

Associations of Gene Expressions in Liver and Blood Cells and Circulating Proteins With Coronary Artery Disease

Tianci Chai

Fujian Medical University Union Hospital

Zhisheng Wan

Fujian Medical University Union Hospital

Xiaojie Yang

Fujian Medical University Union Hospital

Zhihuang Qiu

Fujian Medical University Union Hospital

Liangwan Chen (✉ chenliangwan@fjmu.edu.cn)

Fujian Medical University Union Hospital <https://orcid.org/0000-0002-4211-3842>

Research

Keywords: Low-density lipoprotein cholesterol, Coronary artery disease, Genome-wide association study, Mendelian randomization

Posted Date: March 31st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-365589/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Circulating low-density lipoprotein cholesterol (LDL-C) plays a causal role in coronary artery disease (CAD). However, genetic risk factors of CAD were not yet fully understood. The aim of the study was to identify additional factors that contributed to CAD.

Methods: We conducted integrative analysis on publicly available data from genome-wide association studies and quantitative trait locus studies by employing two-sample Mendelian randomization methods to examine the associations of gene expression and circulating protein levels with LDL-C and CAD.

Results: We found that mRNA expression levels of *SORT1*, *PSRC1* and *CELSR2* in liver cells were significantly associated with both LDL-C and CAD ($P < 5.0 \times 10^{-8}$). Higher expression levels of *SORT1*, *PSRC1* and *CELSR2* in the liver were significantly associated with lower circulating LDL-C level (beta = -0.142, -0.138 and -0.171, respectively) and CAD risk (beta = -0.10, -0.097 and -0.121, respectively). Expression levels of 31 genes in blood cells associated with LDL-C level were found. Expressions of *PSRC1*, *IL6R*, *GGCX*, *VAMP8*, *LIPA*, *NT5C2*, *SWAP70*, *EIF2B2*, *FURIN*, *FES* and *ATP5G1* in blood cells were significantly associated with CAD. Higher circulating granulins and apolipoprotein B levels, which were strongly affected by *PSRC1* and *APOB* variants, were significantly associated with higher LDL-C level (beta = 0.22 and 1.35) and CAD risk, with odds ratios of 1.15 (1.10-1.19) and 1.45 (1.21-1.74), respectively.

Conclusions: This study identified gene expressions in liver and blood cells and circulating granulins and apolipoprotein B that were genetically associated with LDL-C and CAD and provided evidence for exploring the potential causal relationship.

Introduction

Coronary artery disease (CAD) is one of the most common cardiovascular diseases, which is associated with high morbidity and mortality[1]. Special professional groups such as military pilots are at higher risk of CAD than the general population. The pathogenesis of CAD is not fully understood but is known to be the result of the interaction of various genetic and environmental factors. It is a process in which lipoprotein accumulates in the arteries that supply blood to the heart. Dyslipidemia is thought to be a major factor in the development of CAD, and low-density lipoprotein cholesterol (LDL-C) has been shown to be one of the most important causal risk factors. In view of the well-established causal role of LDL-C in CAD, biological molecules associated with LDL-C may be risk factors for CAD.

Large-scale genome-wide association studies (GWAS) have robustly identified large numbers of risk-associated genetic variants implicated in LDL-C[2–4] and CAD[5, 6]. However, the GWAS identified loci were located in complex genomic regions containing multiple genes[7]. It is unclear which of these genes are functionally related to LDL-C and CAD functions. The colocalization of genetic associations for disease traits with those for intermediate molecular phenotypes, such as gene expression and metabolomics, provides powerful evidence to advance hypotheses regarding the genes and pathways

through which these disease-associated variants mediate their effects[8]. Quantitative trait loci (QTLs) for gene expressions and circulating proteins can offer a route to a comprehensive molecular interpretation of the GWAS findings. Therefore, by integrative analysis of omics data have the potential to find pathogenic genes for CAD.

Mendelian randomization (MR) is an analytical technique that assessed the correlation between genetic alternatives to intermediate biomarkers and subsequent diseases that are supposed to be caused by intermediate biomarkers based on the random distribution of genetic variation specific to biomarkers. The recent developed two-sample multi-instrumental MR methods that base on GWAS summary data provide feasible ways to integrate omics data from independent GWAS, including QTL studies on genome-wide mRNA expression and circulating protein levels (eQTL and pQTL, respectively)[9–11]. The two-sample MR methods are data integration methods that have being widely applied in identification of intermediate molecular for disease traits and therefore are effective strategies to explicate the GWAS findings.

In terms of lipoprotein metabolism, the liver is the major site of removal of LDL-C from the circulation. Altered liver expression of genes involved in lipid metabolism[12–14]. Hepatic *de novo* lipogenesis influences hepatic cholesterol content as well as its effects on the circulating lipid levels[15]. Blood cell gene expressions also regulate cholesterol homeostasis[16–18]. In this study, we applied several data integration methods to identify potential risk factors such as gene expressions in liver and blood cells and circulating proteins for LDL-C and CAD by using a combination of data from GWAS, eQTL and pQTL studies (Fig. 1).

Materials And Methods

Data resources

This study used datasets from two LDL-C and one CAD GWAS, four eQTL studies (1 in liver and 3 in whole blood cells), and 3 pQTL studies on circulating protein levels. The two LDL-C datasets were obtained from the lipids GWAS that were conducted by the Global Lipids Genetics Consortium (GLGC). The first GWAS evaluated the associations between almost 2.6 million SNPs and lipid levels in 188,578 Europeans[4]. The dataset containing summary results of the association between almost 2.6 million SNPs and LDL-C level was downloaded at <http://csg.sph.umich.edu/abecasis/public/lipids2013/> and used in our analysis. The second LDL-C GWAS was a meta-analysis of exome-wide association studies that evaluated the association between almost 292,417 variants and LDL-C level in 47,532 East Asians and more than 300,000 individuals primarily (84%) of European descent (<http://csg.sph.umich.edu/abecasis/public/lipids2017EastAsian>)[2].

