Research article

Identification of novel proteins for coronary artery disease by integrating GWAS data and human plasma proteomes

Jiqing Li [a], Jiate Wei [b], Ping Fu [c], Jianhua Gu [a,*]

[a] *Department of Emergency Medicine, Qilu Hospital of Shandong University, Shandong University, Jinan, 250012, Shandong, China*
[b] *Office of Hospital Management Research, Beijing Friendship Hospital, Capital Medical University, Beijing, 100050, China*
[c] *Jinan Center for Disease Control and Prevention, Jinan, 250012, Shandong, China*

ARTICLE INFO

ABSTRACT

Background: Most coronary artery disease (CAD) risk loci identified by genome-wide association studies (GWAS) are located in non-coding regions, hampering the interpretation of how they confer CAD risk. It is essential to integrate GWAS with molecular traits data to further explore the genetic basis of CAD.

Methods: We used the probabilistic Mendelian randomization (PMR) method to identify potential proteins involved in CAD by integrating CAD GWAS data (~76,014 cases and ~264,785 controls) and human plasma proteomes (N = 35,559). Then, Bayesian co-localization analysis, confirmatory PMR analysis using independent plasma proteome data (N = 7752), and gene expression data (N1 = 213, N2 = 670) were performed to validate candidate proteins. We further investigated the associations between candidate proteins and CAD-related traits and explored the rationality and biological functions of candidate proteins through disease enrichment, cell type-specific, GO, and KEGG enrichment analysis.

Results: This study inferred that the abundance of 30 proteins in the plasma was causally associated with CAD ($P < 0.05/4408$, Bonferroni correction), such as PLG, IL15RA, and CSNK2A1. PLG, PSCK9, COLEC11, ZNF180, ERP29, TCP1, FN1, CDH5, IL15RA, MGAT4B, TNFRSF6B, DNM2, and TGF1R were replicated in the confirmatory PMR ($P < 0.05$). PCSK9 (PP.H4 = 0.99), APOB (PP.H4 = 0.89), FN1 (PP.H4 = 0.87), and APOC1 (PP.H4 = 0.78) coding proteins shared one common variant with CAD. MTAP, TCP1, APOC2, ERP29, MORF4L1, C19orf80, PCSK9, APOC1, EPOR, DNM2, TNFRSF6B, CDKN2B, and LDLR were supported by PMR at the transcriptome level in whole blood and/or coronary arteries ($P < 0.05$). Enrichment analysis identified multiple pathways involved in cholesterol metabolism, regulation of lipoprotein levels and telomerase, such as cholesterol metabolism (hsa04979, $P = 2.25E-7$), plasma lipoprotein particle clearance (GO:0034381, $P = 5.47E-5$), and regulation of telomerase activity (GO:0051972, $P = 2.34E-3$).

Conclusions: Our integration analysis has identified 30 candidate proteins for CAD, which may provide important leads to design future functional studies and potential drug targets for CAD.

---

* Corresponding author. Department of Emergency Medicine, Qilu Hospital of Shandong University, 44 Culture West Road, Jinan, 250012, China.
  E-mail address: jianhuagu@yeah.net (J. Gu).

1. Introduction

Coronary artery disease (CAD) stands as a major cardiovascular disease, ranking the top causes of mortality in both developed and developing countries [1]. An individual's risk of developing CAD is influenced by a combination of genetic factors, environment, and lifestyle [2]. Genetic factors are estimated to account for 30 %–60 % of the variation in coronary artery disease risk among individuals [3–8]. Over the past decade, several genome-wide association studies (GWAS) have identified more than 200 risk loci associated with CAD [3–8]. However, the majority of these variants are located outside protein-coding regions. Furthermore, most associations are typically attributed to specific genes primarily due to their proximity to the variant site, which makes it challenging to decipher the biological processes underlying these genetic effects [3]. The significant enrichment of variants in regulatory regions underscores the primary role of CAD risk variants in modulating gene expression [9].Therefore, it is essential to integrate GWAS with molecular traits data to further explore the genetic basis of CAD.

Li et al. have integrated transcriptome data and GWAS to conduct transcription-wide association studies (TWAS) to explore susceptibility genes for CAD [10]. Nevertheless, proteins are the main functional components of biological processes, as well as the final products of transcripts. They often undergo intricate post-translational alterations and refinement, and contain additional information [11,12]. Numerous genetic studies have identified protein quantitative trait loci (pQTLs) in plasma, and demonstrated that pQTLs strengthen the associations found in GWAS across various complex traits [13–16]. Several studies have used pQTLs as instrumental variables to perform Mendelian randomization to identify causative proteins of different phenotypes [17–19]. Currently, few studies have integrated large-scale proteome data with GWAS to identify susceptibility proteins for CAD.

The majority of exposure/molecular traits are influenced by multiple SNPs that may have potential linkage disequilibrium (LD) with each other. Previous studies mainly used Mendelian randomization (MR) or FUSION, PrediXcan, and other TWAS methods for transcription/protein-wide association analysis [10,17–19]. Traditional MR methods are limited to using a single SNP or multiple independent SNPs as instruments. By incorporating multiple correlated SNPs, TWAS methods such as FUSION and PrediXcan improve statistical performance. Unfortunately, many TWAS methods rely on two-stage MR Inference procedures that fail to take into account
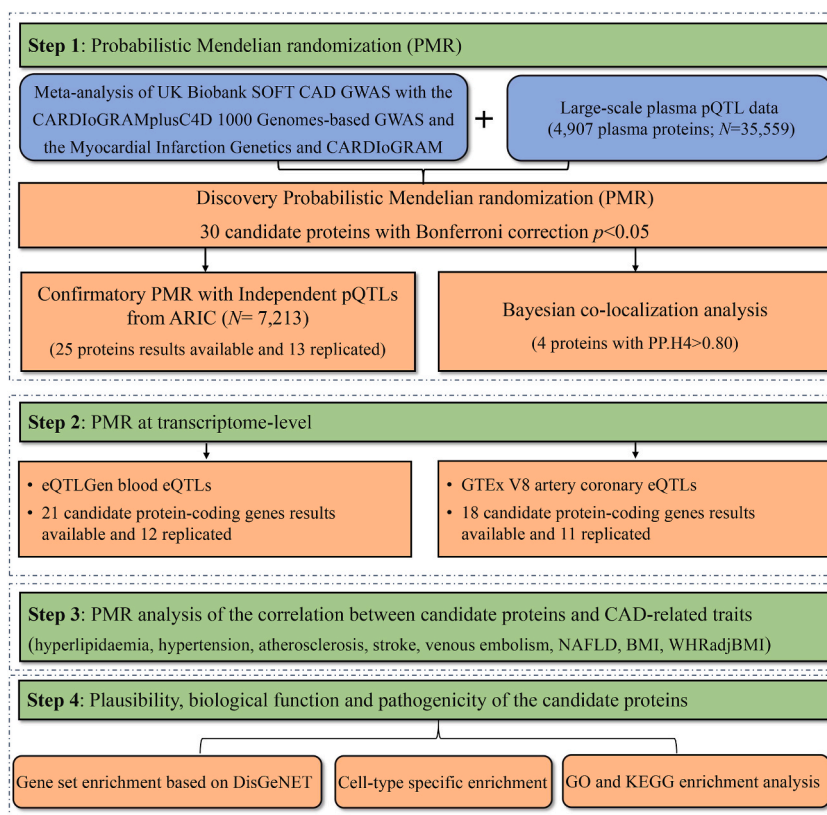
Fig. 1. The overall analysis pipeline for this study. Step 1, we performed discovery PMR to integrate GWAS of CAD and a large-scale plasma pQTL data to identify candidate genes involved in CAD (Bonferroni correction P < 0.05/N); then, we performed confirmatory PMR using an independent pQTL data from ARIC and colocalization analysis; Step 2, we validated the candidate genes at the transcriptome-level; Step 3, we further investigated the associations between candidate genes and CAD related traits; Step 4, Disease enrichment, Cell-type specific enrichment, GO and KEGG enrichment analysis. CAD, coronary artery disease; PMR, probabilistic two-sample Mendelian randomization; pQTL, protein quantitative trait locus; TWAS, transcriptome-wide association studies; eQTL, expression quantitative trait locus; GTEx, Genotype-tissue expression; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

the uncertainty of parameter estimates in exposure studies, which may lead to bias in estimates of causal effects and power loss [20–22]. Therefore, our study adopts a new multi-omics integration analysis method, probabilistic Mendelian randomization (PMR) [23], which incorporates multiple correlated SNPs in the likelihood inference framework to infer the relationship between molecular phenotypes and traits, which greatly improves the statistical power.

