Prioritizing protein targets for dyslipidaemia and cardiovascular diseases using Mendelian randomization in South Asians

Siwei Wu[1,2‡], Alexander Smith[2‡], Jingxian Huang[2], Georg W. Otto[2], Yi-Hsuan Ko[2], James Yarmolinsky[2], Dipender Gill[2], Anand Rohatgi[3], Abbas Dehghan[2,4], Ioanna Tzoulaki[2,4-6,*], Devendra Meena[2,*]

Affiliations

[1] Department of Clinical Nutrition, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine

[2] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

[3] Department of Medicine, Division of Cardiology University of Texas Southwestern Medical Center Dallas TX USA.

[4]UK Dementia Research Institute at Imperial College London, London, UK

[5] British Heart Foundation Centre of Research Excellence, Imperial College London, London, UK

[6] Centre for Systems Biology, Biomedical Research Foundation, Academy of Athens, Athens, Greece

‡Joint first author

*Corresponding authors E-mail: d.meena@imperial.ac.uk (DM) and i.tzoulaki@imperial.ac.uk (IT)

Abstract word count: 250

Additional contents: figures 1-5, supplementary figures 6, supplementary tables 13

Abstract

South Asians are at higher risk of dyslipidaemia–a modifiable risk factor for cardiovascular diseases (CVDs). We aimed to identify protein targets for dyslipidaemia and CVDs in this population.

We used a two-sample Mendelian randomization (MR) approach, supplemented with MR-Egger, weighted median, colocalization, and generalized MR (GMR), to evaluate the effect of 2,800 plasma proteins on high/low/non-high-density lipoprotein cholesterol (HDL-C/LDL-C/nonHDL-C), total cholesterol, and triglycerides. Observational analyses were conducted on MR findings with strong colocalization (posterior probability≥ 80%) and GMR findings. Univariate MR assessed lipid-associated proteins' effect on CVDs. Finally, we compared the potential causal effects of plasma proteins on lipids in South Asians with those in Europeans to study heterogeneity in the MR effects.

We identified 29 genetically proxied proteins potentially causal to at least one lipid measure, 12 of which showed strong colocalization and GMR evidence, including ANGPTL3 and PCSK9. Notably, PCSK9 demonstrated a stronger association with LDL-C in European compared to South Asian ($\beta_{European}$= 0.37; 95% Confidence Interval (CI)= (0.36, 0.38), $\beta_{South\ Asian}$= 0.16; 95% CI= (0.11, 0.21)). Observational analysis suggested significant interaction between PCSK9 levels with LDL-C levels in South Asians with South Asians having a significantly lower effect compared to other ethnicities (PCSK9*South Asian; β= -0.14; 95% CI= (-0.174, -0.107)). Additionally, we showed that CELSR2 is also linked with CAD in South Asians.

Our study highlighted potential causal links between plasma proteins, dyslipidaemia, and CVD in South Asians, with significant heterogeneity across genetic ancestry groups. Larger studies in South Asians are needed to validate these findings.

Introduction

Routinely measured lipid parameters including high/low/non-high-density lipoprotein cholesterol (HDL-C/LDL-C/nonHDL-C), total cholesterol (TC), and triglycerides (TG), are well-established risk factors for cardiovascular diseases (CVDs), including coronary artery disease (CAD), stroke, and heart failure, (1-4) all of which are leading causes of mortality and morbidity worldwide (5). HDL-C and TC have been included in several risk equations like SCORE and PCE as important predictors for CVDs (6, 7). The role of HDL-C and TG as the target for CVD intervention is inconclusive but under investigation (8, 9). LDL-C and nonHDL-C are not only prognostic risk markers but also established therapeutic targets for CVDs (10, 11). Building on this, evidence suggests that individuals of South Asian ancestry, compared with non-Hispanic white population, have a higher prevalence of dyslipidaemia (12) and are more susceptible to cardiometabolic diseases closely related to dyslipidaemia including CAD (13, 14), stroke (15), and type 2 diabetes (16). Therefore, understanding the genomic and proteomic makeup of plasma lipids and identifying causal factors in South Asians is crucial for intervening in dyslipidaemia and preventing lipid-related cardiovascular conditions although the residual risk of lipid modifying medications remains.

Circulating plasma proteins are key to disease mechanisms and are promising drug targets. Previous studies have identified proteins associated with lipids some of which are targeted for dyslipidaemia treatment (17). The most prominent example is Apolipoprotein B (APOB), which is one of the most prognostic and best therapeutic targets for CVDs (18). Despite recent advances, most plasma proteins linked to lipids have been discovered in the European population (19). The proteomic findings in non-European individuals are limited, especially in South Asians, which is the most rapidly growing but neglected population with higher susceptibility to dyslipidaemia (12). We recently reported that certain plasma proteins may exert ancestry-specific causal effects on certain CVDs (20). Therefore, understanding the proteomic features of dyslipidaemia in South Asians is essential for developing more effective disease prevention strategies and drug discovery approaches for this high-risk population.

Recent genome wide association studies (GWAS) involving South Asian populations have made resources on circulating plasma proteins and lipid traits available (21, 22). Furthermore, advances in methodologies for causal inference in epidemiology, including the Mendelian randomization (MR) framework, now allow us to investigate the potential causal relationship between an exposure (e.g. plasma proteins) and an outcome (e.g. lipid traits) for which

GWAS summary statistics are available (23). MR utilises genetic variants as instrument variables (IVs), and is less susceptible to confounders and reverse causation bias than other study designs, and can be applied to for causal inference (24). Such approach was applied to systematically evaluate plasma proteins' effect on lipid traits in EUR, but not yet in SAS (25).

In this study, we aimed to (1) systematically evaluate bi-directional causal effects of plasma proteins on five lipid traits in South Asians using a two-sample proteome-wide MR approach (2) Investigate whether lipid-associated plasma proteins affect CAD and stroke risk in South Asians (3) Compare the effects of plasma proteins on lipid fractions in South Asians and Europeans.

Method

Data source

Genetic associations for genetically predicted plasma proteins

In the UK Biobank Pharma Proteomics Project (UKBPPP), 2,940 probes capturing 2,922 unique proteins were made available (22). We defined each probe as a protein and extracted genetic associations of the 2,940 plasma proteins in individuals of Central/South Asian ancestry (CSA, N=920) and European ancestry (EUR, N=34,557) from the UKBPPP (22). The ancestries were defined with the pan-UK Biobank (UKBB) definitions of genetic ancestry (available in UKBB return dataset 2442). We included the 2,800 plasma proteins as the primary exposures after excluding the ones whose cognate gene is ambiguous (N = 15) and those encoded by genes lying on the X chromosome (due to unavailability of X chromosome data on outcome; N = 88) or the MHC region (due to complex LD in this region; N = 37).

