Original Paper

Weighted Gene Co-Expression Network Analysis Identifies Specific Modules and Hub Genes Related to Hyperlipidemia

Liu Miao^a Rui-Xing Yin^a Shang-Ling Pan^b Shuo Yang^a De-Zhai Yang^c
Wei-Xiong Lin^c

^aDepartment of Cardiology, Institute of Cardiovascular Diseases, The First Affiliated Hospital, Guangxi Medical University, Nanning, Guangxi, ^bDepartment of Pathophysiology, School of Premedical Science, Guangxi Medical University, Nanning, Guangxi, ^cDepartment of Molecular Genetics, Medical Scientific Research Center, Guangxi Medical University, Nanning, Guangxi, PR China

Key Words

Array data • Gene Ontology annotation • Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway • Weighted gene co-expression networks analyzed • Protein-protein interaction (PPI) network • Epigenetic influence

Abstract

Background/Aims: The present study attempted to identify the potential key genes and pathways of hyperlipidemia, and to investigate the possible mechanisms associated with them. **Methods:** The array data of GSE3059 were downloaded, including thirteen samples of hyperlipidemia from the Gene Expression Omnibus (GEO) database. The weighted gene co-expression network analysis (WGCNA) was performed with WGCNA package, and the salmon and midnight blue modules were found as the highest correlation. Gene Ontology annotation and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses for these two modules were performed by cluster Profiler and DOSE package. A protein-protein interaction (PPI) network was established using Cytoscape software, and significant modules were analyzed using Molecular Complex Detection. **Results:** Five genes (histone deacetylase 4, *HDAC4*; F2R like trypsin receptor 1, *F2RL1*; abhydrolase domain containing 2, *ABHD2*; transmembrane 4 L six family member 1, *TM4SF1*; and family with sequence similarity 13-member A, *FAM13A*) were found with a significant meaning. When their expression levels were validated with RT-qPCR, the relative expression levels were lower (*HDAC4*) and higher (*F2RL1*, *ABHD2*, *TM4SF1* and *FAM13A*) in hyperlipidemia than in normal controls ($P < 0.05$ - 0.01). Subgroup analysis showed that the relative expression levels of *HDAC4* were lower, whereas those of *F2RL1* and *ABHD2* were higher in Maonan than in Han ethnic groups ($P < 0.05$). **Conclusion:** Except for genetic factors and environmental exposures, epigenetic influence was another mechanism of hyperlipidemia in our study populations, which needed to further confirm.

© 2018 The Author(s)
Published by S. Karger AG, Basel

Rui-Xing Yin

Dept. of Cardiology, Inst. of Cardiovascular Diseases, The First Affiliated Hospital,
Guangxi Med. Univ. Nanning 530021, Guangxi, (PR China)
E-Mail yinruixing@163.com

Introduction

Cardiovascular disease (CVD) is becoming the major cause of death around the world and imposes a significant health-economic load [1]. We had required risk status estimation of a standard lipid profile to assess risk severity, just as total cholesterol (TC) [2], triglyceride (TG) [3], low-density lipoprotein cholesterol (LDL-C) [4], apolipoprotein (Apo) B [5], high-density lipoproteins cholesterol (HDL-C) [6], ApoA1 [7] and the ratio of ApoA1 to ApoB [8] are recommended methods for cardiovascular risk prediction. Among these indicators, elevated levels of serum TC and/or TG, which have also been defined as hyperlipidemic, could identify as strong contributors to CVD mortality [9]. Besides, potential risk factors of CVD have been reported as diabetes mellitus, hypertension, tobacco use, male gender, age, and family history of atherosclerotic arterial disease, family history of heart attacks, lack of physical activity, alcohol consumption, obesity, stress and personal genetic profile [10-16]. All these risk factors which have been taken to individual are important genetic components.

Over the last 10 years, genome-wide association studies (GWASes) have revolutionized the discovery of common genetic variants associated with a range of serum lipid and most of these studies were performed by microarray analysis [17]. Gene co-expression network, a widely used method, has been carried out in analyzing microarray data, especially for marking functional modules [18]. Weighted Gene Co-expression Network Analysis (WGCNA), as one of the most useful gene co-expression network, may identify potential therapeutic targets via WGCNA data analysis [19]. we detected the mRNA expression profile of hyperlipidemia samples to find out the highly connected hub genes and significant modules to show the potential molecular mechanisms.

China is a multi-ethnic country and Maonan is one of the 55 minorities with a long-time history. Recent phylogenetic and principal component analyses revealed that the Maonan people belong to the Southeastern Asian group and the genetic relationship between Maonan ethnic group and other minorities in Guangxi was much closer than that between Maonan and Han nationalities. In a previous study, we have found that the incidence of hyperlipidemia was distinct between Maonan and Han ethnic groups [20]. In this study, we performed WGCNA to construct the co-expression network and mark significant modules in the network. The outcomes may help us for further elucidating the innate character of hyperlipidemia, and provide new insights to potential gene biomarkers and signaling pathways to treat hyperlipidemia.

Materials and Methods

Microarray data

Microarray data of GSE3059 [21] was downloaded from the National Center For Biotechnology Information (NCBI) Gene Expression Omnibus (GEO, [http:// www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)) database [22], which was based on the platform of GPL570 Affymetrix Human Genome U133 Plus 2.0 Array. GSE3059 contains 13 peripheral blood samples from hyperlipidemia patients (mean age = 50.15, 7 males and 6 females) and 19 peripheral blood samples from control individuals who had never been diagnosed with a chronic illness (mean age = 40.13, 10 males and 9 females). The CEL files were transformed into the expression value matrix using the Affy package in R [23], and the probe information was then transformed into the gene name using Bioconductor in R [24]. If one gene had more than one probe, the mean expression value of this gene was selected.

