Weighted gene co-expression network analysis to identify key modules and hub genes related to hyperlipidaemia

Fu Jun Liao
Affiliated Hospital of Guizhou Medical University
Peng-Fei Zheng
Central Hospital of Shaoyang
Yao-Zong Guan
Guangxi Medical University First Affiliated Hospital
Wei Li (liwei249188@sina.com)
Guizhou Medical University https://orcid.org/0000-0003-1792-8182

Research

Keywords: Weighted gene co-expression network analysis, Hyperlipidaemia, Significant modules, Hub genes

Posted Date: August 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-54056/v1

Version of Record: A version of this preprint was published on March 4th, 2021. See the published version at https://doi.org/10.1186/s12986-021-00555-2.

Abstract

Background: The purpose of this study was to explore the potential molecular targets of hyperlipidaemia and the related molecular mechanisms.

Methods: The microarray data set of GSE66676 obtained from patients with hyperlipidaemia was downloaded. The weighted gene co-expression network (WGCNA) analysis was used to analyze the gene expression profile and royalblue module was considered as the highest correlation. Gene Ontology (GO) functional and Kyoto Encyclopedia of Genes and genomes (KEGG) pathway enrichment analyses were implemented for the identification of genes in the royalblue module using the Database for Annotation, Visualization and Integrated Discovery (DAVID) online tool (version 6.8; http://david.abcc.ncifcrf.gov). A protein-protein interaction (PPI) network was established by using the online STRING tool. Then, several hub genes were identified by the MCODE and cytoHubba plug-ins in Cytoscape software.

Results: The significant module (royalblue) identified was associated with TC, TG and Non-HDL-C. GO and KEGG enrichment analyses revealed that the genes in the royalblue module were associated with carbon metabolism, steroid biosynthesis, fatty acid metabolism and biosynthesis of unsaturated fatty acids pathways. *SQLE* (degree = 17) was revealed as key molecules that associated with hypercholesterolemia (HCH) and *SCD* was revealed as key molecules that associated with hypertriglyceridemia (HTG). Meanwhile, RT-qPCR analysis also confirmed the above results based on our HCH/HTG samples.

Conclusions: *SQLE* and *SCD* are related to hyperlipidaemia, *SQLE/SCD* may be new targets for cholesterol-lowering or triglyceride-lowering therapy, respectively.

Background

Coronary artery disease (CAD) has become a prominent cause of morbidity, mortality, disability, high healthcare costs and functional deterioration and accounts for approximately 30% of all deaths worldwide [1–3]. Hyperlipidaemia is a major risk factor for CAD and its complications. Comprehensive lipid-lowering therapy is recommended for patients with CAD by the 2013 American College of Cardiology (ACC)/American Heart Association (AHA) guidelines for the treatment of blood cholesterol to reduce the risk of cardiovascular events [4]. The guidelines emphasize that lipid-lowering therapy should not focus solely on decreasing low-density lipoprotein cholesterol (LDL-C) levels. Several compelling studies proved that lowering total cholesterol (TC) [5], triglyceride (TG) [5] and LDL-C [6] levels is more effective in reducing cardiovascular risk than lowering LDL-C levels alone [7]. The “6 percent effect” of statins refers to the fact that doubling the dose of statins only decreases LDL-C levels by 6.4%, and PCSK9 inhibitors combined with statins are recommended for acute coronary syndrome (ACS) patients with a high risk of cardiovascular events [8]. Thus, the identification of novel therapeutic targets for hyperlipidaemia is expected to further reduce the risk of cardiovascular disease.

Microarray analysis might serve as a novel and practical approach to identify susceptibility genes associated with hyperlipidaemia [9]. However, the reproducibility and sensitivity of microarray analysis based on differentially expressed genes may be limited [10, 11]. Gene co-expression network-based methods have been widely used in processing microarray data and have especially been used to identify meaningful functional modules [12, 13]. Weighted gene co-expression network analysis (WGCNA) is one of the most effective methods of gene co-expression network analysis. Instead of simply identifying the differentially expressed genes, a scale-free network of gene-gene interactions is generated by WGCNA, and several significant modules comprised of genes with similar functions could be identified by WGCNA; in addition, it can be used to further analyze the correlation between modules and phenotypes or clinical characteristics [14]. Therefore, WGCNA could be utilized to construct a co-expression network and identify significant modules in the network, which may help us to illuminate the intrinsic characteristics of hyperlipidaemia and provide new insights into potential genetic biomarkers, signaling pathways and molecular mechanisms involved in hyperlipidaemia.

Materials And Methods

Hyperlipidaemia microarray data sets

The microarray data set obtained from patients with hyperlipidaemia (GSE66676) was downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database, which is based on the platform of the GPL6244 Affymetrix Human Gene 1.0 ST Array. Gene expression value matrices were obtained from the original files in CEL format after normalizing the expression values by using RMA methods in R software (version 4.0.0). [15]. Then, the Bioconductor package was used to transform the probe identification numbers (IDs) into gene symbols [16]. When multiple probe IDs corresponded to the same gene, the average expression value was used as the expression value.

Construction of the Weighted Gene Co-expression Network

WGCNA is a widely used systems biology method that is usually used to establish a scale-free network based on gene expression data profiles [12]. The co-expression network was constructed by selecting the genes whose variance was greater than all the quartiles of variance. After the sample cluster tree was constructed, cut Height = 35 is used to screen the samples for subsequent studies. To ensure the reliability of the results of the network construction, the outlier samples were eliminated, and the samples in cluster 1 were selected to build the sample dendrogram and trait heatmap. The appropriate soft threshold power (soft power = 9) was selected according to the standard scale-free networks, and the adjacency values between all differentially expressed genes were calculated using a power function. Then, the adjacency values were transformed into a topological overlap matrix (TOM), and the corresponding dissimilarity (1-TOM) values were calculated. 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. To

verify the stability of the identified modules, a module preservation function was used to calculate module preservation and quality statistics in the WGCNA package [17].

Identification of the module of interest and functional annotation

Pearson correlation analysis was used to assess the correlations between modules and clinical characteristics to identify biologically meaningful modules. All genes associated with the significant module were subjected to 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 the gene expression profile and the module eigengenes (Mes). The degree of gene significance (GS) was defined as the absolute value of the correlation between the gene and external traits. In general, modules with increased MS and GS values among all the identified modules were selected for further analysis of their biological function [18]. The protein-protein interaction (PPI) network of genes in the selected module was constructed by the Search Tool for the Retrieval of Interacting Genes database (version 11.0; http://www.string-db.org) [19] and then visualized using Cytoscape software [20]. Molecular complex detection (MCODE) [21] was used to identify the most valuable clustering module. An MCODE score > 4 was the threshold for inclusion in further analysis. CytoHubba, a Cytoscape plugin, was used to identify hub genes in PPI network; it provides 11 methods to explore important nodes in biological networks, of which degree has a better performance [22].

Sample verification and diagnostic criteria

A total of 462 (229 males, 49.57%; 233 females, 50.43%) unrelated participants of normal lipid levels and 485 (236 males, 48.66%; 249 females, 51.34%) unrelated subjects with hypercholesterolemia (HCH, TC > 5.17 mmol/l) and 474 (232 males, 49.16%; 241 females, 50.84%) unrelated participants with hypertriglyceridemia (HTG, TG > 1.70 mmol/l) were arbitrarily chosen based on our previously stratified randomized samples [23]. The age ranged from 24 to 82 years. There was not any difference in age distribution and gender ratio between control and HCH or HTG groups. Patients suffering from HCH did not have a history of HTG and patients suffering from HTG did not have a history of HCH. All participants were basically healthy and had no history of myocardial infarction, CAD, type 2 diabetes mellitus (T2DM) and ischemic stroke. They were not taking any medicines that could alter serum lipid levels. All subjects had signed written informed consent. The research protocol was approved by the Ethics Committee of the First Affiliated Hospital, Guangxi Medical University (No. Lunshen-2014 KY-Guoji-001, Mar. 7, 2014).