In this study, we employed an integrative analytical pipeline to identify proteins that were associated with CAD (Fig. 1). First, we performed a discovery PMR to integrate GWAS of CAD and a large-scale plasma pQTL data to identify susceptibility proteins involved in CAD, protein-coding genes with P value less than 0.05/N after Bonferroni correction were regarded as candidate proteins; Second, we performed confirmatory PMR using independent pQTL data from ARIC and Bayesian co-localization analysis; Third, we explored the association of transcriptome-levels of candidate protein-coding genes with CAD using expression quantitative trait loci (eQTL) data from whole blood and coronary arteries; Fourth, we used PMR to further investigate the associations between candidate proteins and CAD-related traits, such as hyperlipidemia, hypertension, and atherosclerosis; Fifth, Disease enrichment, Cell type specific enrichment, GO and KEGG enrichment analysis were conducted to explore the plausibility, biological function and pathogenicity of the candidate proteins.

2. Materials and method

2.1. CAD GWAS data

We used the summary statistics from a meta-analysis of UK Biobank SOFT CAD GWAS with CARDIoGRAMplusC4D 1000 Genomes-based GWAS and the Myocardial Infarction Genetics and CARDIoGRAM Exome [24], which included ~76,014 cases and ~264785 controls. CAD was defined based on self-reported information (via questionnaire), hospital records, or death registry data, encompassing both fatal and nonfatal myocardial infarction, percutaneous transluminal coronary angioplasty (PTCA), coronary artery bypass grafting (CABG), chronic ischemic heart disease (IHD), and angina. Individuals who did not fulfill the criteria for CAD were designated as controls, after excluding individuals with aneurysms and atherosclerotic cardiovascular disease. Two distinct fixed-effect meta-analyses were performed to integrate the association summary statistics from the UKB with the 1000 Genomes-imputed GWAS outcomes and Exome results [25]. Details information regarding this data resource is available in the original papers [2,24].

2.2. Human plasma pQTLs in the discovery PMR

The human plasma samples were collected from 40,004 Icelanders [26], and were measured with the SomaScan version 4 assay (SomaLogic), which contains 5284 aptamers to recognize specific proteins. After converting protein concentrations into DNA aptamer concentrations, the proteins are quantified using a DNA microarray. After quality control, a total of 4907 aptamers that target a total of 4719 unique proteins remained for the pQTL analysis. A total of 166,281 Icelanders were genotyped using Illumina SNP arrays, resulting in >27 million imputed variants with MAF>0.01 and imputation info score >0.9 after imputation. Of these, 35,559 Icelanders had both genotype and protein measurements. The pQTLs were generated by genome-wide association testing using the linear mixed model implemented in BOLT-LMM [27]. The phenotypes were the standardized residuals after adjusting for rank-inverse normal transformed for age, sex, and sample age. All P values were computed by likelihood-ratio test. More detailed information was provided in the original papers [14].

2.3. Human plasma pQTLs in the confirmatory PMR

The human plasma proteomes used in the confirmatory PMR analysis were obtained from 7213 European American individuals in the Atherosclerosis Risk in Communities (ARIC) study [28]. The relative concentrations of plasma proteins or protein complexes from the blood samples were measured utilizing the V4 platform by SomaLogic Inc. (Boulder, Colorado, USA), employing an aptamer (SOMAmer)-based methodology [29,30]. After quality control, the pQTL analysis included 4657 SOMAmers that measured 4483 unique proteins or protein complexes encoded by 4435 autosomal genes. The Affymetrix 6.0 DNA microarray was used to genotype the ARIC samples, and the TOPMed reference panel was used to perform imputation [31,32]. The SNPs with imputation quality $R^2 \geq 0.8$, call rates ≥ 90 %, Hardy-Weinberg equilibrium p-values $\geq 10^{-6}$, or MAF ≥ 0.01 were included in the pQTL analysis. The association tests were performed by an adaptive permutation approach implemented in QTLtools [33]. More detailed information was provided in the original papers [28].

2.4. Human eQTL data in the PMR

In this study, we used the eQTL summary data from the coronary artery (N = 213) and whole blood (N = 31684) for PMR analysis to explore the association of candidate protein-coding genes expression levels with CAD risk in these tissues. The eQTL summary data of coronary artery was obtained from Genotype-Tissue Expression version 8 (GTEx v8) [34], which examined 15,201 RNA-sequencing samples from 49 tissues of 838 postmortem donors and genotyped 43,066,422 SNPs using whole-genome sequencing (WGS). The mRNA of each of the tissue samples was quantified using the Quant-iTTM RiboGreen®RNA Assay Kit and sequenced to a median depth of 82.6 million reads. STAR v2.5.3a was used to align the RNA-seq data to the human reference genome GRCh38/hg38 [35]. We used two whole blood eQTL summary data, one from GTEx v8 (N = 670) and the other from a large-scale meta-analysis of 31,684 blood samples from 37 eQTLGen Consortium cohorts [36].

2.5. GWAS summary statistics of CAD-related traits

In this study, we investigated the relationship between candidate proteins with CAD-related traits including hypertension, hyperlipidemia, atherosclerosis, stroke, venous thrombosis, Non-alcoholic fatty liver disease (NAFLD), Body Mass Index (BMI), and waist-to-hip ratio (WHR) adjusted for BMI (WHRadjBMI). Among them, the GWAS summary data of hypertension, hyperlipidemia, atherosclerosis, and NAFLD were from the UK Biobank (UKB) data [2,37,38], the stroke GWAS summary data was from the International Stroke Genetics Consortium [39]; the GWAS summary data of venous thrombosis was obtained from a genome-wide meta-analysis including six datasets (CHB-CVDC/DBDS, UKB, FinnGen freeze 6, deCODE, Intermountain Healthcare and MVP) [40]. The GWAS summary data of BMI and WhRadjBMI were both from a meta-analysis of UKB results with publicly available GWAS summary statistics from GIANT [41].

2.6. Probabilistic mendelian randomization

We used a probabilistic Mendelian randomization (PMR) method for proteome-wide association studies (PWAS) applications [23]. PMR-Egger unifies many existing TWAS and MR methods in a MR likelihood framework to test the causal effect of proteins on traits in the presence of horizontal pleiotropy. The approach utilizes multiple related SNPs as instruments, facilitating inference within the likelihood framework and thereby significantly enhancing statistical power. In this study, we extracted cis-acting SNPs that are within 500 kb upstream of the transcription start site and 500 kb downstream of the transcription end site, and used the 1000G LD reference panel in Europeans. To ensure the existence of a solution, we limit the low IV correlation (r2 < 0.1 in 1000G) due to the possibility of decomposition failure caused by the relatively small sample size of the reference panel. In the discovery PMR analysis, P-values were corrected by the Bonferroni method, and p-value <0.05 was defined as statistically significant in the confirmatory PMR. The PMR was conducted by using the R package "PMR" (https://www.xzlab.org./software.html) with the default parameters recommended by the developers.

2.7. Bayesian co-localization analysis

In this study, we used Bayesian co-localization to assess the probability of the same SNP being responsible for both changing the CAD risk and modulating the protein levels [42]. To begin with, it is necessary to set the prior probability of a certain SNP being associated with CAD (P1), the probability of it being a significant pQTL (P2), and the probability of a QTL and GWAS association overlap (P12). We used the default COLOC priors of $P1 = 10^{-4}$, $P2 = 10^{-4}$, and $P12 = 10^{-5}$. There are a total of 5 hypotheses (Supplementary text 1) in Bayesian co-localization analysis, under the assumption that there is at most one causal variant between two traits. Among these five hypotheses, we are mainly interested in the fifth hypothesis, that a shared SNP is both associated with GWAS and pQTL. In addition, we also performed a Sum of Single Effects (SuSiE) regression that allows evidence for association at multiple causal variants to be evaluated simultaneously [43]. As recommended by the developer of this method, posterior probabilities for hypotheses 4 (PP.H4) > 0.80 were regarded as evidence of co-localization. The Bayesian colocalization analysis was conducted by using the R package "coloc" (https://github.com/clagiamba/moloc).

