Universiteit Leiden
The Netherlands

# Metabolic risk and cardiovascular disease: insights from large biobanks with genetic epidemiological approaches

Ao, L.

**Citation**
Ao, L. (2026, January 13). *Metabolic risk and cardiovascular disease: insights from large biobanks with genetic epidemiological approaches*.

| | |
|---|---|
| Version: | Publisher's Version |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | |

**Note:** To cite this publication please use the final published version (if applicable).

# CHAPTER 2

## Potential causal evidence for an ApoB-independent and HDL-related risk profile associated with coronary artery disease

Linjun Ao, Diana van Heemst, J. Wouter Jukema, Patrick C.N. Rensen,
Ko Willems van Dijk, Raymond Noordam

***Journal of Lipid Research.*** 2025;66(4):100778
The supplemental information for this paper is available online at
https://drive.google.com/drive/folders/1F2qSC68lFi9Vwfx5XvhF3fzihKZecBqZ?usp=drive_link

# Abstract

**Background:** Plasma 1-H nuclear magnetic resonance ($^{1}$H-NMR) metabolomic measures have yielded significant insight into the pathophysiology of cardiometabolic disease, but their inter-related nature complicates causal inference and clinical interpretation. This study aimed to investigate the associations of unrelated $^{1}$H-NMR metabolomic profiles with coronary artery disease (CAD) and ischemic stroke (IS).

**Methods:** Principal component analysis was performed on 168 $^{1}$H-NMR metabolomic measures in 56,712 unrelated European participants from UK Biobank to retrieve uncorrelated principal components (PCs), which were used in Cox-proportional hazard models. For each outcome, two-sample Mendelian randomization (MR) analyses were then conducted based on three non-overlapping databases, followed by a meta-analysis.

**Results:** The first six PCs collectively explaining 88% of the total variance were identified. For CAD, results from Cox and MR analyses were generally directionally consistent. The pooled odds ratios (ORs) [95% CI] for CAD per one-SD increase in genetically-influenced PC1 and PC3 (both characterized by distinct ApoB-associated lipoprotein profiles) were 1.04 [1.03, 1.05] and 0.94 [0.93, 0.96], respectively. Besides, the pooled OR [95% CI] for CAD per one-SD increase in genetically-influenced PC4, characterized by simultaneously decreased small HDL and increased large HDL, and independent of ApoB, was 1.05 [1.03, 1.07]. For IS, increases of PC3 and PC5 (characterized by increased amino acids) were associated with a lower risk and a higher risk, respectively.

**Conclusions:** This study confirms associations of ApoB-associated lipoprotein profiles with CAD and IS, and highlights the possible existence of an ApoB-independent lipoprotein profile, characterized by a distinctive HDL sub-particle distribution, driving CAD.

**Keywords**: metabolomic measures; apolipoprotein B; high-density lipoprotein; coronary artery disease; ischemic stroke

# 1. Introduction



The effectiveness of lowering plasma low-density lipoprotein-cholesterol (LDL-C) to reduce coronary artery disease (CAD) risk is beyond any doubt (1). Lowering of plasma triglyceride (TG) levels may also reduce CAD risk on top of LDL-C lowering (2, 3). These observational findings have been attributed to a reduction in apolipoprotein B (ApoB), which has been suggested as the primary marker for cardiovascular disease risk independent of lipid content (cholesterol or TG) and type of ApoB-containing lipoprotein (cholesterol-rich LDL or TG-rich very-low-density lipoprotein [VLDL]) (4, 5). In addition, although an inverse association between high-density lipoprotein-cholesterol (HDL-C) and cardiovascular disease risk has long been established in prospective studies (6), Mendelian randomization (MR) studies (7) and clinical trials (8, 9) have so far failed to convincingly support a causal role for HDL-C in cardiovascular disease.

Importantly, lipoprotein metabolism is a highly dynamic system via which lipids and specific apolipoproteins are passively and actively exchanged between the different lipoprotein classes in the course of their transport and metabolism within the circulation. Considering lipoprotein classes such as LDL or HDL in isolation disregards the intricate interdependence of plasma lipoproteins. It would therefore be more appropriate to consider individual lipoprotein profiles as a whole, characterized by specific distributions of lipids and apolipoproteins over the different lipoprotein classes. Metabolomic platforms based on 1-H nuclear magnetic resonance ($^{1}$H-NMR) imaging of plasma samples provide such individual profiles by generating detailed measures on the composition, size, number and distribution of the different lipoprotein classes in a sample (10). In addition to lipoproteins, other metabolomic measures, such as some amino acids are also comprised in these platforms. Analyses of metabolomic measures as intermediates between exposures and clinical outcomes, is a powerful approach to dissect complex etiologic mechanisms linking metabolic processes to disease (11).

The interrelated nature of the lipoproteins also makes it difficult to identify specific genetic instruments for a single lipid or lipoprotein species without pleiotropic effects on the other lipoprotein subclasses. When performing MR studies, this may lead to biased estimations of health effects (12). Here, we tested the hypothesis that individual $^{1}$H-NMR metabolomics profiles can be grouped into different overall patterns, which are independent from each other and may have differential associations with diseases. To address this hypothesis, we performed principal component analysis (PCA) on $^{1}$H-NMR metabolomic measures from UK Biobank (UKB) participants. The principal components (PCs) can be regarded as uncorrelated traits, characterized by a specific overall metabolomic profile, which were

exploited to determine the associations with CAD and ischemic stroke (IS), and to triangulate findings from observational research through multivariable-adjusted regressions and large-scale multicohort MR analyses (13).

# 2. Materials and Methods

## 2.1 Project design



In the present study, we performed PCA on $^{1}$H-NMR metabolomic measures and conducted prospective multivariable-adjusted regression analyses to investigate the associations between selected PCs and examined diseases in UKB participants. In line with the principles of triangulation (13), i.e., integrating evidence from several different epidemiological approaches that have differing and unrelated key sources of bias, the MR analysis was further conducted to assess potential causal associations of selected PCs with examined cardiovascular diseases.

## 2.2 Study population

A detailed description of the UKB cohort has been presented elsewhere (14). In brief, the UKB cohort recruited 502,628 participants aged 40-70 years across the entire United Kingdom during the baseline survey between 2006 and 2010. Extensive phenotypic and genotypic details of the participants have been collected since the baseline assessment, including sociodemographic data, lifestyle, biological samples, genome-wide genotyping and longitudinal follow-up on a wide range of health-related outcomes. The UKB cohort study was approved by the North-West Multicentre Research Ethics Committee (MREC), and the access for information to invite participants was approved by the Patient Information Advisory Group (PIAG) from England and Wales. The study was conducted in accordance with the principles of the Declaration of Helsinki. All participants provided electronic written informed consent.

The UKB relevant team randomly sampled baseline participants from the full cohort population and released their metabolomic measurements (https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=130). These metabolomic measurements were performed on approximately 121,726 baseline participants between June 2019 and April 2020 (Phase 1) and on approximately 275,000 baseline participants between April 2020 and June 2022 (Phase 2). Participants from Phase 1 were used in the main analysis, and the remaining participants from phase 2 were used for the validation analysis.

Among participants from Phase 1 release data, 110,002 participants with complete metabolomic measures and genetics data were initiated for enrolment in the main analysis.

To minimize ancestry and population stratification bias, we restricted the study population to 71,736 unrelated individuals of European ancestry, based on the estimated kinship coefficients for all pairs and the self-reported ancestral background. A subset of 57,846 participants free from cardiometabolic disease (notably CAD, diabetes, and IS) and without taking cholesterol-lowering medication prior to the baseline survey were then selected for further studies. Finally, a total of 56,712 participants with complete data on covariables, including age, sex, the Townsend deprivation index, smoking status, alcohol consumption, body mass index (BMI), blood pressure lowering medication, and fasting time, were eligible for this study.

