

## Research Article

# Metabolic Pathway Genes Associated with Susceptibility Genes to Coronary Artery Disease

Heng Lu,<sup>1</sup> Yi Chen,<sup>1</sup> and Linlin Li<sup>1,2</sup> 

<sup>1</sup>Department of Cardiovascular Surgery, Union Hospital, Fujian Medical University, Fuzhou, China

<sup>2</sup>Department of Nephrology, Union Hospital, Fujian Medical University, Fuzhou, China

Correspondence should be addressed to Linlin Li; libb1997@sina.com

Received 4 September 2017; Revised 15 November 2017; Accepted 4 December 2017; Published 11 February 2018

Academic Editor: Margarita Hadzopoulou-Cladaras

Copyright © 2018 Heng Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coronary artery disease (CAD) is one of the leading threats to global health. Previous research has proven that metabolic pathway disorders, such as high blood lipids and diabetes, are one of the risk factors that mostly cause CAD. However, the crosstalk between metabolic pathways and CAD was mostly studied on physiology processes by analyzing a single gene function. A canonical correlation analysis was used to identify the metabolic pathways, which were integrated as a unit to coexpress with CAD susceptibility genes, and to resolve additional metabolic factors that are related to CAD. Seven pathways, including citrate cycle, ubiquinone, terpenoid quinone biosynthesis, and N-glycan biosynthesis, were identified as an integrated unit coexpressed with CAD genes. These pathways could not be revealed as a coexpressed pathway through traditional methods as each single gene has weak correlation. Furthermore, sets of genes in these pathways were candidate markers for diagnosis and detection from patients' serum.

## 1. Introduction

Coronary artery disease (CAD) is a leading cause of global death with several risk factors, including smoking, hypertension, diabetes, obesity, high blood lipids, and stress [1–4]. The current risk assessment of CAD is based on the determination of these factors. Moreover, avoiding these risk factors is the most efficient means of preventing CAD [5, 6]. Generally, almost all risk factors of CAD are related to metabolism in views of basic biological research. For example, high blood lipids, which play an important role in causing CAD, is regulated by a complex metabolic network that involves the biosynthesis and degradation of cholesterol, triglycerides, and lipoproteins [7–10]. Another risk factor, diabetes, is a metabolic disease related to glucose [11, 12]. The relationship between diet and risk of CAD also suggests that the metabolic pathway plays a pivotal role in the pathophysiology of CAD. Recent studies revealed that vegetarians as well as Mediterranean diet people have less cases of CAD [13–17]. In general, pathophysiology and dietetics research have determined the close link between metabolism and CAD.

In another side, molecular genetics research also accumulated several evidences which suggest that genetic factors contribute lots to the susceptibility of CAD. Although, in comprehensive studies, susceptible genes to CAD have been revealed to be involved in diverse biological pathways such as inflammation, innate immunity, and cholesterol metabolism. However, genome-wide crosstalks between susceptibility genes and metabolism pathways still remained to be characterized.

To understand the crosstalk between genes in the genome level, the transcriptome network is an efficient way to mining important interaction. The most common model for transcriptome network analysis is coexpression which calculates the correlation between genes in certain conditions. However, the traditional approach is only suitable for a single gene pair. So far, limited transcriptome network analysis works with integrated pathways which is consisted of many genes [18]. Canonical correlation analysis (CCA) is a powerful method to measure the coexpression between two sets of genes [19]. To detect the genetic regulatory variants on the Childhood Asthma Management Program,

researchers used CCA to successfully detect candidate genes. The genes related to the glioma pathway were also identified by CCA from the transcriptome of glioblastoma in patients. The transcriptome plays a key role in studies on AIDS restriction genes using the CCA-determined purine metabolism pathway [20].

In this study, CCA was used to identify the coexpression between integrated metabolic pathways and CAD genes. The most significant metabolic pathways that could be a potential candidate disease marker for CAD diagnosis are discussed.

## 2. Materials and Methods

**2.1. Datasets.** Human genome expression datasets were downloaded from the website of COPRESDB (<http://coxpresdb.jp/>), which contains approximately 4000 experiments and expression data on 20,000 human genes. Metabolic pathway genes were downloaded from KEGG (<http://www.kegg.jp>), which includes 129 typical metabolic pathways with predicted genes. The CAD genes were collected from published literature. Two expression datasets were generated to include the metabolic pathway gene and CAD expression data.

**2.2. CCA.** CCA was performed to dissect the correlations between CAD and metabolic pathway gene expression. In concept, CCA could integrate multiple correlations into an abstract group. The correlation level between two sets of genes can be determined by scoring the relationship between the groups. In statistics, this method generates independent pairs of new variables from the original two sets of variables, namely, the canonical variable. CCA is a method processing cross-covariance matrices. If there are correlations among the variables from two vectors  $\mathbf{X} = (X_1, \dots, X_n)$  and  $\mathbf{Y} = (Y_1, \dots, Y_m)$  of random variables, CCA could detect linear combination of the  $X_i$  and  $Y_j$  with maximum correlation. As a result, the canonical variable is a linear combination of the original variables.

Vector  $\mathbf{A}_i = (a_1, a_2, a_3, \dots, a_n)$  describes the CAD gene  $i$  with different expression levels from different experiments (1 to  $n$ ). Vector  $\mathbf{B}_i = (b_1, b_2, b_3, \dots, b_n)$  describes the metabolic gene  $i$  with different expression levels from different experiments (1 to  $n$ ). Matrix  $\mathbf{M} = (A_1, A_2, A_3, \dots, A_n)$  is for CAD genes, and matrix  $\mathbf{N}_i = (B_1, B_2, B_3, \dots, B_n)$  is for a certain metabolic pathway  $i$ . CCA were calculated between matrix  $\mathbf{M}$  and each  $\mathbf{N}_i$  by following the steps in the *R* program:

- (1) *zscore* ( $M$ ) and *zscore* ( $N_i$ );
- (2)  $r_{11} = \text{corr}(\text{zscore}(M))$ ,  $r_{22} = \text{corr}(\text{zscore}(N_i))$ ;
- (3)  $[AA, BB, r, U, V, \text{stats}] = \text{canoncorr}(\text{zscore}(M), \text{zscore}(N_i))$ ;
- (4)  $r_{1A} = AA^* r_{11}$ ,  $r_{2B} = BB^* r_{22}$ ;
- (5)  $S_a = \text{sum}(r_{1A}^2 / \text{length}(r_{1A}))$ ,  $S_b = \text{sum}(r_{2B}^2 / \text{length}(r_{2B}))$ ;
- (6) Screen the candidate CCA using a certain threshold for  $r$  *P* value,  $S_a$ , and  $S_b$ . (in this study, the thresholds were 0.6, 0.00001, 0.3, and 0.15)

$S_a$  and  $S_b$  were used to validate the percentage of genes in total CADs or a certain metabolic pathway, which could be presented by canonical variables. A high level of  $S_a$  or  $S_b$  indicates that more genes belonging to the group are involved in the correlation. Thus, not only the correlation between gene pairs was identified but also the correlation between sets.

**2.3. Software Tools.** CCA was performed using the *R* platform (<http://www.r-project.org/>). The web-based DAVID tool (<http://david.abcc.ncifcrf.gov/>) was used for gene ontology enrichment analysis for the metabolic pathway. The STRING platform was used for the contraction of the PPI network.

**2.4. Sample Collection.** Fifty serum samples from Chinese patients (Union Hospital of Fujian Medical University) were collected and divided into two groups, namely, CAD (from CAD patients) and control (from healthy people). The blood samples were maintained at 37°C for 1 h and centrifuged at 3000 g/min for 20 min at 4°C, and the supernatant fluid is collected at -80°C until testing. All clinical research experiments were approved by the Medical Ethics Committee of Union Hospital of Fujian Medical University and performed according to the approved guidelines.