2.8. Cell-type expression specificity

We examined the cell-type-specific expression of candidate protein-coding genes utilizing single-cell RNA sequencing (RNA-seq) data derived from atherosclerotic human coronary arteries (GSE131778) [44] and whole blood samples (GSE149938)) [44]. The GSE131778 dataset comprises atherosclerotic coronary artery cells harvested from the excised hearts of four patients undergoing heart transplantation. The proximal-to-mid segment of the right coronary artery was used for RNA-seq. The cells were then loaded onto a microfluidics chip from 10X Genomics and encapsulated within gel beads containing barcoded oligo-dT. Subsequently, the cells were isolated using FACS and captured on the 10X Chromium controller. The GSE149938 dataset offers transcriptome profiles for 7551 human blood cells, encompassing 32 distinct immuno-phenotypic cell types. These cell types encompass hematopoietic stem cells, progenitors, and mature blood cells sourced from 21 healthy donors. The Cell-type Expression Specificity (ES) profiles were generated using the CELLEX tool, which adopts a "wisdom of the crowd" approach. By integrating multiple ES metrics, it produces a comprehensive set of complementary cell-type ES profiles [45].

2.9. Gene set enrichment analysis

We conducted gene ontology (GO) and pathway enrichment analysis on candidate protein-coding genes identified through PMR analysis. The clusterProfiler package was used to identify the GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched by the candidateprotein-coding genes [46]. We also performed disease enrichment analysis based on a database of gene-disease associations (DisGeNET), one of the largest publicly available collections of genes and variants associated with human diseases, integrating data from expert-curated repositories, GWAS catalogs, animal models, and the scientific literature [47]. q-value <0.05 was defined as a criterion for significant enrichment.

2.10. Ethics

All summarized statistics used in this study were generated by previous studies and were public availavle, for which ethical approval and individual consent were obtained for all original studies.

2.11. Role of the funding source

The funding sources had no role in the study design, data collection, analysis, interpretation, and writing of the manuscript.

3. Results

A total of 499 proteins were excluded from PMR analysis due to various reasons such as unsuitable instrument variables, low heritability expression (less than 0.01 %), or inability to decompose the LD matrix, and the remaining 4408 proteins underwent PMR analysis (Fig. 2). PMR identified 30 candidate proteins that were associated with CAD ($P < 0.05/4408$, Table 1). Among them, 14 proteins were positively related to CAD, such as PCSK9, APOB, and LDLR, and 16 proteins were negatively related to CAD, such as PLG, APOC1, and CDKN2B. We obtained the lowest P values for the SNPs within 1 Mb of each of these identified 30 proteins using the summary statistics from the CAD GWAS. Among these, 19 proteins with the most significant P values were less than 5×10^{-8} and the other 11 proteins ranged from 4.16×10^{-4} to 5.77×10^{-8}, including MGAT4B, CSNK2A1, OPCML, TNFRSF6B, COLEC11, SIGLEC7, CDH5, IGF1R, IL15RA, SORCS2, and IL5RA (Supplementary Table S1). We further performed a confirmatory PMR analysis with another independent human plasma pQTL summary data from the ARIC study. C19orf80, CDKN2B, MATP, and LDLR could not be tested due to low heritability (<0.01 %) or the LD matrix failed to decompose, and EPOR was not profiled in the pQTL data. Finally, 25 candidate proteins were tested and 19 proteins of them were associated with CAD ($P < 0.05$) in the confirmatory PMR (Table 1), while the direction of effect of 6 proteins was not consistent with that in the discovery PMR analysis. Thus, a total of 13 proteins were replicated in the confirmatory PMR, including PLG, PSCK9, COLEC11, ZNF180, ERP29, TCP1, FN1, CDH5, IL15RA, MGAT4B, TNFRSF6B, DNM2, and IGF1R.

Bayesian colocalization analysis found that PCSK9 (PP.H4 = 0.99), APOB (PP.H4 = 0.89), and FN1 (PP.H4 = 0.87) each share a variation with CAD. The top shared variants of the above three proteins were rs11591147, rs548145, and rs1250229, respectively (Fig. 3). To relax the single causal variant assumption, we also used SuSiE for colocalization analysis allowing for multiple causal variants. Coloc.SuSiE analysis identified APOC1 with evidence of genetic colocalization based on a PP.H4 $_{SuSiE} = 0.78$.

According to the central dogma of molecular biology that proteins are translated from mRNA, this study further explored the association between the transcription level of protein-coding genes and CAD. A total of 27 identified protein-coding genes were tested in PMR analysis using eQTL summary data from whole blood and coronary artery (Table 2). Among them, genetically regulated mRNA expression levels of 12 and 11 protein-coding genes in whole blood and coronary arteries, respectively, were associated with the risk of coronary artery disease ($P < 0.05$) (Table 2). The transcription levels of 6 protein-coding genes including MTAP, ERP29, MORF4L1, DNM2, CDKN2B, and LDLR in blood and coronary arteries were both related to CAD ($P < 0.05$). Additionally, the transcription levels of IL5RA, SIGLEC7, CSNK2A1, PSMA4, ZNF180, and MAPKAPK5 in blood, as well as TCP1, APOC2, C19orf80, APOC1, and TNFRSF6B in coronary arteries, were also found to be related to CAD ($P < 0.05$). In addition, the effect direction of the transcription levels of protein-coding genes such as PCSK9, PSMA4, EPOR, and LDLR on CAD are inconsistent between whole blood and coronary arteries, which may indicate tissue specificity.

We investigated the association of candidate proteins with CAD-related traits such as hypertension, hyperlipidemia,
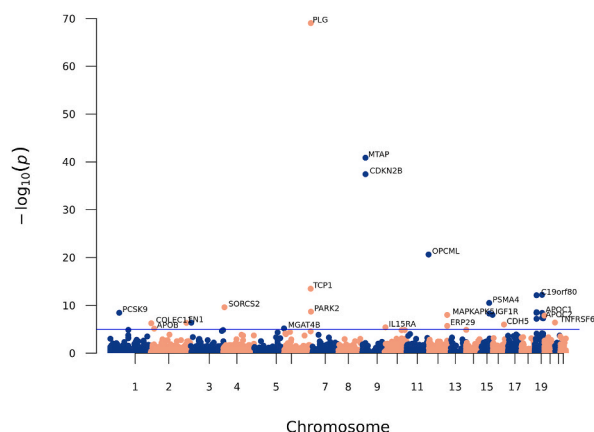
Fig. 2. Manhattan plot for the discovery PMR analysis integrating the CAD GWAS ($N = 336,924$) with the large-scale plasma proteomes ($N = 35,559$). Each point represents a single association test between a gene and CAD ordered by genomic position on the x axis and the association strength on the y axis as the -log10(P). The discovery PWAS identified 30 genes whose *cis*-regulated plasma protein abundance was associated with CVD after Bonferroni correction (P < 1.13E-5, 0.05/4408). The blue horizontal line reflects the significant threshold of the $P < 1.13$E-5.

Table 1
The discovery PMR identified 30 significant proteins, of which, 25 were found in the confirmatory PMR and 13proteins replicated.