We used the summary results from a large-scale CAD GWAS meta-analysis which was conducted by the CARDIoGRAMplusC4D consortium [6]. This study enrolled about 185,000 individuals that were mainly (77%) of European ancestry. The dataset contained the summary statistics of associations between

almost 9.5 million SNPs and CAD tested under the additive model in the initial GWAS (<http://www.cardiogramplusc4d.org/data-downloads/>).

For the four eQTL studies, one eQTL study was conducted by Westra et al. This study performed eQTL meta-analysis in non-transformed peripheral blood samples from 5,311 individuals of European descent [19]. It is the largest eQTL meta-analysis in blood cells to date. The second eQTL study is the genetic architecture of gene expression (GAGE) study that detected eQTLs in whole blood of 2,765 unrelated Europeans [20]. The remaining two eQTL datasets contained the *cis*-eQTL summary level data on mRNA expression levels in liver and whole blood cells from the GTEx project [21]. The summary data of these four eQTL studies was available at <https://cnsgenomics.com/software/smr/#eQTLsummarydata>.

We acquired pQTL summary data from three studies that examined the associations between genome-wide SNPs and circulating protein levels in thousands of individuals. The first pQTL study tested genome-wide associations between 509,946 SNPs and circulating levels of 1,124 proteins in blood samples of 1,000 individuals from the KORA study [22]. The summary data was available at <http://metabolomics.helmholtz-muenchen.de/pgwas/index.php?task=download>. The second pQTL study performed genome-wide testing of 10.6 million imputed autosomal variants against levels of 2,994 circulating proteins in 3,301 individuals of European descent from the INTERVAL study [23]. The summary data was available at <http://www.phpc.cam.ac.uk/ceu/proteins/>. The third pQTL study analyzed 123 metabolites in up to 24,925 individuals [24]. The 123 metabolic traits were quantified by nuclear magnetic resonance spectroscopy in blood samples. The summary data was available at http://www.computationalmedicine.fi/data#NMR_GWAS.

SMR analysis

The summary data-based MR (SMR) approach is a kind of two-sample multi-instrumental MR method that provides feasible ways for integrating GWAS summary data with QTL data [11]. SMR facilitates the goal of integration of summary statistics from large-scale GWAS with transcriptome-wide association data. Besides, in such MR approaches, the exposure and outcome are not necessarily measured in the same samples, and the data are summary statistics rather than the raw data of the individuals [11]. The instrumental variables (SNPs that were both tested in the independent QTL studies and GWAS) provided the summary data (e.g., the regression coefficient beta values and standard error) on the effects of SNPs on the levels of biomolecules and outcomes. By analyzing this summary data SMR estimates the causal effects of the biomolecules on the outcomes.

The SMR software (version 0.712) is a command-line program, which was downloaded from <http://cnsgenomics.com/software/smr/>. The parameters were left at the default setting in the analysis. The outcome data (i.e., SNP rs number, alleles, allele frequency, beta values, standard error, *P* values) required for SMR analysis was collected from the LDL-C and CAD GWAS datasets, and then was organized to a .ma file with 8 columns specific for the SMR analysis by using the R program. The files contained eQTL summary data in binary format for the SMR analysis were available at <http://cnsgenomics.com/software/smr/#DataResource> (described above). We used genotype data of

HapMap r23 CEU as a reference panel to calculate the linkage disequilibrium (LD) correlation matrix for SMR. The genome-wide significance threshold for the SMR analysis was 5.0×10^{-6} . We also performed the heterogeneity in dependent instruments (HEIDI) test to test the ‘no horizontal pleiotropy’ assumption. The HEIDI test was conducted to examine if there is a single causal SNP affecting gene expression and the phenotype. The genes with $P_{\text{HEIDI}} \geq 0.05$ (without heterogeneity) were considered. Functional annotation enrichment analyses were performed for the identified genes by using the DAVID online tools.

pQTL analysis

We searched for pQTLs among SNPs that located in the identified genes. The pQTLs were expected to provide us with clues to find out proteins that were related to LDL-C and CAD risk. We picked out SNPs within the genes of interest according to the UCSC database, then examine whether these SNPs were pQTLs according to the results from the KORA and INTERVAL pQTL studies described above.

MR analysis on proteins

In order to obtain further supporting evidence for proteins identified in pQTL analysis, we employed the inverse-variance weighted (IVW) MR[25], MR-Egger[26] and the MR pleiotropy residual sum and outlier (MR-PRESSO) [27] methods to test for potential causal relationships between circulating protein levels and LDL-C and CAD. The inverse-variance weighted method combines the ratio estimates from each IV in a meta-analysis model[25]. If the associations with circulating protein levels were to lead to horizontal pleiotropy, the intercept from MR-Egger would be expected to differ from zero[26]. In such cases, we interpreted the coefficient from MR-Egger as being the more valid causal estimate. Conversely, in the absence of statistical evidence for horizontal pleiotropy from the intercept on MR-Egger, we used the IVW MR analysis as it retains greater power. The IVW MR and MR-Egger analyses were performed by using the MendelianRandomization R package[28]. We also detected the horizontal pleiotropy and the outlier-corrected causal estimation by using the MR-PRESSO tests[27]. The outlier test in MR-PRESSO is the procedure to test for the MR assumption of no pleiotropy. The source code and documents for MR-PRESSO were available at <https://github.com/rondolab/MR-PRESSO>. The default parameters were used for the MR-PRESSO analysis.