**2.5. RNA Isolation, cDNA Synthesis, and q-PCR Analysis.** The total RNA was extracted using TRIzol reagent (Life Technologies), following the manufacturer's instructions with slight modifications. Up to 1  $\mu\text{g}$  of the total RNA was used by utilizing the First Strand cDNA Synthesis Kit (TOYOBO) to generate cDNA according to the manufacturer's instructions. q-PCR was performed using the TransStart Top Green q-PCR SuperMix (TransGen Biotech) and was conducted using the iQTM5 multicolor real-time PCR detection system (Bio-Rad).

## 3. Results and Discussion

**3.1. Strategy.** In this study, a clustering for the expression pattern of CAD and metabolic pathway genes was first created to estimate the possible gene group, which could provide an evidence for identifying the correlation among gene groups. Next, a coexpression network was constructed among individual CAD and metabolic pathway genes. This step is prepared for comparing with CCA. Then, CCA was performed to identify the canonical variables that delegate the integrated CAD and metabolic pathway genes (Figure 1). In particular, additional genes in a pathway could be regulated by the same factor. Thus, the whole pathway was suggested to be controlled by the same factor. In this study, an entire metabolic pathway was correlated with the entire CAD genes.

**3.2. CAD and Metabolic Pathway Genes.** A published literature shows that SNPs associated with coronary heart disease and mean arterial pressure were collected for further analysis. There are Mendelian forms of CAD that show mutations in genes that are related to low- and high-density lipoprotein metabolism or homeostasis. The heritability of fatal coronary events is estimated at 57% and 38% in monozygotic and dizygotic twin cohorts, respectively. Researchers have long recognized the familial clustering of CAD. The first linkage of

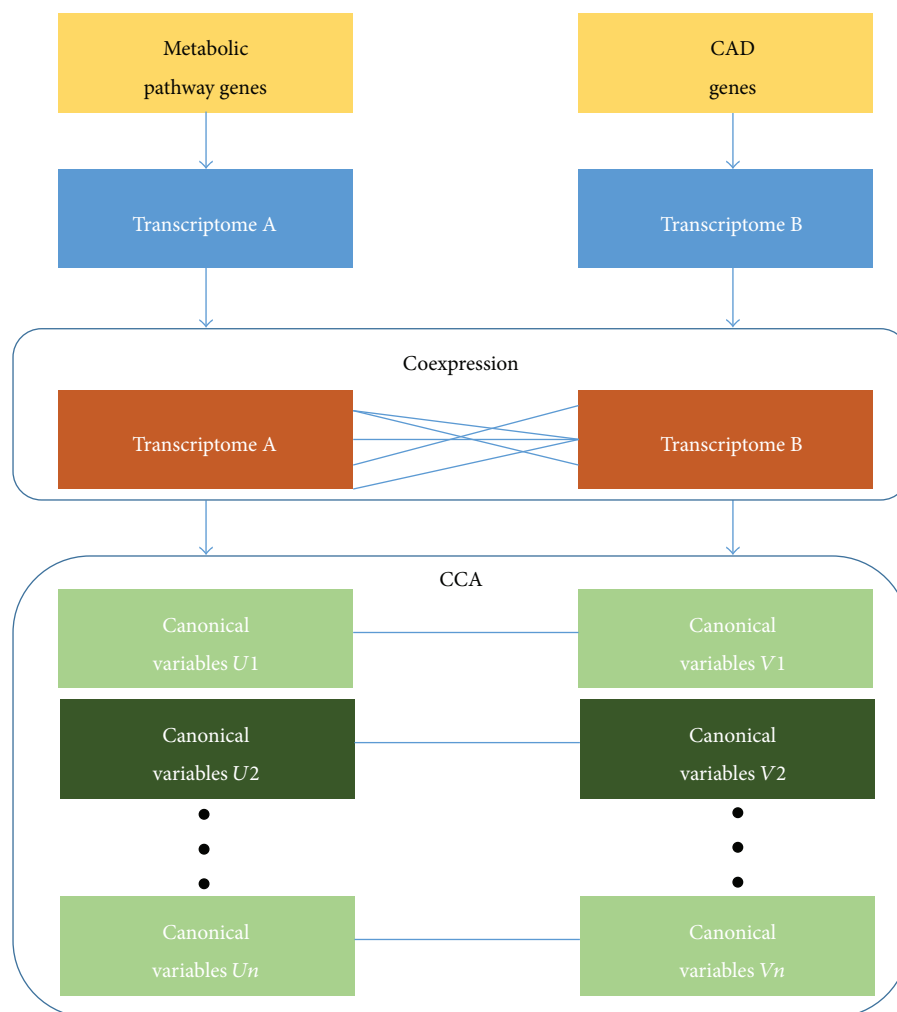


FIGURE 1: Strategy of study for the canonical correlation analysis. Collection of CAD genes and metabolic pathway genes; construction of the coexpression network among individual CAD and metabolic pathway genes; performance of CCA; and analysis of canonical variables.

premature CAD to loci on chromosomes 2q21.2-22 and Xq23-26 was found in Finland [21]. Subsequently, several studies from Germany, the United States, and England identified a novel chromosome locus linked with CAD [22]. Different genes, such as ALOX5AP (5-lipoxygenase-activating protein), are characterized by haplotypes that are associated with CAD. Several genome-wide association studies promoted by the improved genome sequence technology are performed on CAD. A novel replicated locus on chromosome 17 was found by genome-wide mapping of susceptibility to CAD [23]. In Indo-Mauritians, a susceptibility locus on chromosome 16p13 is revealed by a genome-wide scan for CAD [24]. Recently, a new locus on chromosome 10p11.23 for CAD is discovered by a genome-wide association study [25]. In addition, a genome-wide linkage analysis for CAD and its related risk factors is conducted [26].

Eight studies identified 373 SNPs associated with coronary heart disease-related phenotype at a  $10E5$   $P$  level. These SNPs are located at several chromosomes inside coding or noncoding regions. In general, the coding regions match 120 genes involved in diverse biological processes, such as

response to a stimulus, developmental, rhythmic, metabolic, and immune system processes (Figure S1, Table S1). Moreover, the expression of CAD-associated genes was controlled during the progress of the disease. Thus, the expression levels of 128 genes were collected from human transcriptome database with approximately 5000 experiments.

Then, genes that are related to metabolic pathways were collected from the KEGG database. In general, 84 metabolic pathways, including metabolism of carbohydrate, energy, lipid, nucleotide, amino acid, glycan biosynthesis, cofactors and vitamins, terpenoids and polyketides, and biosyntheses of other secondary metabolites (Table S2), cover major metabolic processes in human biology. A total of 300 gene code putative enzymes exist for these pathways according to the current knowledge on the human genome. Currently, certain metabolic pathways, such as D-arginine and D-ornithine metabolism, valine, leucine and isoleucine biosyntheses, lysine biosynthesis, and biotin and lipoic acid metabolism, contain only 1–3 genes. These pathways have a limited value for further CCA. A total of 77 pathways, such as glycolysis/gluconeogenesis, TCA cycle, and pentose phosphate