Protein-coding Genes	Discovery PMR				Confirmtary PMR				Evidence for Replication	
	Beta	Se	P	P_pleiotropy	Beta	Se	P	P_pleiotropy	Effect direction consistency	Significance
PLG	−2.296	0.130	9.19E-70	0.107	−1.097	0.072	3.66E-52	0.278	Yes	Yes
PCSK9	0.044	0.007	3.63E-09	0.481	0.048	0.007	7.35E-12	0.251	Yes	Yes
COLEC11	0.011	0.002	5.67E-07	0.502	0.019	0.004	1.46E-07	0.497	Yes	Yes
ZNF180	−0.907	0.126	6.30E-13	0.715	−0.803	0.153	1.61E-07	0.791	Yes	Yes
ERP29	0.500	0.105	2.01E-06	0.960	0.296	0.058	2.74E-07	0.565	Yes	Yes
TCP1	−1.071	0.141	3.36E-14	0.198	−0.207	0.042	7.71E-07	0.175	Yes	Yes
FN1	−0.061	0.012	8.25E-07	0.055	−0.033	0.009	1.13E-04	0.082	Yes	Yes
CDH5	−0.080	0.016	1.03E-06	0.137	−0.032	0.010	1.50E-03	0.543	Yes	Yes
IL15RA	−0.013	0.003	4.02E-06	0.350	−0.022	0.008	3.44E-03	0.490	Yes	Yes
MGAT4B	0.018	0.004	6.74E-06	0.779	0.013	0.004	3.49E-03	0.443	Yes	Yes
TNFRSF6B	0.116	0.023	3.98E-07	0.453	0.036	0.014	9.78E-03	0.486	Yes	Yes
DNM2	0.647	0.110	3.59E-09	0.033	0.115	0.046	1.15E-02	0.816	Yes	Yes
IGF1R	−0.045	0.008	9.73E-09	0.645	−0.014	0.006	2.63E-02	0.160	Yes	Yes
IL5RA	−0.011	0.002	4.24E-07	0.179	−0.005	0.003	1.02E-01	0.146	Yes	No
OPCML	0.032	0.003	2.32E-21	0.068	0.004	0.013	7.52E-01	0.962	Yes	No
APOB	0.145	0.032	7.50E-06	0.701	0.004	0.016	8.08E-01	0.256	Yes	No
PSMA4	−0.640	0.096	3.16E-11	0.408	1.091	0.243	7.04E-06	0.598	No	Yes
MORF4L1	−0.606	0.104	5.17E-09	0.138	0.167	0.041	4.45E-05	0.376	No	Yes
MAPKAPK5	0.664	0.116	1.02E-08	0.655	−0.405	0.113	3.22E-04	0.072	No	Yes
PARK2	0.071	0.012	2.10E-09	0.617	−0.055	0.017	9.57E-04	0.292	No	Yes
APOC2	−0.228	0.042	4.03E-08	0.023	0.085	0.030	4.59E-03	0.242	No	Yes
SORCS2	0.030	0.005	2.53E-10	0.559	−0.031	0.014	3.22E-02	0.483	No	Yes
CSNK2A1	−0.129	0.023	1.42E-08	0.125	0.035	0.021	8.82E-02	0.225	No	No
APOC1	−0.038	0.006	3.82E-09	0.052	0.013	0.008	9.06E-02	0.659	No	No
SIGLEC7	−0.025	0.005	5.27E-08	0.530	0.004	0.004	2.92E-01	0.231	No	No
LDLR	0.686	0.127	6.52E-08	0.281						
EPOR	0.607	0.102	2.73E-09	0.084						
C19orf80	1.056	0.147	7.80E-13	0.166						
CDKN2B	−1.614	0.125	3.68E-38	0.651						
MTAP	−2.921	0.216	1.35E-41	0.177						

atherosclerosis, stroke, venous thrombosis, NAFLD, BMI, and WhRadjBMI. Among the 30 candidate proteins, two proteinss (PARK2 and MAPKAPK5), 8 proteins (TNFRSF6B, APOC2, MORF4L1, MAPKAPk5, OPCML, ERP29, LDLR, and MGAT4B), two proteinss (OPCML and SIGLEC7), six proteinss (CSNK2A1, MAPKAPK5, ERP29, IGF1R, DNM2, and TCP1), five proteins (LDLR, DNM2, SORCS2, APOC1, and MAPKAPK5), nine proteins (DNM2, ZNF180, FN1, LDLR, C19orf80, APOC2, PSMA4, TNFRSF6B, and PLG), and one
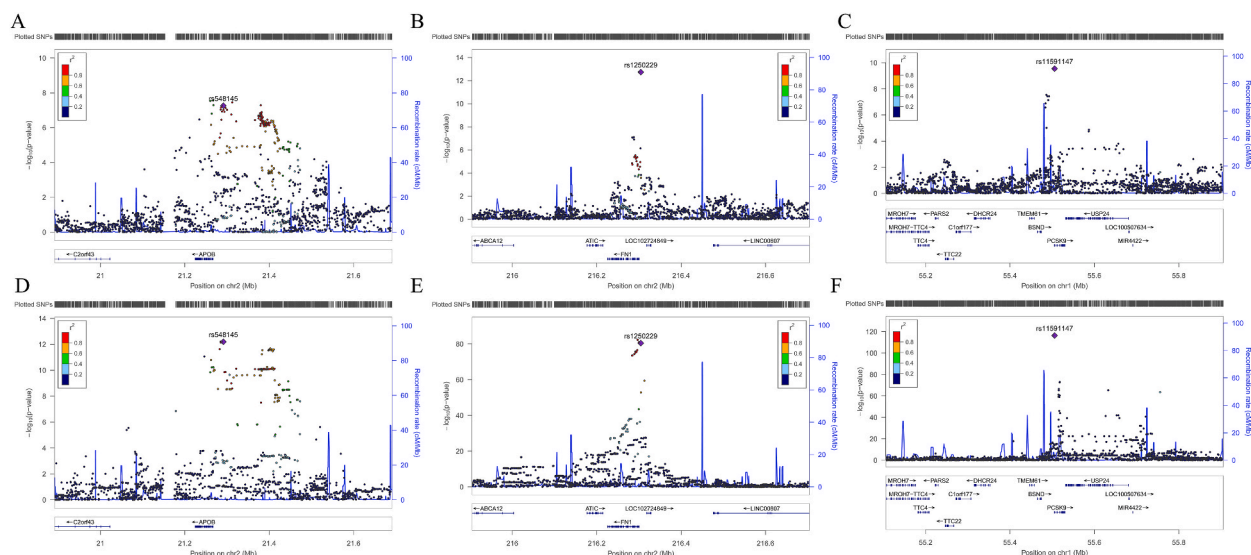
Fig. 3. Locuszoom plots for the colocalization analysis of APOB (A, GWAS; D, pQTL), FN1 (B, GWAS; E, pQTL), and PCSK9 (C, GWAS; F, pQTL).

proteins (OPCML) were associated with stroke, BMI, atherosclerosis, hypertension, venous thrombosis, WhRadjBMI, and dyslipidemia, respectively ($P < 0.05/4408$) (Supplementary Table S2). There were no overlapping candidate proteins between NAFLD and CAD. In addition, we performed disease enrichment analysis and found that candidate protein-coding genes were primarily enriched for CAD, coronary arteriosclerosis, and dyslipidemias (Supplementary Table S3).

We performed cell-type expression specificity analysis in two tissues. In atherosclerotic human coronary arteries, APOC1, APOC2, and SIGLEC7 were only found in macrophages; COLEC11, IL5RA, and OPCML were more abundant in artery cell; the ES of IL15RA, IGF1R, CDH5, and LDLR were relatively high in endothelial cell. The ES of APOB and FN1 in regulatory T cells and vascular smooth muscle cells were relatively high. The highest ES of MGAT4B, TCP1, MTAP, CDKN2B, ERP29, PSMA4, MORF4L1, IGF1R, DNM2, and EPOR in 7 coronary artery cell types were all lower than 0.50 (Fig. 4A, Supplementary Table S4). In whole blood, SIGLEC7 was more abundant in kineNK, nonM, and toxiNK; the expression abundance of APOC1 was relatively high in CMP, MEP, and erythrocytes; the four highest ES for LDLR were in HSC, MPP, claM, and preM. The highest ES of 7 protein-coding genes (TCP1, MAPKAPK5, ERP29, PSMA4, MORF4L1, CSNK2A1, DNM2) in 32 immunophenotypic cell types were all lower than 0.50 (Fig. 4B–Supplementary Table S5). In addition, GO and KEGG enrichment analysis identified 150 GO terms and 1 KEGG pathway (Supplementary Table S6), and the candidate protein-coding genes were mainly enriched in GO terms and pathways involving cholesterol metabolism, regulation of lipoprotein levels, and endocytosis, such as Cholesterol metabolism (hsa04979, $P = 2.25E-7$), plasma lipoprotein particle clearance (GO:0034381, $P = 5.47E-5$), and receptor-mediated endocytosis (GO:0006898, $P = 3.25E-7$). Notably, we also found that candidate protein-coding genes were enriched in the alcohol metabolic process (GO:0006066, $P = 8.01E-5$) and secondary alcohol metabolic process (GO:1902652, $P = 2.01E-3$).

4. Discussion

In this study, we employed a range of analytical techniques to investigate the functional associations between protein biomarkers in blood and CAD risk. Our PMR analysis suggested that the abundance of 30 potential candidate proteins in the blood was causally associated with CAD, 13 of which were validated in a confirmatory PMR analysis which was based on independent pQTL data from ARIC. 19 candidate proteins were supported by PMR at the transcriptional level in whole blood or coronary arteries; The protein levels of PCSK9, APOB, APOC2, and FN1 each share a causal variant with CAD. Furthermore, we also investigated the cell-type expression specificity of candidate protein-coding genes in coronary arteries and whole blood and performed enrichment analysis to explore potential pathways involved in CAD, providing reference information for further mechanism and treatment research.