Epidemiological analysis

Universally standardized methods and protocols were used to conduct the epidemiological survey [24]. Detailed lifestyle and demographic characteristics were collected with a standard set of questionnaires.

Alcohol consumption (0 (non-drinker), < 25 g/day and ≥ 25 g/day) and smoking status (0 (non-smoker), < 20 cigarettes/day and ≥ 20 cigarettes/day) were divided into three different subgroups. Waist circumference, BMI, height, blood pressure and weight were measured as previously described [25].

Biochemical assays

Fasting venous blood samples of 5 ml were collected from each subject. A portion of the sample (2 ml) was placed in a tube and used to measure serum lipid levels. The remaining sample of 3 ml was collected in a glass tube containing anticoagulants (14.70 g/L glucose, 13.20 g/L trisodium citrate, 4.80 g/L citric acid) and utilized to extract deoxyribonucleic acid (DNA). The methods for performing serum ApoA1, HDL-C, ApoB, TG, LDL-C and TC measurements were described in a previous study [26]. All determinations were conducted using an autoanalyzer (Type 7170A; Hitachi Ltd., Tokyo, Japan) in the Clinical Science Experiment Center of the First Affiliated Hospital, Guangxi Medical University [27, 28].

Quantitative real-time PCR

Peripheral blood monocytes (PBMCs) were isolated from blood samples with TRIzol reagent, which was used to extract the total RNA, which was then reverse-transcribed into cDNA by using the PrimeScript RT reagent kit (Takara Bio, Japan). The obtained cDNA was used as a template for RT-qPCR. Table 1 shows that specific primer sequences were used to detect the 2 hub genes, which were designed by Sangon Biotech (Shanghai, China). Quantitative RT-PCR was performed using a Taq PCR Master Mix Kit (Takara) on an ABI Prism 7500 sequence-detection system (Applied Biosystems, USA) using RT Reaction Mix in a total volume of 20 μL with the following reaction conditions: predenaturation at 95 °C for 30 seconds, then 40 cycles of 95 °C for 30 seconds and 60 °C for 30 seconds.

Table 1
PCR primers for quantitative real-time PCR

Gene	Forward primer	Reverse primer
SQLE	TCTGGGGGTTAAGAGCAGTG	GTGTCTACACTTACCATCTGTGGC
SCD	CTTGCGATATGCTGTGGTGC	GGCTCCTAGCCTAATCCCCT
GAPDH	GCAACTAGGATGGTGTGGCT	TCCCATTCCCCAGCTCTCATA

Diagnostic criteria

The values of serum ApoB (0.80–1.05 g/L), HDL-C (1.16–1.42 mmol/L), ApoA1 (1.20–1.60 g/L), TC (3.10–5.17 mmol/L), TG (0.56–1.70 mmol/L), the ApoA1/ApoB ratio (1.00-2.50) and LDL-C (2.70–3.10 mmol/L) were defined as normal at our Clinical Science Experiment Center. The subjects with TG > 1.70 mmol/L were defined as hypertriglyceridemia and TC > 5.17 mmol/L were defined as hypercholesterolemia [29]. The participants with the fasting plasma (blood) glucose value ≥ 7.0 mmol/L were defined as diabetes [30]. The diagnostic criteria of hypertension [31], obesity, normal weight and overweight were referred to our previous study [32].

Statistical analyses

SPSS (Version 22.0) was used to process the research data. The results are presented as the mean ± SD except for TG levels, which are presented as medians and interquartile ranges. The differences in the general characteristics except for TG between HCH/HTG patients and controls were analyzed by independent-samples t tests. Kruskal-Wallis and Mann-Whitney nonparametric tests were used to detected the difference the TG levels between HCH/HTG patients and controls. The chi-square test was utilized to assess the differences in the proportion of smokers, age distribution and alcohol consumption between HCH/HTG patients and controls. The heat mapping of the correlation models and bioinformatic analysis were performed in R software (version 4.0.0). A P value < 0.05 was considered to be statistically significant.

Results

Data preprocessing

Gene expression profiles were obtained after normalization of the data and removing the outliers, and a total of 20284 gene symbols were identified from 67 samples. Additional details about the gene expression profile and the sample phenotypes are presented in **Additional file 1: Tables S1 and S2.**

Weighted gene co-expression networks

The sample cluster tree and sample dendrogram and trait heatmap are shown in the **Additional file 2: Figure S1 and S2.** The gene expression profiles of 42 samples in cluster 1 were selected to build the weighted gene co-expression network. After the soft threshold (β = 9) was determined (Fig. 1), The weighted gene co-expression network was constructed by selecting the genes whose variance was greater than all the quartiles of variance. The adjacency matrix and correlation matrix of the gene expression profile were calculated and then transformed into a topological overlap matrix (TOM), and a clustering tree of genes based on the gene–gene non-ω similarity was obtained (Fig. 2). Combined with the TOM, the gene modules of each gene network were identified by the hierarchical average linkage clustering method, and twenty gene modules were identified by the dynamic tree cut algorithm (cut height = 0.25) (Fig. 3). The gray module contains all the genes that do not belong to the other modules and were excluded from subsequent analysis.

Identification of the modules of interest and functional annotation

The identification of modules that were significantly related to clinical phenotype was of high biological significance. In this study, we noticed that the royalblue module were associated with TC (r^2 = 0.38, P = 0.01), TG (r^2 = 0.41, P = 0.007) and Non-HDL-C (r^2 = 0.32, P = 0.04) and the genes in the royalblue module were studied in the subsequent analyses (Fig. 4). GO and KEGG pathway enrichment analyses were used to further explore the biological functions of the genes in the royalblue module, Furthermore, we noticed that a total of 101 genes in the royalblue module were significantly correlated with the following pathways: hsa01100: Metabolic pathways, hsa01130: Biosynthesis of antibiotics, hsa00100:

Steroid biosynthesis, hsa01212: Fatty acid metabolism, hsa01040: Biosynthesis of unsaturated fatty acids. The cell components, biological processes, molecular functions and KEGG pathway analysis of the royalblue module are also shown in Fig. 5 and more detailed information is presented in **Additional file: Tables S4 and S5**.

PPI network construction and module analysis of DEGs

A PPI network including 93notes and 333 edges was constructed by the STRING online tool. As shown in Fig. 6, the hub genes *SQLE* (degree = 17) and *SCD* (degree = 5) were identified by cytoHubba plug-ins in MCODE1 and MCODE2, respectively. Thus, we speculate that the genes mentioned above may be significantly correlated with blood lipid metabolism.

Validation analysis by RT-qPCR

As shown in Fig. 7A, the results of RT-qPCR revealed that the expression of *SQLE* in HCH and *SCD* in HTG groups were higher than in healthy subjects. At the same time, we also noticed that *SQLE* was positive correlation with TC (Fig. 7C) levels in HCH and *SCD* was positive correlation with TG levels in HTG groups (Fig. 7D).

Demographic and biochemical characteristics

As mentioned in Table 2, the gender ratio, age and height were similar between the controls and HCH/HTG patients. Serum HDL-C and ApoA1 levels and the ApoA1/ApoB ratio were significantly higher, and the proportion of smokers, proportion of drinkers, systolic blood pressure, weight, diastolic blood pressure, glucose level, pulse pressure, body mass index (BMI), and serum LDL-C, ApoB, TG and TC levels were significantly lower in controls than in hyperlipidemic patients.