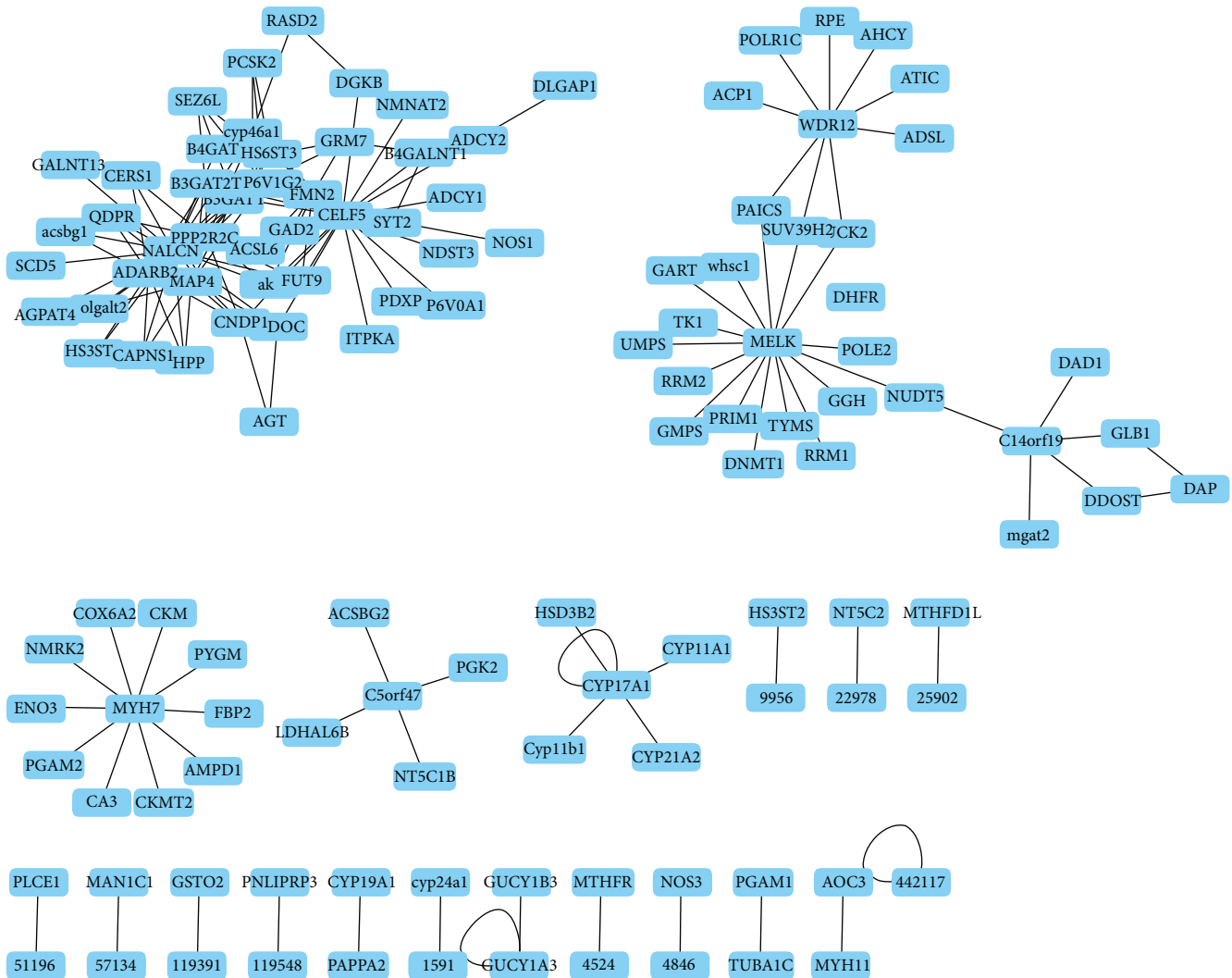


FIGURE 2: Coexpression network of CAD genes and metabolic pathway genes. The connection indicates the Pearson correlation factor  $r > 0.6$ .

pathway (Table S2), have more than five genes in the human genome. These pathways could be subject to CCA because we aim to estimate the correlation between the integrated CAD gene group and metabolic gene pathway, not the individual gene.

**3.3. Coexpression and the PPI Network.** There are huge transcriptome data related to CAD carried out by microarray and next-generation sequencing. For example, researchers establish a mouse heart transcriptomic network sensitivity to various heart diseases using microarray [27]. Cardiac hypertrophy was also studied by transcriptome microarrays to identify long noncoding RNAs [28]. A gene network was established in low-dose, radiation-affected cardiomyocytes [29]. Many transcriptional regulators in human ischemic cardiomyopathy are revealed from the gene expression network [30]. Transcriptome sequencing of different cell lines identifies several candidate genes related to coronary artery calcification [31]. Some potential candidate genes for heart failure were also discovered by transcriptome network

analysis [32]. Transcriptome network analysis is also used to find divergent cardiovascular disease pathways between animal and humans [33].

A coexpression network was constructed using the Pearson correlation to determine the relationship between all CAD susceptibility and metabolic pathway genes. Numerous genes have connections with high  $r$  value ( $r > 0.6$ ). A total of 24 CAD genes, namely, 105 (*ADARB2*), 183 (*AGT*), 1586 (*CYP17A1*), 1611 (*DAP*), 2917 (*GRM7*), 2982 (*GUCY1A3*), 4134 (*MAP4*), 4625 (*MYH7*), 4629 (*MYH11*), 5126 (*PCSK2*), 5522 (*PPP2R2C*), 9229 (*DLGAP1*), 9833 (*MELK*), 23544 (*SEZ6L*), 23551 (*RASD2*), 55017 (*C14orf119*), 55759 (*WDR12*), 56776 (*FMN2*), 60676 (*PAPP2*), 60680 (*CELF5*), 84790 (*TUBA1C*), 127833 (*SYT2*), 133491 (*C5orf47*), and 259232 (*NALCN*), were identified in this coexpression network (Figure 2, Table S3). The network clearly shows several disconnected subnetworks. Six subnetworks, namely, 9833 (maternal embryonic leucine zipper kinase (*MELK*)), 55759 (WD repeat domain 12 (*WDR12*)), 55017 (*C14orf119*), 4625 (myosin, heavy chain 7, cardiac muscle (*MYH7*)), 133491