Construction of weighted gene co-expression network

WGCNA is a widely used system biology method, which is used to construct a scale-free network from gene expression data [18]. To ensure that the results of network construction were reliable, outlier samples were removed. An appropriate soft threshold power (soft power = 3) was selected in accordance with standard scale-free networks, with which adjacencies between all differential genes were calculated by a power function. Then, the adjacency was transformed into a topological overlap matrix (TOM), and

the corresponding dissimilarity (1-TOM) was calculated. Module identification was accomplished with the dynamic tree cut method by hierarchically clustering genes using 1-TOM as the distance measure with a deepsplit value of 2 and a minimum size cutoff of 30 for the resulting dendrogram. Highly similar modules were identified by clustering and then merged together with a height cut-off of 0.4. To test the stability of each identified module, module preservation and quality statistics were computed with the module preservation function implemented in the WGCNA package [25].

Finding module of interest and functional annotation

The module membership (MM) was defined as the correlation of gene expression profile with module eigengenes (Mes). And the gene significance (GS) measure was defined as (the absolute value of) the correlation between gene and external traits. Genes with highest MM and highest GS in modules of interest were natural candidates for further research [26]. Thus, the intramodular hub genes were chosen by external traits based $GS > 0.2$ and $MM > 0.6$ with a threshold of P -value < 0.05 [18]. The gene-gene interaction network was constructed and visualized using Cytoscape software package [27] and Molecular Complex Detection (MCODE) [28] was used to analyze the most notable clustering module. MCODE score > 6 was a threshold for next analysis.

Subjects

Two groups of study population including 833 unrelated participants of Maonan (248 males, 29.78% and 578 females, 70.22%) and 826 unrelated subjects of Han (252 males, 30.50% and 574 females, 69.5%) were randomly selected from our previous stratified randomized samples [29]. All participants were resided in the Huanjiang Maonan Autonomous County in the Northwestern of Guangxi Zhuang Autonomous Region, which is located in Southwestern China. The participants' age ranged from 18 to 80 years with a mean age of 55.98 ± 12.78 years in Han and 56.87 ± 12.12 years in Maonan; respectively. The gender ratio and age distribution were matched between the two groups (Table 1). All participants were essentially healthy with no history of coronary artery disease, stroke, diabetes, hyper- or hypo-thyroids, and chronic renal disease. They were free from medications known to affect lipid profiles.

Epidemiological survey

The epidemiological survey was carried out using internationally standardized method, following a common protocol [30]. Information on demographics, socioeconomic status, and lifestyle factors were collected with standardized questionnaires. Cigarette smoking status was categorized into groups of cigarettes per day: ≤ 20 and > 20 . Alcohol consumption was categorized into groups of grams of alcohol per day: ≤ 25 and > 25 [31]. Several parameters such as blood pressure, height, weight and waist circumference were measured, while body mass index (BMI, kg/m^2) was calculated.

Table 1. Comparison of demographic, lifestyle characteristics and serum lipid levels between the Han and Maonan populations. HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; Apo, Apolipoprotein. ¹Mean $\pm$ SD determined by t-test. ²Because of not normally distributed, the value of triglyceride was presented as median (interquartile range), the difference between the two ethnic groups was determined by the Wilcoxon-Mann-Whitney test

Parameter	Han	Maonan	Test-statistic	P
Number	826	833		
Male/female	248/578	252/582	0.137	0.742
Age (year) ¹	55.98 $\pm$ 12.78	56.87 $\pm$ 12.12	1.672	0.205
Height (cm)	154.02 $\pm$ 7.74	153.53 $\pm$ 8.07	2.495	0.114
Weight (kg)	53.01 $\pm$ 8.92	55.95 $\pm$ 10.60	23.359	1.52E-006
Body mass index (kg/m^2)	22.31 $\pm$ 3.22	23.36 $\pm$ 3.70	6.730	0.037
Waist circumference (cm)	77.13 $\pm$ 7.81	76.34 $\pm$ 9.21	24.311	1.78E-007
Smoking status [n (%)]				
Non-smoker	620(75.00)	657(78.84)		
≤ 20 cigarettes/day	182(22.11)	155(18.66)		
> 20 cigarettes/day	24(2.89)	21(2.40)	5.378	0.068
Alcohol consumption [n (%)]				
Non-drinker	666(80.69)	622(74.65)		
≤ 25 g/day	81(9.66)	107(12.90)		
> 25 g/day	79 (9.65)	104(12.45)	8.485	0.014
Systolic blood pressure (mmHg)	129.26 $\pm$ 19.28	135.89 $\pm$ 24.76	69.976	2.33E-016
Diastolic blood pressure (mmHg)	81.55 $\pm$ 11.46	83.47 $\pm$ 12.55	12.250	1.41E-005
Pulse pressure (mmHg)	47.71 $\pm$ 15.29	52.42 $\pm$ 18.56	50.587	3.63E-015
Glucose (mmol/L)	6.18 $\pm$ 1.91	6.15 $\pm$ 1.43	21.278	1.21E-006
Total cholesterol (mmol/L)	4.97 $\pm$ 1.05	5.11 $\pm$ 1.09	6.922	0.031
Triglyceride (mmol/L) ²	1.49(0.48)	1.63(0.71)	7.471	0.029
HDL-C (mmol/L)	1.75 $\pm$ 0.50	1.81 $\pm$ 0.60	0.198	0.657
LDL-C (mmol/L)	2.88 $\pm$ 0.85	2.89 $\pm$ 0.79	12.497	2.03E-005
ApoA1 (g/L)	1.35 $\pm$ 0.26	1.39 $\pm$ 0.32	0.361	0.548
ApoB (g/L)	0.84 $\pm$ 0.19	0.88 $\pm$ 0.20	1.484	0.223
ApoA1/ApoB	1.67 $\pm$ 0.50	1.66 $\pm$ 0.57	0.095	0.758
Hyperlipidemia [n (%)]	401(48.55)	429(51.50)	11.343	1.89E-005
Diabetes [n (%)]	97(11.01)	124(14.89)	9.431	0.011
Hypertension [n (%)]	383(46.37)	411(49.34)	8.452	0.016

