The impact of fatty acids biosynthesis on the risk of cardiovascular diseases in Europeans and East Asians:

A Mendelian randomization study

Maria-Carolina Borges,[1,2,*]

Phillip Haycock,[1,2]

Jie Zheng,[1,2]

Gibran Hemani,[1,2]

Laurence J Howe,[1,2]

A Floriaan Schmidt,[3,4]

James R Staley,[1,2]

R Thomas Lumbers,[5,6,7]

Albert Henry,[3,5,7]

Rozenn N Lemaitre,[8]

Tom R Gaunt,[1,2]

Michael V Holmes,[9,10]

George Davey-Smith,[1,2]

Aroon D Hingorani,[3,6,7,#]

Deborah A Lawlor,[1,2,11,#]

[#] These authors contributed equally to this work

1 MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, UK.

2 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2PN, UK.

3 Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, WC1E 6DD, UK.

4 Department of Cardiology, Division Heart and Lungs, UMC Utrecht, Utrecht, 3584 CX, Netherlands.

5 Institute of Health Informatics, University College London, London, NW1 2DA, UK.

6 Health Data Research UK London, University College London, NW1 2DA, UK.

7 UCL British Heart Foundation Research Accelerator, London, NW1 2DA, UK.

8 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, WA 98101, USA.

9 Medical Research Council Population Health Research Unit, University of Oxford, Oxford, OX3 7LF, UK.

10 Clinical Trial Service and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK.

11 NIHR Bristol Biomedical Research Centre, Bristol, BS8 2BN, UK.

*** Corresponding author details:**

Address: Oakfield Grove, Clifton, Bristol BS8 2BN.

Telephone: +44 (0)117 331 0098

Email: m.c.borges@bristol.ac.uk.

Twitter handle: @MCarol_Borges

Short title: Mendelian randomization, fatty acids and cardiovascular diseases

ABSTRACT

Despite early interest, the evidence linking fatty acids to cardiovascular diseases remains controversial. We used Mendelian randomization to explore the involvement of polyunsaturated (PUFA) and monounsaturated (MUFA) fatty acids biosynthesis in the aetiology of several cardiovascular disease endpoints in up to 1,153,768 European (maximum 123,668 cases) and 212,453 East Asian (maximum 29,319 cases) ancestry individuals. As instruments, we selected single nucleotide polymorphisms (SNP) mapping to genes with well-known roles in PUFA (i.e. *FADS1/2* and *ELOVL2*) and MUFA (i.e. *SCD*) biosynthesis. Our findings suggest that higher PUFA biosynthesis rate (proxied by rs174576 near *FADS1/2*) is related to higher odds of multiple cardiovascular diseases, particularly ischemic stroke, peripheral artery disease and venous thromboembolism, whereas higher MUFA biosynthesis rate (proxied by rs603424 near *SCD*) is related to lower odds of coronary artery disease among Europeans. Results were unclear for East Asians as most effect estimates were imprecise. By triangulating multiple approaches (i.e. uni-/multi-variable Mendelian randomization, a phenome-wide scan, genetic colocalization and within-sibling analyses), our results are compatible with higher low-density lipoprotein (LDL)-cholesterol (and possibly glucose) being a downstream effect of higher PUFA biosynthesis rate. Our findings indicate that PUFA and MUFA biosynthesis are involved in the aetiology of cardiovascular diseases and suggest LDL-cholesterol as a potential mediating trait between PUFA biosynthesis and cardiovascular diseases risk.

INTRODUCTION

Fatty acids constitute the main components of dietary fats and are required in human nutrition as a source of energy and for metabolic and structural activities (1). They are capable of influencing a wide range of cell signalling pathways and have been implicated in the regulation of several processes involved in the aetiology of cardiovascular diseases, including lipid metabolism (2-4), glucose homeostasis (5, 6), blood pressure (7-9), inflammatory response (10-12), and endothelial function (9, 13). Fatty acids are commonly subdivided into broad classes according to the degree of unsaturation (i.e., number of carbon-carbon double bonds) into saturated (SFA), monounsaturated (MUFA) and polyunsaturated (PUFA) fatty acids, the latter being classified as omega-3 or omega-6 PUFA depending on the position of the first double bond from the terminal methyl group.

Some fatty acids can be synthesized endogenously by fatty acid synthase or taken up by diet and further elongated and desaturated into longer chain fatty acids by fatty acid elongases and desaturases, respectively (14). Genome-wide association studies (GWAS) have reported that circulating fatty acids are strongly influenced by genetic variants near genes coding fatty acid elongases and desaturases: fatty acid desaturase 1 (*FADS1* - ENSG00000149485), fatty acid desaturase 2 (*FADS2* - ENSG00000134824), elongase 2 (*ELOVL2* - ENSG00000197977), and stearoyl-CoA desaturase (*SCD* - ENSG00000099194) (15-21). The chemical reactions and pathways catalysed by the enzymes encoded by *FADS1*, *FADS2*, *ELOVL2*, and *SCD* are summarised in **Figure 1**.

Mendelian randomization uses genetic variants associated with putative risk factors as instruments to assess their involvement in disease aetiology (22-24). The use of human genetics to explore the effect of modifiable risk factors on cardiometabolic diseases, such as in

Mendelian randomization, has proven valuable to (de)prioritise targets for intervention and to assess potential target-mediated adverse effects reducing late-stage failures in RCTs due to lack of efficacy or from target-mediated adverse reactions (25).

Genetic variants affecting the expression or activity of genes encoding for fatty acid elongases and desaturases (e.g. *FADS1/2*, *ELOVL2*, and *SCD*) can be used as causal anchors in Mendelian randomization studies investigating the involvement of fatty acids in the development of cardiovascular diseases. Most previous Mendelian randomization studies investigating the role of fatty acids on the risk of cardiovascular diseases have solely or heavily relied on genetic variants within the locus harbouring *FADS1* and *FADS2*, which are involved in PUFA synthesis by encoding the enzymes delta-5 desaturase (D5D) and delta-6 desaturase (D6D), respectively. Overall, these studies have reported that shorter chain PUFA (e.g. α-linolenic acid (ALA) and linoleic acid (LA)) and longer chain PUFA (e.g. arachidonic acid (AA)) are associated with lower and higher risk of cardiovascular diseases, respectively (26-32).

These studies potentially strengthen the evidence on the involvement of fatty acids in the development of cardiovascular diseases given the well-established link of D5D/D6D with PUFA biosynthesis. However, such studies suffer from a critical limitation given *FADS1/2* variants are not reliable instruments for individual fatty acids. First, *FADS1/2* variants will affect multiple fatty acids on the same pathway and, in some cases, on different pathways with reactions catalysed by D5D/D6D (**Figure 1**). Second, these studies have not extensively explored whether the association of *FADS1/2* variants with cardiovascular diseases risk could be explained by biological pathways independent of fatty acids (e.g. if variants simultaneously influence the expression of other genes in the region that affect cardiovascular diseases) or due to confounding by *linkage disequilibrium* (LD), population stratification or other familial mechanisms.

The aim of this study was to use Mendelian randomization to explore the effect of fatty acids biosynthesis on a wide range of cardiovascular disease end-points in up to 1,153,768 European and 212,453 East Asian ancestry individuals. We extend work in previous studies by using genetic variants regulating multiple rate-limiting enzymes in fatty acids biosynthesis (i.e. D5D/D6D, ELOVL2 and SCD), comparing findings between Europeans and East Asians and extensively exploring the key scenarios that could lead to spurious findings in this and previous Mendelian randomization studies.

RESULTS

Genetic instruments indexing fatty acids biosynthesis

We selected genetic variants mapping to genes with a well-known role in fatty acids biosynthesis (i.e. *FADS1/2*, *ELOVL2*, and *SCD*). To circumvent limitations from previous studies, we used genetic variants to instrument for enzyme activity in a given fatty acids biosynthesis pathway (rather than for individual fatty acids) by deriving the ratio between fatty acids that are the product and the substrate of a reaction catalysed by the corresponding enzyme. This allows harnessing the advantages of cis-acting variants in the vicinity of genes coding for key enzymes in fatty acids biosynthesis pathways and can provide more credible evidence on the likely therapeutic effect of targeting such proteins in preventing cardiovascular diseases (33).