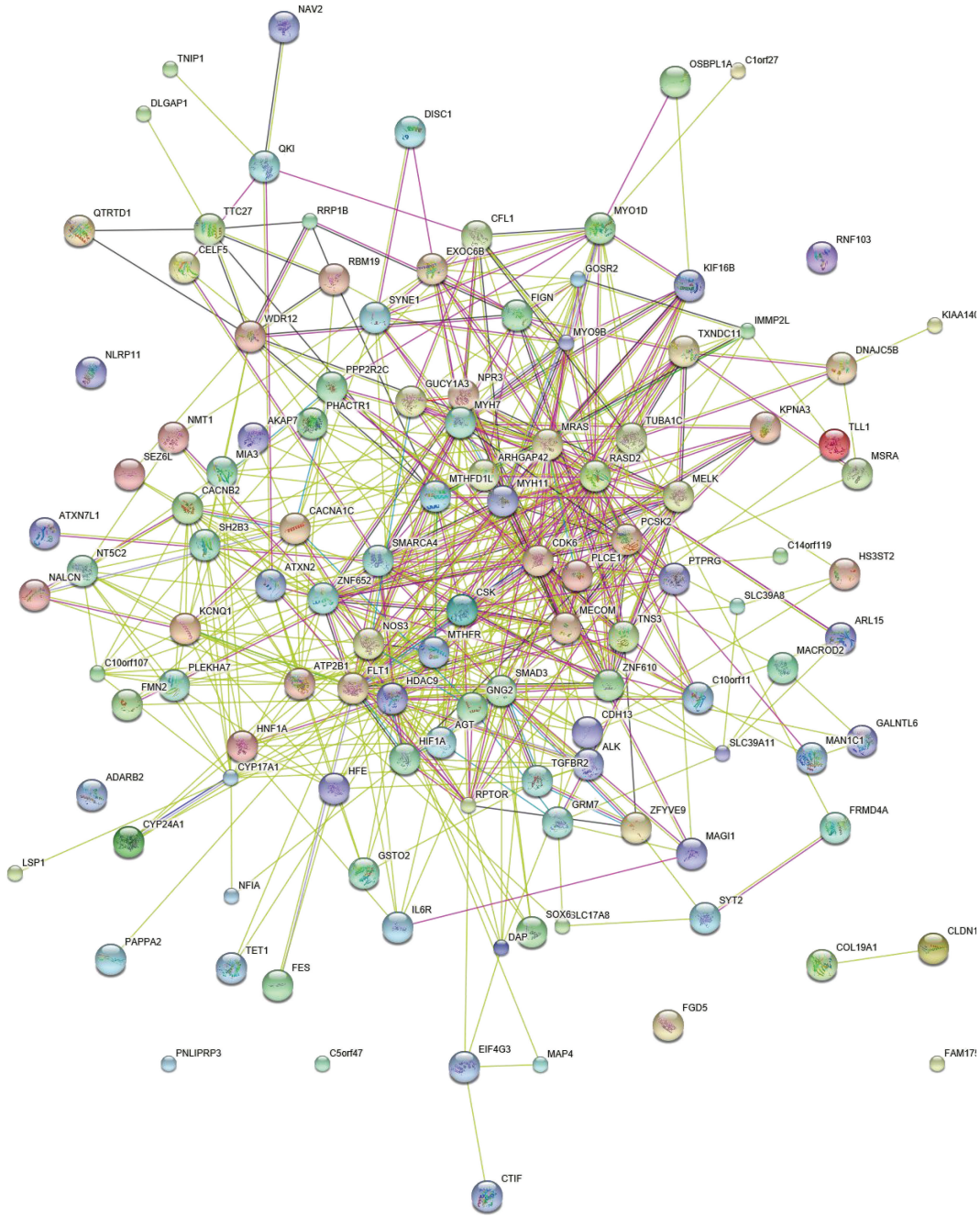


FIGURE 3: PPI network of CAD genes. Note that no metabolic pathway was enriched in the PPI network.

(C5orf47), and 1586 (cytochrome P450, family 17, subfamily A, polypeptide 1 (CYP17A1)) (Figure 2), have different genes at the center. In addition, one subnetwork has multiple genes at the center. In this coexpression network, CAD genes were connected to 80 metabolic pathway genes (Table S4). CAD genes were coexpressed with the metabolic pathway. These metabolic pathway genes were enriched in a diverse metabolic pathway, such as purine metabolism, biosynthesis of antibiotics, pyrimidine metabolism, and glycolysis/gluconeogenesis ( $P$  value <  $10E-03$ ). Furthermore, several

metabolic pathways with less than four genes coexpressed with CAD genes. For example, 23205, 81616, and 23305 belonging to fatty acid biosynthesis exist in this network. The coexpression network between CAD and metabolic pathway genes could not identify the connection between the integrated pathways but simply link from point to point. Furthermore, a PPI network was constructed using CAD genes as baits with low constraints by the STRING platform (Figure 3). Several interactions between CAD genes were observed. Moreover, many interaction proteins were identified through CAD baits. However, the KEGG

TABLE 1: Enriched metabolic pathway in the coexpression network.

Term	Count	Pop	<i>P</i> value	Genes
hsa00230: purine metabolism	17	176	8.46E-11	5427, 9533, 2618, 6240, 5557, 11164, 6241, 26289, 270, 158, 93034, 8833, 10606, 471, 108, 107, 2983
hsa01130: biosynthesis of antibiotics	14	212	6.52E-07	6120, 5224, 5232, 5223, 230, 2618, 26289, 270, 158, 8789, 92483, 10606, 471, 2027
hsa00240: pyrimidine metabolism	10	104	2.40E-06	93,034, 5427, 9533, 7372, 7371, 6240, 5557, 7298, 6241, 7083
hsa00010: glycolysis/gluconeogenesis	7	67	1.06E-04	92483, 5224, 5232, 5223, 230, 2027, 8789
hsa00670: one carbon pool by folate	4	20	0.001425311	1719, 471, 2618, 7298
hsa01230: biosynthesis of amino acids	6	74	0.001495954	6120, 5224, 5232, 5223, 230, 2027
hsa01200: carbon metabolism	7	113	0.001764573	6120, 5224, 5232, 5223, 230, 2027, 8789
hsa04913: ovarian steroidogenesis	5	49	0.002269906	1583, 108, 107, 3284, 1588
hsa00330: arginine and proline metabolism	5	50	0.002446109	84735, 1158, 4842, 57571, 1160
hsa00140: steroid hormone biosynthesis	5	58	0.004206383	1583, 1589, 3284, 1588, 1584
hsa00410: beta-alanine metabolism	4	31	0.005135035	84735, 2572, 57571, 8639
hsa00514: other types of O-glycan biosyntheses	4	31	0.005135035	10690, 135152, 27087, 23127
hsa00061: fatty acid biosynthesis	3	13	0.00927769	23205, 81616, 23305
hsa00790: folate biosynthesis	3	14	0.010744279	5860, 1719, 8836
hsa04925: aldosterone synthesis and secretion	5	81	0.013544109	1583, 108, 107, 1589, 3284
hsa00983: drug metabolism—other enzymes	4	46	0.015345429	8833, 7372, 7371, 7083
hsa01212: fatty acid metabolism	4	48	0.017202919	23205, 79966, 81616, 23305
hsa04922: glucagon signaling pathway	5	99	0.026291005	92483, 5224, 5223, 5837, 108
hsa00534: glycosaminoglycan biosynthesis—heparan sulfate/heparin	3	24	0.030276378	222537, 9348, 266722
hsa03320: PPAR signaling pathway	4	67	0.040871389	23205, 79966, 81616, 23305
hsa00030: pentose phosphate pathway	3	29	0.042939357	6120, 230, 8789
hsa00760: nicotinate and nicotinamide metabolism	3	29	0.042939357	93034, 27231, 23057

enrichment analysis showed that all these genes only were enriched in adrenergic signaling in cardiomyocytes, tight junction, and glutamatergic synapse. No metabolic pathway was found while enriched KEGG is processed in the PPI network.

**3.4. CCA of CAD Susceptibility and Metabolic Pathway Genes.** CCA was used to resolve canonical variables, correlation coefficient, and standard deviation to determine the significant canonical correlations between CAD and metabolic pathway genes in the transcriptome. The canonical variables are a new set of statistically independent variable pairs generated from the original variables, which describe the level of gene expression in different experiments. The canonical variables, which are independent, could be linearly combined to create original variables and could be considered as a component of the original variables. We set a standard to evaluate the canonical variables, which could delegate an integrated pathway at a maximum. Thus, we isolated the canonical variables with large  $S_a$ ,  $S_b$  ( $>0.2$ ),  $r$  values ( $>0.5$ ), and minimal  $P$  values ( $<0.001$ ). The standard deviations indicated the extent of coverage of canonical variables over the entire pathway. Thus, we selected metabolic pathways with more than 10 genes in the genome for further analysis to identify the correlation among the integrated pathways or among single genes. Table 1 indicated that only seven sets of

canonical variables that satisfy all our requirements were detected. These variables are related to seven metabolic pathways, including TCA cycle, ubiquinone and other terpenoid quinone biosyntheses, N-glycan biosynthesis, other glycan degradation, glycosaminoglycan degradation, and glycosylphosphatidylinositol (GPI) anchor biosynthesis and glycosphingolipid biosynthesis—ganglioseries. None of these metabolic pathways were identified by the Pearson correlation or PPI networks. The correlation coefficient factors were more than 0.9 (Figure 4(a)).  $S_a$  indicated that the standard deviation for metabolic pathways had a maximum of 0.3 Table 1.  $S_b$  indicated that the standard deviation of CAD genes varies from 0.19 to 0.93 Table 1. The results suggested that the integrated coexpression in metabolic pathways is not at high level. However, CAD genes were highly integrated to coexpress with metabolic pathways.  $S_b$  was equal to 0.93 for coexpression between CAD genes and ubiquinone and other terpenoid quinone biosyntheses, indicating that 93% of CAD genes can be delegated by canonical variables (Figure 4(a)–4(c)). The results were discussed with published literature by showing the interaction between metabolic pathways and CAD genes.