The associations identified by PMR provide a substantial step toward gene prioritization of GWAS loci. Our study identified multiple proteins that have been extensively studied in previous CAD studies, such as FN1 and PCSK9. FN1 encodes fibronectin, mainly including two isoforms: plasma fibronectin 1 (pFN, present in plasma as a soluble dimer) and cellular fibronectin 1 (cFN, as a dimer or multimer present on the cell surface and in the extracellular matrix) [48]. cFN is deposited as insoluble fibers and is involved in cell adhesion. pFN constitutes a significant fraction of arterial FN and may be involved in normal tissue remodeling, and angiogenesis [49–51]. Many studies have found that multiple SNPs of FN1 were associated with CAD risk, such as rs1250229 and rs1250259 [52–55]. Our study also found that the rs1250229 was a shared variant between CAD and FN1 encoded proteins through coloc analysis. PSCK9 encodes a member of the subtilisin-like proprotein convertase family, which is mainly expressed in the liver, kidney, and intestinal tissues, and can promote the degradation of LDL receptors in the liver, thereby reducing the uptake of LDL in plasma by the liver, leading to hyperlipidemia [56,57]. PCSK9 inhibitors (PCSKi) have been used in the prevention of CAD [58,59]. The shared

Table 2
Results of PMR analysis at the transcriptome level.

Gene Symbol	eQTLGen Blood eQTL				GTEx Blood eQTL				GTEx Artery Cononary eQTL				Evidence tier[a]
	Beta	Se	P	P_pleiotropy	Beta	Se	P	P_pleiotropy	Beta	Se	P	P_pleiotropy	
MTAP	0.042	0.007	5.97E-10	0.913	0.321	0.044	5.13E-13	0.849	0.444	0.024	1.51E-73	0.157	3
ERP29	0.124	0.02	6.71E-10	0.273	0.099	0.026	1.85E-04	0.628	0.225	0.070	1.22E-03	0.088	3
MORF4L1	−0.130	0.021	7.90E-10	0.355	−0.024	0.006	2.05E-04	0.014	−0.037	0.007	1.48E-07	0.742	3
TCP1	−0.001	0.003	7.58E-01	0.605	0.126	0.020	5.42E-10	0.369	0.200	0.039	3.42E-07	0.489	2
CDKN2B	−0.057	0.005	9.50E-26	0.002	−0.007	0.004	5.47E-02	0.002	−0.264	0.022	2.39E-33	0.104	2
LDLR	0.180	0.045	7.51E-05	0.424	−0.009	0.005	8.63E-02	0.000	0.096	0.030	1.53E-03	0.939	2
DNM2	0.092	0.016	1.33E-08	0.847	−0.021	0.011	4.59E-02	0.577	−0.121	0.032	1.29E-04	0.794	2
C19orf80	–	–	–	–	0.581	0.159	2.67E-04	0.000	0.301	0.047	1.94E-10	0.305	2
APOC1	–	–	–	–	0.019	0.005	3.40E-04	0.064	0.008	0.004	4.58E-02	0.764	2
APOC2	–	–	–	–	0.033	0.007	1.95E-06	0.749	0.010	0.004	1.83E-02	0.017	2
EPOR	−0.005	0.006	4.43E-01	0.826	0.011	0.005	1.82E-02	0.744	−0.009	0.011	4.23E-01	0.855	1
MAPKAPK5	−0.037	0.012	2.47E-03	0.007	0.033	0.018	5.78E-02	0.402	0.011	0.019	5.47E-01	0.003	1
PSMA4	−0.031	0.007	2.64E-06	0.553	−0.004	0.004	3.33E-01	0.070	0.011	0.009	2.10E-01	0.594	1
ZNF180	0.257	0.03	1.92E-17	0.651	−0.003	0.006	6.02E-01	0.840	−0.002	0.005	6.79E-01	0.923	1
IL5RA	−0.036	0.009	1.14E-04	0.078	–	–	–	–	–	–	–	–	1
SIGLEC7	−0.02	0.005	1.39E-05	0.354	–	–	–	–	–	–	–	–	1
CSNK2A1	−0.044	0.009	2.16E-06	0.172	–	–	–	–	–	–	–	–	1
PCSK9	–	–	–	–	0.009	0.002	2.77E-04	0.918	−0.004	0.003	2.16E-01	0.921	1
TNFRSF6B	–	–	–	–	−0.005	0.003	5.45E-02	0.278	−0.008	0.004	2.49E-02	0.622	1
FN1	0.001	0.006	9.21E-01	0.008	0.003	0.006	6.12E-01	0.526	−0.001	0.005	8.18E-01	0.406	0
IGF1R	−0.005	0.005	3.54E-01	0.356	–	–	–	–	–	–	–	–	0
IL15RA	0.004	0.003	2.01E-01	0.598	–	–	–	–	–	–	–	–	0
MGAT4B	−0.007	0.006	2.76E-01	0.921	–	–	–	–	–	–	–	–	0
COLEC11	0.008	0.007	2.42E-01	0.506	–	–	–	–	–	–	–	–	0
SORCS2	−0.006	0.005	2.16E-01	0.287	–	–	–	–	–	–	–	–	0
PARK2	−0.059	0.031	5.81E-02	0.862	–	–	–	–	–	–	–	–	0
CDH5	–	–	–	–	0.000	0.005	9.61E-01	0.563	−0.003	0.007	0.723	0.833	0

[a] Tier 3: the transcription level of the protein-coding gene is associated with CAD in all three-transcriptome data and the effect direction is consistent; Tier 2: the transcription level of the protein-coding gene is associated with CAD in coronary arteries and one of the whole blood transcriptomes and the effect direction is consistent; Tier 1: the transcription level of the protein-coding gene is associated with CAD in only one tissue transcriptome; Tier 0: Non-association evidence. PMR, Probabilistic two sample mendelian randomization; eQTL, expression quantitative trait loci.
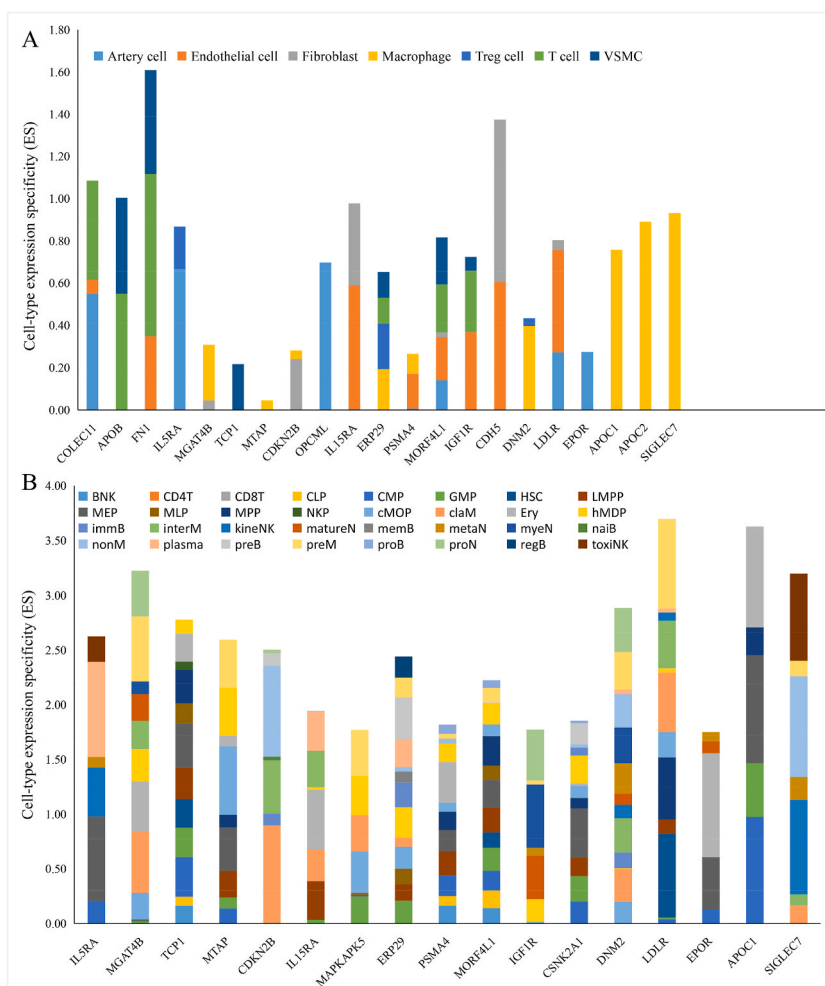
Fig. 4. Single-cell-type expression of the potentially CAD-risk genes in atherosclerotic human coronary arteries (A) and whole blood (B). Bar graph of single-cell-type enrichment for risk genes in CAD from the discovery PMR. The diagram depicts Cell-type expression-specificity (y axis) for each gene (x axis), with evidence of substantial enrichment within a specific cell type (histogram of the bar). We used the "wisdom of the crowd" technique to assess enrichment based on gene expression in one cell type against all other cell types. CAD, coronary artery disease; PMR, probabilistic two-sample Mendelian randomization; VSMC, Vascular smooth muscle cell; Treg, Regulatory T cell. MEP, Megakaryocytic-Erythroid Progenitors; NKP, NK cell precursors; BNK, Blood Natural Killer Cells; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte/macrophage progenitor; HSC, hematopoietic stem cell; LMPP, Lymphoid-primed multipotent progenitor; MEP, Megakaryocyte/erythrocyte progenitor; MLP, Multilymphoid Progenitor; MPP, Multipotent Progenitors and Plasma Cells; NKP, NK cell precursors; cMOP, common monocyte progenitor; claM, classical monocle; Ery, Erythrocytes; hMDP, human monocyte-dendritic progenitor; immB, Immature B; interM, Intermediate monocyle; kineNK, cytokine NK; matureN, mature neutrophil; memB, memory B; metaN, Meta-myelocyte; myeN, myelocyte neutrophil; naiB, Naive B; nonM, non-classical monocyle; preB, preB; preM, Pre-monocyle; proB, proB; proN, Pro-myelocyte; regB, Regulatory B; toxiNK, cytotoxic NK.