Data used in these MR analyses were the pQTL data from the three studies and LDL-C (2013 GWAS)[4] and CAD[6] GWAS data that have been described above. The data required in the MR analysis (i.e., the SNP rs number, beta values, standard errors, and the P values) was extracted from each of the LDL-C and CAD GWAS and pQTL datasets. Then we used the “merge” function of the R program to transform the data into a specific file (an ordinary document with 7 columns) for each protein-trait pair. In the pQTL summary data, SNPs with P value less than 1.0×10^{-4} were selected to be potential instrumental variables, except for the NMR_GWAS pQTL study (5.0×10^{-8}). The selection criterion was set to 1.0×10^{-4} for the two pQTL studies because 5.0×10^{-8} would lead to too few instrumental variables. We clumped SNPs ($LD\ r^2 < 0.01$ within 10,000 kb) based on data of the Europeans from the 1000 Genomes project using the “clump_data” function on the R package of TwoSampleMR to select independent instrumental

variables. The effect allele of each SNP in the LDL-C and CAD GWAS and pQTL studies was manually checked for consistency.

Results

Gene expressions in liver cells associated with LDL-C and CAD

We carried out an SMR analysis to find gene expressions in liver cells that were associated with circulating LDL-C level and CAD risk by integrating eQTL data from the GTEx project with large-scale GWAS data. The mRNA expression levels of a total of 21,032 genes were analyzed, and expression levels of 15 genes (*RHD*, *RHCE*, *ANGPTL3*, *CELSR2*, *PSRC1*, *SORT1*, *SYPL2*, *ATXN7L2*, *DNAH11*, *FADS3*, *ST3GAL4*, *NYNRIN*, *CETP*, *EFCAB13* and *SPTLC3*) in the liver were significantly associated with LDL-C ($P < 5.0 \times 10^{-6}$) (Table 1, Supplemental Table S1). Ten of them pass the HEIDI test. These genes were enriched in gene ontology (GO) biological process of cellular lipid metabolic process ($P = 5.0 \times 10^{-3}$).

Table 1
Liver cell gene expressions associated with circulating LDL-C level

Gene	CHR	Probe ID	Beta	SE	p_SMR	p_HEIDI
RHD	1	ENSG00000187010.14	-0.0260	0.0053	7.68E-07	6.51E-04
RHCE	1	ENSG00000188672.12	0.0515	0.0113	5.08E-06	8.17E-02
ANGPTL3	1	ENSG00000132855.4	0.1423	0.0305	3.05E-06	1.79E-01
CELSR2	1	ENSG00000143126.7	-0.1714	0.0179	8.39E-22	6.12E-02
PSRC1	1	ENSG00000134222.12	-0.1380	0.0128	2.88E-27	1.48E-01
SORT1	1	ENSG00000134243.7	-0.1420	0.0120	3.80E-32	9.62E-03
SYPL2	1	ENSG00000143028.7	-0.0335	0.0054	4.59E-10	7.39E-05
ATXN7L2	1	ENSG00000162650.11	-0.0510	0.0094	6.00E-08	1.58E-03
DNAH11	7	ENSG00000105877.13	0.0396	0.0082	1.45E-06	2.33E-01
FADS3	11	ENSG00000221968.4	0.1049	0.0223	2.44E-06	2.35E-01
ST3GAL4	11	ENSG00000110080.14	0.0870	0.0180	1.39E-06	6.85E-02
NYNRIN	14	ENSG00000205978.5	0.0672	0.0141	1.82E-06	1.00E-01
CETP	16	ENSG00000087237.6	0.1815	0.0398	4.99E-06	NA
EFCAB13	17	ENSG00000178852.11	0.0448	0.0083	5.83E-08	8.05E-01
SPTLC3	20	ENSG00000172296.8	-0.0542	0.0107	3.75E-07	3.81E-01

CHR: Chromosome; SE: Standard error; SMR: summary data-based Mendelian randomization; HEIDI: Heterogeneity in dependent instruments.

The mRNA expression of 3 genes (*CELSR2*, *PSRC1* and *SORT1*) in the liver were significantly associated with CAD risk ($P < 5.0 \times 10^{-6}$) (Table 2, Supplemental Table S2). Therefore, the associations of mRNA expression levels of *SORT1*, *PSRC1* and *CELSR2* in liver with both LDL-C and CAD passed the threshold of 5.0×10^{-8} (Fig. 2). Higher expression levels of *SORT1*, *PSRC1* and *CELSR2* in liver cells were associated with lower LDL-C levels (beta= -0.142, -0.138 and - 0.171, respectively) and CAD risk (beta= -0.10, -0.097 and - 0.121, respectively).

Table 2
Liver cell gene expressions associated with CAD risk

Gene	CHR	Probe ID	Beta	SE	p_SMR	p_HEIDI
CELSR2	1	ENSG00000143126.7	-0.1209	0.0172	2.35E-12	8.48E-01
PSRC1	1	ENSG00000134222.12	-0.0973	0.0131	9.46E-14	3.68E-02
SORT1	1	ENSG00000134243.7	-0.1001	0.0129	9.32E-15	3.32E-02

CHR: Chromosome; SE: Standard error; SMR: summary data-based Mendelian randomization; HEIDI: Heterogeneity in dependent instruments.