**3.5. TCA Cycle.** Table 1 indicates that the canonical variable represents 20% of the variability in the original expression pattern of TCA cycle pathway genes and 19% of the

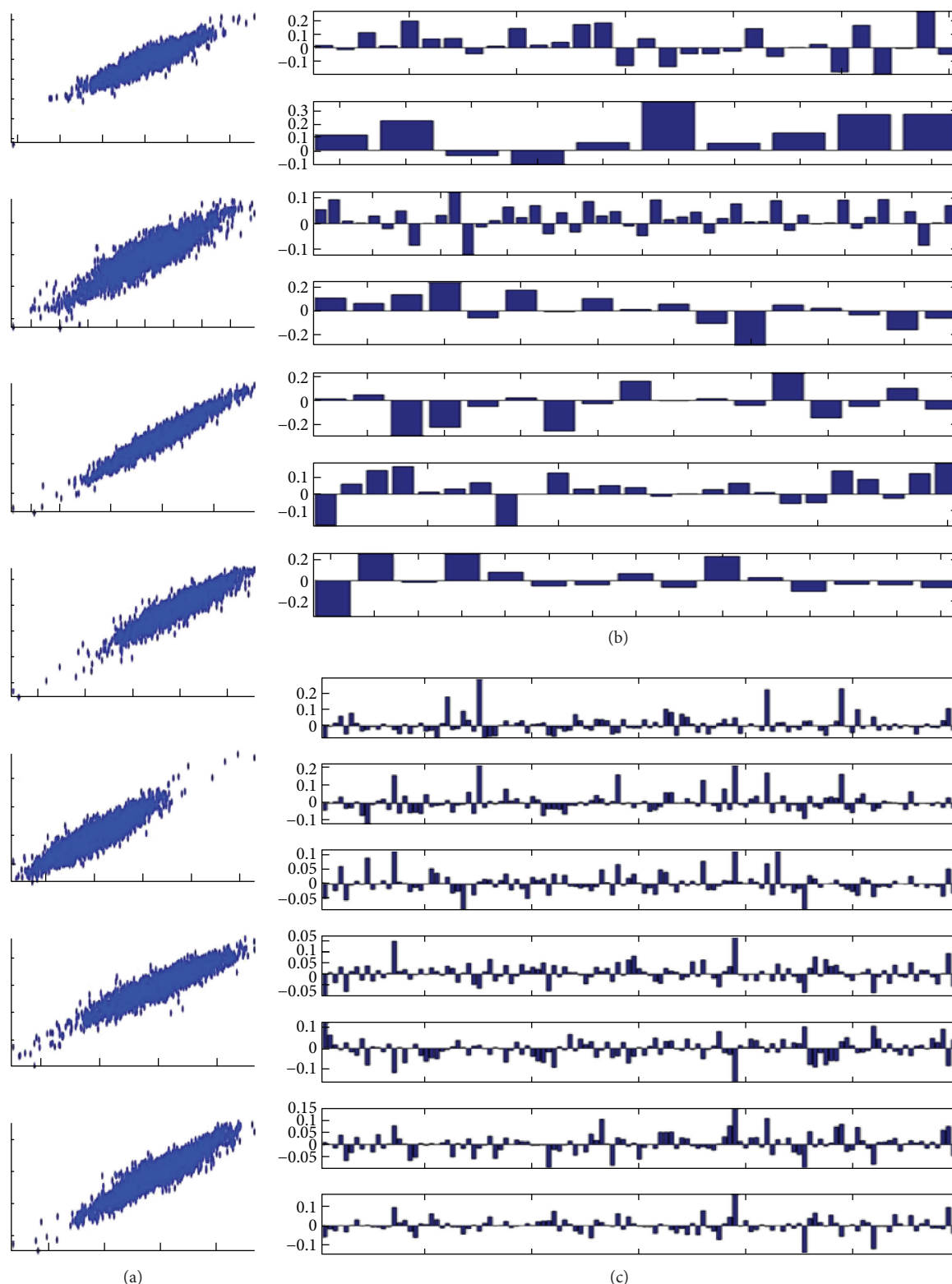


FIGURE 4: Characteristic of canonical variables indicating integrate pathway. A threshold was set to identify the CCA which cover whole pathway ( $S_a, S_b > 0.2$ ,  $r$  values  $> 0.5$ , and  $P$  values  $< 0.001$ ). (a) Scatter plots of seven CCA pairs with top  $r$  values. Note that all identified CCA were highly correlated. (b) The coefficient of metabolic pathway genes on seven canonical variables. Each metabolic pathway contains different number of genes. The genes with big values contributed more in CCA. (c) The coefficient of CAD genes on seven canonical variables. The genes with big values contributed more in CCA.

variability in the expression pattern of CAD genes. The top 6 genes with big absolute value of coefficient factor is selected as representative genes for this canonical variable (Figure 4(b), Table 1). These genes were 8803 (*SUCLA2*), 1737 (*DLAT*), 8801 (*SUCLG2*), 4191 (*MDH2*), 6392 (*SDHD*), and 4190 (*MDH1*).

In the last decades, several studies provide evidence that supports the interaction between CAD and TCA cycle. For example, the degradation of myocardial aspartate and glutamate is induced by the complete citric acid cycle, ischemia, and hypoxia. This process resulted in the production of succinate in isolated hearts [34]. Also, it has been revealed the ATP production in the absence of oxygen through this channeling which could provide cardioprotection [35]. In another case, the accumulation of TCA cycle intermediate fumarate is beneficial for stabilizing hypoxia-inducible factor 1 $\alpha$  (HIF-1 $\alpha$ ), which is a key transcription factor that controls the response to hypoxia and myocardial protection [36]. Together, it is proposed that manipulation of the citrate cycle could increase cardioprotection.

**3.6. Ubiquinone and Other Terpenoid Quinone Biosyntheses.** Table 1 presents that the canonical variable comprises 30% of the variability in the original expression pattern of ubiquinone and other terpenoid quinone biosynthesis pathway genes and 93% of the variability in the expression pattern of CAD genes. The top three genes with a high absolute value of coefficient factor, namely, 27235 (*COQ2*), 84274 (*COQ5*), and 79001 (*vkorc1*), are selected to represent the genes for this canonical variable (Figure 4(b), Table 1).

The key role of ubiquinone in CAD has been accepted by several researchers extensively. Coenzyme Q10, which is a kind of ubiquinone, is supplemented to improve diastolic heart functions, especially in CAD patients. Moreover, it is the major method of treating current diseases [37]. Coenzyme Q10, which could reduce oxidative stress and induce antioxidant activity in CAD patients, is the proposed mechanism [38]. This method is helpful for low-density lipoprotein cholesterol and vitamin E. In particular, this mitochondrial coenzyme is a vital factor for producing ATP [39]. These functions of coenzyme Q10 in antioxidant and energy production imply its role in coronary revascularization. The research demonstrates a significant correlation between the risk of CAD and levels of coenzyme Q10 [40]. Many patients with cardiovascular diseases, such as CAD, cardiomyopathy, congestive heart failure, angina pectoris, and hypertension, showed a deficiency in coenzyme Q10. The deficiency of coenzyme Q10 is a risk factor for cardiovascular disorder, especially increasing the early mortality in myalgic encephalomyelitis/chronic fatigue syndrome [41]. Current studies have proposed that low plasma coenzyme Q10 should be considered a risk factor for CAD and a marker for chronic fatigue in depression [42]. Coenzyme Q10 has been regarded as an independent factor of mortality in congestive heart failure. Furthermore, coenzyme Q10 improves the immune system in vertigo and Meniere' disease-like syndrome [43]. In general, coenzyme Q10 is significant in clinical therapy and diagnosis of CAD patients.