In individuals of European ancestry, the selected genetic variants were rs174546 (*FADS1*, chr11q13.3), rs174576 (*FADS2*, chr11q13.3), rs3734398 (*ELOVL2*, chr6q15), and rs603424 (*SCD*, chr10q22.1), which explained a proportion of the variance in the corresponding marker of enzyme activity of 32.6% (F = 4174) for AA:DGLA (i.e. ratio

between AA and dihomo-γ-linolenic acid (DGLA)), 6.3% (F = 580) for GLA:LA (i.e. the ratio between γ-linolenic acid (GLA) and LA), 2.4% (F = 218) for DHA:n-3 DPA (i.e. the ratio between docosahexaenoic acid (DHA) and omega-3 docosapentaenoic acid (DPA)), and 1.1% (F = 100) for POA:PA (i.e. the ratio between palmitoleic acid (POA) and palmitic acid (PA)), respectively (**Supplementary table 1**). The *FADS1/2* SNPs (i.e. rs174546 and rs174576) were in strong LD (R^2 = 0.93 1000 Genomes European population), and, therefore, only the SNP more strongly associated with the corresponding marker of enzyme activity was used in subsequent analyses (i.e. rs174546).

In individuals of East Asian ancestry, the top variant in the *FADS1/2* locus was palindromic and, therefore, was replaced by rs174546 (i.e. LD R^2 = 0.93 in 1000G East Asian population), which explained 8.4% (F = 125) of the variance in DGLA:LA, a marker of D6D activity (**Supplementary table 1**). No genetic variants were associated with markers of D5D, ELOVL2 or SCD activity and, therefore, rs174546 was the only genetic variant eligible for further analyses in East Asians.

Impact of genetic instruments on circulating fatty acids

Overall, the effect of genetic variants on the fatty acids pool was replicable across independent samples and between Europeans and East Asians. As expected, the genetic variants impact on the fatty acids pool was consistent with their predicted function on fatty acids biosynthesis. The *FADS1/2* SNP (rs174546) was associated with a lower concentration of shorter chain omega-3 (e.g. ALA) and omega-6 (e.g. LA) fatty acids and higher concentration of longer chain omega-3 (e.g. DHA) and omega-6 (e.g. AA) fatty acids. The *ELOVL2* SNP (rs3734398) was mostly associated with higher concentration of DHA and lower concentration of eicosapentaenoic acid (EPA) and n-3 DPA, whereas the *SCD* SNP (rs603424)

was related to lower SFA, particularly PA, and higher MUFA, particularly POA (**Supplementary figures 1** and **2**).

Relation between fatty acids biosynthesis and risk of cardiovascular diseases

We used two-sample Mendelian randomization to probe the lifelong effect of fatty acids biosynthesis on cardiovascular diseases risk and risk factors in individuals of European and East Asian ancestry. Confounding by LD is a key source of bias in Mendelian randomization analyses using one or few independent genetic variants (**Figure 2**) and occurs when the selected genetic instrument is correlated (i.e. in LD) with another genetic variant influencing the outcome independently. Therefore, we used genetic colocalization to tease apart whether results from Mendelian randomization analyses were compatible with a shared variant between enzyme activity markers and cardiovascular disease outcomes or with confounding by LD.

FADS locus: D5D activity in Europeans

Mendelian randomization analyses in European ancestry individuals suggested that higher D5D activity (proxied by increase in AA:DGLA in standard deviation (SD) units) was related to higher odds of multiple cardiovascular diseases, such as coronary artery disease (OR = 1.02; 95% CI: 1.01, 1.03; p-value = 0.006), ischemic stroke (OR = 1.03; 95% CI: 1.01, 1.05; p-value = 0.004), heart failure (OR = 1.02; 95% CI: 1.01, 1.04; p-value = 0.008), atrial fibrillation (OR = 1.02; 95% CI: 1.00, 1.03; p-value = 0.04), peripheral artery disease (OR = 1.08; 95% CI: 1.04, 1.12; p-value = $1 \, x \, 10^{-5}$), venous thromboembolism (OR = 1.07; 95% CI: 1.05, 1.09; p-value = $5 \, x \, 10^{-9}$), and aortic valve stenosis (OR = 1.08; 95% CI: 1.01, 1.15; p-value = 0.02). Only results for ischemic stroke, peripheral artery disease, and venous thromboembolism passed our threshold for multiple testing correction (**Figure 3**). Overall,

results were consistent across studies, except for coronary artery disease, for which the estimated effect was attenuated in UK Biobank compared to other studies, and for aortic aneurysm, for which the estimated effect was in different directions between UK Biobank and other studies (**Supplementary figure 3**). There was strong evidence of genetic colocalization between D5D activity and risk for venous thromboembolism as evidenced by a posterior probability of association (PPA) of 85% for a shared variant. For other cardiovascular disease outcomes, PPA was 0%-27% for a shared variant accompanied by PPA of 60%-100% for the variant being associated with D5D activity only, which could be a result of limited statistical power (**Table 1** and **Figure 4**).

Mendelian randomization analyses indicated that higher D5D activity was related to higher LDL-cholesterol, fasting glucose, and type 2 diabetes risk, but lower triglycerides and diastolic blood pressure among individuals of European ancestry (**Figure 5**). Genetic colocalization provided evidence for a shared variant between D5D activity and LDL-cholesterol (PPA for shared variant = 87%) but not for systolic, diastolic blood pressure, and triglycerides (PPA for distinct variants = 90-100%). Evidence was less conclusive for glucose glucose (PPA for shared variant = 64%) (**Table 2** and **Figure 6**). In sensitivity analyses, support for a shared variant did not increase after conditioning the outcome genetic association data on the top genetic variant for the outcome (**Supplementary table 2**).

FADS locus: D6D activity in East Asians

In East Asian ancestry individuals, there was limited evidence from Mendelian randomization supporting a relationship between D6D activity (proxied by DGLA:LA in SD units) and the odds of cardiovascular endpoints. However, statistical power was substantially lower for analyses in East Asian individuals and, therefore, some findings could be compatible

with higher D6D activity being related to higher odds of disease, such as for atrial fibrillation (OR = 1.02; 95% CI: 0.92, 1.13; p-value = 0.68), or to lower odds of diseases, such as for coronary artery disease (OR = 0.96; 95% CI: 0.91, 1.00; p-value = 0.04) and haemorrhagic stroke (OR = 0.85; 95% CI: 0.75, 0.97; p-value = 0.01) (**Figure 3**). Higher D6D activity was related to higher LDL-cholesterol, fasting glucose, and type 2 diabetes risk, but lower triglycerides and diastolic blood pressure among individuals of East Asian ancestry (**Figure 5**). We could not assess confounding by LD since genetic colocalization assumes that samples are drawn from independent populations of similar allele frequencies and LD pattern, which was not the case for East Asians in our analyses as genetic association data for fatty acids and cardiovascular disease data were derived from Singaporean Chinese and Japanese individuals, respectively.

ELOVL2 locus: ELOVL2 activity in Europeans

Mendelian randomization analyses did not support a relationship between higher ELOVL2 activity (proxied by increase in DHA:DPAn-3 in SD units) and cardiovascular endpoints. However, some results were imprecisely estimated and, therefore, we cannot rule out the presence of potentially important effects, particularly for haemorrhagic stroke (OR = 0.86; 95% CI: 0.71, 1.05; p-value = 0.14) and aortic valve stenosis (OR = 1.17; 95% CI: 0.92, 1.48; p-value = 0.20) (**Figure 3** and **Supplementary figure 3**). Higher ELOVL2 activity was not related to cardiovascular risk factors at p-value < 0.00625 (**Figure 5**).

SCD locus: SCD activity in Europeans

Higher SCD activity (proxied by increase in POA:PA in SD units) was related to lower odds of coronary artery disease (OR = 0.82; 95% CI: 0.76, 0.88; p-value = 1×10^{-7}) in

Mendelian randomization analyses (**Figure 3**), which was consistent across studies (**Supplementary figure 3**). There was limited evidence supporting a relationship between higher SCD activity and other cardiovascular endpoints, although some of these results were imprecisely estimated and, therefore, we cannot rule out the presence of potentially important effects, particularly for peripheral artery disease (OR = 0.85; 95% CI: 0.68, 1.06; p-value = 0.14), aortic aneurysm (OR = 1.16; 95% CI: 0.88, 1.52; p-value = 0.30), and aortic valve stenosis (OR = 0.85; 95% CI: 0.60, 1.22; p-value = 0.38) (**Figure 3** and **Supplementary figure 3**). There was strong evidence that SCD activity colocalised with odds of coronary artery disease (PPA = 99% for a shared variant) (**Table 1** and **Figure 4**).