Biochemical Measurements

Venous blood samples were obtained from all subjects after at least 12 h of fasting. The levels of serum TC, TG, HDL-C and LDL-C in samples were determined by enzymatic methods with commercially available kits, Tcho-1, TG-LH (RANDOX Laboratories Ltd., Ardmore, Diamond Road, Crumlin Co., Antrim, UK, BT29 4QY), Cholestest HDL, and Cholestest LDL (Daiichi Pure Chemicals Co., Ltd., Tokyo, Japan), respectively. Serum ApoA1 and ApoB levels were detected by the immunoturbidimetric immunoassay (RANDOX Laboratories Ltd.). All determinations were performed with an autoanalyzer (Type 7170A; Hitachi Ltd., Tokyo, Japan) in the Clinical Science Experiment Center of the First Affiliated Hospital, Guangxi Medical University [32].

Diagnostic criteria

The normal values of fasting serum TC, TG, HDL-C, LDL-C, ApoA1 and ApoB levels, as well as the ApoA1/ApoB ratio in our Clinical Science Experiment Center were 3.10–5.17, 0.56–1.70, 1.16–1.42, 2.70–3.10 mmol/L, 1.20–1.60, 0.80–1.05 g/L, and 1.00–2.50; respectively. The individuals with TC > 5.17 mmol/L, and/or TG > 1.70 mmol/L were defined as hyperlipidemic [33].

Quantitative real-time RT-qPCR

According to the microarray results, the 5 most dysregulated mRNAs were chosen for further validation by RT-qPCR in hyperlipidemia patients vs. healthy controls. Blood samples were collected and PBMCs were isolated. Total RNAs were isolated from PBMCs using Trizol reagent (Invitrogen), and complementary DNA was synthesized using the TransScript R First-Strand cDNA Synthesis SuperMix (Transgen, China) according to the manufacturer's instructions. Primer sets for selected genes were designed by Sangon Biotech (Shanghai, China); their sequences and reaction conditions are available in Table 2. Each sample was run in triplicate in 96-well plates using LightCycler R 96 and FastStart Essential DNA Green Master (Roche Diagnostics GmbH, Germany). Quantification cycles (Cq) were calculated using the fit point method (LightCycler R 96 Software, Version 1.1 provided by Roche). The expression data were normalized to the reference glyceraldehyde-3-phosphate dehydrogenase (GAPDH). All experiments (sample collection, preparation and storage, primer design, qPCR normalization) were performed according to the MIQE guidelines.

Statistical Analysis

All statistical analyses were performed using the statistical software package SPSS 21.0 (SPSS Inc. Chicago, IL, USA). A chi-square analysis was used to evaluate the difference of the rate between the groups. Continuous data were presented as means ± SD. Comparisons between groups for continuous data were made using the Kruskal-Wallis and Mann-Whitney nonparametric tests. A raw *P* value of < 0.05 was considered nominally significant. The heat-map of correlation models was measured by R software (version 3.3.0).

Results

Data preprocessing

We analyzed the GSE3059 and obtained 54560 expression probes separately from each gene expression profile. After data preprocessing, the expression matrices of 21194 genes were obtained from the 13 samples.

Weighted gene co-expression networks

We selected a suitable weighted parameter of adjacency function, which is the soft-threshold β , before construction the weighted co-expression network. After the calculation, we selected the correlation coefficient close to 0.8, soft-threshold $\beta = 3$ to construct gene modules using the WGCNA package (Fig. 1). After determining the soft threshold, all of

Table 2. PCR primers for quantitative real-time PCR

Gene	Forward primer	Reverse primer
HDAC4	CAGGAGATGCTGGCCATGAA	GCACTCTCTTTGCCCTTCTC
F2RL1	GGCCAAGGAGGGCTCTTATC	GCGATCGCAGGTCAAACATC
ABHD2	CAGTACCTGGCACAGGGTT	AGAGTTGTACAGTTGGGCCG
FAM13A	GAGCAGAAGGAGTAGGGTGC	CTGTCTCTTCCACGGCAAGT
TM4SF1	GCTGCTCTCACCAACAGCAA	CACAAGGCAGTTCAAAGGGC
GAPDH	AGAGAGAGGCCCTCAGTTGCT	TTGTGAGGGAGATGCTCAGTGT

genes were used to construct weighted gene co-expression networks. According to the basic idea of WGCNA, we calculated the correlation matrix and adjacency matrix of the gene expression profile of the hyperlipidemia groups, and then transformed them into a topological overlap matrix (TOM) and obtained a system clustering tree of genes on the basis of gene-gene non- ω similarity (Fig. 2). Together with the TOM, we performed the hierarchical average linkage clustering method to identify the gene modules of each gene network (deepsplit = 2, cut height = 0.4) (Fig. 3). In hyperlipidemia groups, a total of six gene modules were recognized by the dynamic tree cut (Fig. 4). Genes that did not belong to any modules were housed in the gray module. The gray gene modules were ignored in this study.