Table 2
Comparison of demographic, lifestyle characteristics and serum lipid levels of the participants

Characteristic	Control(n = 462)	HCH (n = 485)	HTG (n = 474)	$P_{HCH\ vs.\ controls}$	$P_{HTG\ vs.\ controls}$
Male/female	229/233	236/249	232/241	0.780	0.870
Age (years)	57.60 ± 8.81	58.13 ± 9.69	57.10 ± 7.61	0.379	0.359
Height (cm)	159.83 ± 8.20	160.66 ± 7.93	159.92 ± 8.11	0.114	0.771
Weight (kg)	58.98 ± 9.90	62.66 ± 10.09	60.52 ± 11.08	1.97E-8	0.021
Body mass index (kg/m²)	23.05 ± 3.32	24.21 ± 3.12	23.61 ± 3.67	3.59E-8	0.014
Waist circumference	74.50 ± 8.47	78.37 ± 8.76	81.10 ± 9.21	1.00E-11	4.10E-28
Smoking, *n* %					
Non-smoker	355	331	349		0.892
≤ 20 cigarettes/day	98	114	73		
> 20 cigarettes/day	9	40	52	1.73E-5	4.42E-8
Alcohol, *n* %					
Non-drinker	377	354	363		
≤ 25 g/day	45	55	44		
> 25 g/day	40	76	67	0.002	0.031
SBP (mmHg)	135.49 ± 22.58	139.52 ± 22.56	141.15 ± 20.42	0.006	1.08E-4
DBP (mmHg)	82.45 ± 12.42	84.01 ± 11.71	85.15 ± 11.72	0.047	0.001
PP (mmHg)	53.04 ± 17.77	55.51 ± 18.19	56.00 ± 14.43	0.035	0.007

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.

1 Mean ± SD determined by t-test.

2 Median (interquartile range) tested by the Wilcoxon-Mann-Whitney test.

3 The rate or constituent ratio between the different groups was analyzed by the chi-square test.

Characteristic	Control(n = 462)	HCH (n = 485)	HTG (n = 474)	$P_{HCH\ vs.\ controls}$	$P_{HTG\ vs.\ controls}$
Glu (mmol/L)	6.14 ± 1.42	6.44 ± 1.58	6.35 ± 1.32	0.002	0.015
TC (mmol/L)	4.37 ± 0.64	5.80 ± 0.50	4.46 ± 0.37	9.42E-47	0.007
TG (mmol/L)	0.99(0.53)	1.15(0.45)	2.30(1.03)	4.46E-9	1.69E-84
HDL-C (mmol/L)	1.64 ± 0.48	1.48 ± 0.43	1.44 ± 0.46	2.56E-8	7.22E-11
LDL-C (mmol/L)	2.50 ± 0.55	3.52 ± 0.80	2.79 ± 0.86	4.89E-32	1.86E-9
ApoA1 (g/L)	1.33 ± 0.24	1.23 ± 0.25	1.25 ± 0.24	3.49E-11	4.66E-8
ApoB (g/L)	0.98 ± 0.17	1.02 ± 0.18	1.07 ± 0.21	0.002	1.06E-11
ApoA1/ApoB	1.39 ± 0.33	1.25 ± 0.40	1.22 ± 0.39	1.91E-8	7.62E-12
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.					
1 Mean ± SD determined by t-test.					
2 Median (interquartile range) tested by the Wilcoxon-Mann-Whitney test.					
3 The rate or constituent ratio between the different groups was analyzed by the chi-square test.					

Discussion

Several recent researches showed that hypertension, smoking, obesity, age, dyslipidemia, lack of exercise, gender and diabetes mellitus are common risk factors for cardiovascular disease [33, 34]. A comprehensive understanding of the potential molecular mechanisms involved in the pathogenesis of hyperlipidaemia is helpful for its prevention and treatment. As a novel and practical approach to the identification of hyperlipidaemia susceptibility genes, a microarray analysis using WGCNA may be helpful for the diagnosis of hyperlipidemia [14]. WGCNA could be used to build a scale-free co-expression network of lipids-associated genes by detecting gene-to-gene interactions rather than simply focusing on the differentially expressed genes (DEGs). Co-expressed genes were enriched in different modules by hierarchical average linkage cluster analysis. In the present research, we analyzed the dataset from hyperlipidaemia patients (GSE66676) by using WGCNA analysis identified the royalblue module was significantly associated with TC, TG and Non-HDL. Furthermore, KEGG enrichment analyses of the genes in the royalblue module indicated that the enriched genes in this module might have significant potential biological functions that are closely related to metabolic pathways, steroid biosynthesis, fatty acid metabolism and biosynthesis of unsaturated fatty acids. Two hub genes (*SQLE* and *SCD*) were identified in the royalblue module that were detected by MCODE analysis. Moreover, the verification results were highly consistent with the above findings, and we found that the expression of the *SQLE* gene in patients

with HCH and *SCD* gene in patients with HTG were higher than those in healthy controls, respectively. Therefore, the identified *SQLE* gene was associated with the onset of HCH and *SCD* gene was associated with the onset of HTG, and the underlying molecular mechanisms of these genes might be slightly different.

Fatty acids and cholesterol are essential lipids involved in lots of crucial biological processes, however, excessive free fatty acids and free cholesterol are major risk factors for type 2 diabetes and atherosclerosis [35]. Previous studies on intermediate metabolites in cholesterol biosynthesis have showed that squalene epoxidase (*SQLE*) acts as a crucial regulator downstream HMG-CoA reductase in cholesterol synthesis and the first oxygenation step in cholesterol biosynthesis catalyzed by *SQLE* [36]. Meanwhile, *SQLE* is suggested as the second rate-limiting enzyme in cholesterol synthesis [37, 38]. Inhibition of *SQLE* expression could effectively reduce cholesterol synthesis [39, 40] and the cholesterol-lowering effect is caused by the combination of multiple levels. First, *SQLE* acts as a direct target of *SREBP2* transcription factor, plays a crucial regulatory role in most genes in cholesterol biosensors[41]. Second, the N-terminal of *SQLE* protein may contain a cholesterol-sensitive region that mediates the protease degradation of *SQLE* in a cholesterol-dependent method through relying on an E3 ubiquitin ligase such as *MARCH* [42]. Interestingly, oleate acts as one of unsaturated fatty acids can stabilize *SQLE* by blocking *MARCH6*-mediated degradation [43]. In addition, Masanori Honsho et al. also noticed that inhibition of *SQLE* expression through elevating the plasmalogens levels may be a novel and alternative potential method to reduce cholesterol synthesis [44]. Similarly, the KEGG analyses in the current study indicated that *SQLE* was mainly involved in metabolic pathways and steroid biosynthesis.

Metabolic risk factors such as insulin resistance, obesity, hypertension and dyslipidemia are correlated with each other, so their combination is generally referred to as "metabolic syndrome". Abnormal Stearoyl-coenzyme A desaturase (*SCD*) expression/activity has been noticed in metabolic syndrome subjects indicating that *SCD* may be related to the pathogenesis of metabolic syndrome. By querying the GENE database in NCBI, we noticed that *SCD* (also known as *SCD1*; *FADS5*; *SCDOS*; *hSCD1*; *MSTP008*; gene ID: 6319, HGNC: 10571, OMIM: 604031) is positioned on chromosome 10q24.31 (exon count: 6) and encodes a biological synthase, which mainly involved in the metabolism of fatty acids, especially oleic acid, this protein is an intact membrane protein located in the endoplasmic reticulum and is a member of the fatty acid desaturase family. *SCD* can convert different saturated fatty acids into monounsaturated fatty acids [45]. Both animal and human studies have shown that *SCD* is associated with obesity and insulin resistance[46, 47]. Mice with the *SCD* gene destroyed reduced diet-induced weight gain and improved insulin resistance compared to wild-type controls [48]. Deletion of *SCD1* gene product in mice could effectively improve insulin sensitivity, reduce plasma non-HDL-cholesterol and triglyceride levels and liver lipid accumulation as well as increase beneficial HDL-cholesterol levels [49]. Daniel Castellano-Castillo et al. also found that a negative relationship between *SCD* DNA methylation and BMI and the MetS index [50]. In the current study, we also noticed that *SCD* was mainly involved in fatty acid metabolism and biosynthesis of unsaturated fatty acids pathways.