Genetic associations for genetically predicted lipid traits and CVDs

Five lipid traits including HDL-C, LDL-C, nonHDL-C, TC, and TG were included as the primary outcome in MR analysis. Ancestry-specific (N_{SAS} = up to 40,963, N_{EUR} = 1,320,016) genetic associations of the 5 traits were publicly available from the Global Lipids Genetics Consortium (21). The GWAS statistics for CAD were obtained from the East London Genes & Health (ELGH) study for SAS (26) and from the CARDIoGRAMplusC4D Consortium for EUR (27). The GWAS summary data on stroke for EUR and SAS were sourced from the GIGASTROKE consortium (28).

Proteome-wide MR and colocalization analysis on lipid fractions in SAS

Instrument selection and MR:

To obtain genetic instruments for plasma proteins, biallelic single nucleotide polymorphisms (SNP) lying within +/- 500 kilobase (KB) from the coding gene (defined as cis-acting SNPs), with minor allele frequency (MAF) > 0.05, and reaching genome wide significant level ($P < 5 \times 10^{-8}$) were extracted. The SNPs were further harmonized to the 5 lipid fractions and clumped to an $r^2 < 0.001$ to ensure independence across the instruments. Subsequently, a F-statistic was calculated for each SNP and SNPs with F-statistic < 10 were excluded to avoid weak instrument bias (29). We also applied the Steiger filtering and excluded the SNPs with potential reverse causality (30). After applying these filters, we ended up with 708 proteins

with at least 1 instrument available which were carried forward for the downstream analysis (Figure S1). The Wald ratio method was applied to proteins with 1 single SNP as the instrument while the inverse-variance weighted (IVW) model was applied to proteins which were instrumented by 2 or more SNPs (23). To test for horizontal pleiotropy and to ensure robustness of the proteome-wide MR findings, we applied MR-Egger and Weighted median to the significant associations where at least 3 instruments were available (31, 32). For multiple testing correction, a false discovery rate (FDR) was calculated by applying the Benjamini-Hochberg (BH) adjustment to each lipid fraction (33) and an FDR < 0.05 was defined to be significant.

To avoid bias due to sample overlap between UKBPPP and GLGC, we also performed MR using GLGC data excluding UKBB participants. A correlation analysis was performed to compare the beta estimates derived from GLGC data with or without UKBB individuals.

Bayesian colocalization: A Bayesian colocalization analysis was conducted on all protein-lipid associations with FDR-corrected P < 0.05 to determine if they shared the same causal variant (34). The colocalization enabled us to minimise horizontal pleiotropy caused by linkage disequilibrium (LD) where the plasma protein levels and lipid traits were influenced by 2 distinct variants in LD with each other (35). Colocalization was performed on the same window as the previous proteome-wide MR (within +/- 500KB of the cognate gene), with rare variants (MAF < 0.05) residing in the window dropped. Default priors as described in the original paper were applied (34). A posterior probability of colocalization (PPH4) ≥ 80% indicated strong colocalization, while 60% ≤ PPH4 < 80% was considered suggestive evidence for colocalization.

MR generalized to correlated instruments (GMR): Due to stringent genetic instrument selection, most proteins had few pQTLs, which could bias the MR estimates due to unknown pleiotropy and make MR-Egger and WM unapplicable. Therefore, to improve robustness, we included additional instruments at $p < 1 \times 10^{-6}$ and clumped them with $r^2 = 0.4$ (36). To account for correlation between instruments, a generalized inverse variance weighted regression (gIVW) was applied as the primary method (37). Where applicable, we also performed the MR-Egger generalized to correlated variants (gEgger) (38) and weighted median (32).

Causal effects of lipid-associated proteins on CAD and stroke

A *cis*-MR analysis was conducted with lipid-associated plasma proteins (P-FDR< 0.05) as exposures and CAD and stroke (any stroke, any ischemic stroke, large artery stroke, cardioembolic stroke, and small vessel stroke) as outcomes. An FDR correction was applied separately for each outcome while colocalization and supplemental MR methods (gIVW, gEgger, and weighted median) were performed to validate the genetically proxied associations surviving the FDR correction (32-34, 37, 38). The instrument selection criteria and parameters for MR and colocalization in this step were set as described above.

Additionally, univariate MR was performed to assess whether the lipids measures show similar causal effects on CVDs in South Asians compare to Europeans. Genetic instruments for lipid fractions were extracted from the whole genome (autosomes only), and other MR criteria and methods were the same as those applied in *cis*-MR (29, 30). Where applicable, MR-Egger and weighted median were applied as well (31, 32).

Furthermore, for plasma proteins associated with both lipids and CVDs, we evaluated with multi-trait colocalization whether the 3 traits have a shared causal variant in the corresponding gene region (39). The colocalization analysis was applied on plasma protein, lipid fractions, and the CVD on the genomic region +/- 500KB extended from the cognate gene (39). Rare variants with MAF < 0.05 were dropped. Priors for multi-trait colocalization were set to default values: the probability of any SNP within the colocalization window being exclusively associated with one of the three traits was 1×10^{-6}, with two traits was 1×10^{-7}, and with all three traits was 1×10^{-8} (39).

Reverse MR with lipid fractions as exposures and plasma proteins as outcomes in SAS

Furthermore, to understand the potential causal effects of dyslipidaemia on plasma protein abundance, we conducted reverse MR using lipid fractions (exposure) and plasma proteins (outcomes). Genetic instruments were selected from the 22 autosomes using the filtering criteria as previously described.

Comparison with European population

For plasma proteins with potential effects on lipid fractions in SAS, we estimate their effects in EUR using a two-sample MR approach, applying the same genetic instrument criteria used in SAS. We checked the consistency of the MR estimates and applied a correlation analysis between EUR and SAS population estimates for proteins that were consistent and significant

in both groups. We also compared the 95% confidence intervals, defining a significant difference in the genetically proxied MR estimates when the intervals did not overlap.

Observational associations of protein levels with lipid fractions

For plasma proteins associated to LDL-C and HDL-C with all the MR, strong colocalization, and GMR evidence, we conducted observational analysis using linear regression for each protein and its associated lipid fraction. Each regression model was adjusted for age, sex, Townsend deprivation, BMI, HbA1c, cholesterol medication, smoking, systolic blood pressure, blood pressure medication, and ethnicity. To evaluate potential effect modification by ancestry, we included an interaction term between a binary variable for South Asian ancestry and protein level. All continuous variables and lipid fractions were standardised before modelling and p values were adjusted for multiple testing using FDR at 5%.

Result

An overview of the study design is shown in **Figure 1**.