Finding module of interest and functional annotation

It is a hugely valued biological significance to find out modules most significantly associated with clinical features. The highest association in the Module-feature relationship was found between midnight blue module and glucose ($r^2 = 0.67$, $P = 0.01$), salmon module and WBC ($r^2 = 0.60$, $P = 0.03$) and salmon module and neutrophil ($r^2 = 0.78$, $P = 0.002$), which were selected as module of interest and clinical feature to be studied in subsequent analyses (Fig. 5). The other modules without enough relationship or statistical significance were not further considered. In order to explore biological relevance of midnight blue module and salmon module, 280 genes in midnight blue module and 312 genes in salmon module were respective subjected to Gene Ontology (GO) functional and KEGG pathway enrichment

Fig. 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).

Fig. 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.

Fig. 3. Relationships among modules. On top was hierarchical clustering of module eigengenes that summarize the modules found in the clustering analysis. Branches of the dendrogram (the meta-modules) group together eigengenes that are positively correlated. Below was heatmap plot of the adjacencies in the eigengene network. Each row and column in the heatmap corresponds to one module eigengene (labeled by color). In the heatmap, red represents high adjacency, while blue color represents low adjacency. Squares of red color along the diagonal are the meta-modules.

Fig. 4. 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. And, below module colors is the phenotype.

analyses by R cluster Profile package [34]. Biological processes of salmon module were found to focus on neutrophil mediated immunity ($P = 4.77 \times 10^{-7}$), positive regulation of defense response ($P = 1.19 \times 10^{-6}$), positive regulation of inflammatory response ($P = 4.57 \times 10^{-6}$), and regulation of innate immune response ($P = 6.26 \times 10^{-5}$). At the same time, biological processes of midnight blue module were found to focus on energy reserve metabolic process ($P = 4.10 \times 10^{-5}$), and positive regulation of β cell activation ($P = 5.67 \times 10^{-5}$). However, in KEGG pathway analysis, salmon module was found to focus on leukocyte trans-endothelial migration ($P = 3.70 \times 10^{-5}$), and neurotrophin signaling pathway ($P = 1.43 \times 10^{-4}$). At the same time, midnight blue module was found to focus on insulin signaling pathway ($P = 3.64$

Fig. 5. Module-feature associations. Each row corresponds to a module Eigengene and each column to a clinical feature. 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.

Fig. 6. GO functional and KEGG pathway enrichment analyses for genes in the object module. The x-axis shows the ratio 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): salmon module; (B): midnight blue module.

$\times 10^{-6}$), and insulin resistance ($P = 3.64 \times 10^{-6}$; Fig. 6). The top 6 of the databases were shown in Table 3 and 4.

Protein-protein interaction (PPI) network construction and identify hub genes

When the STRING database [35] was analyzed, a total of 48 nodes and 266 protein pairs were got with a combined weight score > 0.25 in salmon module (Fig. 7A). After analysis in sub-module, only one module with score > 6 was detected by MCODE. As shown in Fig. 7A-1, the top 10 high degree genes, including *PACSIN2* (degree = 27), *HDAC4* (degree = 24), *C11orf21* (degree = 22), *RAB35* (degree = 21), *UPF1* (degree = 20), *NADK* (degree = 18), *FFAR2* (degree = 17), *NRARP* (degree = 16), *F2RL1* (degree = 11) and *ABHD2* (degree = 10) were hub nodes with higher node degrees in this module. However, there were altogether 70 nodes and 232 protein pairs were got with a combined weight score > 0.42 in midnight blue module (Fig. 7B), after analysis in degree, the top 5 high degree genes, including *PDZRN4*

(degree = 45), *TM4SF1* (degree = 42), *SCG5* (degree = 36), *FAM13A* (degree = 34) and *GOT1L1* (degree = 33), could be further considered. When combined with GO functional, KEGG pathway enrichment and PPI analysis, we hypothesized that *HDAC4*, *F2RL1*, *ABHD2*, *TM4SF1* and *FAM13A* as the hub genes were closely relevant to hyperlipidemia.