Unhealthy lifestyle factors such as excessive drinking and cigarette smoking have been linked to hyperlipidaemia [51, 52]. In the present study, we found that the percentage of the participants who smoked was greater in the hyperlipidaemic group than in the normal group. In recent years, the influence of smoking on hyperlipidaemia has attracted increasing attention. Several recent studies have indicated the existence of lower HDL-C and higher TC, LDL-C and TG levels in smokers compared to non-smokers [52]. Moderate drinking reduced the incidence of cardiovascular events, the potential mechanism may be related to increased HDL-C and ApoA1 levels [53]. However, frequent binge drinking was correlated with an increased risk of CAD mortality because it will lead to a number of serious health problems including dyslipidaemia, abnormal liver function and myocardial infarction [54]. Therefore, the preventive effect of healthy lifestyle on hyperlipidemia should not be ignored when exploring new therapeutic targets for hyperlipidemia

Conclusions

WGCNA analysis identified the royalblue module was significantly associated with TC, TG and Non-HDL. GO and KEGG enrichment analyses revealed that the hub genes of *SQLE* was associated with TC and *SCD* was associated with TG metabolism. The verification results of RT-qPCR reveled that the expression of *SQLE* in hypercholesterolemia and SCD in hypertriglyceridemia were higher than in normal controls, which further increased the credibility of the conclusion. Thus, we speculated that *SQLE* may be a novel target for cholesterol-lowering therapy and *SCD* may be a novel target for triglyceride-lowering therapy.

Declarations

Acknowledgements

We thank all the participants of this study.

Authors' contributions

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

Funding

This study was supported by the National Natural Science Foundation of China (No. 81460169).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study design was approved by the Ethics Committee of the First Affiliated hospital, Guangxi Medical University (No: Lunshen-2011-KY-Guoji-001; March 7, 2011). Informed consent was obtained from all participants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Department of Cardiology, Affiliated Hospital of Guizhou Medical University, 28 Guyi Street, Guiyang 550002, Guizhou, People's Republic of China

[2] Department of Cardiology, ShaoYang Central Hospital, 36 QianYuan lane, Shaoyang 422000, Hunan, People's Republic of China

[3] Department of Cardiology, Institute of Cardiovascular Diseases, the First Affiliated Hospital, Guangxi Medical University, 6 Shuangyong Road, Nanning 530021, Guangxi, People's Republic of China

Abbreviations

WGCNA: weighted gene co-expression network analysis; CAMP: cathelicidin antimicrobial peptide; HCH: hypercholesterolemia; HTG: hypertriglyceridemia; SQLE: squalene epoxidase; SCD: stearoyl-CoA desaturase; DAVID: Database for Annotation, Visualization and Integrated Discovery; T2DM: Type 2 diabetes mellitus; GO: Gene Ontology; HDL-C: High-density lipoprotein cholesterol; IS: Ischemic stroke; KEGG: Kyoto Encyclopedia of Genes and genomes; LDL-C: Low-density lipoprotein cholesterol; MCODE: Molecular Complex Detection; Apo: Apolipoprotein; PPI: Protein-protein interaction; GEO: Gene Expression Omnibus; BMI: Body mass index; TG: Triglyceride; RT-qPCR: Quantitative real time polymerase chain reaction; TC: Total cholesterol.

References

1. 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. 2018;12(3):85–108.
2. Yokokawa H, Yasumura S, Tanno K, Ohsawa M, Onoda T, Itai K, et al. Serum low-density lipoprotein to high-density lipoprotein ratio as a predictor of future acute myocardial infarction among men in a

2.7-year cohort study of a Japanese northern rural population. J Atheroscler Thromb. 2011;18(2):89–98.

3. Finegold JA, Asaria P, Francis DP. Mortality from ischaemic heart disease by country, region, and age: statistics from World Health Organisation and United Nations. Int J Cardiol. 2013;168(2):934–45.
4. Ray KK, Kastelein JJ, Boekholdt SM, Nicholls SJ, Khaw KT, Ballantyne CM, et al. The ACC/AHA 2013 guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular disease risk in adults: the good the bad and the uncertain: a comparison with ESC/EAS guidelines for the management of dyslipidaemias 2011. Eur Heart J. 2014;35(15):960–8.
5. Chapman MJ, Ginsberg HN, Amarenco P, Andreotti F, Borén J, Catapano AL, et al. Triglyceride-rich lipoproteins and high-density lipoprotein cholesterol in patients at high risk of cardiovascular disease: evidence and guidance for management. Eur Heart J. 2011;32(11):1345–61.
6. Can M, Acikgoz S, Mungan G, Ugurbas E, Ankarali H, Sumbuloglu V, et al. Is direct method of low density lipoprotein cholesterol measurement appropriate for targeting lipid lowering therapy? Int J Cardiol. 2010;142(1):105–7.
7. Ferrieres J, Amber V, Crisan O, Chazelle F, Junger C, Wood D. Total lipid management and cardiovascular disease in the dyslipidemia international study. Cardiology. 2013;125(3):154–63.
8. Gencer B, Koskinas KC, Räber L, Karagiannis A, Nanchen D, Auer R, et al. Eligibility for PCSK9 Inhibitors According to American College of Cardiology (ACC) and European Society of Cardiology/European Atherosclerosis Society (ESC/EAS) Guidelines After Acute Coronary Syndromes. J Am Heart Assoc. 2017;6(11):e006537.
9. Ma J, Dempsey AA, Stamatiou D, Marshall KW, Liew CC. Identifying leukocyte gene expression patterns associated with plasma lipid levels in human subjects. Atherosclerosis. 2007;191(1):63–72.
10. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. Lancet. 2003;362(9394):1439–44.
11. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? Bioinformatics. 2005;21(2):171–8.
12. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol. 2008;4(8):e1000117.
13. Ruan J, Dean AK, Zhang W. A general co-expression network-based approach to gene expression analysis: comparison and applications. BMC Syst Biol. 2010;4:8.
14. Miao L, Yin RX, Pan SL, Yang S, Yang DZ, Lin WX. Weighted Gene Co-Expression Network Analysis Identifies Specific Modules and Hub Genes Related to Hyperlipidemia. Cell Physiol Biochem. 2018;48(3):1151–63.
15. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20(3):307–15.
16. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80.

17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
18. Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm Genome. 2007;18(6–7):463–72.
19. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447–52.
20. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011;27(3):431–2.
21. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2.
22. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8(Suppl 4(Suppl 4):11.
23. Pan Y, Yu C, Huang J, Rong Y, Chen J, Chen M. 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. 2020;19(1):54–4.
24. An epidemiological study of. cardiovascular and cardiopulmonary disease risk factors in four populations in the People's Republic of China. Baseline report from the P.R.C.-U.S.A. Collaborative Study. People's Republic of China–United States Cardiovascular and Cardiopulmonary Epidemiology Research Group. Circulation. 1992;85(3):1083–96.
25. Guo T, Yin RX, Li H, Wang YM, Wu JZ, Yang DZ. Association of the Trp316Ser variant (rs1801690) near the apolipoprotein H (beta2-glycoprotein-I) gene and serum lipid levels. Int J Clin Exp Pathol. 2015;8(6):7291–304.
26. Sun JQ, Yin RX, Shi GY, Shen SW, Chen X, Bin Y, et al. Association of the ARL15 rs6450176 SNP and serum lipid levels in the Jing and Han populations. Int J Clin Exp Pathol. 2015;8(10):12977–94.
27. Guo T, Yin RX, Nie RJ, Chen X, Bin Y, Lin WX. Suppressor of cytokine signaling 3 A + 930–>G (rs4969168) polymorphism is associated with apolipoprotein A1 and low-density lipoprotein cholesterol. Int J Clin Exp Pathol. 2015;8(6):7305–17.
28. Zeng XN, Yin RX, Huang P, Huang KK, Wu J, Guo T, et al. Association of the MLXIPL/TBL2 rs17145738 SNP and serum lipid levels in the Guangxi Mulao and Han populations. Lipids Health Dis. 2013;12:156.
29. Zheng P-F, Yin R-X, Liu C-X, Deng G-X, Guan Y-Z, Wei B-L. SYNE1-QK1 SNPs, G × G and G × E interactions on the risk of hyperlipidaemia. J Cell Mol Med. 2020;24(10):5772–85.
30. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med. 1998;15(7):539–53.
31. Whitworth JA. 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. J Hypertens 2003, 21(11):1983–1992.