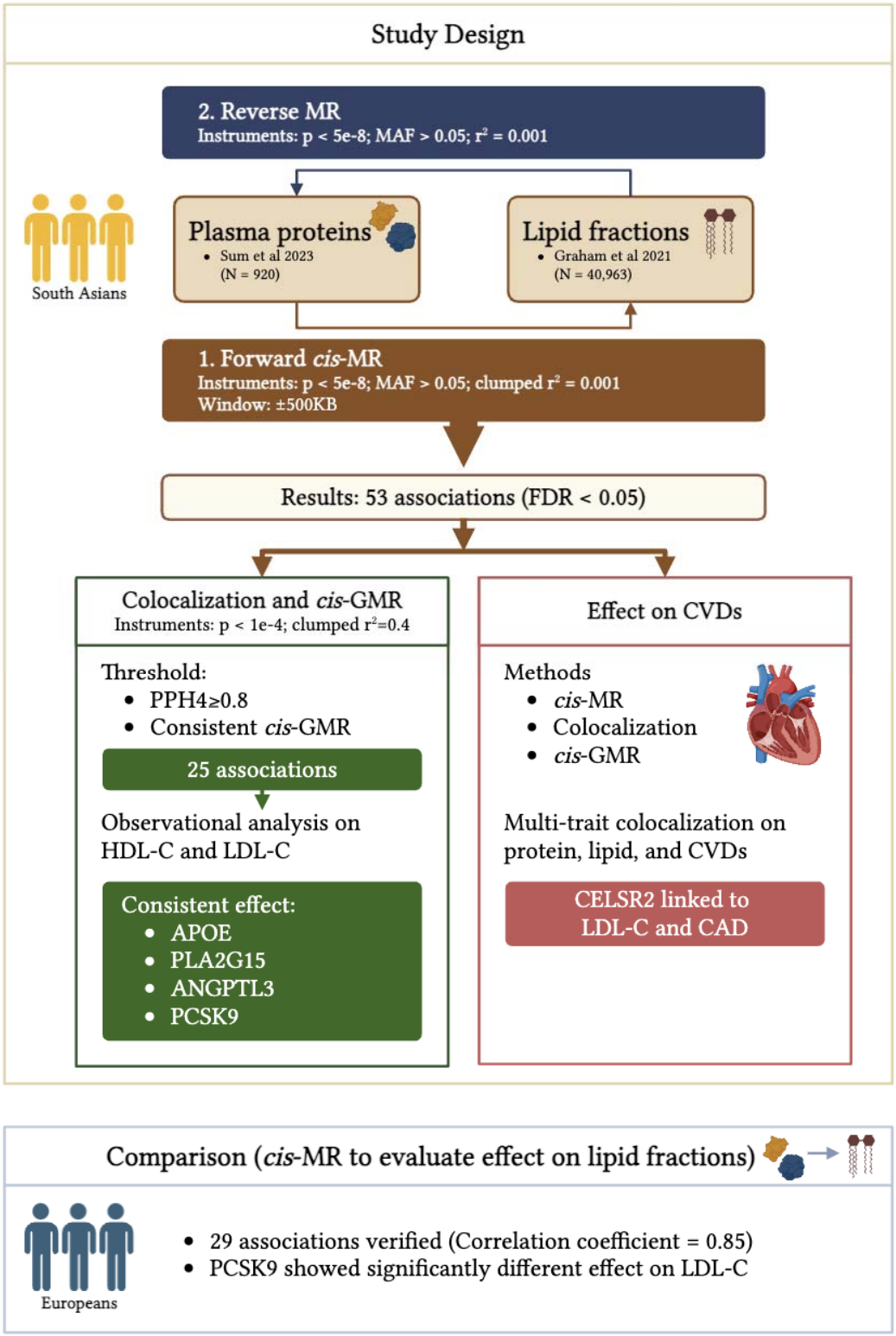

Figure 1. Overview of the study design. MR; Mendelian randomization, FDR; false discovery rate, GMR; generalized MR, PPH4; posterior probability of hypothesis 4, HDL-C and LDL-C; high- and low-density lipoprotein cholesterol

Proteome-wide MR identified 29 plasma proteins associated with lipid fractions in SAS

Excluding proteins with ambiguous cognate genes and those encoded by genes located on the X-chromosome or within the *MHC* region (**Table S1**), 2,800 proteins were included in our study. Of these, 708 had at least one genetic instrument variable available (**Figure S1**) and were carried forward for the proteome-wide MR study. Using the Wald ratio (IV=1) or IVW (IVs≥2) approach as the primary MR method, 186 plasma proteins showed a potential causal effect on at least one of the lipid fractions (309 associations in total; $P < 0.05$; **Table S2**). After adjusting for multiple testing, a total of 29 genetically proxied plasma proteins showed potential causal effect on at least one of the 5 lipid fractions (53 associations in total; FDR < 0.05, **Figure 2, Table S2**).

To avoid bias due to sample overlap between UKBPPP and GLGC, we performed MR using GLGC data without UKBB participants and the MR estimates were largely consistent across GLGC with and without UKBB (Pearson $r^2 = 0.93$, $P< 0.001$; **Figure S2A, Figure S2B**).

Among these 29 lipids-associated proteins (53 associations), 12 (25 associations) were supported by strong colocalization evidence (PPH4 > 80%) with one of the lipids fractions including ANGPTL3, APOE, CELSR2, EPPK1, GAS6, GSTA1, GSTA3, HGFAC, LPA, PCSK9, PLA2G15, and PLTP (**Table S3**). Particularly, we identified strong colocalization of ANGPTL3 with all lipid fractions except HDL, CELSR2 with all lipid fractions except TG, LPA with LDL-C and TC, and PCSK9 with HDL, TC, and non-HDL. Additionally, 7 plasma proteins (7 associations in total) showed suggestive colocalization evidence (80% > PPH4 ≥ 60%) with the tested lipid fractions (**Figure 2**, **Figure S3, Table S3**).

To further validate our findings in the proteome-wide cis-MR, we included moderately correlated SNPs ($r^2 < 0.4$) that are adequately associated with the plasma proteins ($p < 1 \times 10^{-4}$) and applied gIVW (36). Out of the 53 associations of lipid fractions with plasma proteins, gIVW produced consistent estimates with FDR corrected *P* value < 0.05 for 43 associations. Subsequently, gEgger and Weighted median were applied to the 42 associations with≥ 3 SNP instruments. Weighted median produced highly consistent estimates for all 42 associations. The gEgger detected no horizontal pleiotropy (FDR corrected *P* for intercept < 0.05) but

derived inconsistent estimates for the association of TNFSF13 with TC and nonHDL-C, THOP1 with HDL-C, and EPPK1 with TC (**Figure 2**, **Table S4**).

Altogether, 30 associations of 14 genetically proxied proteins with lipid fractions were identified by proteome-wide MR, further validated by colocalization (either strong or suggestive) and the subsequent MR analysis generalized to correlated instruments (GMR). The top findings include CELSR2 associated with all lipid fractions except TG, PCSK9 and HGFAC with TC, LDL-C, and non-HDL-C, LPA with LDL-C and TC, and ANGPTL3 with all lipid fractions (**Figure 2**).