Validation with RT-qPCR and the relationship between genes and lipid parameter

We performed RT-qPCR to validate the hypothesized data. To verify the main conclusion drawn from the microarray results for peripheral blood samples, the relative expression levels of genes coding for *HDAC4*, *F2RL1*, *ABHD2*, *TM4SF1* and *FAM13A* were determined. Overall, the RT-qPCR results were qualitatively consistent with the results of the microarray analysis. However, the RT-qPCR analysis has shown that the *F2RL1*, *ABHD2*, *TM4SF1* and *FAM13A* relative expression tended to give higher up-regulation levels in hyperlipidemia group than

Table 3. GO and KEGG enrichment analysis in salmon modules (top 6 significantly enriched biology terms). GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes pathway; BP, biological process

ID	Category	Description	Q-value	Gene ID	Count
BP	GO:0002446	neutrophil mediated immunity	1.094E-03	10097/101/10092/427/10576/960/1476/1727/2150/2665/2799/3326/3689/3916/64386/11333/5175/5660/5795/5873/9545/387/9342/54472	25
BP	GO:0031349	regulation of defense response	1.094E-03	101/2113/2150/8772/2867/3329/3572/51135/83737/3689/3916/4214/5603/57506/118788/5294/5734/5817/8780/293/6777/7099	22
BP	GO:0042119	neutrophil activation	1.094E-03	10097/101/10092/427/10576/960/1476/1727/2150/2665/2799/3326/3689/3916/64386/11333/5175/5660/5795/5873/9545/387/9342/54472	24
BP	GO:0036230	granulocyte activation	1.094E-03	10097/101/10092/427/10576/960/1476/1727/2150/2665/2799/3326/3689/3916/64386/11333/5175/5660/5795/5873/9545/387/9342/54472	24
BP	GO:0002283	neutrophil activation involved in immune response	1.657E-03	10097/101/10092/427/10576/960/1476/1727/2665/2799/3326/3689/3916/64386/11333/5175/5660/5795/5873/9545/387/9342/54472	23
BP	GO:0043312	neutrophil degranulation	1.657E-03	10097/101/10092/427/10576/960/1476/1727/2665/2799/3326/3689/3916/64386/11333/5175/5660/5795/5873/9545/387/9342/54472	23
KEGG	hsa05130	Pathogenic Escherichia coli infection	3.348E-03	7430/10092/9181/387/7099/10094	6
KEGG	hsa04076	trans endothelial migration	4.173E-03	7430/9759/3689/5175/5603/5294/83593/50848	8
KEGG	hsa05131	Shigellosis	4.173E-03	10092/1399/5603/10094/199746/960	6
KEGG	hsa05142	Chagas disease (American trypanosomiasis)	6.305E-03	8772/8837/7099/5603/7048/5294/51135	7
KEGG	hsa03018	RNA degradation	6.305E-03	10140/25804/23019/84186/29883/3329	6
KEGG	hsa05152	Tuberculosis	6.305E-03	3111/8772/387/7099/3689/5603/3916/51135/3329	9

Table 4. GO and KEGG enrichment analysis in midnight blue modules (top 6 significantly enriched biology terms). GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes pathway; BP, biological process

ID	Category	Description	Q-value	Gene ID	Count
BP	GO:0006112	energy reserve metabolic process	7.522E-03	112609/9365/3953	3
BP	GO:0006898	receptor-mediated endocytosis	7.522E-03	2348/3512/3537/283316/8909/28823/28442	7
BP	GO:0009887	organ morphogenesis	7.522E-03	3265/2273/8829/2263/333	5
BP	GO:0000187	activation of MAPK activity	7.522E-03	5562/60370/5148/7057/983	5
BP	GO:0007411	axon guidance	7.522E-03	3265/8029/6710/10512/6708/123016	6
BP	GO:0050871	positive regulation of β cell activation	8.140E-03	3537/100423062/28442	3
KEGG	hsa04144	Endocytosis	1.094E-03	2348/3265/26052/23362/2263/2321/2352	7
KEGG	hsa04910	Insulin signaling pathway	1.657E-03	3265/10891/5562/23265/57521	5
KEGG	hsa04931	Insulin resistance	1.657E-03	5524/10891/5562/27330	4

Fig. 7. The Protein-protein interaction analysis of the differentially expressed genes. (A and B): Protein-protein interaction network of the module genes (A shown as salmon and B as midnightblue module). Edge stands for the interaction between two genes. A degree was used for describing the importance of protein nodes in the network, red shows a high degree and green presents a low degree. (A1-3): The significant modules identified from the protein-protein interaction network using the molecular complex detection method with a score of > 6.0. MCODE_{A-1} score = 9.6458, MCODE_{A-2} score = 4.2253 and MCODE_{A-3} score = 2.2171.

those in normal controls, besides, *F2RL1* and *ABHD2* relative expression levels in Maonan ethnic group were higher than those in Han group (Fig. 8A). As shown in Fig. 8B and 8C, the relationship between genes and lipid parameter and *HDAC4* was negative correlation with other genes and lipid parameter.

Discussion

TC, LDL-C, HDL-C, and TG are correlated phenotypes inherited as complex traits and hyperlipidemia as a complex and multifactorial disease caused by an interaction between genetic and environmental factors [36]. In our study, we have identified 5 genes which may modify serum lipid levels, but, the reason for causing the change of serum lipid levels was slightly different.

Histone deacetylases (HDACs) are enzymes that regulate gene expression by removing acetyl residues from target histones or non-histone proteins and consequently modifying chromatin structure. HDACs are divided into four classes (class I: HDAC1, 2, 3 and 8; class II: HDAC4, 5, 6, 7, 9 and 10; class III: SIRT1-7; class IV: HDAC11) [37]. HDAC4, one member of the tissue-specific Class II HDACs, is highly expressed in neurons [38] and bone mass, and plays an essential role in maintaining neuronal survival [39] and chondrocyte hypertrophy [40]. When we analysis the other 4 genes in UCSC, we could only find that the *F2RL1* and *ABHD2* concentrate peak appeared in *H3K27AC* before gene transaction, but nothing can

Fig. 8. Validation with RT-qPCR and the relationship between genes and lipid parameter. (A): Several hub genes identified in microarray data are dysregulation in hyperlipidemia patients. The mRNA expression of hub genes identified in microarray data validated by RT-qPCR is shown. Total RNAs were isolated from PBMCs or hyperlipidemia patients and normal donors. Reverse-transcribed to cDNA and used as template for RT-qPCR analysis. Relative Expression of each gene in PBMCs from healthy donors were considered as 1. (B): Hyperlipidemia patients. (C): Normal donors.