Gene expressions in blood cells associated with LDL-C and CAD

By integrating LDL-C GWAS data with eQTL data, we found that expression levels of 31 genes were significantly associated with LDL-C level, including expression level of *PSRC1* (Fig. 2), without significant heterogeneity (Supplemental Table S3). We then performed SMR analysis for CAD. By integrating CAD GWAS data with data from eQTL studies, we found that mRNA levels of 17 probes of 11 genes (*PSRC1*, *IL6R*, *GGCX*, *VAMP8*, *LIPA*, *NT5C2*, *SWAP70*, *EIF2B2*, *FURIN*, *FES* and *ATP5G1*) were significantly associated with CAD (Table 3), including expression level of *PSRC1* (Fig. 2), without significant heterogeneity ($P_{HEIDI} > 0.05$). These genes were enriched in GO biological process of myeloid leukocyte mediated immunity ($P = 9.8 \times 10^{-4}$). We also found additional 88 genes that were nominally associated with CAD in blood cells (Supplemental Table S4).

Table 3
Blood cell gene expressions associated with CAD risk

Gene	CHR	Studies	Probe ID	Beta	SE	p_SMR	p_HEIDI
PSRC1	1	CAGE	ILMN_1671843	-0.2078	0.0242	8.53E-18	3.57E-01
PSRC1	1	CAGE	ILMN_2315964	-0.4152	0.0644	1.15E-10	6.57E-01
PSRC1	1	WhB	ENSG00000134222.12	-0.3423	0.0711	1.46E-06	1.68E-01
IL6R	1	Westra	ILMN_1696394	0.2204	0.0472	2.98E-06	4.09E-01
GGCX	2	WhB	ENSG00000115486.7	0.1279	0.0233	3.84E-08	6.91E-02
VAMP8	2	Westra	ILMN_2190084	-0.0763	0.0126	1.42E-09	5.90E-02
VAMP8	2	CAGE	ILMN_2190084	-0.0707	0.0117	1.68E-09	5.95E-02
LIPA	10	CAGE	ILMN_1718063	0.0657	0.0097	1.25E-11	9.51E-01
LIPA	10	WhB	ENSG00000107798.13	0.1386	0.0227	1.07E-09	9.09E-01
NT5C2	10	Westra	ILMN_1682165	-0.1418	0.0282	4.97E-07	1.84E-01
NT5C2	10	CAGE	ILMN_1682165	-0.1080	0.0219	8.14E-07	8.67E-01
SWAP70	11	Westra	ILMN_1785175	-0.1477	0.0287	2.74E-07	6.39E-01
SWAP70	11	CAGE	ILMN_1785175	-0.1058	0.0201	1.43E-07	7.81E-01
EIF2B2	14	CAGE	ILMN_1713380	0.0593	0.0129	3.94E-06	6.76E-01
FURIN	15	Westra	ILMN_1790228	-0.2324	0.0498	3.06E-06	2.40E-01
FES	15	Westra	ILMN_1693650	-0.1309	0.0247	1.11E-07	7.41E-02
ATP5G1	17	Westra	ILMN_1712430	0.1695	0.0363	3.03E-06	3.29E-01

CHR: Chromosome; SE: Standard error; SMR: summary data-based Mendelian randomization; HEIDI: Heterogeneity in dependent instruments.

pQTLs in the identified genes

We selected 45 genes of interest according to the shared genes. In UCSC database we found 7,290 SNPs in these genes. In data from the KORA and INTERVAL pQTL studies we found 1,238 pQTLs (Supplemental table S5). Many of these pQTLs were significantly associated with LDL-C or CAD. For example, the pQTL rs599839 in *PSRC1* was significantly associated with LDL-C ($P = 2.75 \times 10^{-268}$) and CAD ($P = 5.20 \times 10^{-17}$). This SNP was strongly associated with circulating levels of granulins in both of the two pQTL studies (beta = -0.7453 and 0.8071, $P = 3.32 \times 10^{-51}$ and 1.00×10^{-200} , respectively). Indeed, twenty SNPs in *PSRC1* were found to be associated with circulating level of granulins. Fourteen SNPs in *APOB* and five SNPs in *PSRC1* were strongly associated with circulating level of apolipoprotein

B. Among these SNPs, rs1367117 was associated with both LDL-C ($P= 1.45 \times 10^{-282}$) and CAD ($P= 1.10 \times 10^{-4}$); rs693 ($P= 3.16 \times 10^{-5}$) and rs10199768 ($P= 3.10 \times 10^{-3}$) were associated with LDL-C; rs2337383 was associated with CAD ($P= 3.70 \times 10^{-4}$). A total of 81 SNPs in *ERGIC3* were significantly associated with circulating level of copine 1, and 19 and 5 of them were associated with LDL-C and CAD, respectively. In addition, we found that the associations of some SNPs with circulating levels of granulins, apolipoprotein B and copine 1 were replicated in the two pQTL studies.

Proteins causally associated with CAD

We tested that if the three proteins, i.e., copine 1, granulins and apolipoprotein B, were genetically associated with CAD using several MR methods. The results were presented in Table 4. We first examined the associations between circulating levels of these proteins and LDL-C and found that the associations between circulating levels of granulins and apolipoprotein B and LDL-C were significant in every MR analyses using data from KORA and INTERVAL studies. In datasets available from the NMR_GWAS, we only found data for apolipoprotein B among the tested metabolites. By using this data, the association between circulating level of apolipoprotein B and LDL-C was validated. Circulating copine 1 was found to be significantly associated with LDL-C by using KORA data, but not in MR-PRESSO analysis using the INTERVAL data.