## 2.3 Profiling of metabolomic measures

The blood sample collection of UKB participants was undertaken at baseline between 2006 and 2010, and the blood sample handling and storage protocol has been previously described (15). The details of obtaining the NMR-based metabolomics data for UKB participants, including quality control, have been clearly described elsewhere (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=220) and in previous studies (16, 17). In brief, EDTA plasma samples collected at baseline recruitment were measured using a high throughput 1H-NMR-metabolomics platform (Nightingale Health, Helsinki, Finland). The NMR-based metabolomics platform has been broadly applied in large-scale epidemiologic studies, providing additional evidence to reveal underlying biological mechanisms (10, 18). This study included 168 direct metabolomic measures, which were listed in Table S1, along with their concentrations in the study population.

## 2.4 Principal component analysis

PCA is a method used for dimension reduction by projecting each data point onto a new orthogonal coordinate system while capturing as much of the variation as possible (19). The uncorrelated patterns identified by the PCA method from interrelated risk factors are also known as PCs. In this study, all 168 metabolomic measures from 56,712 participants were first transformed to approximate a normal distribution by inverse rank-based normal transformation and standardized with standard deviation (SD) one and mean zero. PCA was then performed as a singular value decomposition of the 168 standardized data matrix. The correlations between the metabolomic measures and each PC could be represented by loadings, defined as the eigenvector scaled up by the square roots of the eigenvalues of the respective PC (19). For individual participant, each PC score was calculated by summing the standardized measures weighted by the corresponding eigenvectors (19).

## 2.5 Prospective analyses

Outcome diagnoses were coded according to the International Classification of Diseases edition 10 (ICD-10) and were based on the date of the first occurrence. CAD is defined as angina pectoris (I20), myocardial infarction (MI) (I21 and I22), and acute and chronic ischemic heart disease (IHD) (I24 and I25); IS is defined as cerebral infarction (I63). These variables have been generated by the UKB data management team through a standard algorithm (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/first_occurrences_outcomes.pdf), combining self-reported health conditions from baseline and linked data from hospital admissions, primary care, and death registers. The linked data and its sources were presented here (https://biobank.ndph.ox.ac.uk/showcase/exinfo.cgi?src=Data_providers_and_dates). Outcomes in this prospective analysis were incident diseases during the time period from recruitment to January 1st, 2021. Follow-up time is computed from the baseline visit to the diagnosis of incident disease, loss-to-follow-up or death, or the end of the study period, whichever came first.

The prospective analyses were performed for the included 56,712 UKB participants. Three multivariable-adjusted Cox proportional hazards models were fitted to estimate hazard ratios (HRs) and corresponding 95% confidence intervals (95% CI) for the associations of PCs with incident CAD and IS: Model 1 was adjusted for age, sex, and the Townsend deprivation index; Model 2 was additionally adjusted for smoking status, alcohol consumption frequency, BMI, and blood pressure lowering medication; Model 3 was additionally adjusted for fasting time. To assess non-linear associations between PCs and outcomes, a penalized cubic regression spline was applied, adjusted for the same covariates in Model 3. We also performed multicollinearity diagnosis for Model 3 based on the variance inflation factor (VIF), which is equal to or greater than one (absence of multicollinearity). As a rule of thumb, a VIF value that exceeds 5 indicates a problematic amount of multicollinearity (20). To test the interaction effects with sex and age, multiplicative interaction terms between PCs and age and sex were added to Model 3. Sex-stratified (in women and men) and age-stratified (in 40-49, 50-59, and 60-70 years) analyses were further conducted using the fully-adjusted regression Model 3.

## 2.6 Mendelian randomization

MR analysis uses genetic variants, typically single-nucleotide polymorphisms (SNPs), as instrumental variables (21). This study used the two-sample MR method, which requires that groups of participants in the gene-exposure association analysis and gene-outcome association analysis do not overlap (22, 23). Gene-exposure association was based on

performing genome-wide association studies (GWAS) on each PC. Gene-outcome association was estimated or extracted from three large databases for each outcome. The overall workflow of MR analyses was presented in **Figure 1**.

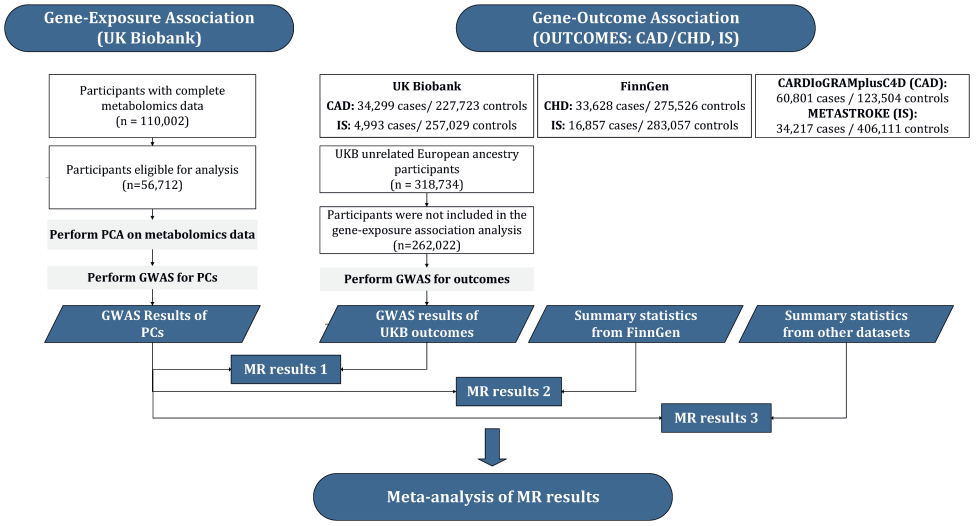



**Figure 1. The overall workflow of Mendelian randomization analyses in the present study.** CAD: coronary artery disease; CHD: coronary heart disease; IS: ischemic stroke; PCA: principal component analysis; GWAS: genome-wide association study; MR: Mendelian Randomization.

### *Genotyping and genetic imputations*

UK Biobank has undertaken a project to genotype all 500,000 participants. The majority of participants (~450k) are genotyped on the UK Biobank Axiom® Array, with 50,000 participants genotyped on the Affymetrix UK BiLEVE Axiom® array. The quality control of the genetic data was undertaken at the Wellcome Trust Centre for Human Genetics (WTCHG). Further details of the array design, genotyping and imputation procedures have been described elsewhere (24). In addition, UK Biobank resources performed centralized imputations on the autosomal SNPs using computationally efficient methods combined with the Haplotype Reference Consortium reference panels (25), UK10K haplotype (26), and 1000 Genomes Phase 3 (27). Autosomal SNPs were pre-phased using SHAPEIT3 and imputed using IMPUTE4. In total, ~96 million SNPs were imputed.

### *Associations of genetic variants with exposure*

Using the software program GEM (version 1.4.2) (28), the GWAS was performed for each PC with the regression model adjusted for age, sex, first 10 genetic PCs, and fasting time. SNPs with a minor allele frequency below 0.001 were removed. For each PC, genome-wide significant SNPs ($P < 5*10^{-8}$) were selected, and then pruned to obtain independent instrumental variables by the *TwoSampleMR* and *ieugwasr* packages, which use the PLINK clumping method with a clumping window of 10Mb and linkage disequilibrium $r^2 < 0.001$ (29). To avoid potential residual pleiotropic effects, only SNPs without overlap among PCs were selected and subsequently used to extract the gene-outcome associations. The proportion of exposure variation explained by the genetic variants was depicted by the $R^2$ statistic (30), and the potentially weak instrument bias was examined by the *F*-statistic, for which a threshold greater than 10 is conventionally considered sufficient for MR analysis (23).