Fig. 9. Annotation of genes in UCSC. (A): TM4SF1; (B): FAM13A; (C): F2RL1; (D): ABHD2.

be found in *TM4SF1* and *FAM13A* (Fig. 9). Braun *et al.* found that DNA methylation was a key epigenetic mechanism that is suggested to be associated with blood lipid levels [41]. Maybe histone modification is another key epigenetic mechanism for lipid modified.

Besides, nuclear HDAC4 distribution was enhanced in Purkinje neurons from Atm-deficient mice after lipopolysaccharides (LPS) stimulation, and Atm was identified to be involved in ataxia-telangiectasia characterized by immune deficiency [42], indicating that HDAC4 may directly or indirectly regulate inflammation genes. Recently, novel data have emerged that immune responses and lipid metabolism interact in a unique metabolic pathway underlying atherosclerosis [43]. Mihaylova *et al.* have suggested that class I/II HDACs may be influenced blood glucose and would be potential therapeutics for metabolic syndrome [44]. In our current study, *F2RL1*, *ABHD2*, *TM4SF1* and *FAM13A* expression tended to give higher up-regulation levels in hyperlipidemia group than those in normal controls and *HDAC4* was negatively correlated with other genes. These results are consistent with the previous findings.

Except for genetic factors, environmental exposures were another aspect which cannot be ignored. Maonan people like to pickle sour meat, snails and vegetables. They get meat mainly from poultries and livestock, such as pigs, oxen, chickens, ducks and so on. A typical food, Minglun Sliced Pig is a well-known dish of the Maonan ethnic group. Most of the Maonan people like to eat food which is cooked half ripe. In addition, they also like to eat beef, pork and /or animal offals in a hot pot which contain abundant saturated fatty acid [20]. Long-term high saturated fat diet is an important risk factor for obesity, dyslipidemia, atherosclerosis, and hypertension [45]. The major dietary saturated long-chain fatty acids such as myristic acid (14:0) and palmitic acid (16:0) have been associated with deleterious effects on blood lipid metabolism, especially due to their influence on plasma TC and TG levels [46].

Recently, HDAC inhibitors in turn regulate the activity of HDACs, and have been widely used as therapeutics in psychiatry and neurology, in which a number of adverse outcomes are associated with aberrant HDAC function. At the same time, dietary HDAC inhibitors have been shown to have a regulatory effect similar to that of pharmacological HDAC inhibitors without the possible side-effects [47]. Wood and Yang *et al.* have found that a small number of dietary compounds appear to have dual effects on histone acetylation status and excessive alcohol consumption and cigarette smoking have been associated with histone acetylation status modified [48, 49]. In the present study, we showed that the percent of alcohol consumption were higher in Maonan than in Han groups. Most of the local adult men of the Maonan people like to drink. They even have the custom that it would be considered to be impolite to treat their guests without wines. Some families made wines themselves using grain sorghums and corns. Excessive alcohol consumption, just as HDAC inhibitors, removal of acetyl moiety alters the histone configuration, once more returning the chromatin to the closed form and influence transaction. That would be an explanation for the reason why the relative expression of *F2RL1* and *ABHD2* were higher in Maonan than in Han groups.

In conclusion, thirteen hyperlipidemia microarray datasets from GEO series had been systematic analyzed. After finished weighted gene co-expression networks analysis, we found salmon and midnight blue module with the highest correlation. We performed GO functional, KEGG pathway enrichment and protein-protein interactions analyses for these two modules, five genes (*HDAC4*, *F2RL1*, *ABHD2*, *TM4SF1* and *FAM13A*) were found with a significant meaning. After validation with RT-qPCR, we found that the relative expression of *HDAC4* was lower, but *F2RL1*, *ABHD2*, *TM4SF1* and *FAM13A* were higher in hyperlipidemia than in normal controls. Subgroup analysis showed that the relative expression of *HDAC4* was lower, but, *F2RL1* and *ABHD2* expression was higher in Maonan than in Han ethnic groups. Except for genetic factors and environmental exposures, epigenetic influence was another mechanism which cannot be ignored, but, it was still needed to further confirm.

Acknowledgements

L.M. conceived the study, participated in the design, undertook genotyping, performed the statistical analyses, and drafted the manuscript. R.-X.Y. conceived the study, participated in the design, carried out the epidemiological survey, collected the samples, and helped to draft the manuscript. S.Y. collaborated to the genotyping. S.-L.P., D.-Z.Y. and W.-X.L. carried out the epidemiological survey and collected the samples. All authors read and approved the final manuscript. The authors acknowledge the essential role of the funding of the National Natural Science Foundation of China (No: 81460169) and the Innovation Project of Guangxi Graduate Education in this motif.

Disclosure Statement

The authors have no potential conflicts of interest to report.