Higher SCD activity was related to lower LDL-cholesterol, triglycerides, systolic and diastolic blood pressure (**Figure 5**); however, colocalization analyses supported distinct genetic variants between POA:PA and these endpoints (PPA for distinct variants = 86-100%) (**Table 2** and **Figure 6**). In sensitivity analyses, support for a shared variant did not increase after conditioning the outcome genetic association data on the top genetic variant for the outcome in the genomic region, except for DBP (PPA = 99%) (**Supplementary table 2**).

Exploring bias in Mendelian randomization analyses

Apart from confounding by LD, other key sources of bias could invalidate inferences from this and previous Mendelian randomization studies as detailed in **Figure 2**, including: horizontal pleiotropy, where the genetic variant influences the outcome via a different biological pathway; confounding by population stratification, assortative mating or indirect genetic effects, which could create a spurious association between genetic variant and outcome in samples of unrelated individuals; and selection bias, where the genetic variant (or, more likely, its downstream traits) and the outcome affect selection into the sample resulting in a

194 spurious association. We conducted extensive sensitivity analyses to explore the presence of
195 such biases in our findings as detailed below.

196

197 Horizontal pleiotropy

198 Horizontal pleiotropy is one of the main threats to the validity of Mendelian
199 randomization studies since it is a widespread biological phenomenon and cannot be
200 empirically verified. We used two approaches to explore the plausibility that our results are
201 explained by horizontal pleiotropy: (i) a phenome-wide scan of the selected genetic variants
202 using data from European and East Asian ancestry individuals and (ii) multivariable Mendelian
203 randomization to estimate the direct effect of *FADS1*, *ELOVL2* and *SCD* expression on
204 cardiovascular outcomes after accounting for the potential effect of other genes expressed in
205 the corresponding genomic region using data from European ancestry individuals only (tissue-
206 specific gene expression data was not available for East Asians).

207 In the phenome-wide scan, the *FADS1/2* variant (rs174546) was related not only to fatty
208 acids but also to numerous non-fatty acid traits such as lipid, glycaemic, blood cell traits,
209 physical measures (e.g. pulse, heart rate and height), immune-related disorders (e.g. asthma,
210 hypothyroidism, Crohn's disease, inflammatory bowel disease) and several biomarkers (e.g.
211 total bilirubin, insulin growth factor-1, cystatin C, alkaline phosphatase and urate) among
212 individuals of European ancestry (**Figure 7** and **Supplementary table 3**). The pleiotropic
213 associations of the *FADS1/2* variant (rs174546) were also seen in East Asians in relation to
214 lipid, glycaemic, blood cell traits (**Figure 7** and **Supplementary table 4**). The *ELOVL2* variant
215 (rs3734398) was related to levels of an unknown metabolite X-12627 and DHA and the *SCD*
216 variant (rs603424) was related to multiple SFA/MUFA, as well as to bone mineral density and
217 blood cell-related traits (**Figure 7** and **Supplementary table 3**).

Overall, the selected genetic instruments were strongly associated with the tissue expression of the target genes in the expected direction (**Supplementary figure 4**), except for *FADS1* in whole blood, and genetic colocalization supported a shared variant between the enzyme activity proxies and the expression of the target gene in key tissues (**Supplementary figure 5** and **Supplementary table 5**). As an example, there was strong evidence that SCD activity (proxied by POA:PA) colocalised with *SCD* expression in adipose tissues (PPA = 100% for a shared variant), which are key tissues for *de novo* lipogenesis. The selected genetic variants were also associated with the expression of nearby non-target genes, which was particularly the case for the *FADS* variant (**Supplementary figure 4**). The association of the genetic variants with the expression of non-target genes could bias our analyses if the proteins encoded by these genes directly influence cardiovascular diseases. To explore that, we used multivariable Mendelian randomization, which supported a direct effect of the target genes (i.e. *FADS1*, *ELOVL2* and *SCD*) on cardiovascular diseases and risk factors in individuals of European ancestry (**Supplementary figure 6** and **7**). The conditional F statistics for these analyses ranged from 18 to 433 and 5 to 50 in unadjusted and adjusted models, respectively (**Supplementary table 6**).

Confounding by population stratification, assortative mating and indirect genetic effects

Mendelian randomization studies generally assume that genetic association estimates reflect the direct effect of a genetic variant on a phenotype, i.e., the downstream effect of inheriting an allele. However, there is growing evidence that genetic association estimates obtained from samples of unrelated individuals may also capture non-direct sources of association relating to population stratification, assortative mating and indirect genetic effects of parents (34). Of particular concern for this study, there is evidence that the *FADS1/2* locus was under important selection pressure in different populations and at different times, possibly

as a response to dietary changes and the need for adequate supply of essential long-chain PUFA from precursors (35). Despite attempts to control for population stratification in genetic association data (e.g. by adjusting for genomic principal components), there could still be residual population structure as has recently been shown in UK Biobank (36). In addition, there is a possibility that indirect genetic effects of parents bias studies among unrelated individuals given the literature suggesting that maternal genotype for *FADS1/2* variants might indirectly influence offspring outcomes via intrauterine effects and/or breastfeeding (37).

We used two approaches to explore the potential impact of confounding by population stratification, assortative mating and indirect genetic effects in our analyses: (i) testing for the association of the selected genetic instruments with two negative control outcomes (i.e. skin colour and ease of skin tanning), and (ii) comparing within-sibling associations of the selected genetic variants with cardiovascular risk factors with estimates obtained from unrelated individuals.

The *FADS1/2* SNP (rs174546) was associated with both negative control outcomes among Europeans: skin colour [mean change of 0.005 unit per C allele (p-value = 5×10^{-5})] and ease of skin tanning [mean change of -0.006 unit per C allele (p-value = 0.007)], while the *SCD* SNP was associated with ease of skin tanning [mean change of -0.006 unit increase per C allele (p-value = 0.037)] (**Supplementary table 7**). Since these traits could not conceivably be affected by fatty acids biosynthesis, evidence for an association between genetic variants and these negative control outcomes is indicative of residual population stratification.

We compared within-sibship associations of the selected genetic variants with cardiovascular risk factors with estimates obtained from unrelated individuals. Estimates were broadly consistent indicating that our findings are unlikely to be substantially biased by population stratification, assortative mating or indirect genetic effects (**Figure 8**). As an example, for the *FADS1/2* SNP (rs174546), each C allele was related to a mean LDL-

cholesterol increase of 0.041 (95% CI: 0.025; 0.057) and 0.036 (95% CI: 0.025; 0.046) standard deviation units within-siblings and in unrelated individuals, respectively.

Selection bias

Several processes of sample selection, occurring from study design to data analyses, can result in selected samples not representative of their target populations, which may bias causal inference, including when using Mendelian randomization (38). We were particularly concerned about selection due to ascertainment of cardiovascular disease status as detailed in Materials and Methods. To explore whether these processes of sample selection could bias our findings, we adopted a positive exposure control approach in which we used Mendelian randomization to estimate the effect of well-established cardiovascular risk factors (i.e. LDL-cholesterol, triglycerides, systolic, diastolic blood pressure, glucose, type 2 diabetes, smoking, and body mass index) on the risk of cardiovascular diseases. If the effects estimated in the positive control analyses were compatible with what expected and were comparable across data sources, such analyses would argue against selection being a major source of bias.

Overall, we observed the expected effect of well-established risk factors on the development of cardiovascular diseases across studies (**Supplementary figure 8**). Systolic, diastolic blood pressure and body mass index were related to higher odds of all cardiovascular disease outcomes in studies of European ancestry individuals (i.e. UK Biobank, genetic association metanalyses and FinnGen) and higher odds of most cardiovascular outcomes (except for coronary and peripheral artery disease) in Biobank Japan. Higher LDL-cholesterol and triglycerides, and liability to type 2 diabetes were related to higher odds of coronary artery disease and peripheral artery disease across studies for both ancestries, while glucose and smoking were related to higher odds of peripheral artery disease in Europeans and East Asians.

There were a few instances where these risk factors were related to lower odds of disease, such as type 2 diabetes liability with hemorrhagic stroke (Biobank Japan) and LDL-cholesterol with hemorrhagic stroke (UK Biobank and Biobank Japan).