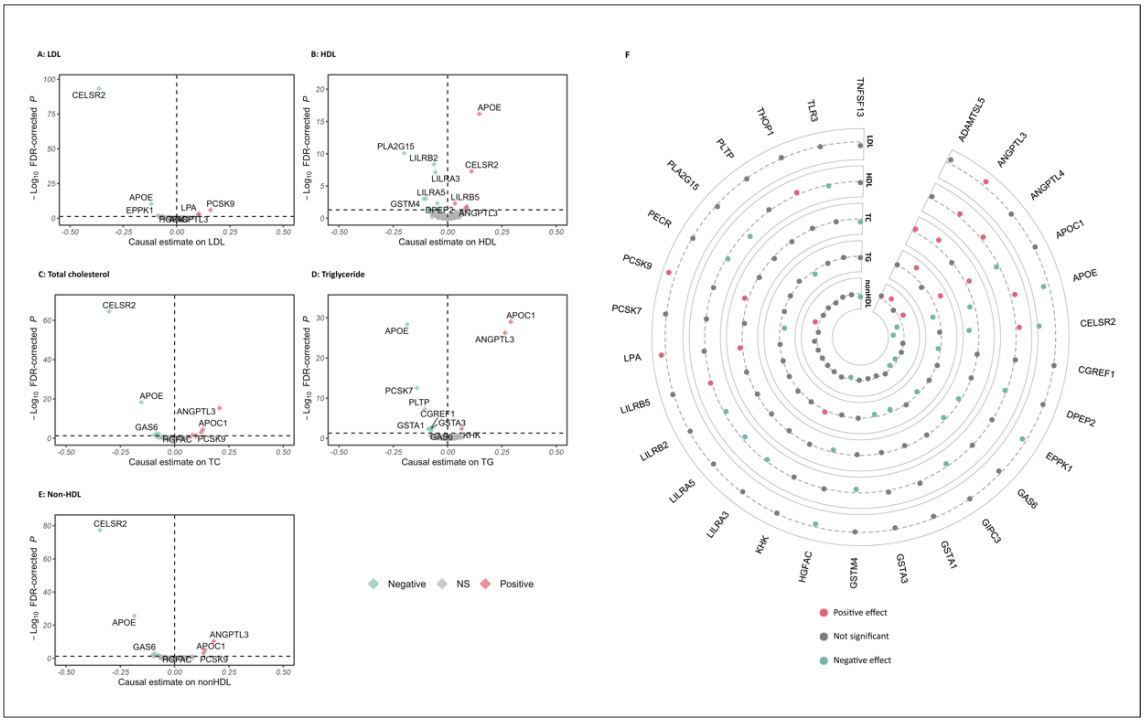

Figure 2. Volcano plots showing the causal effect of circulating plasma protein on the 5 lipid traits A) LDL-C; B) HDL-C; C) TC; D) TG; E) non-HDL-C. Each dot indicates a plasma protein with the x-axis showing the Wald ratio or IVW while the y-axis showing -log$_{10}$ FDR-corrected P from the MR analysis. F) Circular plot showing the overlap of plasma proteins with the 5 lipid traits tested.

Potential causal effects of lipid-associated proteins on CVD outcomes

Subsequently, we investigated whether lipid-associated proteins (N=29) identified by proteome-wide MR have potential causal effects on risk of CAD and stroke in SAS. After correcting for multiple testing, only genetically predicted CELSR2 had a causal association with CAD (Odds ratio (OR) = 0.64, 95% Confidence Interval (CI) = (0.50, 0.81), FDR = 0.003; **Figure 3A, Table S5**) which was also supported by strong colocalization evidence (PPH4 = 93.6%, **Figure 3B**, **Table S6**). Additionally, gIVW, gEgger, and weighted median

produced consistent estimates and no pleiotropy was detected (**Figure 3C**, **Table S7**). Notably, ANGPTL3 and LPA showed suggestive associations with CVDs (P< 0.05) but did not pass the 5% FDR threshold (**Figure 3A**).

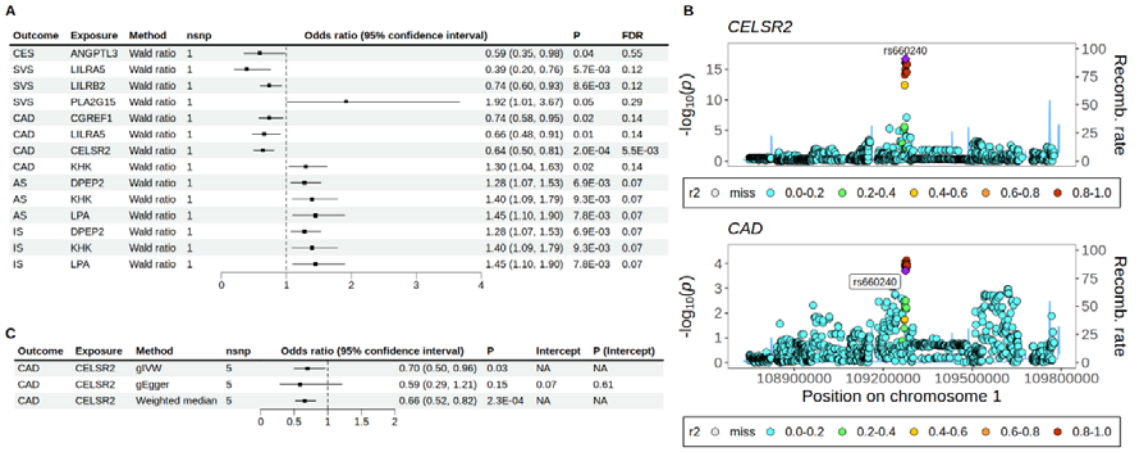

Figure 3. Effect of lipid-associated plasma proteins on CVDs A) Forest plot showing all plasma proteins that were nominally associated with CVDs (P< 0.05). B) Stacked genomic plot showing evidence for colocalization between CELSR2 and CAD; and C) Forest plot showing the effect of CELSR2 on CAD estimated by gIVW, gEgger, and weighted median.

Since the efficacy of lipid-associated proteins may be prioritized depending on the association of lipid fractions with CVDs, we assessed the causal effects of five lipid fractions on CVDs to indicate proteins more promising to CVD treatment. Our univariable MR identified four associations reaching nominal significance (P < 0.05), including LDL-C with CAD (β_{IVW} = 1.64; 95% CI = (1.03, 2.62) and cardioembolic stroke (β_{IVW} = 1.78; 95% CI = (1.01, 3.15); **Table S8**). The MR-Egger and weighted median methods produced estimates consistent in direction with the inverse variance weighted method (P< 0.05).

Since CELSR2 showed potential causal effects on both LDL-C and CAD while genetically proxied LDL-C was also causally associated with CAD risk, a multi-trait colocalization was performed on the 3 traits in the genomic region +/- 500KB extended from the CELSR2 gene. The multi-trait colocalization produced a posterior probability of 70.0% that CELSR2, LDL-C, and CAD colocalized in this region (**Figure 4**, **Table S9**). The posterior probabilities for all 15 scenarios of multi-trait colocalization were presented in **Table S9**.