### *Associations of genetic variants with outcome*

Associations of the identified exposure-related SNPs with each outcome were estimated or extracted from three large databases. Specifically, summary statistics for CAD were extracted from UKB, FinnGen study (31) and CARDIoGRAMplusC4D (32). Summary statistics for IS were extracted from UKB, FinnGen study (31) and MEGASTROKE (33). The descriptions of the large databases and the detailed process for extracting the gene-outcome associations are presented in Supplementary eMethods.

### *Estimation of the associations between principal components and outcome*

The associations of PCs with each outcome from each database were estimated by the inverse-variance weighted (IVW) method, which combines the Wald ratio estimates (the estimated association of the genetic variant with outcome divided by the estimated association of the genetic variant with exposure) for individual genetic variants by a fixed-effect meta-analysis with inverse-variants weights (23). Those estimates were expressed as log odds ratios (ORs) for each PC per one-SD increase. We subsequently conducted meta-analyses for each PC to pool the estimates from the three outcome databases. The heterogeneity of the estimated ORs from three databases for each PC was represented by $I^2$, and detected by the Cochran Q test.

Given that the IVW method assumes all genetic instruments are valid (e.g., no horizontal pleiotropy), we conducted sensitivity analyses using the weighted-median estimator and the MR-Egger method to assess whether IVW analyses were biased due to horizontal pleiotropy (34, 35). Rather than taking a weighted mean of the ratio estimates as in the IVW method, the weighted-median estimator could still provide a consistent estimate of the causal effect

even when up to 50% of the identified genetic variants are invalid IVs (34). In contrast to the IVW method, the MR-Egger method does not require a zero horizontal pleiotropy effect, and could detect pleiotropy by the intercept term (under the InSIDE assumption), which when different from zero indicates a bias in the IVW estimation (35).

## 2.7 Validation

After excluding participants already included in the main analysis, and in line with the exclusion criteria in the main analysis, a total of 77,212 European-ancestry and unrelated participants from UKB phase 2 release were included, who were free of cardiometabolic disease (notably CAD, diabetes, and IS) and not taking cholesterol-lowering medication prior to the baseline survey.

The loadings (the correlations between the 168 metabolomic measures and each PC) in the validation analysis were derived using the same methodology as in the main analysis (including inverse rank-based normal transformation and standardization). To compare the PCA results from the validation analysis with the PCA results from the main analysis, we further calculated the correlations between the loadings of each PC obtained from the main and validation samples, expressed as an R-square. In addition, using the independent SNPs and the corresponding weights identified in the main MR analysis, the genetic risk score for each PC was calculated in the validation sample. We subsequently assessed the associations between genetic risk scores and the corresponding PCs from the validation data, and examined the extent to which the genetic risk scores explained the variance in the corresponding PCs.

All statistical analyses described above were performed in the R (version 4.0.2) software, with ‘prcomp’, ‘survival’ and ‘TwoSampleMR’ packages for PCA, Cox regression analyses and MR analyses, respectively.

**Table 1. Baseline characteristics of the study population for analysing metabolomic measures**

| | Main sample | | | Validation sample | | |
|---|---|---|---|---|---|---|
| | **Overall** | **Women** | **Men** | **Overall** | **Women** | **Men** |
| n | 56712 | 32341 | 24371 | 77212 | 44327 | 32885 |
| Age, years (median [IQR]) | 56 [49, 62] | 57 [49, 62] | 56 [49, 62] | 56 [49, 62] | 57 [49, 62] | 56 [49, 62] |
| Townsend deprivation index (mean (SD)) | -1.55 (2.94) | -1.55 (2.91) | -1.55 (2.98) | -1.63 (2.90) | -1.66 (2.85) | -1.60 (2.96) |
| Smoking status, n (%) | | | | | | |
| Never | 31935 (56.3) | 19198 (59.4) | 12737 (52.3) | 43816 (56.7) | 26643 (60.1) | 17173 (52.2) |
| Previous | 18986 (33.5) | 10312 (31.9) | 8674 (35.6) | 25543 (33.1) | 13842 (31.2) | 11701 (35.6) |
| Current | 5791 (10.2) | 2831 (8.8) | 2960 (12.1) | 7853 (10.2) | 3842 ( 8.7) | 4011 (12.2) |
| Alcohol consumption status, n (%) | | | | | | |
| Never | 3441 (6.1) | 2354 (7.3) | 1087 (4.5) | 4636 (6.0) | 3141 (7.1) | 1495 (4.5) |
| Special occasions only | 5659 (10.0) | 4176 (12.9) | 1483 (6.1) | 8033 (10.4) | 5946 (13.4) | 2087 (6.3) |
| One to three times a month | 6536 (11.5) | 4340 (13.4) | 2196 (9.0) | 8744 (11.3) | 5862 (13.2) | 2882 (8.8) |
| Once or twice a week | 15200 (26.8) | 8638 (26.7) | 6562 (26.9) | 21102 (27.3) | 12098 (27.3) | 9004 (27.4) |
| Three or four times a week | 14079 (24.8) | 7317 (22.6) | 6762 (27.7) | 19062 (24.7) | 9790 (22.1) | 9272 (28.2) |
| Daily or almost daily | 11797 (20.8) | 5516 (17.1) | 6281 (25.8) | 15635 (20.2) | 7490 (16.9) | 8145 (24.8) |
| BMI, kg/m$^2$ (mean (SD)) | 26.94 (4.55) | 26.60 (4.87) | 27.39 (4.04) | 26.95 (4.55) | 26.64 (4.89) | 27.37 (4.01) |
| Blood pressure medication = Yes, n (%) | 6463 (11.4) | 3501 (10.8) | 2962 (12.2) | 8731 (11.3) | 4944 (11.2) | 3787 (11.5) |

Abbreviations: BMI, body mass index; IQR, interquartile range; SD, Standard deviation.

# 3. Results

The present study included 56,712 unrelated European-ancestry participants (57% women, a median age of 56 [IQR: 49, 62] years) from the UKB cohort in the main analyses. The baseline characteristics of the included participants are presented in **Table 1** stratified by sex. During the 13.7-year follow-up period, 3,253 participants developed CAD, and 640 participants developed IS, with incidence rates of 501 [95% CI: 484, 519] and 97 [89, 104] per 100,000 person-years, respectively.

## 3.1 Principal component analysis

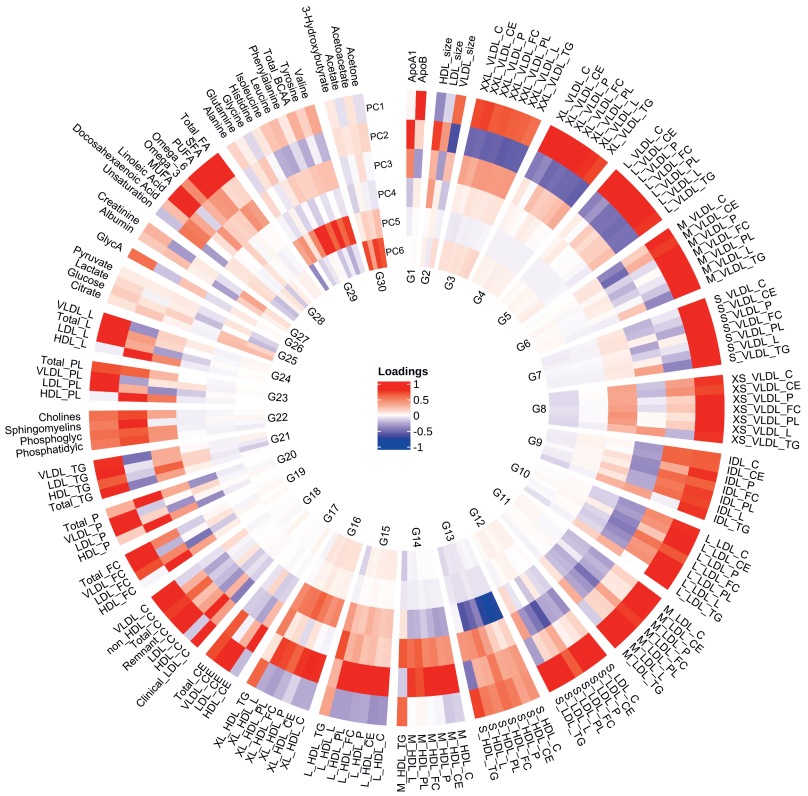
Considering the combination of the eigenvalues-greater-than-one rule (36), explained variance and loading interpretability, the first six PCs with a cumulative explained variance of 88% were selected for further analyses. The detailed eigenvalues and explained variances of all PCs are presented in Table S2.