References

- Pandya A, Gaziano TA, Weinstein MC, Cutler D: More americans living longer with cardiovascular disease will increase costs while lowering quality of life. *Health Aff (Millwood)* 2013;32:1706-1714.
- Matthan NR, Zhu L, Pencina M, D'Agostino RB, Schaefer EJ, Lichtenstein AH: Sex-specific differences in the predictive value of cholesterol homeostasis markers and 10-year cardiovascular disease event rate in Framingham Offspring Study participants. *J Am Heart Assoc* 2013;2:e005066.
- Lindman AS, Veierod MB, Tverdal A, Pedersen JI, Selmer R: Nonfasting triglycerides and risk of cardiovascular death in men and women from the Norwegian Counties Study. *Eur J Epidemiol* 2010;25:789-798.
- Orehov AN, Bobryshev YV, Sobenin IA, Melnichenko AA, Chistiakov DA: Modified low density lipoprotein and lipoprotein-containing circulating immune complexes as diagnostic and prognostic biomarkers of atherosclerosis and type 1 diabetes macrovascular disease. *Int J Mol Sci* 2014;15:12807-12841.
- Furtado JD, Campos H, Appel LJ, Miller ER, Laranjo N, Carey VJ, Sacks FM: Effect of protein, unsaturated fat, and carbohydrate intakes on plasma apolipoprotein B and VLDL and LDL containing apolipoprotein C-III: results from the OmniHeart Trial. *Am J Clin Nutr* 2008;87:1623-1630.
- Chapman MJ, Ginsberg HN, Amarenco P, Andreotti F, Boren J, Catapano AL, Descamps OS, Fisher E, Kovanen PT, Kuivenhoven JA, Lesnik P, Masana L, Nordestgaard BG, Ray KK, Reiner Z, Taskinen MR, Tokgozoglu L, Tybjaerg-Hansen A, Watts GF, European Atherosclerosis Society Consensus P: 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;32:1345-1361.
- Hamon SC, Kardina SL, Boerwinkle E, Liu K, Klos KL, Clark AG, Sing CF: Evidence for consistent intragenic and intergenic interactions between SNP effects in the APOA1/C3/A4/A5 gene cluster. *Hum Hered* 2006;61:87-96.
- Tamang HK, Timilsina U, Singh KP, Shrestha S, Raman RK, Panta P, Karna P, Khadka L, Dahal C: Apo B/Apo A-I Ratio is Statistically A Better Predictor of Cardiovascular Disease (CVD) than Conventional Lipid Profile: A Study from Kathmandu Valley, Nepal. *J Clin Diagn Res* 2014;8:34-36.
- Prospective Studies C, Lewington S, Whitlock G, Clarke R, Sherliker P, Emberson J, Halsey J, Qizilbash N, Peto R, Collins R: Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55, 000 vascular deaths. *Lancet* 2007;370:1829-1839.
- Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L, Investigators IS: Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;364:937-952.
- Merry AH, Boer JM, Schouten LJ, Feskens EJ, Verschuren WM, Gorgels AP, van den Brandt PA: Smoking, alcohol consumption, physical activity, and family history and the risks of acute myocardial infarction and unstable angina pectoris: a prospective cohort study. *BMC Cardiovasc Disord* 2011;11:13.

- 12 Kessler T, Erdmann J, Schunkert H: Genetics of coronary artery disease and myocardial infarction--2013. *Curr Cardiol Rep* 2013;15:368.
- 13 Heidemann C, Hoffmann K, Klipstein-Grobusch K, Weikert C, Pischon T, Hense HW, Boeing H: Potentially modifiable classic risk factors and their impact on incident myocardial infarction: results from the EPIC-Potsdam study. *Eur J Cardiovasc Prev Rehabil* 2007;14:65-71.
- 14 Kathiresan S, Srivastava D: Genetics of human cardiovascular disease. *Cell* 2012;148:1242-1257.
- 15 Abdulla J, Kober L, Abildstrom SZ, Christensen E, James WP, Torp-Pedersen C: Impact of obesity as a mortality predictor in high-risk patients with myocardial infarction or chronic heart failure: a pooled analysis of five registries. *Eur Heart J* 2008;29:594-601.
- 16 Labounty TM, Gomez MJ, Achenbach S, Al-Mallah M, Berman DS, Budoff MJ, Cademartiri F, Callister TQ, Chang HJ, Cheng V, Chinnaiyan KM, Chow B, Cury R, Delago A, Dunning A, Feuchtnier G, Hadamitzky M, Hausleiter J, Kaufmann P, Kim YJ, Leipsic J, Lin FY, Maffei E, Raff G, Shaw LJ, Villines TC, Min JK: Body mass index and the prevalence, severity, and risk of coronary artery disease: an international multicentre study of 13, 874 patients. *Eur Heart J Cardiovasc Imaging* 2013;14:456-463.
- 17 Jeemon P, Pettigrew K, Sainsbury C, Prabhakaran D, Padmanabhan S: Implications of discoveries from genome-wide association studies in current cardiovascular practice. *World J Cardiol* 2011;3:230-247.
- 18 Horvath S, Dong J: Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol* 2008;4:e1000117.
- 19 Zhao H, Cai W, Su S, Zhi D, Lu J, Liu S: Screening genes crucial for pediatric pilocytic astrocytoma using weighted gene coexpression network analysis combined with methylation data analysis. *Cancer Gene Ther* 2014;21:448-455.
- 20 Miao L, Yin RX, Yang S, Pan SL, Lin WX, Yang DZ: BRCA2 rs9534275 polymorphism and serum lipid traits in the Maonan and Han populations. *Int J Clin Exp Pathol* 2017;10:3163-3178.
- 21 Ma J, Dempsey AA, Stamatou D, Marshall KW, Liew CC: Identifying leukocyte gene expression patterns associated with plasma lipid levels in human subjects. *Atherosclerosis* 2007;191:63-72.
- 22 Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R: NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* 2007;35:D760-765.
- 23 Gautier L, Cope L, Bolstad BM, Irizarry RA: affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004;20:307-315.
- 24 Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J: Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- 25 Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- 26 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;18:463-472.
- 27 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-2504.
- 28 Bader GD, Hogue CW: An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2.
- 29 Miao L, Yin RX, Pan SL, Yang S, Yang DZ, Lin WX: Association between the MVK and MMAB polymorphisms and serum lipid levels. *Oncotarget* 2017;8:70378-70393.
- 30 Miao L, Yin RX, Wu JZ, Yang S, Lin WX, Pan SL: The SRGAP2 SNPs, their haplotypes and G × E interactions on serum lipid traits. *Sci Rep* 2017;7:11626.
- 31 Zhang QH, Yin RX, Gao H, Huang F, Wu JZ, Pan SL, Lin WX, Yang DZ: Association of the SPTLC3 rs364585 polymorphism and serum lipid profiles in two Chinese ethnic groups. *Lipids Health Dis* 2017;16:1.
- 32 Miao L, Yin RX, Huang F, Chen WX, Cao XL, Wu JZ: The effect of MVK-MMAB variants, their haplotypes and GxE interactions on serum lipid levels and the risk of coronary heart disease and ischemic stroke. *Oncotarget* 2017;8:72801-72817.
- 33 Miao L, Yin RX, Yang S, Huang F, Chen WX, Cao XL: Association between single nucleotide polymorphism rs9534275 and the risk of coronary artery disease and ischemic stroke. *Lipids Health Dis* 2017;16:193.