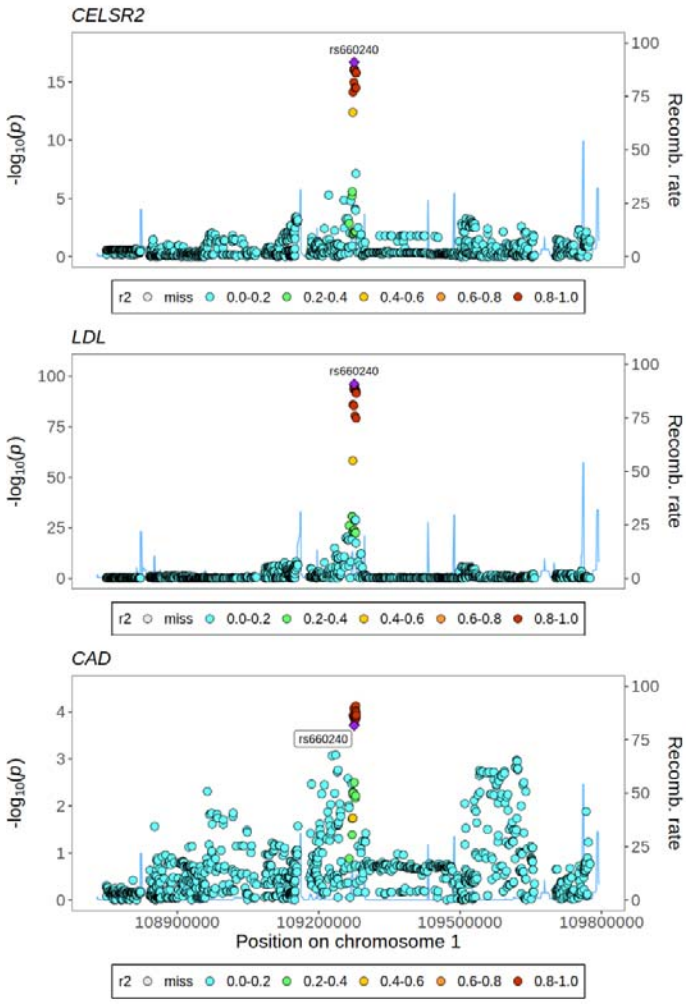

Figure 4. Stacked regional genomic plot from multi-trait colocalization showing the colocalized genetic variant rs660240 across LDL-C, CAD at the CELSR2 locus in SAS.

Reverse MR with lipid fractions as exposures and plasma proteins as outcomes

To understand the plasma proteins modified by lipid fractions, a reverse MR was performed using lipid fractions as exposures and plasma proteins as outcomes. This analysis identified genetically proxied associations between TG and LDLR, and HDL-C with 3 proteins: APOA1, MENT, and FGFBP2 (β_{range} = 0.55 to 0.61, FDR-corrected P< 0.05**, Table S10, Figure S4**). Subsequent MR-Egger and weighted median produced consistent estimates and detected no horizontal pleiotropy (**Table S11, Figure S5**).

Comparison with EUR population

Among the 53 proteome-wide MR identified associations in SAS, 29 of them had consistent beta estimates in EUR (FDR-corrected P < 0.05). The correlation analysis on the 30 beta estimates in SAS against those in EUR produced a correlation coefficient of 0.85 **(***P* = 6.5 ×10 $^{-9}$, **Figure S6).** Out of the 30 associations in SAS with MR, colocalization, and GMR evidence, 22 were verified in EUR population (**Table S12** and **Figure 5**). Proteome-wide MR, colocalization, and GMR identified 6 proteins associated with HDL-C in SAS, but only ANGPTL3 and APOE were also significant in EUR (**Figure 5**). Of the 6 proteins linked to LDL-C in SAS, ANGPTL3, CELSR2, LPA, and PCSK9 were verified in EUR, with PCSK9 showing a stronger effect in EUR (β_{EUR} = 0.37; 95% CI = (0.36, 0.38), β_{SAS} = 0.16; 95% CI = (0.11, 0.21); **Figure 5**). Among the remaining 18 associations with nonHDL-C, TC, or TG, only the association of EPPK1 with nonHDL-C and HGFAC with TC were not significant in EUR.

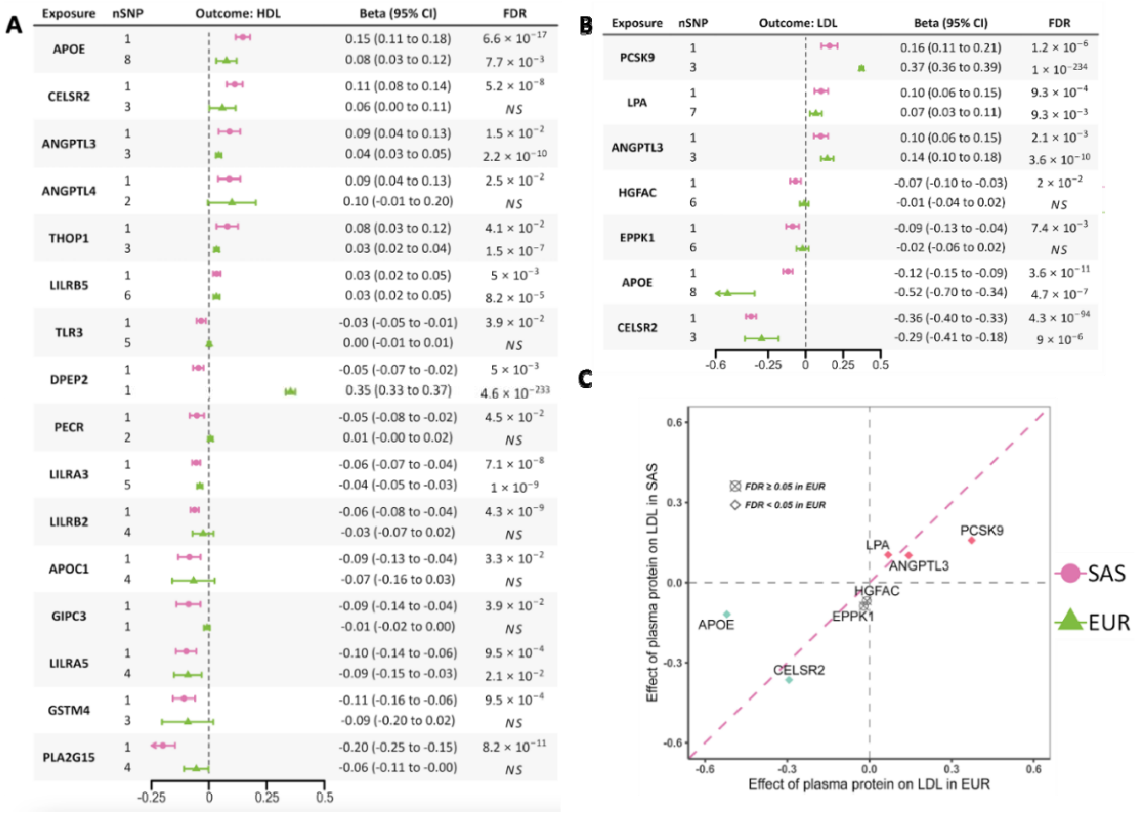

Figure 5. Effect of plasma proteins on lipid fractions using cis-MR in SAS and EUR on A) HDL-C; B) LDL-C; and C) Scatter plot for comparison of causal effect estimates from MR between EUR and SAS for LDL-C.