DISCUSSION

Main findings

In Europeans, our findings indicate that higher PUFA biosynthesis (proxied by *FADS1*/D5D activity) is related to higher risk of several cardiovascular diseases (and risk factors), while higher MUFA biosynthesis (proxied by *SCD*/SCD activity) is related to lower risk of coronary artery disease. In addition, despite the strong LD in the *FADS1/2* region, our results indicate that the relation between PUFA biosynthesis and cardiovascular diseases is driven by changes in *FADS1* (not *FADS2*) expression among Europeans. In East Asians, the same *FADS1/2* variant was related to similar pleiotropic effects on the phenome (e.g. lipid, glycaemic, blood cell traits) compared to Europeans, although the relation with cardiovascular diseases was unclear as most effect estimates were either imprecisely estimated (e.g. atrial fibrillation) or, for coronary artery disease, in the opposite direction in East Asians compared to results in Europeans.

By triangulating multiple approaches, our results are compatible with higher LDL-cholesterol (and possibly glucose) being a downstream effect of higher D5D activity (coded by *FADS1*) instead of explained by confounding by LD or by the co-expression of other genes in the region. Given the well-established role of *FADS1*/D5D activity in PUFA biosynthesis and the well-known involvement of LDL-cholesterol in the aetiology of multiple cardiovascular

diseases, this strengthens the evidence for a causal relationship and provides a putative mediating pathway for the effect of PUFA biosynthesis on the risk of cardiovascular diseases.

Previous literature

The relation between fatty acids and cardiovascular diseases has been explored in classical observational studies, randomized controlled trials and Mendelian randomization studies. Most previous studies have focused on coronary artery disease and, to a lesser extent, on stroke; therefore, other types of cardiovascular disease endpoints, such as heart failure and atrial fibrillation, remained under explored

Previous meta-analyses of classical observational studies indicate that higher circulating long-chain omega-3 and omega-6 PUFA are either not associated or are associated with lower risk of coronary artery disease and stroke (39-43), whereas higher circulating MUFA and SFA are either not associated or are associated with higher risk of coronary artery disease and stroke (39, 40). Recent systematic reviews of randomized controlled trials of dietary advice or supplementation of omega-3 and omega-6 PUFA have suggested little to no benefit in reducing the risk of cardiovascular diseases (44-46). However, most studies included in these systematic reviews were at moderate to high risk of bias and there is large uncertainty on the evidence linking PUFA to some cardiovascular outcomes (44-46). It is important to emphasise that comparing our findings to previous classical observational and randomized controlled trials deserves caution as our genetic instruments have a broad impact on the fatty acids pool and, therefore, cannot be used to make inferences about individual fatty acids/fatty acids classes. As an example, higher D5D activity (instrumented by rs174546) is related to higher longer chain omega-3 and omega-6 PUFA (e.g. AA, EPA and DHA) but lower shorter chain omega-3 and omega-6 PUFA (e.g. LA and ALA).

Several GWAS have reported that SNPs within the *FADS1/2* locus are associated with cardiovascular risk factors (e.g. LDL-cholesterol and triglycerides) (19, 47, 48) and previous Mendelian randomization studies have reported that longer and shorter chain PUFA are related to risk of cardiovascular diseases in contrasting directions among Europeans, including coronary artery disease in CARDIoGRAMplusC4D and UK Biobank (26, 27, 49), ischemic stroke in MEGASTROKE and UK Biobank (28, 29), and venous thromboembolism in UK Biobank (29). Our findings expand on previous Mendelian randomization studies by implicating higher D5D activity in the development of a wide range of cardiovascular diseases among Europeans in the largest available samples to date (up to 1,153,768 individuals). In addition, to our knowledge, this is the first Mendelian randomization study to report a potential protective effect of higher SCD activity on coronary artery disease among Europeans and to explore the relation between D6D activity (coded by *FADS2*) and cardiovascular diseases among East Asians.

Plausibility of Mendelian randomization assumptions

A major challenge in Mendelian randomization studies is the unprovable assumption that the estimated effect of the genetic instrument on the outcome is mediated by the exposure, and not biased by horizontal pleiotropy, population stratification, assortative mating, indirect genetic effects or selection bias (38, 50, 51). We assessed the plausibility that our findings were explained by these sources of bias through a series of sensitivity analyses.

To mitigate bias due to horizontal pleiotropy (i.e. the genetic instrument influences exposure and outcome via independent pathways), we have restricted our analyses to genetic variants near genes with well-established role in fatty acids biosynthesis. We have confirmed that these variants have the expected impact on the circulating fatty acids pool and on the

363 expression of the target genes in key tissues (except for *FADS1* in whole blood). Previous
364 evidence confirms that the selected *FADS1/2* and *SCD* variants, or variants in high LD, are
365 related to changes in fatty acids composition across multiple sites, including adipose tissue (52-
366 54), brain (55) and liver (56). Genetic colocalization and multivariable Mendelian
367 randomization (adjusting for co-expressed genes in the region) supported a causal relation
368 between D5D activity, venous thromboembolism and LDL-cholesterol, and between SCD
369 activity and coronary artery disease. It is important to note that we were likely underpowered
370 to test colocalization between fatty acids biosynthesis and cardiovascular disease outcomes.
371 Where there was evidence that the selected genetic variant was associated with the expression
372 of non-target genes in the region in a given tissue, findings from multivariable Mendelian
373 randomization were consistent with expression of the target gene (i.e. *FADS1*) having direct
374 effects on the outcome.

375 Confounding could be introduced in Mendelian randomization studies due to
376 population stratification, assortative mating and indirect genetic effects. Of these factors,
377 population stratification is likely to be of the most concern for this study (51). Despite attempts
378 to control for population structure in genetic association data, there could still be residual
379 population structure (36). We showed that within-sibship associations of *FADS1/2*, *ELOVL2*,
380 and *SCD* variants with established cardiovascular risk factors were broadly similar to estimates
381 from unrelated individuals, suggesting that our results are unlikely to be affected by population
382 stratification, assortative mating or indirect genetic effects of parents.

383 Non-random sample selection may introduce bias in Mendelian randomization studies
384 especially if the mechanism of selection depends on the exposure and/or outcome (38, 57).
385 Using a positive control approach, we were able to identify the expected effect of well-
386 established risk factors on cardiovascular diseases across studies contributing with data on
387 cardiovascular disease endpoints, which is reassuring given our concerns that case-control

388 ascertainment could introduce bias in the analyses. Although results from the positive control
389 approach argues against selection being a major source of bias in this study, we cannot fully
390 rule out that selection might have introduced some bias in our analyses as bias due to selection
391 will depend on context-specific causal structures underlying the data under consideration.

392

393 *Implications*

394 Our findings are supportive of the involvement of fatty acids biosynthesis, especially
395 D5D and SCD activity, in the aetiology of cardiovascular diseases. Further work is needed to
396 understand the precise underlying mechanism(s).

397 The relation between D5D activity and cardiovascular diseases is plausibly mediated
398 by one or more fatty acids involved in the PUFA biosynthesis pathway. Given the ubiquitous
399 impact of higher D5D activity on the circulating PUFA pool, we cannot pinpoint which specific
400 fatty acids are driving these effects. For illustration, higher D5D activity decreases LA and
401 ALA (and other omega-3 and omega-6 PUFA upstream of the reaction catalysed by D5D).
402 Lower LA may relate to unfavourable metabolic changes, such as higher plasma LDL-
403 cholesterol, apolipoprotein B, and triglycerides, and haemoglobin A1c (2, 58) and, therefore,
404 is a plausible mediator of the relation between higher D5D activity and higher cardiovascular
405 diseases risk. On the other hand, higher D5D activity increases long-chain PUFA such as AA,
406 which influences key membrane/tissue functions, such as membrane fluidity, the activity of
407 membrane-bound receptors, transport proteins and signal transmission (59), and is a precursor
408 for eicosanoids (e.g. prostaglandins, leukotrienes, and thromboxane), which are involved in
409 inflammation, platelet aggregation and vascular remodeling (60).

410 The putative mechanisms underpinning the relation SCD activity and coronary artery
411 disease in humans are unclear. *Scd-1* deficient rodents are protected against diet-induced

obesity, insulin resistance, and hepatic steatosis (61-63), but show increased inflammation and atherogenesis (63, 64). The putative protective effect of higher SCD activity on coronary artery disease might be related to lower availability of palmitic acid and consequent lower production of its toxic metabolites, such as ceramides (65).