Table 4
The causation between circulating protein levels and LDL-C and CAD

Protein levels		Trait/	IVW/MR-Egger			MR-PRESSO		
Study	Protein	Disease	Beta	SE	P value	Beta	SE	P value
1. KORA	Copine 1	LDL-C	0.022	0.005	1.08E-05	0.014	0.003	2.60E-03
2. INTERVAL	Copine 1	LDL-C	0.038	0.012	1.54E-03	0.015	0.011	0.213
1. KORA	Granulins	LDL-C	0.262	0.019	2.95E-43	0.123	0.027	4.49E-04
2. INTERVAL	Granulins	LDL-C	0.217	0.020	1.99E-27	0.127	0.022	3.87E-05
1. KORA	Apolipoprotein B	LDL-C	0.526	0.202	9.22E-03	0.261	0.113	4.37E-02
2. INTERVAL	Apolipoprotein B	LDL-C	1.401	0.229	4.89E-10	0.676	0.123	3.89E-08
3. NMR_GWAS	Apolipoprotein B	LDL-C	1.348	0.163	1.34E-16	1.023	0.049	1.40E-09
1. KORA	Copine 1	CAD	0.004	0.004	0.359	0.004	0.004	0.360
2. INTERVAL	Copine 1	CAD	0.013	0.015	0.378	0.002	0.008	0.821
1. KORA	Granulins	CAD	0.122	0.028	1.32E-05	0.052	0.015	1.57E-03
2. INTERVAL	Granulins	CAD	0.138	0.020	5.20E-12	0.079	0.016	4.60E-05
1. KORA	Apolipoprotein B	CAD	0.010	0.018	0.574	0.010	0.018	0.604
2. INTERVAL	Apolipoprotein B	CAD	0.289	0.105	5.92E-03	0.225	0.089	0.052
3. NMR_GWAS	Apolipoprotein B	CAD	0.373	0.093	6.05E-05	0.339	0.034	5.36E-09

CAD, Coronary artery disease; IVW, inverse-variance weighted; LDL-C, Low-density lipoprotein cholesterol; MR, Mendelian randomization; MR-PRESSO, MR pleiotropy residual sum and outlier; SE: Standard error.

Because circulating levels of granulins and apolipoprotein B were significantly associated with LDL-C, we thought that these proteins may be risk factors of CAD. As expected, significant associations between granulins and apolipoprotein B levels and CAD were found (Table 4). These associations were validated

in several MR analyses on data from two pQTL studies. The effect of protein granulins levels on CAD risk was relatively small, with an odds ratio (OR) of about 1.15 (95% confidence interval: 1.10–1.19). The effect of protein apolipoprotein B levels on CAD risk was larger than that of granulins, with an OR of about 1.45 (95% confidence interval: 1.21–1.74). As it is known, LDL-C level was causally associated with CAD. Therefore, granulins, apolipoprotein B, LDL-C and CAD were genetically correlated with each other.

Discussion

In this study we took the advantage of MR approaches to find out potential causal factors (e.g., gene expressions in liver and blood cells and circulating protein levels) for CAD by integrating data from GWAS. We found many genes that may play causal roles in LDL-C metabolism and CAD. We also found that circulating levels of granulins and apolipoprotein B were causally associated with LDL-C and CAD.

Until now, GWAS have successfully identified over 175 genetic loci for lipids. Some of these loci involved therapeutic targets such as *HMGCR* (statins), *PCSK9* (antibodies), and *NPC1L1* (ezetimibe). The causal effect of LDL-C on CAD has been well established. Identification of related factors for LDL-C may help to understand the etiology of CAD and develop novel therapies. Elucidation of the causal factors underlying GWAS signals remains challenging due to the complexities of the genomic loci (e.g., LD) and the interactions. It is difficult to determine the most functionally relevant genes for LDL-C and CAD only based on the genomic data. On the other hand, previous case-control studies have found gene expressions in blood cells implicated in dyslipidemia and CAD, however, these observational studies are subject to confounding and reverse causations. And there was not liver gene expression profile study with large samples. MR is an analytical technique that assess the correlation between genetic alternatives to phenotypes and outcomes based on the random distribution of genetic variation specific to biomarkers, and thus makes up the shortages of traditional epidemiology study [29, 30]. By applying the two-sample MR approach we identified the most relevant genes in the complex genomic loci for LDL-C and CAD. The findings of this study also indicated that the GWAS-identified loci contained causal factors for LDL-C and CAD, e.g., expressions of the genes in the loci.

The expression level of *PSRC1* (Proline and Serine Rich Coiled-Coil 1) was associated with LDL-C and CAD in liver and blood cells. *PSRC1* codes a proline-rich protein, which is a target for regulation by the tumor suppressor protein p53. The genetic associations between *PSRC1* variants and LDL-C and CAD have been well established but the role of this gene in LDL-C and CAD has not been well discussed. Our analysis suggested that SNPs in *PSRC1* were strongly associated with gene expression, LDL-C and CAD, and furthermore, these SNPs were significantly associated with circulating levels of granulins and apolipoprotein B. Moreover, we demonstrated that circulating levels of these proteins were causally associated with LDL-C and CAD. Circulating progranulin is a dimer of high-density lipoprotein (HDL) [31]. HDL/apolipoprotein A-I was known to bind to macrophage-derived progranulin and suppress its conversion into proinflammatory granulins [32]. However, how *PSRC1* interacts with *GRN* is unclear but we found that these genes were connected via other genes that were related to cardiovascular diseases (e.g., *APP*, *YY1*) (Supplemental Fig. 1). Based on the multiple sources of evidence, the identified genes

were suggested to be potential functional candidate genes. The interaction among these factors probably point to a pathway for LDL-C and CAD.