Observational associations of protein levels with lipid fractions

There were 3 proteins with strong evidence (MR, strong colocalization, and GMR) of association HDL-C (APOE, CELSR2, and PLA2G15) and 6 proteins with LDL-C (ANGPTL3, CELSR2, EPPK1, HGFAC, LPA, and PCSK9) that were tested in observational analysis. After adjusting for multiple testing, all 9 proteins had a significant association with their genetically associated lipid fraction (FDR adjusted $P < 0.05$) but only 4 had a consistent direction of effect with MR estimates **(Table S13)**; APOE ($\beta = 0.078$; 95% CI = (0.075, 0.082)) and PLA2G15 ($\beta = -0.067$; 95% CI = (-0.071, -0.063)) with HDL, and ANGPTL3 ($\beta = 0.180$; 95% CI = (0.175, 0.184)) and PCSK9 ($\beta = 0.196$; 95% CI = (0.192, 0.200)) with LDL. Out of these four proteins, there was a significant interaction between ANGPTL3 and PCSK9 levels with LDL-Cin South Asians with South Asians having a significantly lower effect compared to other ancestries (ANGPLT3*South Asian; $\beta = -0.072$; 95% CI = (-0.103, -0.040), PCSK9*South Asian; $\beta = -0.140$; 95% CI = (-0.174, -0.107)).

Discussion

Key findings in this study

Here, we performed a bidirectional proteome-wide MR on five lipid fractions, and linked lipid-related proteins to cardiovascular outcomes in SAS. Our study confirmed key proteins (PCSK9, ANGPTL3, LPA), identified novel targets (GSTA1, GSTA3, EPPK1, PECR, and PLA2G12), and strengthened evidence for CELSR2 and GAS6 in dyslipidaemia. Notably, our results highlight significant heterogeneity in MR estimates across genetic ancestry groups, particularly for the effect of PCSK9 on LDL-C. We also report CELSR2 with evidence for its effect on LDL-C and CAD risk. Reverse MR identified LDLR as modifiable by TG, and APOA1, MENT, and FGFBP2 by HDL-C.

Enhanced role of CELSR2 and GAS6 in lipid metabolism and cardiovascular outcomes

We found an inverse association of genetically proxied CELSR2 with LDL-C and CAD risk in SAS. CELSR2 is a transmembrane protein belonging to the flamingo family of cadherin superfamily (40). Although the biological function of CELSR2 is not well understood, the role of CELSR2 in lipid metabolism was indicated by some previous studies. A locus in the vicinity of CELSR2, rs599839 (in LD with rs660240, the instrument of CELSR2 in this study (r^2 = 0.989 in SAS, r^2 = 0.871 in EUR)), was first reportedly associated with CAD, LDL-C and TC by 2 European ancestral GWAS (41-43). rs660240, a 3' UTR variant, is an eQTL for *CELSR2*, *PSRC1*, and *SORT1* in liver tissue (Open Target Genetics). It shows slight variations in allele frequencies between South Asians, Europeans, and East Asians, which could have implications for studies related to disease susceptibility and treatment response. Furthermore, a transcriptomic study revealed the risk allele of rs599839 to CAD and high LDL-C also suppressed the expression of *CELSR2* gene in liver (44). Extending to non-European populations, the association of *CELSR2* variants with lipid fractions and CAD risk was also verified in the South Asian population (45, 46). However, although the effect of CELRS2 on lipid metabolism was indicated by genetic and transcriptomic studies, the underlying mechanism is less clear. One study demonstrated that CELSR2 deficiency can elevate reactive oxygen species of hepatocytes, which impairs lipid homeostasis and physiological unfolded protein response(47). In conclusion, our result is consistent with the significant role of CELSR2 in CAD and lipid metabolism suggested in earlier studies. Our study also extends the generalizability of CELSR2's function to the South Asian population.

We identified an inverse association of genetically proxied GAS6 with TC and TG. GAS6 is a ligand for TAM receptor protein tyrosine kinases including AXL, TYRO3 and MER. The GAS6 - TAM pathway was found implicated in carcinoma, inflammation, and haemostasis and has been targeted for the treatment of carcinoma (48-50). In recent years, there is also evidence that GAS6 plays a role in regulating obesity and lipid metabolism (51, 52) as plasma gamma-glutamyl carboxylated GAS6 (Gla-GAS6) was found significantly lower in hyperlipidaemic individuals compared with healthy controls (51). The subsequent experiment showed that higher Gla-GAS6 expression induced by vitamin K in plasma and hepatocyte could reduce the plasma lipid level in hyperlipidaemic mice (51). The Gla-GAS6 takes effect by regulating the AMPK/SREBP1/PPARα signalling pathways of hepatic lipid metabolism (51). Our study supports this hypothesis by providing genetic evidence for GAS6 as a potential regulator of lipid metabolism.

Known targets and novel findings in lipid metabolism

In this study, we replicated previously established protein associations with lipid metabolism, including ANGPTL3, APOE, LPA, PCSK9, and PLTP, some of which are drug targets for dyslipidemia treatment (53). However, despite the concordance in direction and significance for the association of PCSK9 on LDL-C in both SAS and EUR, we identified a significantly reduced effect in SAS as supported by both MR and observational analysis. Since drug responses to lipid-lowering therapy can vary across ethnicities, this finding may have important clinical implications. Previous studies suggest that atorvastatin and simvastatin have similar lipid-lowering effects in SAS patients compared to those in EUR (54). Therefore, PCSK9, as a novel target for lipid-lowering medication, warrants further investigation for their effect in more diverse ethnicity. Mechanisms accounting for the ancestral heterogeneity in PCSK9's effect are limited. However, a recent study sequencing PCSK9 gene in Indians indicated difference in prevalence of mutation in SAS. The by-ancestry heterogeneous mutation pattern can result in heterogeneity of PCSK9 structure and activity, which may modify the effect of PCSK9 abundance (55).

Additionally, our study identified novel associations, including HGFAC, GIPC3, EPPK1, GSTA1, GSTA3, PECR, and PLA2G12. HGFAC activates hepatocyte growth factor (HGF) by converting it to a heterodimer, which then binds to the MET receptor to activate downstream signalling. A previous study linked a putative HGFAC loss-of-function variant to elevated serum TG and LDL-C (56). In another animal experiment, an increase of circulating TG was present in both male and female HGFAC-KO mice while higher level of

circulating TC was present in male HGFAC-KO mice (57). The association of GIPC3 with HDL-C was firstly identified in our study. However, a previous study reported that GIPC1, another subtype of GIPC family, is involved in lipid metabolism by regulating the SR-B1 expression (58). Lastly, we also identified GSTA1, GSTA3, EPPK1, PECR, and PLA2G12 associated with lipid traits, all of which are novel findings.