**3.7. Glycosaminoglycan Degradation.** As shown in Table 1, the canonical variable represents 22% of the variability in the original expression pattern of glycosaminoglycan degradation pathway genes and 69% of the variability in the expression pattern of CAD genes. The top four genes with a high absolute value of coefficient factor, namely, 2720 (*GLB1*), 3074 (*HEXB*), 6677 (*SPAM1*), and 2799 (*GNS*), are selected to represent the genes for this canonical variable (Figure 4(b), Table 1).

Studies on the link between glycosaminoglycan and CAD were limited in the last decades. However, the clues were already revealed by research on certain kinds of glycosaminoglycan with antioxidative activity. A pilot study has been conducted with sulodexide, a glycosaminoglycan that primarily consists of heparin, which mainly causes oxidative stress [44]. Sulodexide treatment could decrease only the level of plasma 8-isoprostane but not LDL cholesterol, triglycerides, fibrinogen, and C-reactive protein. However, a significant reduction in oxidative stress in CAD patients administered with sulodexide was observed, implying the important role of plasma 8-isoprostane and glycosaminoglycan. A study showed that another long-chain glycosaminoglycan-hyaluronan could interact with versican to create matrices, which are required for arterial smooth muscle growth during the vascular disease [45]. The histopathological studies on coronary arteries revealed that abundant glycosaminoglycan is associated with percutaneous transluminal coronary angioplasty [46].

**3.8. Glycosphingolipid Biosynthesis–Ganglioseries.** In Table 1, the canonical variable accounts for 22% of the variability in the original expression pattern of glycosphingolipid biosynthesis–ganglioseries pathway genes and 92% of the variability in the expression pattern of CAD genes. The top three genes with a high absolute value of coefficient factor, namely, 2583 (*B4GALNT1*), 2720 (*GLB1*), and 3074 (*HEXB*), are selected to represent genes for this canonical variable (Figure 4(b), Table 1).

The role of glycosphingolipid biosynthesis–ganglioseries in CAD disease was also revealed by recent studies. Glycosphingolipids act as messengers that are integrated into the cell membrane to transduce the growth factor. High levels of cholesterol in the blood are proven to be associated with glycosphingolipid. Recent studies concluded that glycosphingolipids were involved in the major lipoprotein classes in normal and dyslipoproteinemic sera [47]. Cholesterol efflux could be inhibited by the accumulation of glycosphingolipid [48]. The cellular cholesterol homeostasis is induced by glycosphingolipid storage [49]. Studies found that arterial stiffness and atherosclerosis could be ameliorated by inhibiting glycosphingolipid synthesis in mice and rabbit models given the pivotal role of cholesterol in CAD [50].

**3.9. Other Pathways without Literature.** In Table 2, the canonical variable comprises 23% of the variability in the original expression pattern of N-glycan biosynthesis pathway genes and 28% of the variability in the expression pattern of CAD genes. The top 11 genes with a high absolute value of coefficient factor, namely, 4247 (*mgat2*), 4248



TABLE 2: The canonical variable delegate integrates metabolic pathways.

Metabolic pathways	Gene number	$r$	$S_a$	$S_b$	Represent gene ID
Citrate cycle (TCA cycle)	30	0.90	0.20	0.19	8803*, 1737*, 8801*, 4191*, 6392*, 4190*
Ubiquinone and other terpenoid quinone biosyntheses	10	0.90	0.30	0.93	27235, 84274, 79001
N-glycan biosynthesis	48	0.96	0.23	0.28	4247, 4248, 85365, 1650, 11282, 79868, 57134, 8813, 4122, 146664, 29929
Other glycan degradation	17	0.92	0.27	0.62	10825, 2720, 3074, 129807, 2519
Glycosaminoglycan degradation	17	0.91	0.22	0.69	2720, 3074, 6677, 2799
Glycosylphosphatidylinositol (GPI) anchor biosynthesis	25	0.93	0.21	0.43	9091, 284098, 2822, 5281, 5279
Glycosphingolipid biosynthesis–ganglioseries	15	0.93	0.22	0.92	2583, 2720, 3074

$S_a$  and  $S_b$  indicate the extent of canonical variables covering the whole pathway. The representative genes were selected based on the percentage of covering (\*gene number in pathways  $S_b$ ).

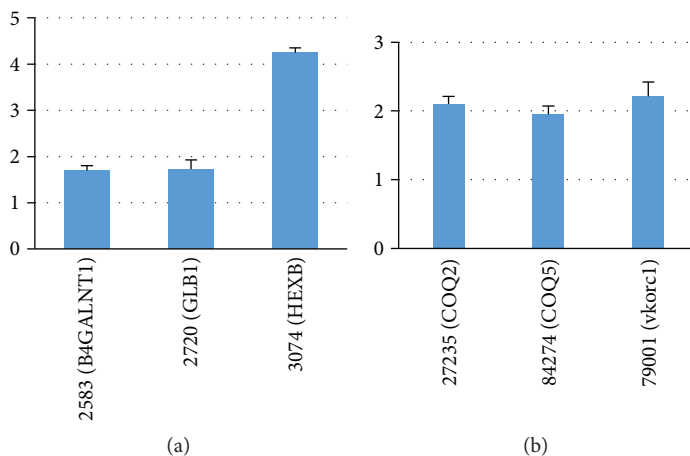


FIGURE 5: qRT-PCR of metabolic pathway genes in CAD patients. (a) 2583 (*B4GALNT1*), 2720 (*GLB1*), and 3074 (*HEXB*) in glycosphingolipid biosynthesis. (b) 27235 (*COQ2*), 84274 (*COQ5*), and 79001 (*vkorc1*) in ubiquinone and other terpenoid quinone biosyntheses.

(*MGAT3*), 85365 (*ALG2*), 1650 (*DDOST*), 11282 (*MGAT4B*), 79868 (*ALG13*), 57134 (*MAN1C1*), 8813 (*DPM1*), 4122 (*MAN2A2*), 146664 (*MGAT5B*), and 29929 (*ALG6*), are selected to represent the genes for this canonical variable (Figure 4(b), Table 1).

As shown in Table 1, the canonical variable explains 27% of the variability in the original expression pattern of other glycan degradation pathway genes and 62% of the variability in the expression pattern of CAD genes. The top five genes with a high absolute value of coefficient factor, namely, 10825 (*NEU3*), 2720 (*GLB1*), 074 (*HEXB*), 129807 (*NEU4*), and 2519 (*FUCA2*), are selected to represent the genes for this canonical variable (Figure 4(b), Table 1).

Table 1 also shows that the canonical variable represents 21% of the variability in the original expression pattern of the GPI anchor biosynthesis pathway genes and 43% of the variability in the expression pattern of CAD genes. The top five genes with a high absolute value of coefficient factor, namely, (*PIGQ*), 284098 (*PIGW*), 2822 (*GPLD1*), 5281 (*PIGF*), and 5279 (*PIGC*), are selected to represent the genes for this canonical variable (Figure 4(b), Table 1).