The loadings for the first six PCs are shown in **Figure 2** and summarized in **Table 2**. In detail, PC1 (46.7% variance explained) is mainly characterized by higher levels of ApoB, ApoB-containing lipoproteins and fatty acids. PC2 (22.5% variance explained) is mainly characterized by higher levels of apolipoprotein A1 (ApoA1), HDL particles and lower levels of VLDL particles. PC3 (9.4% variance explained) is characterized by lower levels of most ApoB-containing lipoproteins, but higher levels of ApoA1 and HDL particles. PC4 (5.0% variance explained) is characterized by lower levels of small HDL particles and higher levels of very large HDL particles, independent of ApoB. PC5 (2.7% variance explained) is characterized by higher levels of amino acids, and PC6 (1.6% variance explained) is characterized by higher ketone body levels.

**Table 2. The main loadings of the first six principal components (PCs)**

| PCs | Explained variance | Meaning of the loadings |
|---|---|---|
| PC1 | 46.7% | Positively correlated with ApoB, ApoB-containing lipoproteins and fatty acids |
| PC2 | 22.5% | Positively correlated with ApoA1 and HDL particles and negatively correlated with VLDL particles |
| PC3 | 9.4% | Negatively correlated with most ApoB-containing lipoproteins, and positively correlated with ApoA1 and HDL particles |
| PC4 | 5.0% | Negatively correlated with small HDL particles and positively with large and very large HDL particles |
| PC5 | 2.7% | Positively correlated with amino acids |
| PC6 | 1.6% | Positively correlated with ketone bodies |

Abbreviations: ApoB, apolipoprotein B; HDL, high-density lipoprotein; PCs, principal components; VLDL, very-low-density lipoprotein.




**Groups:**

| | | | | |
|---|---|---|---|---|
| G1 Apolipoproteins | G7 Small VLDL | G13 Small HDL | G19 Free cholesterol | G25 Glycolysis related metabolites |
| G2 Lipoprotein particle sizes | G8 Very small VLDL | G14 Medium HDL | G20 Lipoprotein particle concentrations | G26 Inflammation |
| G3 Extremely large VLDL | G9 IDL | G15 Large HDL | G21 Triglycerides | G27 Fluid balance |
| G4 Very large VLDL | G10 Large LDL | G16 Very large HDL | G22 Other lipids | G28 Fatty acids |
| G5 Large VLDL | G11 Medium LDL | G17 Cholesteryl esters | G23 Phospholipids | G29 Amino bodies |
| G6 Medium VLDL | G12 Small LDL | G18 Cholesterol | G24 Total lipids | G30 Ketone bodies |

**Figure 2. A circular loading plot of the six principal components.** From the outside to the inside, each circle represents the correlation between the respective principle components (PC) and 168 metabolomic measures. All metabolomic measures are divided into 30 groups and shown in clockwise order. Red or blue colour indicates the increase or decrease of metabolomic measures in the six PCs.

## 3.2 Prospective multivariable-adjusted regression analyses

The VIF values of PCs and included covariates in model 3 were all less than 5 and most of them were even less than 1.5, indicating absence of multicollinearity (Table S3). According to the curves based on the spline models (Figure S1 and Figure S2), the linearity assumption for the associations of PCs with CAD and IS was appropriate. The estimated multivariable-adjusted associations of the six PCs with the incident diseases are presented in **Figure 3**. For the risk of CAD, the HRs [95% CI] for per one-SD increase in PC1, PC2, PC3 and PC4 were 1.02 [1.02, 1.03], 0.98 [0.97, 0.99], 0.98 [0.97, 0.99], and 1.02 [1.01, 1.03], respectively. For the risk of IS, the HRs [95% CI] for per one-SD increase in PC4 and PC6 were 1.04 [1.01, 1.07] and 1.08 [1.03, 1.34], respectively. In addition, for the CAD risk, PC1 showed interaction ($P_{interaction}$ = 3.26e-11) with age, while PC2 ($P_{interaction}$ = 0.004) and PC3 ($P_{interaction}$ = 0.01) showed interactions with sex. No interactions with age or sex were observed with IS. The estimated risk directions in the stratified analyses were similar to the main analyses. Consistent with the interaction tests, PC1-associated CAD risk attenuated with increasing age, and PC2 was more strongly associated with CAD in women, while PC3 was more strongly associated with CAD in men (Figure S3).

## 3.3 Mendelian randomization

A total of 150 independent SNPs were found to be significantly associated with the PCs, presented in Table S4 with the corresponding mapping genes. In detail, 41 SNPs, 37 SNPs, 31 SNPs, 22 SNPs, 11 SNPs, and 8 SNPs explained 7.3%, 9.0%, 7.5%, 9.0%, 0.9%, and 1.1% of the variation in PC1, PC2, PC3, PC4, PC5 and PC6, respectively. All *F* statistics were larger than 10. Supplementary tables provide the details of the independent and non-overlapping genetic variants, including their position, gene-exposure associations and corresponding $R^2$ statistics and *F* statistics, and gene-outcome associations (Table S5-10).

For the association of each PC with each outcome, Cochran Q statistics detected no heterogeneity (*P* values > 0.05) in the estimated ORs across the three outcome databases (Table S11). **Figure 3** shows the estimated associations between each PC and each outcome from each database by the inverse-variance weighted (IVW) method, and their pooled estimates across the three databases. For the risk of CAD, the pooled estimated ORs [95% CI] per one-SD increase in PC1, PC3 and PC4 were 1.04 [1.03, 1.05], 0.94 [0.93,0.96] and 1.05 [1.03, 1.07], respectively. For the risk of IS, the pooled estimated ORs [95% CI] per one-SD increase in PC3, PC5 and PC6 were 0.97 [0.96,0.99], 1.12 [1.07, 1.18] and 0.91 [0.84, 0.99], respectively.

The estimated ORs based on the weighted-median estimator analyses were similar to those from IVW analyses (Table S11). No horizontal pleiotropic effect was detected according to the tests for intercept terms by MR-Egger analyses, except for the effect of PC5 on CHD from the FinnGen database with *P* value of 0.04 (Table S12).

## 3.4 Validation

Similar to the main analysis, the top six PCs from the validation analysis have a cumulative explained variance of 87.75%. The detailed loadings of each PC were presented in Figure S4. The correlations between the loadings of each PC from the main and validation analyses were greater than 0.99 (**Figure 4**). In addition, in the validation data, the associations between the genetic risk scores and the corresponding PC1, PC2, PC3, PC4, PC5 and PC6 were 0.23, 0.25, 0.26, 0.28, 0.08, and 0.09, respectively, with *P* values less than 2.2e-16. Similar to the explained variance found in the MR study, the genetic risk scores explained 5.5%, 6.5%, 6.7%, 7.7%, 0.68% and 0.79% of the variance in PC1, PC2, PC3, PC4, PC5, and PC6 respectively.