32. Khounphinith E, Yin R-X, Cao X-L, Huang F, Wu J-Z, Li H. TIMD4 rs6882076 SNP Is Associated with Decreased Levels of Triglycerides and the Risk of Coronary Heart Disease and Ischemic Stroke. Int J Med Sci. 2019;16(6):864–71.
33. Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR, Garside DB, et al. Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. JAMA. 2003;290(7):891–7.
34. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937–52.
35. Wang Y-J, Bian Y, Luo J, Lu M, Xiong Y, Guo S-Y, et al. Cholesterol and fatty acids regulate cysteine ubiquitylation of ACAT2 through competitive oxidation. Nat Cell Biol. 2017;19(7):808–19.
36. Honsho M, Abe Y, Fujiki Y. Dysregulation of Plasmalogen Homeostasis Impairs Cholesterol Biosynthesis. J Biol Chem. 2015;290(48):28822–33.
37. Hidaka Y, Satoh T, Kamei T. Regulation of squalene epoxidase in HepG2 cells. J Lipid Res. 1990;31(11):2087–94.
38. Gonzalez R, Carlson JP, Dempsey ME. Two major regulatory steps in cholesterol synthesis by human renal cancer cells. Arch Biochem Biophys. 1979;196(2):574–80.
39. Hidaka Y, Hotta H, Nagata Y, Iwasawa Y, Horie M, Kamei T. Effect of a novel squalene epoxidase inhibitor, NB-598, on the regulation of cholesterol metabolism in Hep G2 cells. J Biol Chem. 1991;266(20):13171–7.
40. Sawada M, Matsuo M, Hagihara H, Tenda N, Nagayoshi A, Okumura H, et al. Effect of FR194738, a potent inhibitor of squalene epoxidase, on cholesterol metabolism in HepG2 cells. Eur J Pharmacol. 2001;431(1):11–6.
41. Sharpe LJ, Brown AJ. Controlling cholesterol synthesis beyond 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR). J Biol Chem. 2013;288(26):18707–15.
42. Zelcer N, Sharpe LJ, Loregger A, Kristiana I, Cook EC, Phan L, et al. The E3 ubiquitin ligase MARCH6 degrades squalene monooxygenase and affects 3-hydroxy-3-methyl-glutaryl coenzyme A reductase and the cholesterol synthesis pathway. Mol Cell Biol. 2014;34(7):1262–70.
43. Stevenson J, Luu W, Kristiana I, Brown AJ. Squalene mono-oxygenase, a key enzyme in cholesterol synthesis, is stabilized by unsaturated fatty acids. Biochem J. 2014;461(3):435–42.
44. Honsho M, Abe Y, Fujiki Y. Dysregulation of Plasmalogen Homeostasis Impairs Cholesterol Biosynthesis. J Biol Chem. 2015;290(48):28822–33.
45. Ntambi JM, Miyazaki M. Regulation of stearoyl-CoA desaturases and role in metabolism. Prog Lipid Res. 2004;43(2):91–104.
46. Rahman SM, Dobrzyn A, Lee SH, Dobrzyn P, Miyazaki M, Ntambi JM. Stearoyl-CoA desaturase 1 deficiency increases insulin signaling and glycogen accumulation in brown adipose tissue. Am J Physiol Endocrinol Metab. 2005;288(2):E381–7.

47. García-Serrano S, Moreno-Santos I, Garrido-Sánchez L, Gutierrez-Repiso C, García-Almeida JM, García-Arnés J, et al. Stearoyl-CoA desaturase-1 is associated with insulin resistance in morbidly obese subjects. Mol Med. 2011;17(3–4):273–80.
48. Ntambi JM, Miyazaki M, Stoehr JP, Lan H, Kendziorski CM, Yandell BS, et al. Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. Proc Natl Acad Sci U S A. 2002;99(17):11482–6.
49. MacDonald MLE, Singaraja RR, Bissada N, Ruddle P, Watts R, Karasinska JM, et al. Absence of stearoyl-CoA desaturase-1 ameliorates features of the metabolic syndrome in LDLR-deficient mice. J Lipid Res. 2008;49(1):217–29.
50. Castellano-Castillo D, Moreno-Indias I, Sanchez-Alcoholado L, Ramos-Molina B, Alcaide-Torres J, Morcillo S, et al. Altered Adipose Tissue DNA Methylation Status in Metabolic Syndrome: Relationships Between Global DNA Methylation and Specific Methylation at Adipogenic, Lipid Metabolism and Inflammatory Candidate Genes and Metabolic Variables. J Clin Med. 2019;8(1):87.
51. Waskiewicz A, Sygnowska E. Alcohol intake and cardiovascular risk factor profile in men participating in the WOBASZ study. Kardiol Pol. 2013;71(4):359–65.
52. Rao Ch S, Subash YE. The effect of chronic tobacco smoking and chewing on the lipid profile. J Clin Diagn Res. 2013;7(1):31–4.
53. Ruixing Y, Shangling P, Hong C, Hanjun Y, Hai W, Yuming C, et al. Diet, alcohol consumption, and serum lipid levels of the middle-aged and elderly in the Guangxi Bai Ku Yao and Han populations. Alcohol. 2008;42(3):219–29.
54. Pai JK, Mukamal KJ, Rimm EB. Long-term alcohol consumption in relation to all-cause and cardiovascular mortality among survivors of myocardial infarction: the Health Professionals Follow-up Study. Eur Heart J. 2012;33(13):1598–605.