variant rs11571147 between CAD and PCSK9 has also been reported by other studies on its association with CAD [60]. The above findings also indicate the effectiveness and robustness of the analytical methods used in this study to a certain extent.

Our study also identified several novel candidate proteins, such as IL15RA, IL5RA, and CSNK2A1. Our study indicated that both IL15RA and IL5RA protein abundance were associated with decreased CAD risk. Both IL15RA and IL5RA encode interleukin receptors. Some studies have shown that IL-15 is upregulated in human and animal atherosclerotic lesions, which may lead to the recruitment and activation of T cells in the process of atherosclerosis [61,62]. On the other hand, IL-15 protects cardiomyocytes from oxidative stress through the PI3K/ERK1/2 pathway [63], supplementing IL-15 can improve the survival rate of cardiomyocytes under oxidative stress and can improve cardiac function recovery after myocardial infarction [64,65]. The role of IL-15 in atherosclerosis and cardiac outcomes needs to be further explored [66]. Several studies have shown that IL5 is athero-protective, supported by that it is primarily produced by anti-inflammatory T helper 2 (Th2) cells [67–70]. CSNK2A1 encodes the catalytic subunit of a highly conserved serine-threonine kinase. Song et al. have reported that CSNK2 is involved in the regulation of the phosphorylation of the autophagy gene ATG16L1, which is related to various diseases including cardiovascular diseases [71,72]. Furthermore, CSNK2A1 plays a key role

in the regulation of DNA damage repair and metabolic processes and promotes glycerolipid metabolism [73].

In the confirmatory PMR analysis of 25 candidate proteins, only 13 proteins showed confirmatory evidence with a $P < 0.05$ and consistent effect direction. This could be due to several reasons such as differences in age, gender, and disease history between the two proteome data source populations, potential methodological differences for the two pQTL data in terms of tissue collection, sample processing, and analysis platform, and limited and non-overlapping genotypes of the two pQTL data, resulting in differences in the instrumental variables of the two PMR analyzes. Also, 6 proteins with $P < 0.05$ had inconsistent effect direction with the discovery PMR analysis, which may imply a complex relationship between these proteins and CAD rather than a simple linear relationship. We further performed PMR analysis at the transcriptome level to explore the GWAS-related transcriptional regulation mechanism. Among the 27 candidate proteins, 19 proteins reached a consensus at the transcript level, and the other 8 proteins had no statistical association with CAD at the transcript levels of whole blood and coronary arteries. This may be due to the inconsistency of mRNA expression and protein levels. Previous studies have shown that the expression of many genes was poorly correlated with protein levels, proteins with distant pQTLs (pQTLs not within ±10 Mb of the DNA sequence encoding the protein) were poorly correlated with mRNA expression, which may be partly attributable to post-transcriptional mechanisms, such as protein translation and degradation [74–76]. A weak correlation between expression and protein means that the potential impact of eQTLs on downstream phenotypes is usually weakened or buffered.

Enrichment analysis showed that candidate protein-coding genes were mainly enriched in pathways involving cholesterol, lipid, and lipoprotein metabolism and their regulation. Seven protein-coding genes, including ZNF180, TNFRSF6B, MGAT4B, C19orf80, PARK2, OPCML, and SIGLEC7, were not included in the identified GO terms and KEGG pathways, indicating that they may affect CAD through other pathways that may not yet be identified, which requires further investigation. It is worth noting that this study also found that multiple pathways involving telomerase and telomeres may also be involved in the pathogenesis of CAD, such as the positive regulation of telomere maintenance by telomerase and the regulation of telomerase activity. Telomerase consists of a telomerase RNA component (TERC) and a catalytic component, telomerase reverse transcriptase (TERT), which uses TERC as a template to synthesize new telomere DNA repeats at single-stranded overhangs to maintain telomeres. Several studies have observed shorter telomeres in endothelial cells and vascular smooth muscle cells (VSMCs) in atherosclerotic artery walls, consistent with various regions of the human vasculature [77–79]. In the cell-type expression specific analysis, we also found that TCP1, a gene enriched in telomerase and telomere pathway, was specifically expressed in coronary artery VSMCs. Hoffmann et al. suggested that the antioxidant function of mitochondrial telomerase reverse transcriptase may have athero-protective effects, making it a potential target for clinical trials [80].

The study has certain limitations. Firstly, the proteome data utilized in the study only covers specific proteomes and not the entire proteome. This may have restricted the scope of the study to reveal more associations. Further studies that employ alternative detection techniques and biological samples may uncover additional associations. Secondly, the PMR method requires the use of the 1000 Genomes Project data as the LD reference panel. Consequently, only SNPs overlapping with 1000 G data can be selected as instrumental variables, which may have reduced the statistical power of the study to a certain extent. Third, the diagnosis of CAD in GWAS is derived from various sources, which may lead to heterogeneity in the results due to different diagnostic criteria. To enhance accuracy and identify more specific disease-related markers, future GWAS studies should consider using more detailed diagnostic criteria. Fourth, hundreds of proteins were not tested using PMR due to their low expression heritability or failure of LD matrix decomposition, resulting in the loss of potential candidate proteins. Furthermore, this study only mapped GWAS signals at the protein level, and the biological mechanisms of some candidate proteins and pathways are still unclear. More mechanism studies and data such as mQTL and single-cell sequencing are required to gain a better understanding of the role of candidate protein-coding genes associated with CAD.

In conclusion, we found strong evidence supporting the association of multiple protein biomarkers in blood with CAD, among which multiple proteins such as TCP1 and PCSK9 were further validated at the transcriptional level. We also performed cell-type expression specificity and enrichment analysis. These findings provide information about the genetic and physiological processes of CAD, providing clues for further exploration of CAD pathogenesis and new therapeutic targets. Future studies utilizing more extensive large-scale molecular datasets from CAD-associated tissues are warranted, which could further help generate new insights into the genetic and functional mechanisms of CAD, as well as drug targets.

Funding

Data availability statement

All data relevant to the study are public available. The data generated in this study will be available from the corresponding author upon reasonable request. GWAS summary statistics from these analyses are available at CARDIoGRAMplusC4D Consortium (http://www.cardiogramplusc4d.org/). The statistical code needed to reproduce the results in the article is available upon request.

CRediT authorship contribution statement

Jiqing Li: Writing – review & editing, Writing – original draft, Software, Methodology, Data curation. **Jiate Wei:** Writing – original draft. **Ping Fu:** Formal analysis. **Jianhua Gu:** Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38036.

References

[1] Global, regional, National age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013, Lancet 385 (9963) (2015) 117–171.

[2] A.K. Malakar, D. Choudhury, B. Halder, P. Paul, A. Uddin, S. Chakraborty, A review on coronary artery disease, its risk factors, and therapeutics, J. Cell. Physiol. 234 (10) (2019) 16812–16823.

[3] A.V. Khera, S. Kathiresan, Genetics of coronary artery disease: discovery, biology and clinical translation, Nat. Rev. Genet. 18 (6) (2017) 331–344.

[4] H. Schunkert, I.R. König, S. Kathiresan, M.P. Reilly, T.L. Assimes, H. Holm, et al., Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease, Nat. Genet. 43 (4) (2011) 333–338.

[5] A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease, Nat. Genet. 43 (4) (2011) 339–344.

