the human frontal cortex during chronological aging: Insights from network-based analysis of transcriptomic studies. *Aging (Albany N. Y.)* **10**, 2816–2831 (2018).
61. Pirooznia, M., Nagarajan, V. & Deng, Y. GeneVenn—A web application for comparing gene lists using Venn diagrams. *Bioinformatics* **1**, 420–422 (2007).
62. Yu, H. *et al.* LEPR hypomethylation is significantly associated with gastric cancer in males. *Exp. Mol. Pathol.* **116**, 104493 (2020).
63. Shen, Q. *et al.* Serum DKK1 as a protein biomarker for the diagnosis of hepatocellular carcinoma: A large-scale, multicentre study. *Lancet Oncol.* **13**, 817–826 (2012).
64. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

## Acknowledgements

The authors acknowledge the essential role of the funding of the Technology Innovation Guidance Program of Hunan Province (No: 2018SK51809 and 2018SK51802).

## Author contributions

P.-F.Z. conceived the study, participated in the design, performed the statistical analyses, and drafted the manuscript. P.L. conceived the study, participated in the design and helped to draft the manuscript. P.-F.Z. and L.-Z.C. carried out the epidemiological survey and collected the samples. Y.-Z.G. performed the statistical analyses. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-86207-0>.

**Correspondence** and requests for materials should be addressed to P.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021