Figures

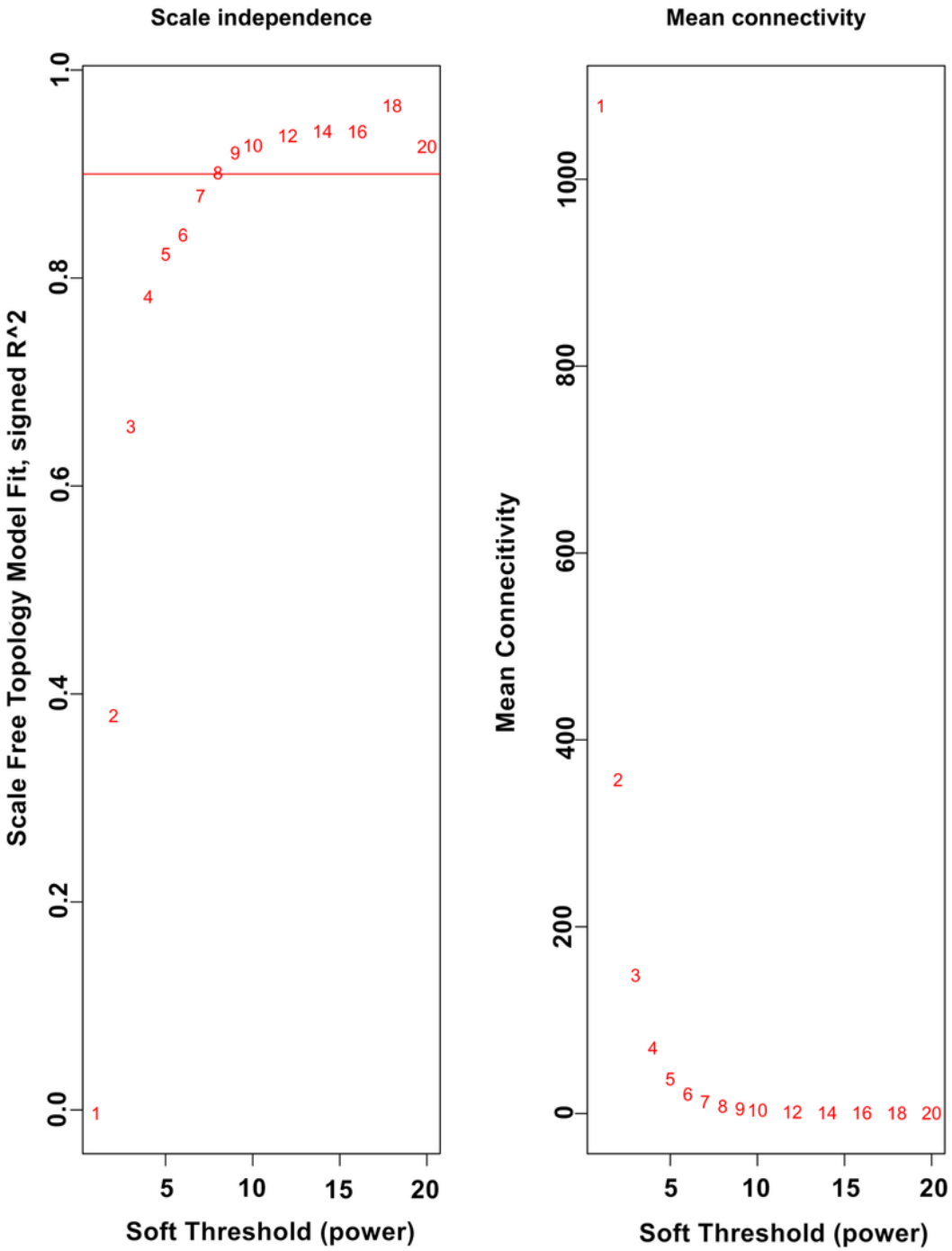

Figure 1

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).

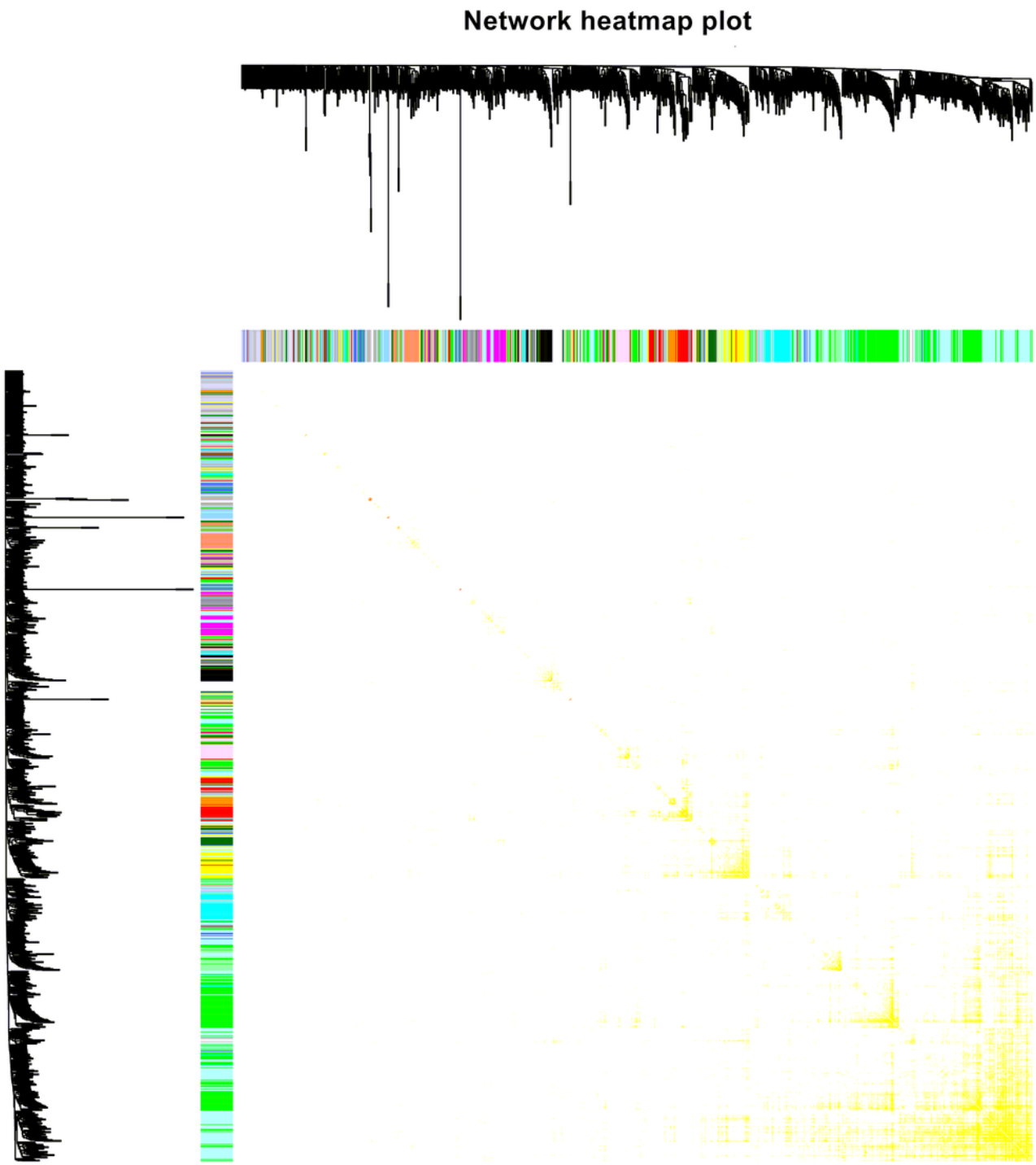

Figure 2

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.