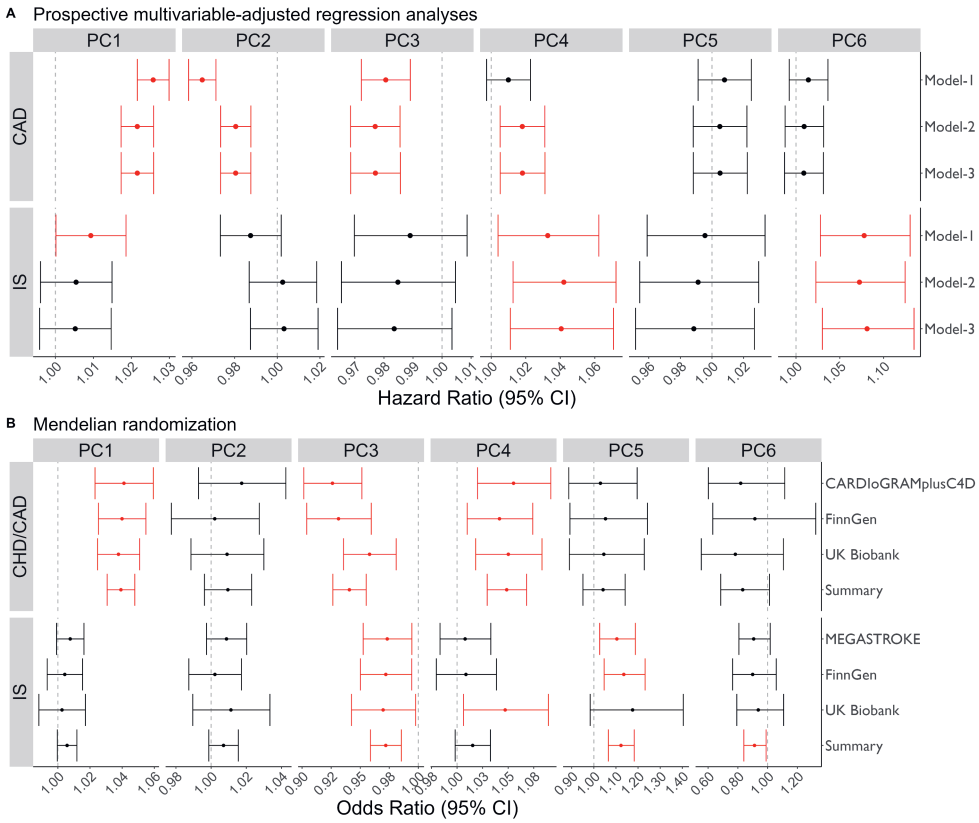



**Figure 3. Associations of the six principal components with the risks of coronary artery disease and ischemic stroke by prospective multivariable-adjusted regression analyses (A) and Mendelian randomization (B).** In subplot A, the x-axis presents the estimated hazard ratios (95% CI) for the incident coronary artery disease (CAD), and ischemic stroke (IS) according to per one-SD increase in each PC. The right y-axis indicates results from specific models. Model-1 was adjusted for sex, age and Townsend index; Model-2 was Model-1 additionally adjusted for smoking status, alcohol consumption frequency, BMI and blood pressure lowering medication; Model-3 was Model-2 additionally adjusted for fasting time. In subplot B, the x-axis presents the estimated odds ratios (95% CI) by inverse-variance weighted method for the risks of CAD and IS per one-SD increase in each PC; the right y-axis is labelled with the names of three data sources of the summary association statistics between genetic variants and outcome, and with 'summary' indicating the fixed-effect meta-analysis. For the two subplots, red lines (but not black lines) indicate significant associations between PCs and outcomes.

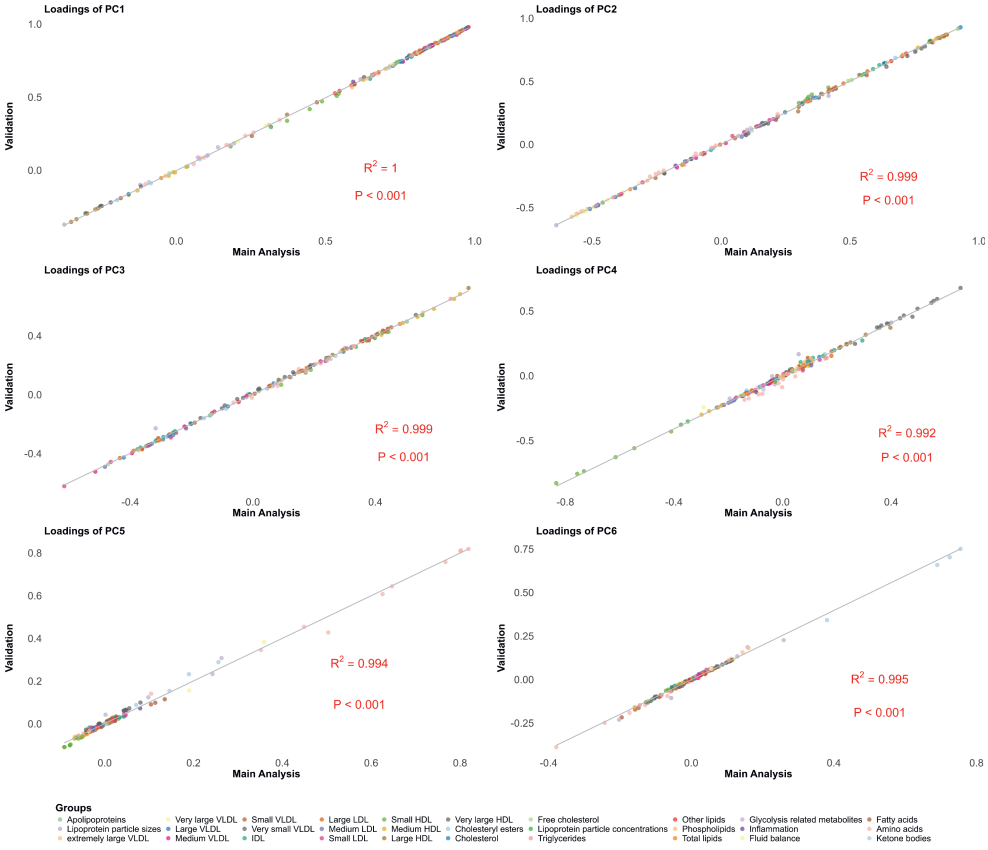



**Figure 4. Correlations of the loading results between main analysis and validation analysis.** X-axis shows the loadings of all 168 metabolomic measures in each PC from the main analysis, and Y-axis shows the corresponding loadings of the 168 metabolomic measures in each PC from the validation. Correlations are presented as $R^2$ values and P values, indicating correlation magnitude and correlation tests, respectively. All 168 metabolomic measures were presented in 30 groups with different colours.

# 4. Discussion



We applied PCA on 168 $^{1}$H NMR-based metabolomic measures in 56,712 UKB participants and identified six main PCs representing uncorrelated metabolomic profiles, with a cumulative explained variance of 88%. Multivariable-adjusted Cox regression and large-scale multicohort MR analyses were used to examine the disease risks associated with PCs. PC1 (characterized by higher levels of ApoB and ApoB-containing lipoproteins) was associated with a higher risk of CAD, and PC3 (characterized by lower levels of ApoB-containing lipoproteins and higher levels of ApoA1 and HDL particles) was associated with a lower risk of CAD. Notably, PC4 (characterized by lower levels of small HDL particles and higher levels of very large HDL particles) was also associated with higher risk of CAD. In addition, PC3 and PC5 (characterized by higher levels of amino acids) were associated with lower and higher IS risk, respectively.

The loadings of each PC represent the contributions of the various lipoprotein subfractions (and other measured metabolites) to the corresponding PC and thus represent specific sub-particle lipoprotein profiles. The notable observation is that, apparently, the combined production, metabolism and clearance of lipoproteins not only result in the overall distribution of lipids and apolipoproteins over major lipoproteins such as HDL, VLDL and LDL, but also in a limited number of independent sub-particle lipoprotein profiles as represented by the PCs. Our MR analyses provided evidence for causality of some of the PCs with specific cardiovascular disease outcomes. However, it should be emphasized that the inherent interconnectivity of the different components in the PCs complicates assigning causality to individual components.