[6] P. Deloukas, S. Kanoni, C. Willenborg, M. Farrall, T.L. Assimes, J.R. Thompson, et al., Large-scale association analysis identifies new risk loci for coronary artery disease, Nat. Genet. 45 (1) (2013) 25–33.

[7] M. Nikpay, A. Goel, H.H. Won, L.M. Hall, C. Willenborg, S. Kanoni, et al., A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease, Nat. Genet. 47 (10) (2015) 1121–1130.

[8] T.R. Webb, J. Erdmann, K.E. Stirrups, N.O. Stitziel, N.G. Masca, H. Jansen, et al., Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease, J. Am. Coll. Cardiol. 69 (7) (2017) 823–836.

[9] M.T. Maurano, R. Humbert, E. Rynes, R.E. Thurman, E. Haugen, H. Wang, et al., Systematic localization of common disease-associated variation in regulatory DNA, Science 337 (6099) (2012) 1190–1195.

[10] L. Li, Z. Chen, M. von Scheidt, S. Li, A. Steiner, U. Güldener, et al., Transcriptome-wide association study of coronary artery disease identifies novel susceptibility genes, Basic Res. Cardiol. 117 (1) (2022) 6.

[11] J. Zhang, D. Dutta, A. Köttgen, A. Tin, P. Schlosser, M.E. Grams, et al., Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies, Nat. Genet. 54 (5) (2022) 593–602.

[12] D. Rolland, V. Basrur, Y.K. Jeon, C. McNeil-Schwalm, D. Fermin, K.P. Conlon, et al., Functional proteogenomics reveals biomarkers and therapeutic targets in lymphomas, Proc. Natl. Acad. Sci. U. S. A. 114 (25) (2017) 6581–6586.

[13] L. Folkersen, S. Gustafsson, Q. Wang, D.H. Hansen, Å.K. Hedman, A. Schork, et al., Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals, Nat. Metab. 2 (10) (2020) 1135–1148.

[14] E. Ferkingstad, P. Sulem, B.A. Atlason, G. Sveinbjornsson, M.I. Magnusson, E.L. Styrmisdottir, et al., Large-scale integration of the plasma proteome with genetics and disease, Nat. Genet. 53 (12) (2021) 1712–1721.

[15] C. Yao, G. Chen, C. Song, J. Keefe, M. Mendelson, T. Huan, et al., Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease, Nat. Commun. 9 (1) (2018) 3268.

[16] B.B. Sun, J.C. Maranville, J.E. Peters, D. Stacey, J.R. Staley, J. Blackshaw, et al., Genomic atlas of the human plasma proteome, Nature 558 (7708) (2018) 73–79.

[17] J. Bourgault, E. Abner, H.D. Manikpurage, N. Pujol-Gualdo, T. Laisk, É. Gobeil, et al., Proteome-wide mendelian randomization identifies causal links between blood proteins and Acute pancreatitis, Gastroenterology 164 (6) (2023) 953, 65.e3.

[18] C. Zhang, F. Qin, X. Li, X. Du, T. Li, Identification of novel proteins for lacunar stroke by integrating genome-wide association data and human brain proteomes, BMC Med. 20 (1) (2022) 211.

[19] A.P. Wingo, Y. Liu, E.S. Gerasimov, J. Gockley, B.A. Logsdon, D.M. Duong, et al., Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis, Nat. Genet. 53 (2) (2021) 143–146.

[20] S. Burgess, D.S. Small, S.G. Thompson, A review of instrumental variable estimators for Mendelian randomization, Stat. Methods Med. Res. 26 (5) (2017) 2333–2355.

[21] S. Burgess, F. Dudbridge, S.G. Thompson, Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods, Stat. Med. 35 (11) (2016) 1880–1906.

[22] H. Zhu, X. Zhou, Transcriptome-wide association studies: a view from Mendelian randomization, Quant Biol 9 (2) (2021) 107–121.

[23] Z. Yuan, H. Zhu, P. Zeng, S. Yang, S. Sun, C. Yang, et al., Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies, Nat. Commun. 11 (1) (2020) 3861.

[24] C.P. Nelson, A. Goel, A.S. Butterworth, S. Kanoni, T.R. Webb, E. Marouli, et al., Association analyses based on false discovery rate implicate new loci for coronary artery disease, Nat. Genet. 49 (9) (2017) 1385–1391.

[25] R. Mägi, A.P. Morris, GWAMA: software for genome-wide association meta-analysis, BMC Bioinf. 11 (2010) 288.

[26] T. Rafnar, S. Thorlacius, E. Steingrimsson, M.H. Schierup, J.N. Madsen, V. Calian, et al., The Icelandic Cancer Project–a population-wide approach to studying cancer, Nat. Rev. Cancer 4 (6) (2004) 488–492.

[27] P.R. Loh, G. Tucker, B.K. Bulik-Sullivan, B.J. Vilhjálmsson, H.K. Finucane, R.M. Salem, et al., Efficient Bayesian mixed-model analysis increases association power in large cohorts, Nat. Genet. 47 (3) (2015) 284–290.

[28] The atherosclerosis risk in Communities (ARIC) study: design and objectives. The ARIC investigators, Am. J. Epidemiol. 129 (4) (1989) 687–702.

[29] L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E.N. Brody, et al., Aptamer-based multiplexed proteomic technology for biomarker discovery, PLoS One 5 (12) (2010) e15004.

[30] S.A. Williams, M. Kivimaki, C. Langenberg, A.D. Hingorani, J.P. Casas, C. Bouchard, et al., Plasma protein patterns as comprehensive indicators of health, Nat Med 25 (12) (2019) 1851–1857.

[31] M.H. Kowalski, H. Qian, Z. Hou, J.D. Rosen, A.L. Tapia, Y. Shan, et al., Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations, PLoS Genet. 15 (12) (2019) e1008500.

[32] B.N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next generation of genome-wide association studies, PLoS Genet. 5 (6) (2009) e1000529.

[33] O. Delaneau, H. Ongen, A.A. Brown, A. Fort, N.I. Panousis, E.T. Dermitzakis, A complete tool set for molecular QTL discovery and analysis, Nat. Commun. 8 (2017) 15452.

[34] The GTEx Consortium atlas of genetic regulatory effects across human tissues, Science 369 (6509) (2020) 1318–1330.

[35] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (1) (2013) 15–21.

[36] U. Võsa, A. Claringbould, H.J. Westra, et al., Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression, Nat. Genet. 53 (9) (2021) 1300–1310.

[37] B.V. Halldorsson, H.P. Eggertsson, K. Moore, H. Hauswedell, O. Eiriksson, M.O. Ulfarsson, et al., The sequences of 150,119 genomes in the UK Biobank, Nature 607 (7920) (2022) 732–740.

[38] G. Sveinbjornsson, M.O. Ulfarsson, R.B. Thorolfsdottir, B.A. Jonsson, E. Einarsson, G. Gunnlaugsson, et al., Multiomics study of nonalcoholic fatty liver disease, Nat. Genet. 54 (11) (2022) 1652–1663.

[39] R. Malik, G. Chauhan, M. Traylor, M. Sargurupremraj, Y. Okada, A. Mishra, et al., Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes, Nat. Genet. 50 (4) (2018) 524–537.

[40] J. Ghouse, V. Tragante, G. Ahlberg, S.A. Rand, J.B. Jespersen, E.B. Leinøe, et al., Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism, Nat. Genet. 55 (3) (2023) 399–409.

[41] L. Yengo, J. Sidorenko, K.E. Kemper, Z. Zheng, A.R. Wood, M.N. Weedon, et al., Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry, Hum. Mol. Genet. 27 (20) (2018) 3641–3649.

[42] C. Giambartolomei, D. Vukcevic, E.E. Schadt, L. Franke, A.D. Hingorani, C. Wallace, et al., Bayesian test for colocalisation between pairs of genetic association studies using summary statistics, PLoS Genet. 10 (5) (2014) e1004383.

[43] C. Wallace, A more accurate method for colocalisation analysis allowing for multiple causal variants, PLoS Genet. 17 (9) (2021) e1009440.

[44] X. Xie, M. Liu, Y. Zhang, B. Wang, C. Zhu, C. Wang, et al., Single-cell transcriptomic landscape of human blood cells, Natl. Sci. Rev. 8 (3) (2021) nwaa180.

[45] P.N. Timshel, J.J. Thompson, T.H. Pers, Genetic mapping of etiologic brain cell types for obesity, Elife 9 (2020) e55851.

[46] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS 16 (5) (2012) 284–287.