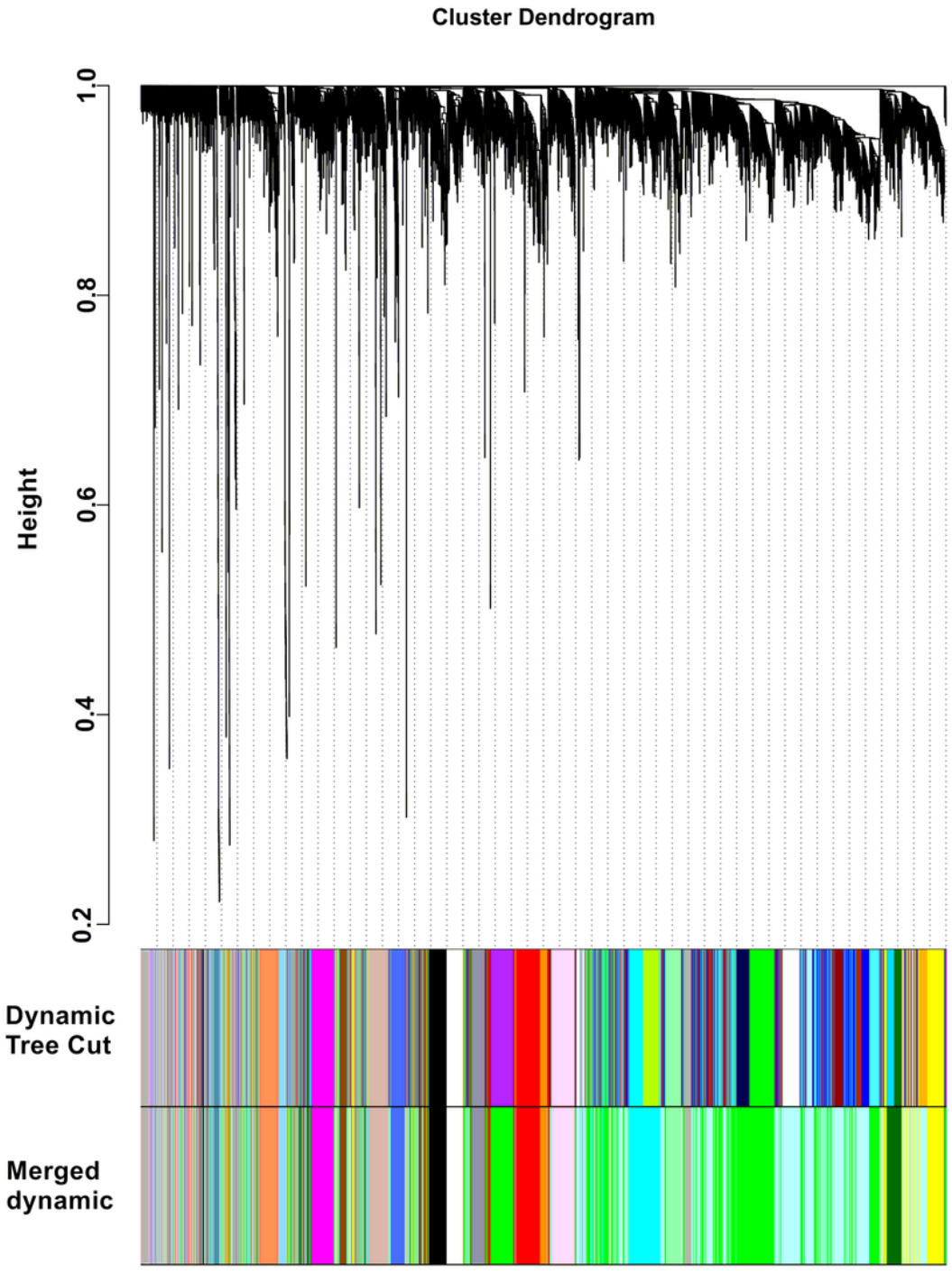

Figure 3

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.

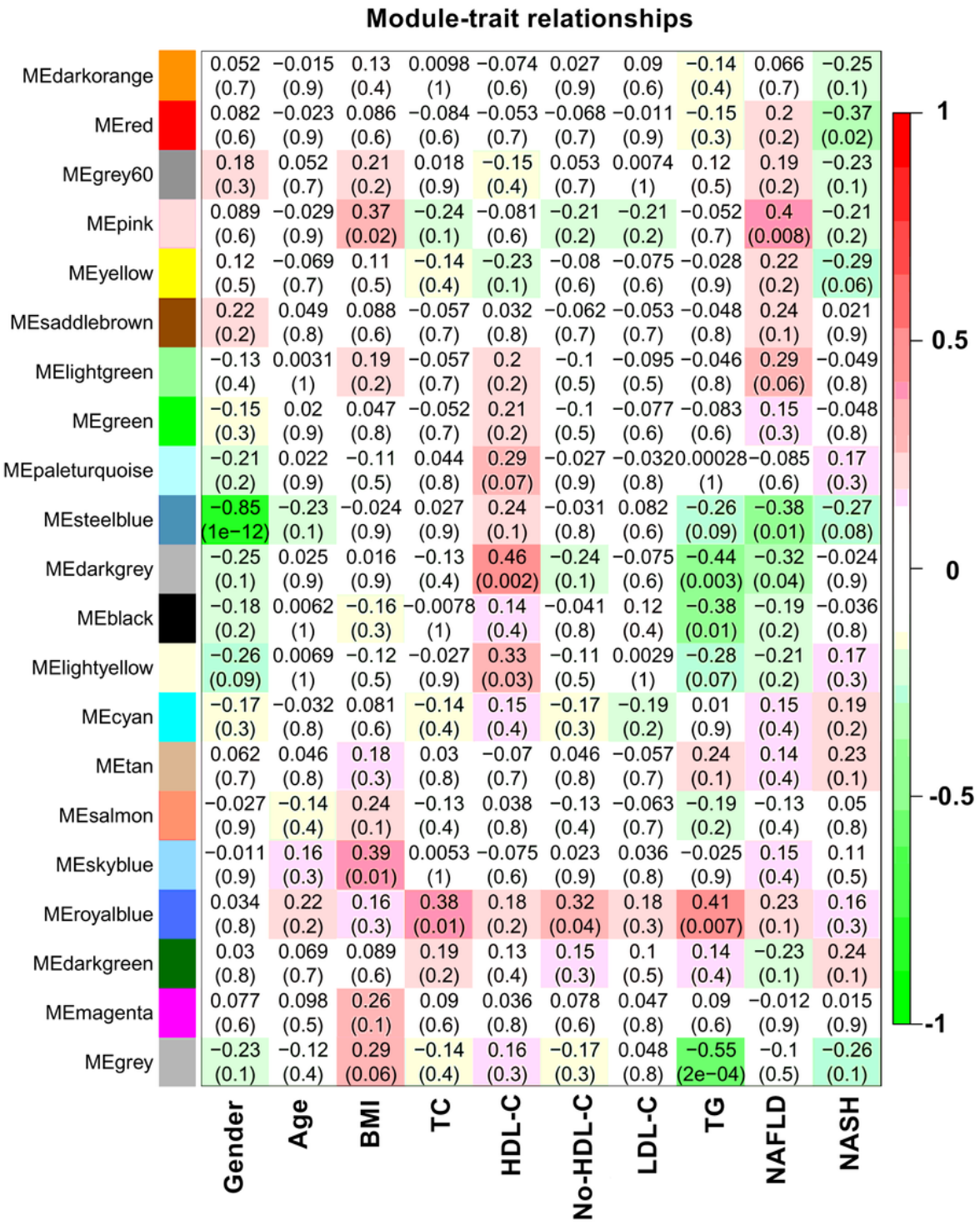

Figure 4

Module-feature associations. Each row corresponds to a modulEigengene 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.

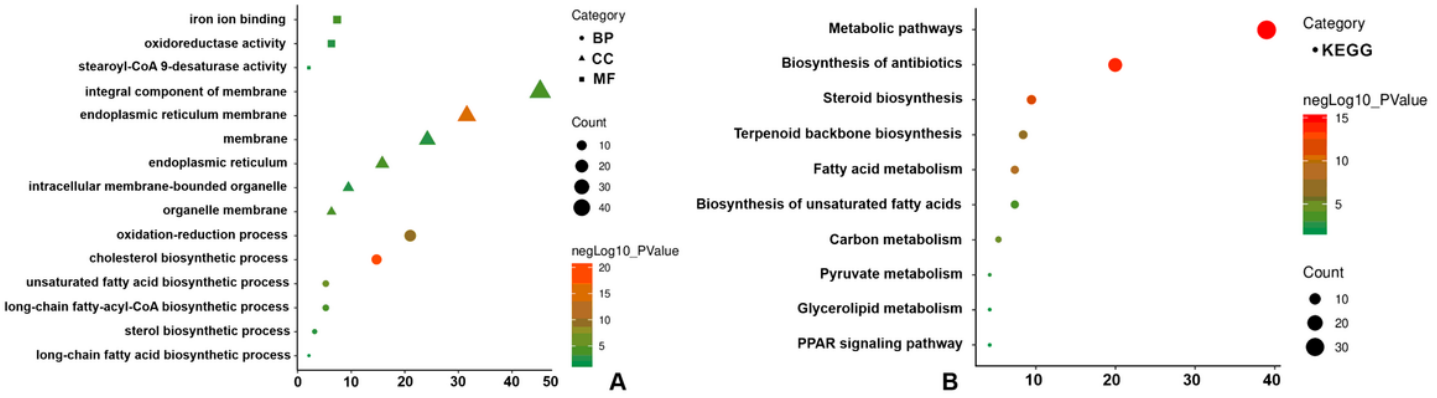

Figure 5

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 -log10 (P-value) of each term is colored according to the legend. (A): GO functional enrichment analysis. (B): KEGG pathway enrichment analysis.