Nevertheless, our findings on the associations of PC1 and PC3, which mainly captured higher and lower ApoB-associated lipoproteins, respectively, with CAD are in line with previous findings that ApoB-containing lipoproteins drive atherogenic cardiovascular disease (4, 5). All ApoB-containing lipoproteins carry a single ApoB molecule and the concentration of ApoB thus directly reflects the number of circulating particles. The concentration of particles is the primary determinant for entry and entrapment in the arterial wall (via osmotic pressure), which is the first step in the development of atherogenesis. This explains why, in large prospective cohort analyses and clinical trials, ApoB is superior over LDL-C or triglycerides, which correlate with, but do not directly reflect, particle number (5, 37).

Our study found that PC4 was also associated with CAD, which seemed independent from ApoB. Based on the loading values (Figure 2), the higher PC4-related risk of CAD is associated with a very specific HDL size distribution characterized by lower levels of small

HDL particles and higher levels of large HDL particles. To assess the contribution of ApoB to CAD in PC4, we further reperformed the MR analyses for the residual PC4, which was derived by regressing PC4 on ApoB. The result (Figure S5) showed that for the risk of CAD, the estimated pooled OR [95% CI] was 1.07 [1.03, 1.11] for per one-SD increase in residual-PC4. These analyses provide evidence for a potential (causal) association of HDL particles/composition with CAD, independent of ApoB.

Given the lack of potential causal evidence between HDL-C levels and CAD, the focus of HDL research has shifted to the assessment of the roles of HDL function in CAD, such as reverse cholesterol transport, and inhibition of inflammation and oxidative stress (38-40). Reverse cholesterol transport, being the most extensively studied function of HDL, includes cellular cholesterol efflux to nascent HDL particles, cholesterol esterification, maturation of HDL and cholesterol clearance via the liver. Previous studies indicated that cholesterol efflux capacity was inversely associated with cardiovascular events, independent of HDL-C (41-43). In addition, cholesterol clearance from HDL was shown to play a role in reverse cholesterol transport and CAD. The scavenger receptor class BI (SR-BI) promotes selective hepatic uptake of cholesterol, primarily from large cholesterol-enriched HDL particles (44), and hepatic SR-BI deficiency was associated with an increased risk of CAD despite increased HDL-C levels (45). Presumably, this increased HDL-C is caused by the accumulation of cholesterol loaded large HDL particles that cannot be cleared via SR-BI by the liver. In line with this interpretation, the present study showed that PC4 indicated higher levels of large HDL particles, and was associated with increased CAD risk.

In addition, previous evidence suggested that the HDL particle profile measured by NMR spectroscopy, especially small HDL particles, should be considered to better stratify CAD risk in the population (46). Notably, in line with our observations, small HDL was found to have atheroprotective effects (47, 48). Very small HDL particle number has been previously found to be strongly associated with lower CAD risk in patients with type 1 diabetes (49). Moreover, a recent study showed that the plasma proteome of the different HDL subspecies differed markedly, in parallel with *in vitro* cholesterol efflux capacity. Interestingly, prominent differences in the protein composition of small (preβ-1) HDL particles and large (α-1) HDL particles were found in CHD patients versus controls (50). Therefore, these studies combined with our data provide evidence for the hypothesis that for HDL-targeted therapy to be effective in prevention of CAD, higher levels of small HDL particles and lower levels of large HDL particles are warranted.

In recent years, HDL has been proposed as therapeutic target to combat atherosclerotic cardiovascular diseases (51). ApoA1 is one of the main proteins in HDLs, and it has long been suggested that increasing the plasma levels of ApoA1 should result in an increased level

of small HDL particles, increased cholesterol efflux capacity and reduced CAD risk (41), in line with the observed PC4 associations in the present study. However, the results from the recently conducted AEGIS-II trial showed that weekly four infusions of CSL112 (composed of human plasma–derived apoA1) did not result in a significant reduction (HR: 0.93 [95% CI: 0.81, 1.05]) in the risk of recurrent cardiovascular events after acute myocardial infarction during 90 days of follow-up (52). Although increased HDL-C was observed in AEGIS-II, the pre- and postinjection cholesterol efflux capacity and HDL particle distribution were not assessed. These data have prompted the question for reconceptualization of HDL function in CAD (53). The short duration of follow-up and the advanced stage of atherosclerotic cardiovascular disease in this secondary prevention trial may explain the negative results of AEGIS-II. Moreover, in line with the observed association with large HDL particles, not only increased efflux of cholesterol from atherosclerotic plaques to HDL may be required for benefit, but also an increased flux of cholesterol from large HDL particles to the liver.

In addition, we observed associations between ApoB-associated lipoprotein profiles and IS risk, with both the prospective multivariable-adjusted regressions and MR analyses showing a trend in higher IS risk for PC1, and a trend in lower IS risk for PC3 (Figure 3). Our findings are supported by previous evidence that strongly suggests a positive association between ApoB and IS risk (54, 55). Notably, previous studies have shown that a higher ApoB/ApoA1 ratio is associated with a higher risk of IS, especially in younger individuals (56, 57). In line with these previous findings, we observed that PC3, which is not only characterized by lower levels of ApoB but also higher levels of ApoA1, associated with lower IS risk. Therefore, our study provides further evidence for the effects of dyslipidemia on the development of IS, especially for the role of ApoB and ApoB/ApoA1 ratio.

We also found that PC5, mainly characterised by higher levels of different amino acids, is a risk factor for IS. Especially the branched chain amino acids (BCAAs), valine, leucine and isoleucine contribute to PC5. In line with our observations, previous studies have shown a positive association between baseline circulating BCAAs and IS risk (58, 59). It is well-known that BCAAs play a role in energy homeostasis through nutrient signalling, and high levels of circulating BCAAs are associated with metabolic disorders, including obesity, insulin resistance and type 2 diabetes (60-62). However, the functional roles and underlying mechanisms of BCAAs in IS remain unclear (63). In addition, as BCAAs can readily pass the blood brain barrier and are required for glutamate synthesis, high levels of BCAA have been implicated in glutamate excitotoxicity, which could trigger oxidative stress, inflammation and endothelial damage (64). In addition to BCAAs, we found that other amino acids, such as glutamine, phenylalanine and tyrosine, also positively contributed to PC5. A previous study summarized the differential and potential roles of these amino acids in IS, and indicated that these might be biomarkers for early diagnosis in IS patients (63). Although we

identified a positive association between PC5 and IS using MR analysis, additional studies are needed to determine which amino acid(s) mainly drive the development of IS.

There is increasing evidence that ketone bodies (specifically, acetoacetate, acetone, and β-hydroxybutyrate [β-OHB]) are protective against cardiovascular disease by providing an important additional source of ATP production and being important signalling molecules (65-67). However, the potential protective effects observed in previous studies are mainly based on experimental or clinical research on patients with cardiovascular disease, whereas epidemiological associations between plasma ketone bodies and cardiovascular events have only been rarely explored and remain unclear. One prospective study found that the positive association of ketone bodies with stroke became null when the regression model was fully adjusted for potential confounding factors (68). Similar to the previous study, we found that the association of PC6 (mainly characterized by higher levels of ketone bodies) with IS attenuated with a HR [95% CI] of 1.05 [0.99, 1.11] if regression Model 3 was additionally adjusted for physical activity.

In addition, it is well-known that IS is a late-life disease that often occurs after a long period of exposure to risk factors. The risk estimation of IS could therefore be confounded by many other unmeasured factors. For the prospective analyses in the present study, the limited age range from 40-70, the low disease incidence and residual confounding effects may explain the null effect of PC5 with IS, and the attenuated association of PC6 with IS after further adjusting for physical activity. Interpretation of stroke risks should always be done with caution, and it is advisable to evaluate and synthesize evidence from different studies, including studies on underlying biological mechanisms.