- 34 Yu G, Wang LG, Han Y, He QY: clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284-287.
- 35 Chen X, Zheng Z, Chen L, Zheng H: MAPK, NFkappaB, and VEGF signaling pathways regulate breast cancer liver metastasis. *Oncotarget* 2017;8:101452-101460.
- 36 Assmann G, Cullen P, Schulte H: Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* 2002;105:310-315.
- 37 Haberland M, Montgomery RL, Olson EN: The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* 2009;10:32-42.
- 38 Darcy MJ, Calvin K, Cavnar K, Ouimet CC: Regional and subcellular distribution of HDAC4 in mouse brain. *J Comp Neurol* 2010;518:722-740.
- 39 Majdzadeh N, Morrison BE, D'Mello SR: Class IIA HDACs in the regulation of neurodegeneration. *Front Biosci* 2008;13:1072-1082.
- 40 Vega RB, Matsuda K, Oh J, Barbosa AC, Yang X, Meadows E, McAnally J, Pomajzl C, Shelton JM, Richardson JA, Karsenty G, Olson EN: Histone deacetylase 4 controls chondrocyte hypertrophy during skeletogenesis. *Cell* 2004;119:555-566.
- 41 Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, consortium B, Hofman A, Hu FB, Franco OH, Dehghan A: Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. *Clin Epigenetics* 2017;9:15.
- 42 Yang Y, Hui CW, Li J, Herrup K: Correction: The Interaction of the Atm Genotype with Inflammation and Oxidative Stress. *PLoS One* 2015;10:e0130645.
- 43 Schaftenaar F, Frodermann V, Kuiper J, Lutgens E: Atherosclerosis: the interplay between lipids and immune cells. *Curr Opin Lipidol* 2016;27:209-215.
- 44 Mihaylova MM, Vasquez DS, Ravnskjaer K, Denechaud PD, Yu RT, Alvarez JG, Downes M, Evans RM, Montminy M, Shaw RJ: Class Ila histone deacetylases are hormone-activated regulators of FOXO and mammalian glucose homeostasis. *Cell* 2011;145:607-621.
- 45 Ruixing Y, Jinzhen W, Yaoheng H, Jing T, Hai W, Muyan L, Yiyang L, Dongmei F, Hanjun Y, Yuming C: Associations of diet and lifestyle with hyperlipidemia for middle-aged and elderly persons among the Guangxi Bai Ku Yao and Han populations. *J Am Diet Assoc* 2008;108:970-976.
- 46 Lottenberg AM, Afonso Mda S, Lavrador MS, Machado RM, Nakandakare ER: The role of dietary fatty acids in the pathology of metabolic syndrome. *J Nutr Biochem* 2012;23:1027-1040.
- 47 Bassett SA, Barnett MP: The role of dietary histone deacetylases (HDACs) inhibitors in health and disease. *Nutrients* 2014;6:4273-4301.
- 48 Wood JG, Rogina B, Lavu S, Howitz K, Helfand SL, Tatar M, Sinclair D: Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* 2004;430:686-689.
- 49 Yang SR, Wright J, Bauter M, Seweryniak K, Kode A, Rahman I: Sirtuin regulates cigarette smoke-induced proinflammatory mediator release via RelA/p65 NF-kappaB in macrophages *in vitro* and in rat lungs *in vivo*: implications for chronic inflammation and aging. *Am J Physiol Lung Cell Mol Physiol* 2007;292:L567-576.