Our study found that apolipoprotein B was causally associated with LDL-C and CAD. It is known that one molecule of apolipoprotein B is required for the proper assembly and secretion of each very low density lipoproteins [33], which are converted into LDL-C after the hydrolysis of their triglyceride moiety [34]. Indeed, apolipoprotein B is a proposed LDL-C-lowering target. Proprotein convertase subtilisin/kexin type 9 (*PCSK9*) regulates circulating LDL-C level. Several studies have shown that PCSK9 regulates apolipoprotein B synthesis and secretion[35–37]. In animals, a single dose of siRNA targeting *PCSK9* resulted in a lowering of circulating PCSK9, apolipoprotein B, and LDL-C, without measurable effects on either high density lipoprotein cholesterol or triglycerides [38]. In the MR analyses in human samples we showed that there was a causal effect of circulating apolipoprotein B on CAD.

This study has some limitations. First, we did not validate the associations found in the SMR analysis in an independent sample due to lacking in data. Second, the identified “candidates” of the causal factors were inferred by statistical models. Additional studies such as experimental studies and randomized intervention trials were needed to strengthen the evidence. Third, although we have identified a set of causal genes, the connections between these genes were unclear. How the lipid metabolism related genes connect to each other and how they affect CAD risk is needed to test in future studies.

Conclusion

This study elucidates the biological basis of genetic linkage and promotes the identification of susceptibility genes. The findings suggested that gene expressions in liver and blood cells and circulating protein levels were genetically associated with LDL-C and CAD and may be novel risk factors for CAD. This study also provided evidence for exploring the potential causal relationship between granulins and apolipoprotein B and LDL-C and CAD. The findings may provide the clues for seeking new therapeutic target for dyslipidemia and CAD.

Declarations

Acknowledgments

None.

Authors' contributions

TC and LC designed this study; TC, ZW, XY, and ZQ collected the data; TC and ZW were responsible for the statistical analysis; TC wrote the draft; ZW, ZW, XY, and LC revised this draft; LC finalized this manuscript. The authors read and approved the final manuscript.

Funding

The study was supported by the National Natural Science Foundation of China [U2005202, 81670438 and 81700418], the Fujian Province Major Science and Technology Program [2018YZ001-1], the Natural Science Foundation of Fujian Province (2020J02056), the Fujian provincial health technology project (2019-ZQN-50) and Natural Science Foundation of Fujian Province [2020J01998].

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Fujian Medical University Union Hospital. Written informed consent was waived by the Ethics Commission of each hospital for emerging infectious.

Consent for publication

No individual participant data is reported that would require consent to publish from the participant (or legal parent or guardian for children).

Competing interests

The authors declare that they have no competing interests.

References

1. GBD_2017_Causes_of_Death_Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1736-88.
2. Lu X, Peloso GM, Liu DJ, Wu Y, Zhang H, Zhou W, et al. Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nat Genet*. 2017;49:1722-30.
3. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-13.
4. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274-83.
5. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25-33.
6. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121-30.

7. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet.* 2012;90:7-24.
8. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet.* 2021;22:19-37.
9. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48:245-52.
10. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015;47:1091-8.
11. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48:481-7.
12. De Ita JR, Castilla-Cortázar I, Aguirre GA, Sánchez-Yago C, Santos-Ruiz MO, Guerra-Menéndez L, et al. Altered liver expression of genes involved in lipid and glucose metabolism in mice with partial IGF-1 deficiency: an experimental approach to metabolic syndrome. *J Transl Med.* 2015;13:326.
13. Rong S, Cortés VA, Rashid S, Anderson NN, McDonald JG, Liang G, et al. Expression of SREBP-1c Requires SREBP-2-mediated Generation of a Sterol Ligand for LXR in Livers of Mice. *Elife.* 2017;6.
14. Horton JD, Goldstein JL, Brown MS. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J Clin Invest.* 2002;109:1125-31.
15. Berger JM, Moon YA. Increased Hepatic Lipogenesis Elevates Liver Cholesterol Content. *Mol Cells.* 2021;44:116-25.
16. Li C, Hu Z, Zhang W, Yu J, Yang Y, Xu Z, et al. Regulation of Cholesterol Homeostasis by a Novel Long Non-coding RNA LASER. *Sci Rep.* 2019;9:7693.
17. Perucha E, Melchiotti R, Bibby JA, Wu W, Frederiksen KS, Roberts CA, et al. The cholesterol biosynthesis pathway regulates IL-10 expression in human Th1 cells. *Nat Commun.* 2019;10:498.
18. Ma X, Bi E, Huang C, Lu Y, Xue G, Guo X, et al. Cholesterol negatively regulates IL-9-producing CD8(+) T cell differentiation and antitumor activity. *J Exp Med.* 2018;215:1555-69.
19. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013;45:1238-43.
20. Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, et al. The Genetic Architecture of Gene Expression in Peripheral Blood. *Am J Hum Genet.* 2017;100:228-37.
21. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature.* 2017;550:204-13.
22. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun.* 2017;8:14357.
23. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature.* 2018;558:73-9.

24. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun*. 2016;7:11122.
25. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37:658-65.
26. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44:512-25.
27. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50:693-8.
28. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol*. 2017;46:1734-9.
29. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23:R89-98.
30. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32:1-22.
31. Nguyen AD, Nguyen TA, Cenik B, Yu G, Herz J, Walther TC, et al. Secreted progranulin is a homodimer and is not a component of high density lipoproteins (HDL). *J Biol Chem*. 2013;288:8627-35.
32. Okura H, Yamashita S, Ohama T, Saga A, Yamamoto-Kakuta A, Hamada Y, et al. HDL/apolipoprotein A-I binds to macrophage-derived progranulin and suppresses its conversion into proinflammatory granulins. *J Atheroscler Thromb*. 2010;17:568-77.
33. Elovson J, Chatterton JE, Bell GT, Schumaker VN, Reuben MA, Puppione DL, et al. Plasma very low density lipoproteins contain a single molecule of apolipoprotein B. *J Lipid Res*. 1988;29:1461-73.
34. Kissebah AH, Alfarsi S, Adams PW. Integrated regulation of very low density lipoprotein triglyceride and apolipoprotein-B kinetics in man: normolipemic subjects, familial hypertriglyceridemia and familial combined hyperlipidemia. *Metabolism*. 1981;30:856-68.
35. Sun H, Samarghandi A, Zhang N, Yao Z, Xiong M, Teng BB. Proprotein convertase subtilisin/kexin type 9 interacts with apolipoprotein B and prevents its intracellular degradation, irrespective of the low-density lipoprotein receptor. *Arterioscler Thromb Vasc Biol*. 2012;32:1585-95.
36. Tavori H, Fan D, Blakemore JL, Yancey PG, Ding L, Linton MF, et al. Serum proprotein convertase subtilisin/kexin type 9 and cell surface low-density lipoprotein receptor: evidence for a reciprocal regulation. *Circulation*. 2013;127:2403-13.
37. Levy E, Ben Djoudi Ouadda A, Spahis S, Sane AT, Garofalo C, Grenier E, et al. PCSK9 plays a significant role in cholesterol homeostasis and lipid transport in intestinal epithelial cells. *Atherosclerosis*. 2013;227:297-306.
38. Frank-Kamenetsky M, Grefhorst A, Anderson NN, Racie TS, Bramlage B, Akinc A, et al. Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proc Natl Acad Sci U S A* 2008;105:11915–20.

Figures

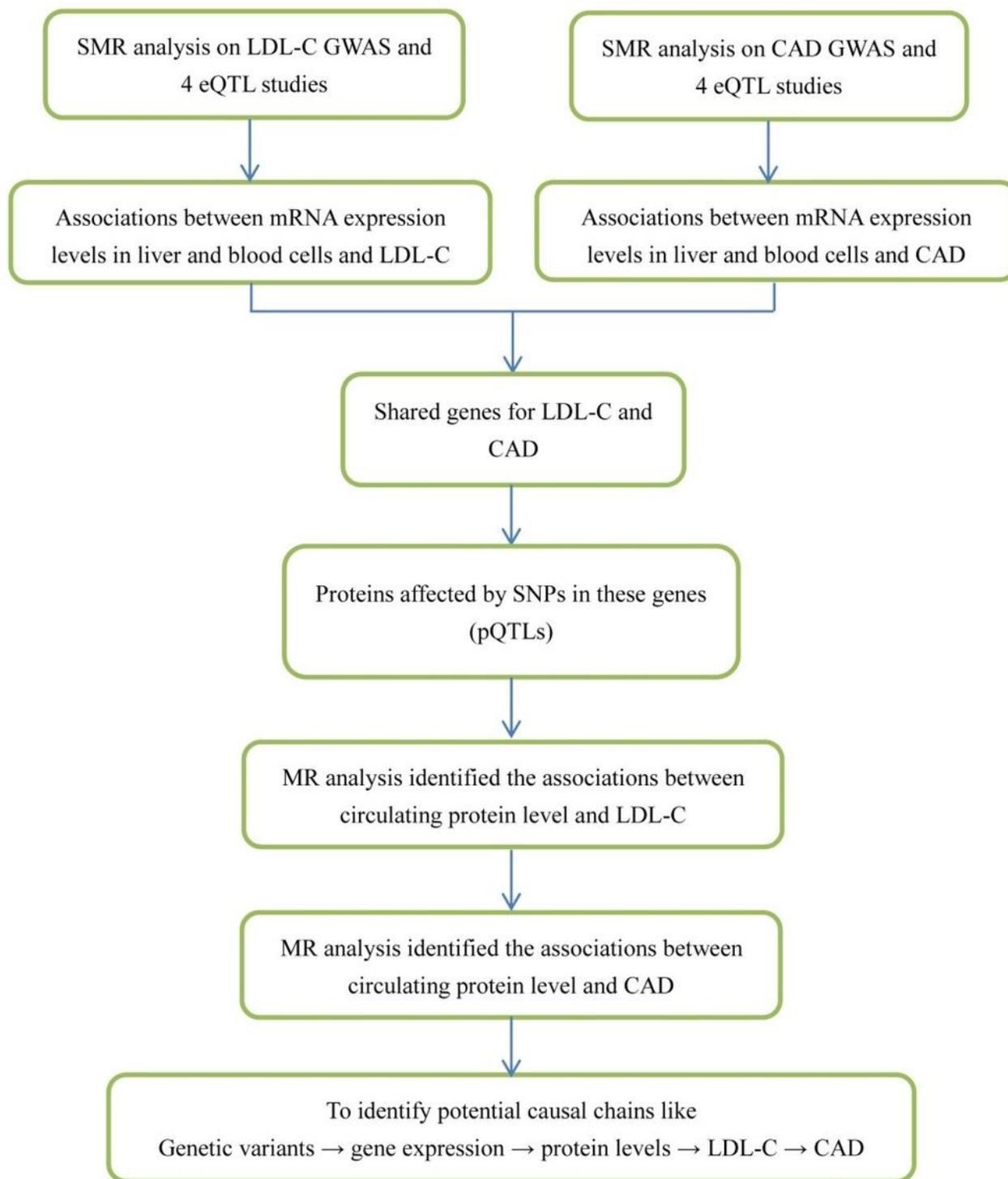


Figure 1

The flowchart of the study design. We designed this study to explore potential risk factors such as SNPs, gene expression in liver and blood cells, and circulating protein levels for CAD. This study comprised multiple steps of analysis. The first step is the SMR analyses. In this step we integrative