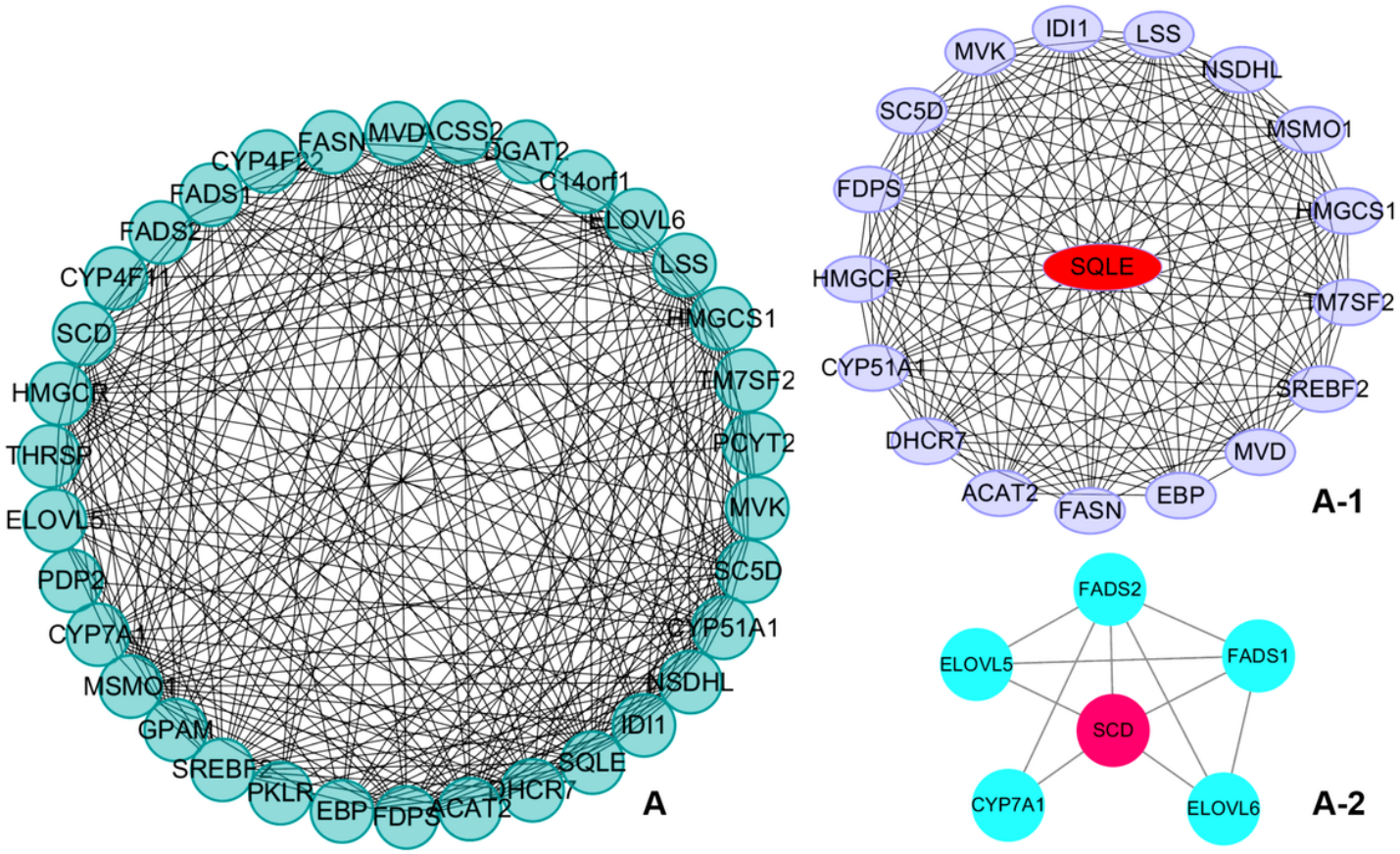

Figure 6

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 > 4.0. (A-1) Molecular-1 with MCODE score = 17.29. (A-2) Molecular-2 with MCODE score = 4.4.

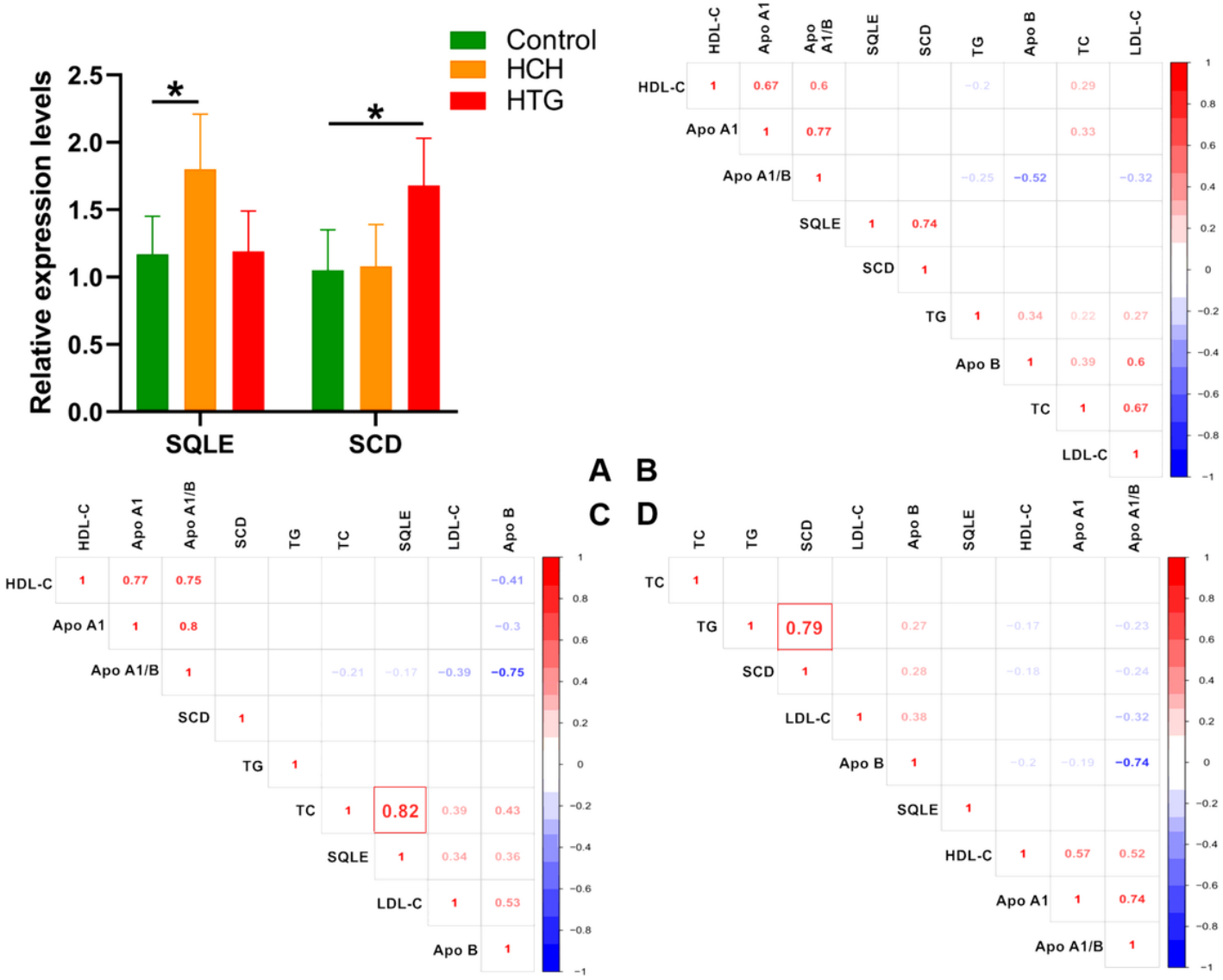

Figure 7

Validation with RT-qPCR (A) and the relationship between genes and lipid parameter in Control (B), HCH (C) and HTG (D). *P < 0.001.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.xlsx
- Additionalfile2.docx