**3.10. Validation of Expression Genes as Candidate Disease Markers for CAD.** To explore the potential application of

these metabolic pathway genes as disease markers for CAD, we used q-PCR for analyzing the expression pattern of CAD genes and its canonical correlated metabolic pathways. As more than 90% of CAD genes are associated with ubiquinone and other terpenoid quinone biosyntheses and glycosphingolipid biosyntheses–ganglioseries, we selected genes from both metabolic pathways for further analysis. Fifty serum samples were collected from the CAD patients and the control group. The expression levels of metabolic genes, namely, 2583 (*B4GALNT1*), 2720 (*GLB1*), and 3074 (*HEXB*), were determined by q-PCR. Except 3074 (*HEXB*) with 4 times induction, 2583 (*B4GALNT1*) and 2720 (*GLB1*) only increased 1.7 times in CAD patients (Figure 5(a)). The metabolic genes showed significant difference between CAD patients and the control group with a  $P$  value of less than  $10E-4$ . Six metabolic genes showed a significant difference between the CAD patients and the control group with a  $P$  value of less than  $10E-4$ . In traditional method, it is difficult to justify these genes as disease markers with less than two times induction. However, three genes together identified from CCA showed a significant  $P$  value. Thus, it suggests that a combination of genes is a potential disease marker for CAD by checking the expression levels rather than single genes. Also, genes 27235 (*COQ2*), 84274 (*COQ5*), and 79001

(*vkorc1*) in ubiquinone and other terpenoid quinone biosynthesis pathways were also determined by qRT-PCR in these samples. A similar result was found as the glycosphingolipid biosynthesis pathway. Each gene was induced only 2 times in CAD patients (Figure 5(b)). However, three genes together showed a very significant *P* value. In statistics, the CCA method indicated that this combination of genes could be potential disease markers.

#### 4. Conclusion

In this study, the canonical correlation between CAD and metabolic pathways gene expressions was analyzed. The results showed that the TCA cycle, ubiquinone and other terpenoid quinone biosyntheses, N-glycan biosynthesis, other glycan degradation, glycosaminoglycan degradation, GPI anchor biosynthesis, and glycosphingolipid biosynthesis–ganglioseries are the most significant metabolic pathways correlated with CAD genes. Furthermore, these metabolic pathway genes are beneficial for the diagnosis and detection of CAD through the quantification of the expression level of patients' serum.

#### Conflicts of Interest

The authors declare no financial interest related to this work.

#### Acknowledgments

The authors would like to thank Fuzhou Shichen Biotheon Technology Co. Ltd. (<http://www.fuzhoushichen.com>) for the technical support.

#### Supplementary Materials

*Supplementary 1.* Figure S1: functional category of CAD genes.

*Supplementary 2.* Table S1: list of CAD genes in this study. Table S2: list of metabolic pathways used in this study. Gene number indicated totally predicted genes in the pathways. Table S3: CAD gene ID identified in the coexpression network. Table S4: metabolic gene ID identified in the coexpression network.

#### References

- [1] R. Calderon Jr., R. H. Schneider, C. N. Alexander, H. F. Myers, S. I. Nidich, and C. Haney, "Stress, stress reduction and hypercholesterolemia in African Americans: a review," *Ethnicity & disease*, vol. 9, no. 3, pp. 451–462, 1998.
- [2] M. Kivimäki, S. T. Nyberg, G. D. Batty et al., "Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data," *The Lancet*, vol. 380, no. 9852, pp. 1491–1497, 2012.
- [3] R. Lozano, M. Naghavi, K. Foreman et al., "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2095–2128, 2013.
- [4] J. A. Finegold, P. Asaria, and D. P. Francis, "Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations," *International Journal of Cardiology*, vol. 168, no. 2, pp. 934–945, 2013.
- [5] H. C. McGill, C. A. McMahan, and S. S. Gidding, "Preventing heart disease in the 21st century: implications of the Pathobiological determinants of atherosclerosis in youth (PDAY) Study," *Circulation*, vol. 117, no. 9, pp. 1216–1227, 2008.
- [6] C. J. McNeal, T. Dajani, D. Wilson, A. E. Cassidy-Bushrow, J. B. Dickerson, and M. Ory, "Hypercholesterolemia in youth: opportunities and obstacles to prevent premature atherosclerotic cardiovascular disease," *Current Atherosclerosis Reports*, vol. 12, no. 1, pp. 20–28, 2010.
- [7] W. B. Kannel and R. S. Vasan, "Triglycerides as vascular risk factors: new epidemiologic insights," *Current Opinion Cardiology*, vol. 24, no. 4, pp. 345–350, 2009.
- [8] J. Danesh, R. Collins, and R. Peto, "Lipoprotein (a) and coronary heart disease: meta-analysis of prospective studies," *Circulation*, vol. 102, no. 10, pp. 1082–1085, 2000.
- [9] B. Smolders, R. Lemmens, and V. Thijs, "Lipoprotein (a) and stroke: a meta-analysis of observational studies," *Stroke*, vol. 38, no. 6, pp. 1959–1966, 2007.
- [10] P. J. Schreiner, J. D. Morrisett, A. R. Sharrett et al., "Lipoprotein[a] as a risk factor for preclinical atherosclerosis," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 13, no. 6, pp. 826–833, 1993.
- [11] A. E. Kitabchi, G. E. Umpierrez, J. M. Miles, and J. N. Fisher, "Hyperglycemic crises in adult patients with diabetes," *Diabetes Care*, vol. 32, no. 7, pp. 1335–1343, 2009.
- [12] V. S. Malik, B. M. Popkin, G. A. Bray, J. P. Despres, and F. B. Hu, "Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk," *Circulation*, vol. 121, no. 11, pp. 1356–1364, 2010.
- [13] X. Wang, Y. Ouyang, J. Liu et al., "Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies," *BMJ*, vol. 349, article g4490, 2014.
- [14] D. Li, "Effect of the vegetarian diet on non-communicable diseases," *Journal of the Science of Food and Agriculture*, vol. 94, no. 2, pp. 169–173, 2014.
- [15] T. Huang, B. Yang, J. Zheng, G. Li, M. L. Wahlqvist, and D. Li, "Cardiovascular disease mortality and cancer incidence in vegetarians: a meta-analysis and systematic review," *Annals of Nutrition & Metabolism*, vol. 60, no. 4, pp. 233–240, 2012.
- [16] C. Walker and B. V. Reamy, "Diets for cardiovascular disease prevention: what is the evidence?," *American Family Physician*, vol. 79, no. 7, pp. 571–578, 2009.
- [17] D. E. Threapleton, D. C. Greenwood, C. E. L. Evans et al., "Dietary fibre intake and risk of cardiovascular disease: systematic review and meta-analysis," *BMJ*, vol. 347, article f6879, 2013.
- [18] C. Chen, T. K. Hyun, X. Han et al., "Coexpression within integrated mitochondrial pathways reveals different networks in normal and chemically treated transcriptomes," *International Journal of Genomics*, vol. 2014, Article ID 452891, 10 pages, 2014.
- [19] M. G. Naylor, X. Lin, S. T. Weiss, B. A. Raby, and C. Lange, "Using canonical correlation analysis to discover genetic regulatory variants," *PLoS One*, vol. 5, no. 5, article e10395, 2010.