The large number of $^{1}$H-NMR-based metabolomic measures in a large number of disease-free individuals, especially various lipids and lipoprotein fractions, enabled thorough description of the interrelationship among metabolomic measures and the identification of specific profiles. For MR analyses, a large-scale multicohort design was used in the present study, which provided ample power and mutual validation. However, there are also some limitations that need to be considered. First, UKB plasma samples used for the NMR measurements suffered from 5-10% dilution, which may slightly affect the absolute concentrations of metabolomic measures but is expected to have limited impact on most epidemiological analyses (17). In addition, the metabolomic measures from UKB are from non-fasting samples, resulting in measurements of lipids and lipoproteins that may not be representative for average daily levels, especially TG. However, recent studies suggested that fasting is not routinely required for relative risk analysis of lipid profiles, and that the measurement of ApoB is stable with or without fasting (69). Finally, although we conducted validation analyses based on an independent sample from the UKB cohort, this does not

represent an independent validation as both samples were recruited according to the same criteria and study protocol. In future studies, additional validation analyses using external data beyond the UKB cohort are needed.

In conclusion, the present study, based on uncorrelated profiles of metabolomic measures, not only confirms the effect of ApoB-containing lipoproteins on CAD, but also provides evidence for the potential role of HDL sub-particle distribution in the development of CAD. Furthermore, our findings support the notion that lipids, lipoproteins and amino acids are important risk factors for the development of IS.

# Reference

1. Ference B. A., Ginsberg H. N., Graham I., Ray K. K., Packard C. J., Bruckert E., et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J. 2017;38(32):2459-72.
2. Ibi D., Blauw L. L., Noordam R., Dolle M. E. T., Jukema J. W., Rosendaal F. R., et al. Triglyceride-lowering LPL alleles combined with LDL-C-lowering alleles are associated with an additively improved lipoprotein profile. Atherosclerosis. 2021;328:144-52.
3. Lotta L. A., Stewart I. D., Sharp S. J., Day F. R., Burgess S., Luan J., et al. Association of Genetically Enhanced Lipoprotein Lipase-Mediated Lipolysis and Low-Density Lipoprotein Cholesterol-Lowering Alleles With Risk of Coronary Disease and Type 2 Diabetes. JAMA Cardiol. 2018;3(10):957-66.
4. Ference B. A., Kastelein J. J. P., Ray K. K., Ginsberg H. N., Chapman M. J., Packard C. J., et al. Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. JAMA. 2019;321(4):364-73.
5. Marston N. A., Giugliano R. P., Melloni G. E. M., Park J. G., Morrill V., Blazing M. A., et al. Association of Apolipoprotein B-Containing Lipoproteins and Risk of Myocardial Infarction in Individuals With and Without Atherosclerosis: Distinguishing Between Particle Concentration, Type, and Content. JAMA Cardiol. 2022;7(3):250-6.
6. The Emerging Risk Factors Collaboration. Major Lipids, Apolipoproteins, and Risk of Vascular Disease. JAMA. 2009;302(18):1993-2000.
7. Voight Benjamin F., Peloso Gina M., Orho-Melander Marju, Frikke-Schmidt Ruth, Barbalic Maja, Jensen Majken K., et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. The Lancet. 2012;380(9841):572-80.
8. Schwartz G. G., Olsson A. G., Abt M., Ballantyne C. M., Barter P. J., Brumm J., et al. Effects of dalcetrapib in patients with a recent acute coronary syndrome. N Engl J Med. 2012;367(22):2089-99.
9. Lincoff A. M., Nicholls S. J., Riesmeyer J. S., Barter P. J., Brewer H. B., Fox K. A. A., et al. Evacetrapib and Cardiovascular Outcomes in High-Risk Vascular Disease. N Engl J Med. 2017;376(20):1933-42.
10. Soininen P., Kangas A. J., Wurtz P., Suna T., Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet. 2015;8(1):192-206.
11. Cavus E., Karakas M., Ojeda F. M., Kontto J., Veronesi G., Ferrario M. M., et al. Association of Circulating Metabolites With Risk of Coronary Heart Disease in a European Population: Results From the Biomarkers for Cardiovascular Risk Assessment in Europe (BiomarCaRE) Consortium. JAMA Cardiol. 2019;4(12):1270-9.
12. Wurtz P., Kangas A. J., Soininen P., Lehtimaki T., Kahonen M., Viikari J. S., et al. Lipoprotein subclass profiling reveals pleiotropy in the genetic variants of lipid risk factors for coronary heart disease: a note on Mendelian randomization studies. J Am Coll Cardiol. 2013;62(20):1906-8.
13. Lawlor D. A., Tilling K., Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2016;45(6):1866-86.
14. Sudlow C., Gallacher J., Allen N., Beral V., Burton P., Danesh J., et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
15. Elliott P., Peakman T. C., Biobank U. K. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. Int J Epidemiol. 2008;37(2):234-44.
16. Ritchie S. C., Surendran P., Karthikeyan S., Lambert S. A., Bolton T., Pennells L., et al. Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. Sci Data. 2023;10(1):64.
17. Julkunen H., Cichonska A., Tiainen M., Koskela H., Nybo K., Makela V., et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. Nat Commun. 2023;14(1):604.
18. Wurtz P., Kangas A. J., Soininen P., Lawlor D. A., Davey Smith G., Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. Am J Epidemiol. 2017;186(9):1084-96.
19. Jolliffe I. T., Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202.
20. Gareth James, Daniela Witten, Trevor Hastie, Tibshirani Robert. An Introduction to Statistical Learning. 2 ed: Springer New York, NY; 2021. XV, 607 p.

21. Smith G. D., Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.
22. Pierce B. L., Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. Am J Epidemiol. 2013;178(7):1177-84.
23. Burgess S., Butterworth A., Thompson S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.
24. Bycroft C., Freeman C., Petkova D., Band G., Elliott L. T., Sharp K., et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
25. McCarthy S., Das S., Kretzschmar W., Delaneau O., Wood A. R., Teumer A., et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-83.
26. Consortium Uk K., Walter K., Min J. L., Huang J., Crooks L., Memari Y., et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82-90.
27. Genomes Project Consortium, Auton A., Brooks L. D., Durbin R. M., Garrison E. P., Kang H. M., et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
28. Westerman Kenneth E., Pham Duy T., Hong Liang, Chen Ye, Sevilla-González Magdalena, Sung Yun Ju, et al. GEM: scalable and flexible gene–environment interaction analysis in millions of samples. Bioinformatics. 2021;37(20):3514-20.
29. Hemani Gibran, Zheng Jie, Elsworth Ben, Wade Kaitlin, Haberland Valeriia, Baird Denis, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018;7:e34408.
30. Shim H., Chasman D. I., Smith J. D., Mora S., Ridker P. M., Nickerson D. A., et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One. 2015;10(4):e0120758.
31. Kurki M. I., Karjalainen J., Palta P., Sipila T. P., Kristiansson K., Donner K. M., et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508-18.
32. Nikpay M., Goel A., Won H. H., Hall L. M., 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.
33. Malik R., Chauhan G., Traylor M., Sargurupremraj M., Okada Y., Mishra A., et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet. 2018;50(4):524-37.
34. Bowden J., Davey Smith G., Haycock P. C., Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.
35. Bowden J., Davey Smith G., Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.
36. Kaiser Henry F. The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement. 1960;20(1):141-51.
37. Glavinovic T., Thanassoulis G., de Graaf J., Couture P., Hegele R. A., Sniderman A. D. Physiological Bases for the Superiority of Apolipoprotein B Over Low-Density Lipoprotein Cholesterol and Non-High-Density Lipoprotein Cholesterol as a Marker of Cardiovascular Risk. J Am Heart Assoc. 2022;11(20):e025858.
38. Rohatgi A., Westerterp M., von Eckardstein A., Remaley A., Rye K. A. HDL in the 21st Century: A Multifunctional Roadmap for Future HDL Research. Circulation. 2021;143(23):2293-309.
39. Linton M. F., Yancey P. G., Tao H., Davies S. S. HDL Function and Atherosclerosis: Reactive Dicarbonyls as Promising Targets of Therapy. Circ Res. 2023;132(11):1521-45.
40. von Eckardstein A., Nordestgaard B. G., Remaley A. T., Catapano A. L. High-density lipoprotein revisited: biological functions and clinical relevance. Eur Heart J. 2022.
41. Rohatgi A., Khera A., Berry J. D., Givens E. G., Ayers C. R., Wedin K. E., et al. HDL cholesterol efflux capacity and incident cardiovascular events. N Engl J Med. 2014;371(25):2383-93.
42. Khera Amit V., Cuchel Marina, de la Llera-Moya Margarita, Rodrigues Amrith, Burke Megan F., Jafri Kashif, et al. Cholesterol Efflux Capacity, High-Density Lipoprotein Function, and Atherosclerosis. New England Journal of Medicine. 2011;364(2):127-35.
43. Soria-Florido M. T., Schroder H., Grau M., Fito M., Lassale C. High density lipoprotein functionality and cardiovascular events and mortality: A systematic review and meta-analysis. Atherosclerosis. 2020;302:36-42.
44. Rosenson R. S., Brewer H. B., Jr., Davidson W. S., Fayad Z. A., Fuster V., Goldstein J., et al. Cholesterol efflux and atheroprotection: advancing the concept of reverse cholesterol transport. Circulation. 2012;125(15):1905-19.