Conclusions

We found supportive evidence for an involvement of PUFA and MUFA biosynthesis in the aetiology of cardiovascular diseases. Our study illustrates the power of integrating multiple approaches to improve causal inference on the role of modifiable risk factors in the development of cardiovascular diseases.

MATERIALS AND METHODS

Data sources

The study included data from multiple consortia of genetic association studies (66-70) and biobanks (71-77).

Genetic associations with cardiovascular diseases

The outcomes of interest were (prevalent/incident) coronary artery disease, ischemic stroke, haemorrhagic stroke, heart failure, atrial fibrillation, peripheral arterial disease, aortic aneurysm, venous thromboembolism, and aortic valve stenosis.

Summary data for the association between genetic variants and these cardiovascular disease endpoints was obtained from UK Biobank, FinnGen (release 4), BioBank Japan and

several large-scale GWAS of cardiovascular disease outcomes. If genetic association data on a cardiovascular endpoint were available from two or more independent datasets of individuals from the same genetic ancestry (i.e. UK Biobank and FinnGen), genetic association estimates were pooled across data sources using fixed-effect meta-analysis with inverse variance weights. Characteristics of studies and criteria for case definition are detailed in **Supplementary table 8** and **Supplementary methods**.

For individuals of European ancestry only/predominantly (i.e. UK Biobank, FinnGen and large-scale genetic association consortia), data were available on all outcomes of interest: coronary artery disease (N cases/controls = 123,668/702,156), ischemic stroke (N cases/controls = 53,395/1,030,253), haemorrhagic stroke (N cases/controls = 4,558/627,188), heart failure (N cases/controls = 64,696/1,089,072), atrial fibrillation (N cases/controls = 77,945/1,067,430), peripheral arterial disease (N cases/controls = 9,836/627,950), aortic aneurysm (N cases/controls = 9,735/730,073), venous thromboembolism (N cases/controls = 25,284/616,235), and aortic valve stenosis (N cases/controls = 2,844/461,776).

For individuals of East Asian ancestry (i.e. BioBank Japan), cardiovascular outcomes data were available for coronary artery disease (N cases/controls = 29,319/183,134), ischemic stroke (N cases/controls = 17,671/192,383), haemorrhagic stroke (N cases/controls = 2,820/192,383), heart failure (N cases/controls = 9,413/203,040), atrial fibrillation (N cases/controls = 8,180/28,612), and peripheral arterial disease (N cases/controls = 3,593/208,860).

Genetic associations with cardiovascular disease risk factors

Other outcomes of interest were eight well-established risk factors for cardiovascular diseases (i.e. LDL-cholesterol, triglycerides, systolic, diastolic blood pressure, fasting glucose,

type 2 diabetes, smoking, and body mass index). Genetic association data for these risk factors were extracted for Europeans from a large-scale GWAS for type 2 diabetes (78) and UK Biobank for the other risk factors and for East Asians from BioBank Japan using the IEU OpenGWAS project database (79).

Genetic associations with circulating fatty acid concentration

For European ancestry individuals, we used genetic association data on circulating fatty acids from *The Cohorts for Heart and Aging Research in Genomic Epidemiology* (CHARGE) consortium, which has high resolution profiling of circulating fatty acids (N = 26 fatty acids measures) measured in 8,631-8,866 individuals (15-17). We also used data from two other genetic association meta-analyses on circulating fatty acids (18, 19) for assessing replication as detailed in 'Data analysis' under 'Assessing the impact of genetic instruments on the fatty acids pool'.

For East Asian ancestry individuals, we used genetic association data on fatty acids from the Singapore Chinese Health Study (SCHS) for circulating PUFA (N = 1,361) (21) and from a metanalysis of the Nutrition and Health of Aging Population in China (NHAPC) and the Chinese ancestry individuals of the Multi-Ethnic Study of Atherosclerosis (MESA) for RBC or circulating SFA and MUFA (N = 3,521) (80, 81).

Characteristics of these studies are detailed in **Supplementary table 9**.

Data analysis

Selection of genetic instruments indexing fatty acids biosynthesis

We selected genetic variants mapping to genes that have well-characterised roles in fatty acids biosynthesis and have been previously reported by GWAS to influence circulating fatty acids (**Figure 1**). In Europeans, three genomic regions were eligible, harbouring *FADS1/2*, *ELOVL2* and *SCD* genes, whereas, in East Asians, only the *FADS1/2* locus was strongly associated with circulating fatty acids, which may be related to the modest sample size available for East Asians (N = 1,361-3,521). *FADS1/2* were considered as one single genomic region since these genes are in close proximity to each other (i.e. 0.8 kb) on the long arm of human chromosome 11.

Genetic variants regulating the expression/activity of *FADS1/2*, *ELOVL2* and *SCD* will affect multiple fatty acids on the same pathway and, in some cases, on different pathways with reactions catalysed by the same enzymes (**Figure 1**). As a result, selecting genetic variants for individual fatty acids can be highly redundant. Instead, we selected the genetic variant (± 500 kB of the target gene) most strongly related (p-value $< 5 \, x \, 10^{-8}$) to a proxy of the enzyme activity (i.e. the ratio between fatty acids that are the product and the substrate of a reaction catalysed by a particular enzyme) within each genomic locus (**Table 3**). As an example, a higher ratio of AA to DGLA would indicate more active conversion due to higher expression/activity of D5D, the enzyme coded by *FADS1*.

For European ancestry individuals, we derived genetic association data for proxies of enzyme activity by applying the GWIS ("Genome-wide Inferred Study") method to genetic association data for circulating fatty acids from the CHARGE consortium (15-17) for the ratios of AA to DGLA (proxy of D5D activity), GLA to LA (proxy of D6D activity), DHA to DPA n-3 (proxy of ELOVL2 activity), and POA to PA (proxy of SCD activity). Briefly, GWIS approximates genetic association estimates for a new variable as a linear function of the allele frequencies, population means of measured traits (assumed to approximate the intercepts of the model) and genetic association estimates of measured traits. Corresponding standard errors can

506 be derived using the Delta-method having obtained the covariance matrix for effect estimates
507 (82).

508 For East Asian ancestry individuals, the original GWAS investigators derived genetic
509 association data for the proxies of enzyme activity from individual level data on circulating
510 fatty acids (21), as follows: AA to DGLA (proxy of D5D activity) and DGLA to LA (proxy
511 of D6D activity).

512 If the selected genetic variant was a palindromic SNP, it was replaced by a non-
513 palindromic proxy variant in strong LD to avoid data harmonisation problems in subsequent
514 analyses. All SNP-trait associations were harmonised so that the allele associated with
515 increasing enzyme activity was the effect allele, indicating more active conversion.

516 We approximated the R^2, a measure of the variance in exposure explained by the genetic
517 variant, and the F-statistics, a measure of instrument strength (83), as follows:

518 $$R^2 = 2 * {\beta_{gx}}^2 * MAF * (1 - MAF)$$

519 $$F = \left(\frac{n - k - 1}{k}\right) * \left(\frac{R^2}{1 - R^2}\right)$$

520 where:

521 β_{gx} = SNP-fatty acid trait association estimate (in standard deviation units)

522 MAF = minor allele frequency

523 n = sample size

524 k = number of SNPs

525

526 *Assessing the impact of genetic instruments on the fatty acids pool*

We assessed the impact of the selected genetic instruments on the circulating fatty acids pool in the discovery samples (i.e. CHARGE in Europeans and SCHS in East Asians) for internal validation. Among European ancestry individuals, we could test for replication in two independent datasets (external validation) (18, 19).

Mendelian randomization analysis

For each cardiovascular outcome, we used the Wald ratio method (84, 85) to estimate the odds ratio of disease for each standard unit increase in the proxy of enzyme activity by dividing estimates for the genetic association with cardiovascular outcome by estimates for the genetic association with enzyme activity as follows:

$$\beta_{MR} = \frac{\beta_y}{\beta_x}$$

and corresponding standard error:

$$SE_{MR} = \frac{SE_y}{\beta_x}$$

Where β_y and β_x are the coefficients for the association of the genetic variant with the outcome (Y) and the exposure (X), respectively, and SE_y is the standard error for the association of the genetic variant with Y.