Reverse effect of lipid fractions on plasma proteins

Conversely, we applied MR to identify plasma proteins modified by lipid fractions. We found that genetically proxied TG levels were associated with increased plasma LDLR. LDLR is a cell membrane glycoprotein that regulates lipid homeostasis by binding and internalizing circulating cholesterol-containing lipoprotein particles, including LDL-C, VLDL-C, and chylomicron remnants (59). Deficiency in LDLR can result in dyslipidaemia. Therefore, we interpreted that the LDLR upregulation triggered by genetically proxied TG is likely to reverse the hyperlipidaemia. In addition, we observed genetically proxied HDL-C level associated with increased plasma APOA1, MENT, and FGFBP2 level. All 3 proteins play important roles in various disorders and mechanisms, including cholesterol transport, angiogenesis, tissue repair, and cellular metabolism. Therefore, further investigation is necessary for detailed biological interpretation of these findings.

Strength and Limitations

Our study has several strengths. First, to the best of our knowledge, this is the first MR study to systematically evaluate the potential causal association between plasma proteins and lipid traits in South Asians, replicating known protein-stroke associations and discovering novel targets. Secondly, we combined MR with colocalization, to reduce bias from LD and reverse causation — a potential limitation of conventional observational studies. Finally, by incorporating GWAS on CVDs, we linked lipid associated proteins to cardiovascular outcomes and identified CELSR2 as a promising target for both LDL-C and CAD.

There are also limitations in our study. First, the genetic associations for plasma proteins in SAS were based on a relatively small sample size, compared with current GWAS standards, so our findings should be interpreted cautiously, and larger GWAS are needed for validation. Second, despite using various MR methods and sensitivity analyses, unaccounted pleiotropy may still bias our results. Third, our analysis assumes the absence of SNP-SNP and

SNP□environment interaction. Lastly, our findings may not be generalizable to ancestry groups other than the ones studied here.

Conclusion

Our comprehensive study triangulated evidence from MR, colocalization, and observational analyses, highlighting several novel proteins associated with lipid fractions in South Asians. Notably, our analysis suggests that the causal effect of PCSK9 on LDL-C may be ancestry-specific. Future studies with larger sample sizes are needed to validate our findings, along with further mechanistic and clinical studies to confirm the role of PCSK9 on LDL-C in South Asians and Europeans.

Contributors: DM, IT and AD contributed to the conception and design of the study. SW and AS contributed to acquisition and statistical analysis of data. DM, IT, AD supervised the project. All authors contributed to the drafting and critical revision of the manuscript.

Acknowledgments: The authors would like to thank the staff and participants who contributed to the UK Biobank, UKBPPP, GLGC, the East London Genes & Health (ELGH) study, the CARDIoGRAMplusC4D consortium, GIGASTROKE consortium for making the resources available.

Funding: This work is supported by the British Heart Foundation Research Excellence Award (4) (RE/24/130023). IT and AR are supported by NIH R01 HL162300 02.

Conflict of interest: A.R. reported consultancies for Raydel, Johnson & Johnson, and JP Morgan, all with modest compensation, and received a significant research grant from CSL Limited but with no salary support included. The remaining authors have no disclosures to report.

Data Availability: The UKBPPP data can be downloaded from http://ukb-ppp.gwas.eu. GLGC dataset can be downloaded from https://csg.sph.umich.edu/willer/public/glgc-lipids2021/results/ancestry_specific/. ELGH data can be downloaded from https://www.genesandhealth.org/research/scientific-data-downloads/gwas-data-genes-health-feb-2020-datafreeze. GWAS summary statistics for stroke and its subtypes for South Asians and Europeans can be obtained from the GWAS catalogue with accession GCST90104559-GCST90104563, and GCST90104539-GCST90104543, respectively.

References

1. Reiner Ž. Hypertriglyceridaemia and risk of coronary artery disease. Nature Reviews Cardiology. 2017;14(7):401-11.
2. Yaghi S, Elkind MSV. Lipids and Cerebrovascular Disease. Stroke. 2015;46(11):3322-8.
3. Holmes MV, Millwood IY, Kartsonaki C, Hill MR, Bennett DA, Boxall R, et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. Journal of the American College of Cardiology. 2018;71(6):620-32.
4. Velagaleti RS, Massaro J, Vasan RS, Robins SJ, Kannel WB, Levy D. Relations of Lipid Concentrations to Heart Failure Incidence. Circulation. 2009;120(23):2345-51.
5. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1151-210.
6. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. European Heart Journal. 2003;24(11):987-1003.
7. Lloyd-Jones DM, Braun LT, Ndumele CE, Smith SC, Sperling LS, Virani SS, et al. Use of Risk Assessment Tools to Guide Decision-Making in the Primary Prevention of Atherosclerotic Cardiovascular Disease: A Special Report From the American Heart Association and American College of Cardiology. Journal of the American College of Cardiology. 2019;73(24):3153-67.
8. Brewer HB. High-Density Lipoproteins: A New Potential Therapeutic Target for the Prevention of Cardiovascular Disease. Arteriosclerosis, Thrombosis, and Vascular Biology. 2004;24(3):387-91.
9. Drexel H, Tamargo J, Kaski JC, Lewis BS, Saely CH, Fraunberger P, et al. Triglycerides revisited: is hypertriglyceridaemia a necessary therapeutic target in cardiovascular disease? European Heart Journal - Cardiovascular Pharmacotherapy. 2023;9(6):570-82.
10. Boekholdt SM, Arsenault BJ, Mora S, Pedersen TR, LaRosa JC, Nestel PJ, et al. Association of LDL Cholesterol, Non–HDL Cholesterol, and Apolipoprotein B Levels With Risk of Cardiovascular Events Among Patients Treated With Statins: A Meta-analysis. JAMA. 2012;307(12):1302-9.
11. Pischon T, Girman CJ, Sacks FM, Rifai N, Stampfer MJ, Rimm EB. Non–High-Density Lipoprotein Cholesterol and Apolipoprotein B in the Prediction of Coronary Heart Disease in Men. Circulation. 2005;112(22):3375-83.
12. Frank ATH, Zhao B, Jose PO, Azar KMJ, Fortmann SP, Palaniappan LP. Racial/Ethnic Differences in Dyslipidemia Patterns. Circulation. 2014;129(5):570-9.
13. Anand SS, Yusuf S, Vuksan V, Devanesen S, Teo KK, Montague PA, et al. Differences in risk factors, atherosclerosis, and cardiovascular disease between ethnic groups in Canada: the Study of Health Assessment and Risk in Ethnic groups (SHARE). Lancet. 2000;356(9226):279-84.
14. Bhopal R, Unwin N, White M, Yallop J, Walker L, Alberti KGMM, et al. Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations: cross sectional study. BMJ. 1999;319(7204):215-20.
15. Wasay M, Khatri IA, Kaul S. Stroke in South Asian countries. Nat Rev Neurol. 2014;10(3):135-43.