45. Zanoni Paolo, Khetarpal Sumeet A., Larach Daniel B., Hancock-Cerutti William F., Millar John S., Cuchel Marina, et al. Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. Science. 2016;351(6278):1166-71.
46. Duparc T., Ruidavets J. B., Genoux A., Ingueneau C., Najib S., Ferrieres J., et al. Serum level of HDL particles are independently associated with long-term prognosis in patients with coronary artery disease: The GENES study. Sci Rep. 2020;10(1):8138.
47. McGarrah R. W., Craig D. M., Haynes C., Dowdy Z. E., Shah S. H., Kraus W. E. High-density lipoprotein subclass measurements improve mortality risk prediction, discrimination and reclassification in a cardiac catheterization cohort. Atherosclerosis. 2016;246:229-35.
48. Wang X., Liu X., Xie Z., Tian J., Huang X., Zhang R., et al. Small HDL subclass is associated with coronary plaque stability: An optical coherence tomography study in patients with coronary artery disease. J Clin Lipidol. 2019;13(2):326-34 e2.
49. Costacou Tina, Vaisar Tomas, Miller Rachel G., Davidson W. Sean, Heinecke Jay W., Orchard Trevor J., et al. High-Density Lipoprotein Particle Concentration and Size Predict Incident Coronary Artery Disease Events in a Cohort With Type 1 Diabetes. Journal of the American Heart Association. 2024;13(14).
50. Vaisar Tomas, Babenko Ilona, Horvath Katalin V., Niisuke Katrin, Asztalos Bela F. Relationships between HDL subpopulation proteome and HDL function in overweight/obese people with and without coronary heart disease. Atherosclerosis. 2024;397.
51. Begue F., Apalama M. L., Lambert G., Meilhac O. HDL as a Treatment Target: Should We Abandon This Idea? Curr Atheroscler Rep. 2023;25(12):1093-9.
52. Gibson C. Michael, Duffy Danielle, Korjian Serge, Bahit M. Cecilia, Chi Gerald, Alexander John H., et al. Apolipoprotein A1 Infusions and Cardiovascular Outcomes after Acute Myocardial Infarction. New England Journal of Medicine. 2024;390(17):1560-71.
53. Ballantyne Christie M., Nambi Vijay. HDL Therapeutics — Time for a Curtain Call or Time to Reconceptualize? New England Journal of Medicine. 2024;0(0).
54. Johannesen Camilla D. L., Mortensen Martin B., Langsted Anne, Nordestgaard Børge G. ApoBandNon-HDLCholesterol VersusLDLCholesterol for Ischemic Stroke Risk. Annals of Neurology. 2022;92(3):379-89.
55. Yuan S., Tang B., Zheng J., Larsson S. C. Circulating Lipoprotein Lipids, Apolipoproteins and Ischemic Stroke. Ann Neurol. 2020;88(6):1229-36.
56. Holme I., Aastveit A. H., Hammar N., Jungner I., Walldius G. Relationships between lipoprotein components and risk of ischaemic and haemorrhagic stroke in the Apolipoprotein MOrtality RISk study (AMORIS). Journal of Internal Medicine. 2009;265(2):275-87.
57. Dong H., Chen W., Wang X., Pi F., Wu Y., Pang S., et al. Apolipoprotein A1, B levels, and their ratio and the risk of a first stroke: a meta-analysis and case-control study. Metab Brain Dis. 2015;30(6):1319-30.
58. Ruiz-Canela M., Toledo E., Clish C. B., Hruby A., Liang L., Salas-Salvado J., et al. Plasma Branched-Chain Amino Acids and Incident Cardiovascular Disease in the PREDIMED Trial. Clin Chem. 2016;62(4):582-92.
59. Holmes M. V., Millwood I. Y., Kartsonaki C., Hill M. R., Bennett D. A., Boxall R., et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. J Am Coll Cardiol. 2018;71(6):620-32.
60. McGarrah R. W., White P. J. Branched-chain amino acids in cardiovascular disease. Nat Rev Cardiol. 2022.
61. Wang T. J., Larson M. G., Vasan R. S., Cheng S., Rhee E. P., McCabe E., et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448-53.
62. Newgard Christopher B., An Jie, Bain James R., Muehlbauer Michael J., Stevens Robert D., Lien Lillian F., et al. A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. Cell Metabolism. 2009;9(4):311-26.
63. Wang Xiang-Ping, Yan Dan, Jin Xia-Ping, Zhang Wen-Yan, Shi Tao, Wang Xiang, et al. The role of amino acid metabolism alterations in acute ischemic stroke: From mechanism to application. Pharmacological Research. 2024;207.
64. Lai T. W., Zhang S., Wang Y. T. Excitotoxicity and stroke: identifying novel targets for neuroprotection. Prog Neurobiol. 2014;115:157-88.
65. Lopaschuk Gary D., Dyck Jason R. B. Ketones and the cardiovascular system. Nature Cardiovascular Research. 2023;2(5):425-37.
66. Gormsen L. C., Svart M., Thomsen H. H., Sondergaard E., Vendelbo M. H., Christensen N., et al. Ketone Body Infusion With 3-Hydroxybutyrate Reduces Myocardial Glucose Uptake and Increases Blood Flow in Humans: A Positron Emission Tomography Study. J Am Heart Assoc. 2017;6(3).
67. Feng G., Wu Z., Yang L., Wang K., Wang H. beta-hydroxybutyrate and ischemic stroke: roles and mechanisms. Mol Brain. 2024;17(1):48.

68. Shemesh E., Chevli P. A., Islam T., German C. A., Otvos J., Yeboah J., et al. Circulating ketone bodies and cardiovascular outcomes: the MESA study. Eur Heart J. 2023;44(18):1636-46.
69. Nordestgaard B. G., Langsted A., Mora S., Kolovou G., Baum H., Bruckert E., et al. Fasting is not routinely required for determination of a lipid profile: clinical and laboratory implications including flagging at desirable concentration cut-points-a joint consensus statement from the European Atherosclerosis Society and European Federation of Clinical Chemistry and Laboratory Medicine. Eur Heart J. 2016;37(25):1944-58.