We used a Bonferroni correction to account for the maximum number of outcomes available (p-value = 0.05/9 outcomes = 0.00556 in Europeans). The same approach was used to estimate the relation between enzyme activity and the eight well-established risk factors for cardiovascular diseases using Bonferroni correction to account for multiple testing (p-value = 0.05/8 risk factors = 0.00625). We use these p-value thresholds simply as a heuristic for highlighting associations worthy of follow-up. Mendelian randomization analyses were

549 performed using R software version 3.6.2 (R Foundation for Statistical Computing) including
550 the TwoSampleMR R package (86).

551

552 *Genetic colocalization*

553 We used *coloc* (87), a method for pairwise genetic colocalization analysis, to test
554 whether the same genetic variant influences fatty acids biosynthesis and cardiovascular
555 diseases risk or risk factor. *Coloc* enumerates all possible configurations of causal variants for
556 each of two traits (e.g. fatty acid- and cardiovascular disease-related traits), and uses a Bayesian
557 approach to calculate support for each causal model (H_1: association with trait 1 only; H_2:
558 association with trait 2 only; H_3: association with both traits due to distinct causal variants; H_4:
559 association with both traits due to a single shared causal variant; H_5: no association). We
560 restricted *coloc* analysis to a genomic region within a 500-kb window around each target gene
561 (*FADS1/2*, *ELOVL2*, and *SCD*) and assumed prior probabilities that any random SNP in the
562 region is associated with trait 1 ($p1 = 1 \times 10^{-4}$), trait 2 ($p2 = 1 \times 10^{-4}$), or both traits ($p12 =$
563 1×10^{-6}). A posterior probability of association (PPA) ≥ 70% for association with both traits
564 due to a single causal variant was considered as strong evidence for a shared genetic variant.

565 *Coloc* assumes a single causal variant in the genomic region, and, as a result, the
566 presence of multiple conditionally independent SNPs within a region can affect the
567 performance of the method. Therefore, where *Coloc* provided some evidence of distinct genetic
568 signals (i.e. PPA for distinct genetic variants > 30%), we also performed approximate
569 conditional analyses using GCTA (88, 89) (adjusting for the top SNP in the outcome dataset in
570 each genomic region) and re-ran *Coloc* using the adjusted association estimates as a sensitivity
571 analyses. Colocalization analyses were restricted to European datasets as the method assumes
572 that samples are drawn from independent populations of similar genetic background (i.e. allele

573 frequencies and LD pattern are identical), which was not the case for East Asians in our
574 analyses since fatty acids and cardiovascular disease data were derived from Singaporean
575 Chinese and Japanese individuals, respectively.

576

577 *Phenome-wide scan*

578 We explored the potential mechanisms that might link the selected genetic variants to
579 cardiovascular diseases by using an automated phenome-wide scan tool from the IEU
580 OpenGWAS project database (79) to test the association of the selected genetic variants with
581 32,534-34,465 (non-unique) traits for European ancestry individuals and 110 traits for Japanese
582 individuals from BioBank Japan. We used a Bonferroni correction to account for multiple
583 testing considering the maximum number of traits included in the phenome-wide scan in
584 Europeans (p-value = 0.05/34,465 = 1.5×10^{-6}) and East Asians (p-value = 0.05/110 = $4.5 \times$
585 10^{-4}).

586

587 *Gene expression and tissue-specific analyses*

588 We explored the influence of higher expression of the target genes (i.e. *FADS1*,
589 *ELOVL2* and *SCD*), and their tissue specificity, on cardiovascular diseases risk and risk factors
590 in individuals of European ancestry by integrating expression quantitative trait loci (eQTL)
591 data from Genotype-Tissue Expression (GTEx) version 8 with genetic association data for
592 cardiovascular traits. We extracted eQTL data from the GTEx for multiple tissues of relevance
593 to cardiovascular diseases — i.e. subcutaneous/visceral adipose tissues, aorta/coronary/tibial
594 arteries, heart, liver, pancreas and whole blood (N = 208-670 individuals per tissue) (90).

595 In step i, we performed a cross-tissue assessment of the association of the selected
596 genetic variants with transcription of any genes in the region (aka cis-genes), defined as genes

for which the transcription start site was 1 Mb away from the genetic variant. Cis-genes were selected for follow-up analysis if the P-value for the SNP-gene expression association was below the 5% false discovery rate (FDR) threshold for each tissue.

In the step ii, we used *coloc* to test whether the same genetic variant influences fatty acids biosynthesis and expression of the target gene across tissues using the same approach described in '*Genetic colocalization*'.

In the step iii, we used multivariable Mendelian randomization to jointly model the expression of a target gene (i.e. *FADS1/2*, *ELOVL2*, and *SCD*) and a co-expressed cis-gene (identified as described in "step i") on cardiovascular outcomes across tissues. This analysis allowed us to estimate the direct contribution of changes in the expression of each target gene where the selected genetic variant was related to co-expression of a non-target cis-gene. We selected independent eQTLs ($P < 5 \times 10^{-5}$; $R^2 < 0.05$ 1000 Genomes EUR reference population) for each combination of target gene (i.e. *FADS1/2*, *ELOVL2*, and *SCD*) and non-target cis-gene and performed multivariable Mendelian randomization using the MVMR R package (91). Multivariable Mendelian randomization models were estimated for each combination of target gene, co-expressed gene, outcome and tissue if (a) the target gene SNP (i.e. rs174546, rs2236212, and rs603424) was related to expression of non-target genes in that tissue (step i), (b) more than two independent SNPs were selected for the analyses, and (c) the conditional F statistics for the target gene expression was equal or higher than 5.

Negative control outcomes

We tested the association between the selected genetic variants with two negative control outcomes – i.e. skin colour and ease of skin tanning, using UK Biobank genetic association data deposited in the IEU Open GWAS Project (79). Since these traits could not

621 conceivably be affected by fatty acids biosynthesis, any evidence for an association between
622 genetic variants and these negative control outcomes would be indicative of residual population
623 stratification (92).

624

625 *Within-sibship analyses*

626 We used data from a recent within-sibship GWAS, including up to 178,076 individuals
627 (77,832 sibling pairs) from 23 cohorts, to evaluate if our findings are sensitive to population
628 stratification, assortative mating, and indirect genetic effects of parents. Within-family designs,
629 such as parent-offspring trio or within-sibship models, control for variation in parental
630 genotypes, and so are not affected by these potential biases (51, 93, 94).

631 We compared the within-sibship association of the selected genetic variants with
632 cardiovascular risk factors (LDL-cholesterol, triglycerides, systolic blood pressure, glycated
633 haemoglobin, smoking, and body mass index) with estimates from standard GWAS models in
634 unrelated individuals (sample size ranging from 50,361 for glycated haemoglobin to 155,457
635 for body mass index). Data on cardiovascular disease endpoints and other risk factors (i.e.
636 diastolic blood pressure, fasting glucose, and type 2 diabetes) were not available.

637

638 *Positive control exposures*

639 Several processes of sample selection, occurring from study design to data analyses,
640 can result in selected samples not representative of their target populations, which may bias
641 causal inference, including when using Mendelian randomization (38). We were particularly
642 concerned about selection due to ascertainment of cardiovascular disease status. As an
643 example, BioBank Japan is a hospital-based study, in which cases for cardiovascular diseases,
644 except atrial fibrillation, were compared to a control group including a mixture of hospital-

based (i.e. individuals diagnosed at health centres with other diseases) and community-based (i.e. individuals from population-based cohorts) controls as previously described (74). In addition, UK Biobank has a response rate of 5.5% and its participants have fewer self-reported health conditions and are more likely to be older, female, wealthier, leaner, non-smokers, non-drinkers than the general UK population (95).

To explore whether these processes of sample selection could bias our findings, we adopted a positive exposure control approach in which we used Mendelian randomization to estimate the effect of well-established cardiovascular risk factors (i.e. LDL-cholesterol, triglycerides, systolic, diastolic blood pressure, glucose, type 2 diabetes, smoking, and body mass index) on the risk of cardiovascular diseases. If the effects estimated in the positive control analyses were compatible with what expected and were comparable across data sources, such analyses would argue against selection being a major source of bias.

ACKNOWLEDGEMENTS

MCB was supported by a UK Medical Research Council (MRC) Skills Development Fellowship (MR/P014054/1). MCB and DAL are supported by the British Heart Foundation (AA/18/7/34219) and DAL is supported by a British Heart Foundation Chair (CH/F/20/90003). MCB, PH, JZ, GH, LJH, TRG, GDS, and DAL work in a Unit receives funding from the University of Bristol and UK Medical Research Council (MRC) (MC_UU_00011/1, MC_UU_00011/4 and MC_UU_00011/6). PCH was supported by Cancer Research UK (C52724/A20138 & C18281/A29019). MVH works in a unit that receives funding from the UK Medical Research Council and is supported by a British Heart Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. The funders had no role in the design, analyses,

interpretation of results or writing of the paper. The views expressed in this paper are those of the authors and not necessarily any of the funders.