16. Kanaya AM, Herrington D, Vittinghoff E, Ewing SK, Liu K, Blaha MJ, et al. Understanding the High Prevalence of Diabetes in U.S. South Asians Compared With Four Racial/Ethnic Groups: The MASALA and MESA Studies. Diabetes Care. 2014;37(6):1621-8.
17. Hegele RA, Tsimikas S. Lipid-Lowering Agents. Circulation Research. 2019;124(3):386-404.
18. De Oliveira-Gomes D, Joshi PH, Peterson ED, Rohatgi A, Khera A, Navar AM. Apolipoprotein B: Bridging the Gap Between Evidence and Clinical Practice. Circulation. 2024;150(1):62-79.
19. Figarska SM, Gustafsson S, Sundström J, Ärnlöv J, Mälarstig A, Elmståhl S, et al. Associations of Circulating Protein Levels With Lipid Fractions in the General Population. Arteriosclerosis, Thrombosis, and Vascular Biology. 2018;38(10):2505-18.
20. Wu S, Meena D, Yarmolinsky J, Gill D, Smith A, Dib MJ, et al. Mendelian Randomization and Bayesian Colocalization Analysis Implicate Glycoprotein VI as a Potential Drug Target for Cardioembolic Stroke in South Asian Populations. Journal of the American Heart Association. 2024;13(16):e035008.
21. Graham SE, Clarke SL, Wu KH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. Nature. 2021;600(7890):675-9.
22. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622(7982):329-38.
23. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. Genetic Epidemiology. 2013;37(7):658-65.
24. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008;27(8):1133-63.
25. Kim MS, Song M, Kim B, Shim I, Kim DS, Natarajan P, et al. Prioritization of therapeutic targets for dyslipidemia using integrative multi-omics and multi-trait analysis. Cell Reports Medicine. 2023;4(9).
26. Finer S, Martin HC, Khan A, Hunt KA, MacLaughlin B, Ahmed Z, et al. Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. International Journal of Epidemiology. 2019;49(1):20-1i.
27. 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(10):1121-30.
28. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. Nature. 2022;611(7934):115-23.
29. Pierce BL, Ahsan H, VanderWeele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. International Journal of Epidemiology. 2011;40(3):740-52.
30. Li J, Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLOS Genetics. 2017;13(11).

31. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology. 2015;44(2):512-25.
32. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic Epidemiology. 2016;40(4):304-14.
33. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289-300.
34. Williams SM, Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genetics. 2014;10(5).
35. Zuber V, Grinberg NF, Gill D, Manipur I, Slob EAW, Patel A, et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. The American Journal of Human Genetics. 2022;109(5):767-82.
36. Gordillo-Marañón M, Zwierzyna M, Charoen P, Drenos F, Chopade S, Shah T, et al. Validation of lipid-related therapeutic targets for coronary heart disease prevention using human genetics. Nature Communications. 2021;12(1).
37. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Statistics in Medicine. 2015;35(11):1880-906.
38. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. European Journal of Epidemiology. 2017;32(5):377-89.
39. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics. 2018;34(15):2538-45.
40. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
41. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nature Genetics. 2008;40(2):189-97.
42. Samani NJ, Braund PS, Erdmann J, Götz A, Tomaszewski M, Linsel-Nitschke P, et al. The novel genetic variant predisposing to coronary artery disease in the region of the PSRC1 and CELSR2 genes on chromosome 1 associates with serum cholesterol. J Mol Med (Berl). 2008;86(11):1233-41.
43. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007;357(5):443-53.
44. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. 2008;6(5):e107.
45. Saleheen D, Soranzo N, Rasheed A, Scharnagl H, Gwilliam R, Alexander M, et al. Genetic determinants of major blood lipids in Pakistanis compared with Europeans. Circ Cardiovasc Genet. 2010;3(4):348-57.
46. Walia GK, Gupta V, Aggarwal A, Asghar M, Dudbridge F, Timpson N, et al. Association of Common Genetic Variants with Lipid Traits in the Indian Population. PLOS ONE. 2014;9(7):e101688.

47. Tan J, Che Y, Liu Y, Hu J, Wang W, Hu L, et al. CELSR2 deficiency suppresses lipid accumulation in hepatocyte by impairing the UPR and elevating ROS level. Faseb j. 2021;35(10):e21908.
48. Linger RM, Keating AK, Earp HS, Graham DK. TAM receptor tyrosine kinases: biologic functions, signaling, and potential therapeutic targeting in human cancer. Adv Cancer Res. 2008;100:35-83.
49. van der Meer JH, van der Poll T, van 't Veer C. TAM receptors, Gas6, and protein S: roles in inflammation and hemostasis. Blood. 2014;123(16):2460-9.
50. Wu G, Ma Z, Hu W, Wang D, Gong B, Fan C, et al. Molecular insights of Gas6/TAM in cancer development and therapy. Cell Death Dis. 2017;8(3):e2700.
51. Bordoloi J, Ozah D, Bora T, Kalita J, Manna P. Gamma-glutamyl carboxylated Gas6 mediates the beneficial effect of vitamin K on lowering hyperlipidemia via regulating the AMPK/SREBP1/PPARα signaling cascade of lipid metabolism. J Nutr Biochem. 2019;70:174-84.
52. Wu KS, Hung YJ, Lee CH, Hsiao FC, Hsieh PS. The Involvement of GAS6 Signaling in the Development of Obesity and Associated Inflammation. Int J Endocrinol. 2015;2015:202513.
53. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. Nucleic Acids Res. 2023;51(D1):D1353-d9.
54. Gupta M, Braga MF, Teoh H, Tsigoulis M, Verma S. Statin effects on LDL and HDL cholesterol in South Asian and white populations. J Clin Pharmacol. 2009;49(7):831-7.
55. Reddy LL, Shah SAV, Ponde CK, Dalal JJ, Jatale RG, Dalal RJ, et al. Screening of PCSK9 and LDLR genetic variants in Familial Hypercholesterolemia (FH) patients in India. Journal of Human Genetics. 2021;66(10):983-93.
56. (AMP) AMP. Common Metabolic Diseases Knowledge Portal: rs3748034 [Available from: https://hugeamp.org/variant.html?variant=4%3A3446091%3AG%3AT.
57. Sargsyan A, Doridot L, Hannou SA, Tong W, Srinivasan H, Ivison R, et al. HGFAC is a ChREBP-regulated hepatokine that enhances glucose and lipid homeostasis. JCI Insight. 2023;8(1).
58. Zhang Z, Zhou Q, Liu R, Liu L, Shen WJ, Azhar S, et al. The adaptor protein GIPC1 stabilizes the scavenger receptor SR-B1 and increases its cholesterol uptake. J Biol Chem. 2021;296:100616.
59. Go GW, Mani A. Low-density lipoprotein receptor (LDLR) family orchestrates cholesterol homeostasis. Yale J Biol Med. 2012;85(1):19-28.