[47] J. Piñero, J.M. Ramírez-Anguita, J. Saüch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, et al., The DisGeNET knowledge platform for disease genomics: 2019 update, Nucleic Acids Res. 48 (D1) (2020) D845–845D855.

[48] W.S. To, K.S. Midwood, Plasma and cellular fibronectin: distinct and independent functions during tissue repair, Fibrogenesis Tissue Repair 4 (2011) 21.

[49] F.A. Moretti, A.K. Chauhan, A. Iaconcig, F. Porro, F.E. Baralle, A.F. Muro, A major fraction of fibronectin present in the extracellular matrix of tissues is plasma-derived, J. Biol. Chem. 282 (38) (2007) 28057–28062.

[50] H. Kumra, L. Sabatier, A. Hassan, T. Sakai, D.F. Mosher, J. Brinckmann, et al., Roles of fibronectin isoforms in neonatal vascular development and matrix integrity, PLoS Biol. 16 (7) (2018) e2004812.

[51] H.Y. Chiang, V.A. Korshunov, A. Serour, F. Shi, J. Sottile, Fibronectin is an important regulator of flow-induced vascular remodeling, Arterioscler. Thromb. Vasc. Biol. 29 (7) (2009) 1074–1079.

[52] S. Soubeyrand, P. Lau, M. Nikpay, A.T. Dang, R. McPherson, Common polymorphism that protects from cardiovascular disease increases fibronectin processing and secretion, Circ Genom Precis Med 15 (2) (2022) e003428.

[53] M.M. Page, K.L. Ellis, D.C. Chan, J. Pang, A.J. Hooper, D.A. Bell, et al., A variant in the fibronectin (FN1) gene, rs1250229-T, is associated with decreased risk of coronary artery disease in familial hypercholesterolaemia, J Clin Lipidol 16 (4) (2022) 525–529.

[54] P. van der Harst, N. Verweij, Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease, Circ. Res. 122 (3) (2018) 433–443.

[55] H.R. Warren, E. Evangelou, C.P. Cabrera, H. Gao, M. Ren, B. Mifsud, et al., Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk, Nat. Genet. 49 (3) (2017) 403–415.

[56] M.S. Sabatine, Advances in the treatment of dyslipidemia, Cleve. Clin. J. Med. 83 (3) (2016) 181–186.

[57] F. Rodriguez, R.A. Harrington, Cholesterol, cardiovascular risk, statins, PCSK9 inhibitors, and the future of LDL-C lowering, JAMA 316 (19) (2016) 1967–1968.

[58] R. Chaudhary, J. Garg, N. Shah, A. Sumner, PCSK9 inhibitors: a new era of lipid lowering therapy, World J. Cardiol. 9 (2) (2017) 76–91.

[59] F. Ferrari, R. Stein, M.T. Motta, E.H. Moriguchi, PCSK9 inhibitors: clinical relevance, molecular mechanisms, and safety in clinical practice, Arq. Bras. Cardiol. 112 (4) (2019) 453–460.

[60] N. Shyamala, K.K. Gundapaneni, R.K. Galimudi, M.A. Tupurani, C. Padala, K. Puranam, et al., PCSK9 genetic (rs11591147) and epigenetic (DNA methylation) modifications associated with PCSK9 expression and serum proteins in CAD patients, J. Gene Med. 23 (8) (2021) e3346.

[61] D.M. Wuttge, P. Eriksson, A. Sirsjö, G.K. Hansson, S. Stemme, Expression of interleukin-15 in mouse and human atherosclerotic lesions, Am. J. Pathol. 159 (2) (2001) 417–423.

[62] C. Gokkusu, M. Aydin, E. Ozkok, F. Tulubas, A. Elitok, B. Pamukcu, et al., Influences of genetic variants in interleukin-15 gene and serum interleukin-15 levels on coronary heart disease, Cytokine 49 (1) (2010) 58–63.

[63] Y. Yeghiazarians, H. Honbo, I. Imhof, B. Woods, V. Aguilera, J. Ye, et al., IL-15: a novel prosurvival signaling pathway in cardiomyocytes, J. Cardiovasc. Pharmacol. 63 (5) (2014) 406–411.

[64] K. Ameri, D. Bayardorj, R. Samurkashian, M. Fredkin, E. Fuh, V. Nguyen, et al., Administration of interleukin-15 peptide improves cardiac function in a mouse model of myocardial infarction, J. Cardiovasc. Pharmacol. 75 (1) (2020) 98–102.

[65] N.G. Barra, S. Reid, R. MacKenzie, G. Werstuck, B.L. Trigatti, C. Richards, et al., Interleukin-15 contributes to the regulation of murine adipose tissue and human adipocytes, Obesity 18 (8) (2010) 1601–1607.

[66] L. Guo, M.F. Liu, J.N. Huang, J.M. Li, J. Jiang, J.A. Wang, Role of interleukin-15 in cardiovascular diseases, J. Cell Mol. Med. 24 (13) (2020) 7094–7101.

[67] C.J. Binder, K. Hartvigsen, M.K. Chang, M. Miller, D. Broide, W. Palinski, et al., IL-5 links adaptive and natural immunity specific for epitopes of oxidized LDL and protects from atherosclerosis, J. Clin. Invest. 114 (3) (2004) 427–437.

[68] M. Sämpi, O. Ukkola, M. Päivänsalo, Y.A. Kesäniemi, C.J. Binder, S. Hörkkö, Plasma interleukin-5 levels are related to antibodies binding to oxidized low-density lipoprotein and to decreased subclinical atherosclerosis, J. Am. Coll. Cardiol. 52 (17) (2008) 1370–1378.

[69] S. Taleb, A. Tedgui, Z. Mallat, Adaptive T cell immune responses and atherogenesis, Curr. Opin. Pharmacol. 10 (2) (2010) 197–202.

[70] A. Silveira, O. McLeod, R.J. Strawbridge, K. Gertow, B. Sennblad, D. Baldassarre, et al., Plasma IL-5 concentration and subclinical carotid atherosclerosis, Atherosclerosis 239 (1) (2015) 125–130.

[71] C. Borgo, C. D'Amore, S. Sarno, M. Salvi, M. Ruzzene, Protein kinase CK2: a potential therapeutic target for diverse human diseases, Signal Transduct. Targeted Ther. 6 (1) (2021) 183.

[72] H. Song, J. Pu, L. Wang, L. Wu, J. Xiao, Q. Liu, et al., ATG16L1 phosphorylation is oppositely regulated by CSNK2/casein kinase 2 and PPP1/protein phosphatase 1 which determines the fate of cardiomyocytes during hypoxia/reoxygenation, Autophagy 11 (8) (2015) 1308–1325.

[73] T.M. Onorato, S. Chakraborty, D. Haldar, Phosphorylation of rat liver mitochondrial glycerol-3-phosphate acyltransferase by casein kinase 2, J. Biol. Chem. 280 (20) (2005) 19527–19534.

[74] J.M. Chick, S.C. Munger, P. Simecek, E.L. Huttlin, K. Choi, D.M. Gatti, et al., Defining the consequences of genetic variation on a proteome-wide scale, Nature 534 (7608) (2016) 500–505.

[75] A. Battle, Z. Khan, S.H. Wang, A. Mitrano, M.J. Ford, J.K. Pritchard, et al., Genomic variation. Impact of regulatory variation from RNA to protein, Science 347 (6222) (2015) 664–667.

[76] C. Vogel, E.M. Marcotte, Insights into the regulation of protein abundance from proteomic and transcriptomic analyses, Nat. Rev. Genet. 13 (4) (2012) 227–232.

[77] K. Okuda, M.Y. Khan, J. Skurnick, M. Kimura, H. Aviv, A. Aviv, Telomere attrition of the human abdominal aorta: relationships with age and atherosclerosis, Atherosclerosis 152 (2) (2000) 391–398.

[78] E. Chang, C.B. Harley, Telomere length and replicative aging in human vascular tissues, Proc. Natl. Acad. Sci. U. S. A. 92 (24) (1995) 11190–11194.

[79] J. Huzen, W. Peeters, R.A. de Boer, F.L. Moll, L.S. Wong, V. Codd, et al., Circulating leukocyte and carotid atherosclerotic plaque telomere length: interrelation, association with plaque characteristics, and restenosis after endarterectomy, Arterioscler. Thromb. Vasc. Biol. 31 (5) (2011) 1219–1225.

[80] J. Hoffmann, G. Richardson, J. Haendeler, J. Altschmied, V. Andrés, I. Spyridopoulos, Telomerase as a therapeutic target in cardiovascular disease, Arterioscler. Thromb. Vasc. Biol. 41 (3) (2021) 1047–1061.