- [20] H. Ye, J. Yuan, Z. Wang et al., "A canonical correlation analysis of AIDS restriction genes and metabolic pathways identifies purine metabolism as a key cooperater," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 2460184, 10 pages, 2016.
- [21] P. Pajukanta, M. Cargill, L. Viitanen et al., "Two loci on chromosomes 2 and X for premature coronary heart disease identified in early- and late-settlement populations of Finland," *American Journal of Human Genetics*, vol. 67, no. 6, pp. 1481–1493, 2000.
- [22] Y. Guo, F. Wang, L. Li et al., "Genome-wide linkage analysis of large multiple multigenerational families identifies novel genetic loci for coronary artery disease," *Scientific Reports*, vol. 7, no. 1, p. 5472, 2017.
- [23] M. Farrall, F. R. Green, J. F. Peden et al., "Genome-wide mapping of susceptibility to coronary artery disease identifies a novel replicated locus on chromosome 17," *PLoS Genetics*, vol. 2, no. 5, article e72, 2006.
- [24] S. Francke, M. Manraj, C. Lacquemant et al., "A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27," *Human Molecular Genetics*, vol. 10, no. 24, pp. 2751–2765, 2001.
- [25] J. Erdmann, C. Willenborg, J. Nahrstaedt et al., "Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11. 23," *European Heart Journal*, vol. 32, no. 2, pp. 158–168, 2010.
- [26] U. Broeckel, C. Hengstenberg, B. Mayer et al., "A comprehensive linkage analysis for myocardial infarction and its related risk factors," *Nature Genetics*, vol. 30, no. 2, pp. 210–214, 2002.
- [27] S. E. Hong, I. Park, H. Cha et al., "Identification of mouse heart transcriptomic network sensitive to various heart diseases," *Biotechnology Journal*, vol. 3, no. 5, pp. 648–658, 2008.
- [28] L. Zhang, E. A. Hamad, M. Vausort et al., "Whole transcriptome microarrays identify long non-coding RNAs associated with cardiac hypertrophy," *Genomics Data*, vol. 5, pp. 68–71, 2015.
- [29] M. A. Coleman, S. P. Sasi, J. Onufrak et al., "Low-dose radiation affects cardiac physiology: gene networks and molecular signaling in cardiomyocytes," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 309, no. 11, pp. H1947–H1963, 2015.
- [30] I. Herrero, E. Roselló-Lletí, A. Ortega et al., "Gene expression network analysis reveals new transcriptional regulators as novel factors in human ischemic cardiomyopathy," *BMC Medical Genomics*, vol. 8, no. 1, pp. 1–14, 2015.
- [31] S. K. Sen, J. J. Barb, P. F. Cherukuri et al., "Identification of candidate genes involved in coronary artery calcification by transcriptome sequencing of cell lines," *BMC Genomics*, vol. 15, no. 1, p. 198, 2014.
- [32] J. Chen, H. Y. Wang, and C. Y. Zeng, "Transcriptome network analysis of potential candidate genes for heart failure," *Genetics and Molecular Research*, vol. 12, no. 4, pp. 4687–4697, 2013.
- [33] Y. Zhao, Z. Sheng, and J. Huang, "A systematic analysis of heart transcriptome highlights divergent cardiovascular disease pathways between animal models and humans," *Molecular BioSystems*, vol. 8, no. 2, pp. 504–510, 2012.
- [34] G. Czibik, V. Steeples, A. Yavari, and H. Ashrafian, "Citric acid cycle intermediates in cardioprotection," *Circulation Cardiovascular Genetics*, vol. 7, no. 5, pp. 711–719, 2014.
- [35] H. Taegtmeier, "Metabolic responses to cardiac hypoxia. Increased production of succinate by rabbit papillary muscles," *Circulation Research*, vol. 43, no. 5, pp. 808–815, 1978.
- [36] H. Ashrafian, G. Czibik, M. Bellahcene et al., "Fumarate is cardioprotective via activation of the Nrf2 antioxidant pathway," *Cell Metabolism*, vol. 15, no. 3, pp. 361–371, 2012.
- [37] S. Balta, S. Demirkol, and T. Celik, "Coenzyme Q10 supplementation may improve diastolic heart functions especially coronary artery disease patients," *Hemodialysis International*, vol. 17, no. 3, pp. 467–468, 2013.
- [38] B. J. Lee, Y. C. Huang, S. J. Chen, and P. T. Lin, "Coenzyme Q10 supplementation reduces oxidative stress and increases antioxidant enzyme activity in patients with coronary artery disease," *Nutrition*, vol. 28, no. 3, pp. 250–255, 2012.
- [39] A. Kumar, H. Kaur, P. Devi, and V. Mohan, "Role of coenzyme Q10 (CoQ10) in cardiac disease, hypertension and Meniere-like syndrome," *Pharmacology & Therapeutics*, vol. 124, no. 3, pp. 259–268, 2009.
- [40] B. J. Lee, C. H. Yen, H. C. Hsu, J. Y. Lin, S. Hsia, and P. T. Lin, "A significant correlation between the plasma levels of coenzyme Q10 and vitamin B-6 and a reduced risk of coronary artery disease," *Nutrition Research*, vol. 32, no. 10, pp. 751–756, 2012.
- [41] M. Maes, I. Mihaylova, M. Kubera, M. Uytterhoeven, N. Vrydags, and E. Bosmans, "Coenzyme Q10 deficiency in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is related to fatigue, autonomic and neurocognitive symptoms and is another risk factor explaining the early mortality in ME/CFS due to cardiovascular disorder," *Neuroendocrinology Letters*, vol. 30, no. 4, pp. 470–476, 2008.
- [42] M. Maes, I. Mihaylova, M. Kubera, M. Uytterhoeven, N. Vrydags, and E. Bosmans, "Lower plasma coenzyme Q 10 in depression: a marker for treatment resistance and chronic fatigue in depression and a risk factor to cardiovascular disorder in that illness," *Neuroendocrinology Letters*, vol. 30, no. 4, pp. 462–469, 2009.
- [43] R. B. Singh, M. A. Niaz, V. Rastogi, and S. S. Rastogi, "Coenzyme Q in cardiovascular disease," *The Journal of the Association of Physicians of India*, vol. 46, no. 3, pp. 299–306, 1998.
- [44] M. Bilinska, J. Wolszakiewicz, M. Duda, J. Janas, A. Beresewicz, and R. Piotrowicz, "Antioxidative activity of sulodexide, a glycosaminoglycan, in patients with stable coronary artery disease: a pilot study," *Medical Science Monitor*, vol. 15, no. 12, pp. CR618–CR623, 2009.
- [45] T. N. Wight and M. J. Merrilees, "Proteoglycans in atherosclerosis and restenosis," *Circulation Research*, vol. 94, no. 9, pp. 1158–1167, 2004.
- [46] M. Ueda, A. E. Becker, and T. Fujimoto, "Pathological changes induced by repeated percutaneous transluminal coronary angioplasty," *Heart*, vol. 58, no. 6, pp. 635–643, 1987.
- [47] G. Dawson, A. W. Kruski, and A. M. Scanu, "Distribution of glycosphingolipids in the serum lipoproteins of normal human subjects and patients with hypo- and hyperlipidemias," *Journal of Lipid Research*, vol. 17, no. 2, pp. 125–131, 1976.
- [48] S. Chatterjee and P. O. Kwiterovich, "Glycosphingolipids of human plasma lipoproteins," *Lipids*, vol. 11, no. 6, pp. 462–466, 1976.

- [49] E. N. Glaros, W. S. Kim, C. M. Quinn et al., "Glycosphingolipid accumulation inhibits cholesterol efflux via the ABCA1/apolipoprotein A-I pathway: 1-phenyl-2-decanoylamino-3-morpholino-1-propanol is a novel cholesterol efflux accelerator," *The Journal of Biological Chemistry*, vol. 280, no. 26, pp. 24515–24523, 2005.
- [50] S. Chatterjee, D. Bedja, S. Mishra et al., "Inhibition of glycosphingolipid synthesis ameliorates atherosclerosis and arterial stiffness in apolipoprotein E<sup>-/-</sup> mice and rabbits fed a high-fat and -cholesterol diet," *Circulation*, vol. 129, no. 23, pp. 2403–2413, 2014.