analyzed data from 4 eQTL studies (1 in liver and 3 in whole blood cells) and data from 2 LDL-C and 1 CAD GWAS to identify mRNAs that were associated with LDL-C and CAD. Second, we found out the shared genes for LDL-C and CAD. Third, we looked for pQTLs inside the genes identified in the previous steps using data from 3 pQTL studies on circulating protein levels. Finally, for circulating protein levels that were significantly affected by SNPs (i.e., pQTLs) in the identified genes, we applied 3 MR methods (IVW, MR-Egger, MR-PRESSO) to examine whether they were genetically associated with LDL-C and CAD by integrating data from the 3 pQTL studies and LDL-C and CAD GWAS.

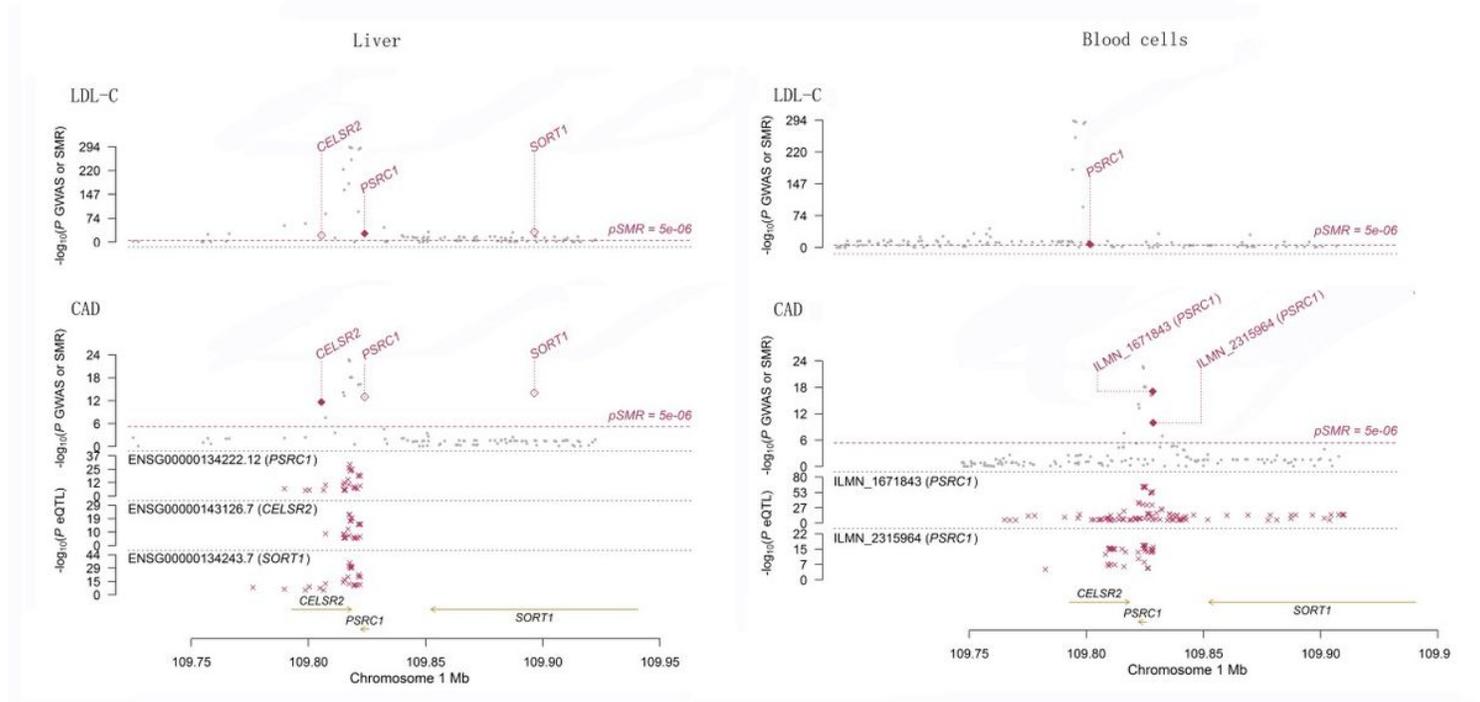


Figure 2

The association between mRNA expression levels of 1p13.3 genes and LDL-C and CAD in liver and blood cells. The two panels present the results of the associations between mRNA expression levels of 1p13.3 genes and LDL-C and CAD in liver (left) and blood cells (right). Each panel consists of two parts. The x-axis represents the genomic position (GRCh37.p13). The lower part of each panel shows the results of eQTL. The y-axis represents $-\log_{10}(P \text{ eQTL})$. In this part we can see that SNPs in 1p13.3 were strongly associated with mRNA expression levels of SORT1, PSRC1 or CELSR2. The upper part of each panel shows the results of GWAS (gray dots) and SMR (red rhombus) analysis. The y-axis represents $-\log_{10}(P \text{ GWAS or SMR})$. In this part we can see that SNPs in 1p13.3 were strongly associated with LDL-C and CAD. According to the SMR analysis, mRNA expression levels of SORT1, PSRC1 and CELSR2 in liver cells were significantly associated with LDL-C and CAD, and mRNA expression level of PSRC1 in blood cells was significantly associated with LDL-C and CAD.

Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [Supplementaryfigures.pdf](#)
- [Supplementarytables.xlsx](#)