We want to acknowledge participants and investigators from FinnGen study, UK Biobank, Biobank Japan, and the multiple large-scale GWAS consortia which made summary data available, including CARDIoGRAMplusC4D, MEGASTROKE, HERMES, atrial fibrillation GWAS metanalysis, aortic aneurysm GWAS metanalysis, CHARGE, and the Within Family Consortium. This work was carried out using UK Biobank project 15825. Data on coronary artery disease have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG'. The MEGASTROKE project received funding from sources specified at http://www.megastroke.org/acknowledgments.html (a list of all MEGASTROKE authors can be found in the appendix A - Supplementary methods). This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bristol.ac.uk/acrc/. Quality Control filtering of the UK Biobank data was conducted by R.Mitchell, G.Hemani, T.Dudding, L.Corbin, S.Harrison, L.Paternoster as described in the published protocol (doi: 10.5523/bris.1ovaau5sxunp2cv8rcy88688v). The MRC IEU UK Biobank GWAS pipeline was developed by B.Elsworth, R.Mitchell, C.Raistrick, L.Paternoster, G.Hemani, T.Gaunt (doi: 10.5523/bris.pnoat8cxo0u52p6ynfaekeigi).

CONFLICT OF INTEREST STATEMENT

DAL receives support from several national and international government and charitable research funders, as well as from Medtronic Ltd and Roche Diagnostics for research unrelated to that presented here. TRG receives funding from Biogen for work unrelated to that

695 presented here. MVH has consulted for Boehringer Ingelheim, and in adherence to the
696 University of Oxford's Clinical Trial Service Unit & Epidemiological Studies Unit (CSTU)
697 staff policy, did not accept personal honoraria or other payments from pharmaceutical
698 companies.

699 All other authors declare no competing interests.

REFERENCES

1 Ratnayake, W.M. and Galli, C. (2009) Fat and fatty acid terminology, methods of analysis and fat digestion and metabolism: a background review paper. *Ann Nutr Metab*, **55**, 8-43.

2 Mensink, R.P., Zock, P.L., Kester, A.D. and Katan, M.B. (2003) Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am J Clin Nutr*, **77**, 1146-1155.

3 Balk, E.M., Lichtenstein, A.H., Chung, M., Kupelnick, B., Chew, P. and Lau, J. (2006) Effects of omega-3 fatty acids on serum markers of cardiovascular disease risk: a systematic review. *Atherosclerosis*, **189**, 19-30.

4 Hooper, L., Summerbell, C.D., Thompson, R., Sills, D., Roberts, F.G., Moore, H.J. and Davey Smith, G. (2012) Reduced or modified dietary fat for preventing cardiovascular disease. *Cochrane Database Syst Rev*, in press., CD002137.

5 Sanders, T.A. (2009) Fat and fatty acid intake and metabolic effects in the human body. *Ann Nutr Metab*, **55**, 162-172.

6 Schwingshackl, L., Strasser, B. and Hoffmann, G. (2011) Effects of monounsaturated fatty acids on glycaemic control in patients with abnormal glucose metabolism: a systematic review and meta-analysis. *Ann Nutr Metab*, **58**, 290-296.

7 Miller, P.E., Van Elswyk, M. and Alexander, D.D. (2014) Long-chain omega-3 fatty acids eicosapentaenoic acid and docosahexaenoic acid and blood pressure: a meta-analysis of randomized controlled trials. *Am J Hypertens*, **27**, 885-896.

8 Shah, M., Adams-Huet, B. and Garg, A. (2007) Effect of high-carbohydrate or high-cis-monounsaturated fat diets on blood pressure: a meta-analysis of intervention trials. *Am J Clin Nutr*, **85**, 1251-1256.

9 Mozaffarian, D. and Wu, J.H. (2011) Omega-3 fatty acids and cardiovascular disease: effects on risk factors, molecular pathways, and clinical events. *J Am Coll Cardiol*, **58**, 2047-2067.

10 Masson, C.J. and Mensink, R.P. (2011) Exchanging saturated fatty acids for (n-6) polyunsaturated fatty acids in a mixed meal may decrease postprandial lipemia and markers of inflammation and endothelial activity in overweight men. *J Nutr*, **141**, 816-821.

11 Schwartz, E.A., Zhang, W.Y., Karnik, S.K., Borwege, S., Anand, V.R., Laine, P.S., Su, Y. and Reaven, P.D. (2010) Nutrient modification of the innate immune response: a novel mechanism by which saturated fatty acids greatly amplify monocyte inflammation. *Arterioscler Thromb Vasc Biol*, **30**, 802-808.

12 Sampath, H. and Ntambi, J.M. (2005) Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annu Rev Nutr*, **25**, 317-340.

13 Pase, M.P., Grima, N.A. and Sarris, J. (2011) Do long-chain n-3 fatty acids reduce arterial stiffness? A meta-analysis of randomised controlled trials. *Br J Nutr*, **106**, 974-980.

14 Jakobsson, A., Westerberg, R. and Jacobsson, A. (2006) Fatty acid elongases in mammals: their regulation and roles in metabolism. *Prog Lipid Res*, **45**, 237-249.

15 Lemaitre, R.N., Tanaka, T., Tang, W., Manichaikul, A., Foy, M., Kabagambe, E.K., Nettleton, J.A., King, I.B., Weng, L.C., Bhattacharya, S. *et al.* (2011) Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet*, **7**, e1002193.

16 Guan, W., Steffen, B.T., Lemaitre, R.N., Wu, J.H., Tanaka, T., Manichaikul, A., Foy, M., Rich, S.S., Wang, L., Nettleton, J.A. *et al.* (2014) Genome-wide association study of plasma N6 polyunsaturated fatty acids within the cohorts for heart and aging research in genomic epidemiology consortium. *Circ Cardiovasc Genet*, **7**, 321-331.

17 Wu, J.H., Lemaitre, R.N., Manichaikul, A., Guan, W., Tanaka, T., Foy, M., Kabagambe, E.K., Djousse, L., Siscovick, D., Fretts, A.M. *et al.* (2013) Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de novo

lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet*, **6**, 171-183.
18 Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P. *et al.* (2014) An atlas of genetic influences on human blood metabolites. *Nat Genet*, **46**, 543-550.
19 Kettunen, J., Demirkan, A., Wurtz, P., Draisma, H.H., Haller, T., Rawal, R., Vaarhorst, A., Kangas, A.J., Lyytikainen, L.P., Pirinen, M. *et al.* (2016) Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun*, **7**, 11122.
20 Tintle, N.L., Pottala, J.V., Lacey, S., Ramachandran, V., Westra, J., Rogers, A., Clark, J., Olthoff, B., Larson, M., Harris, W. *et al.* (2015) A genome-wide association study of saturated, mono- and polyunsaturated red blood cell fatty acids in the Framingham Heart Offspring Study. *Prostaglandins Leukot Essent Fatty Acids*, **94**, 65-72.
21 Dorajoo, R., Sun, Y., Han, Y., Ke, T., Burger, A., Chang, X., Low, H.Q., Guan, W., Lemaitre, R.N., Khor, C.C. *et al.* (2015) A genome-wide association study of n-3 and n-6 plasma fatty acids in a Singaporean Chinese population. *Genes Nutr*, **10**, 53.
22 Ebrahim, S. and Davey Smith, G. (2008) Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet*, **123**, 15-33.
23 Hingorani, A. and Humphries, S. (2005) Nature's randomised trials. *Lancet*, **366**, 1906-1908.
24 Smith, G.D., Lawlor, D.A., Harbord, R., Timpson, N., Day, I. and Ebrahim, S. (2007) Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med*, **4**, e352.
25 Holmes, M.V., Ala-Korpela, M. and Smith, G.D. (2017) Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol*, **14**, 577-590.
26 Jager, S., Cuadrat, R., Hoffmann, P., Wittenbecher, C. and Schulze, M.B. (2020) Desaturase Activity and the Risk of Type 2 Diabetes and Coronary Artery Disease: A Mendelian Randomization Study. *Nutrients*, **12**.
27 Liao, L.Z., Li, W.D., Liu, Y., Li, J.P., Zhuang, X.D. and Liao, X.X. (2020) Exploring the causal pathway from omega-6 levels to coronary heart disease: A network Mendelian randomization study. *Nutr Metab Cardiovasc Dis*, **30**, 233-240.
28 Yuan, T., Si, S., Li, Y., Li, W., Chen, X., Liu, C., Li, J., Wang, B., Hou, L., Liu, Y. *et al.* (2020) Roles for circulating polyunsaturated fatty acids in ischemic stroke and modifiable factors: a Mendelian randomization study. *Nutr J*, **19**, 70.
29 Yuan, S., Back, M., Bruzelius, M., Mason, A.M., Burgess, S. and Larsson, S. (2019) Plasma Phospholipid Fatty Acids, FADS1 and Risk of 15 Cardiovascular Diseases: A Mendelian Randomisation Study. *Nutrients*, **11**.
30 Zhang, T., Zhao, J.V. and Schooling, C.M. (2021) The associations of plasma phospholipid arachidonic acid with cardiovascular diseases: A Mendelian randomization study. *EBioMedicine*, **63**, 103189.
31 Chen, H.Y., Cairns, B.J., Small, A.M., Burr, H.A., Ambikkumar, A., Martinsson, A., Theriault, S., Munter, H.M., Steffen, B., Zhang, R. *et al.* (2020) Association of FADS1/2 Locus Variants and Polyunsaturated Fatty Acids With Aortic Stenosis. *JAMA Cardiol*, **5**, 694-702.
32 Yuan, S. and Larsson, S.C. (2019) Plasma Phospholipid Fatty Acids and Risk of Atrial Fibrillation: A Mendelian Randomization Study. *Nutrients*, **11**.
33 Holmes, M.V., Richardson, T.G., Ference, B.A., Davies, N.M. and Davey Smith, G. (2021) Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat Rev Cardiol*, **18**, 435-453.
34 Young, A.I., Benonisdottir, S., Przeworski, M. and Kong, A. (2019) Deconstructing the sources of genotype-phenotype associations in humans. *Science*, **365**, 1396-1400.
35 Ameur, A., Enroth, S., Johansson, A., Zaboli, G., Igl, W., Johansson, A.C., Rivas, M.A., Daly, M.J., Schmitz, G., Hicks, A.A. *et al.* (2012) Genetic adaptation of fatty-acid metabolism: a human-

specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am J Hum Genet*, **90**, 809-820.
36 Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., Hemani, G., Davey Smith, G. and Timpson, N.J. (2020) Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Hum Genet*, **139**, 23-41.
37 Hartwig, F.P., Davies, N.M., Horta, B.L., Ahluwalia, T.S., Bisgaard, H., Bonnelykke, K., Caspi, A., Moffitt, T.E., Poulton, R., Sajjad, A. *et al.* (2019) Effect modification of FADS2 polymorphisms on the association between breastfeeding and intelligence: results from a collaborative meta-analysis. *Int J Epidemiol*, **48**, 45-57.
38 Hughes, R.A., Davies, N.M., Davey Smith, G. and Tilling, K. (2019) Selection Bias When Estimating Average Treatment Effects Using One-sample Instrumental Variable Analysis. *Epidemiology*, **30**, 350-357.
39 Borges, M.C., Schmidt, A.F., Jefferis, B., Wannamethee, S.G., Lawlor, D.A., Kivimaki, M., Kumari, M., Gaunt, T.R., Ben-Shlomo, Y., Tillin, T. *et al.* (2020) Circulating Fatty Acids and Risk of Coronary Heart Disease and Stroke: Individual Participant Data Meta-Analysis in Up to 16 126 Participants. *J Am Heart Assoc*, **9**, e013131.
40 Chowdhury, R., Warnakula, S., Kunutsor, S., Crowe, F., Ward, H.A., Johnson, L., Franco, O.H., Butterworth, A.S., Forouhi, N.G., Thompson, S.G. *et al.* (2014) Association of dietary, circulating, and supplement fatty acids with coronary risk: a systematic review and meta-analysis. *Ann Intern Med*, **160**, 398-406.
41 Chowdhury, R., Stevens, S., Gorman, D., Pan, A., Warnakula, S., Chowdhury, S., Ward, H., Johnson, L., Crowe, F., Hu, F.B. *et al.* (2012) Association between fish consumption, long chain omega 3 fatty acids, and risk of cerebrovascular disease: systematic review and meta-analysis. *BMJ*, **345**, e6698.
42 Iso, H., Sato, S., Umemura, U., Kudo, M., Koike, K., Kitamura, A., Imano, H., Okamura, T., Naito, Y. and Shimamoto, T. (2002) Linoleic acid, other fatty acids, and the risk of stroke. *Stroke*, **33**, 2086-2093.
43 Marklund, M., Wu, J.H.Y., Imamura, F., Del Gobbo, L.C., Fretts, A., de Goede, J., Shi, P., Tintle, N., Wennberg, M., Aslibekyan, S. *et al.* (2019) Biomarkers of Dietary Omega-6 Fatty Acids and Incident Cardiovascular Disease and Mortality. *Circulation*, **139**, 2422-2436.
44 Abdelhamid, A.S., Martin, N., Bridges, C., Brainard, J.S., Wang, X., Brown, T.J., Hanson, S., Jimoh, O.F., Ajabnoor, S.M., Deane, K.H. *et al.* (2018) Polyunsaturated fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*, **7**, CD012345.
45 Hooper, L., Al-Khudairy, L., Abdelhamid, A.S., Rees, K., Brainard, J.S., Brown, T.J., Ajabnoor, S.M., O'Brien, A.T., Winstanley, L.E., Donaldson, D.H. *et al.* (2018) Omega-6 fats for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*, **7**, CD011094.
46 Abdelhamid, A.S., Brown, T.J., Brainard, J.S., Biswas, P., Thorpe, G.C., Moore, H.J., Deane, K.H., Summerbell, C.D., Worthington, H.V., Song, F. *et al.* (2020) Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*, **3**, CD003177.
47 Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T. *et al.* (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet*, **41**, 47-55.
48 Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. *et al.* (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*, **41**, 56-65.
49 Park, S., Lee, S., Kim, Y., Lee, Y., Kang, M., Kim, K., Kim, Y., Han, S., Lee, H., Lee, J. *et al.* (2021) Causal Effects of Serum Levels of n-3 or n-6 Polyunsaturated Fatty Acids on Coronary Artery Disease: Mendelian Randomization Study. *Nutrients*, **13**.
50 Hemani, G., Bowden, J. and Davey Smith, G. (2018) Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet*, **27**, R195-R208.

51 Morris, T.T., Davies, N.M., Hemani, G. and Smith, G.D. (2020) Population phenomena inflate genetic associations of complex social traits. *Sci Adv*, **6**, eaay0328.

52 Baylin, A., Ruiz-Narvaez, E., Kraft, P. and Campos, H. (2007) alpha-Linolenic acid, Delta6-desaturase gene polymorphism, and the risk of nonfatal myocardial infarction. *Am J Clin Nutr*, **85**, 554-560.

53 Vaittinen, M., Walle, P., Kuosmanen, E., Mannisto, V., Kakela, P., Agren, J., Schwab, U. and Pihlajamaki, J. (2016) FADS2 genotype regulates delta-6 desaturase activity and inflammation in human adipose tissue. *J Lipid Res*, **57**, 56-65.

54 Marklund, M., Morris, A.P., Mahajan, A., Ingelsson, E., Lindgren, C.M., Lind, L. and Riserus, U. (2018) Genome-Wide Association Studies of Estimated Fatty Acid Desaturase Activity in Serum and Adipose Tissue in Elderly Individuals: Associations with Insulin Sensitivity. *Nutrients*, **10**.

55 Freemantle, E., Lalovic, A., Mechawar, N. and Turecki, G. (2012) Age and haplotype variations within FADS1 interact and associate with alterations in fatty acid composition in human male cortical brain tissue. *PLoS One*, **7**, e42696.

56 Wang, L., Athinarayanan, S., Jiang, G., Chalasani, N., Zhang, M. and Liu, W. (2015) Fatty acid desaturase 1 gene polymorphisms control human hepatic lipid composition. *Hepatology*, **61**, 119-128.

57 Gkatzionis, A. and Burgess, S. (2019) Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int J Epidemiol*, **48**, 691-701.

58 Imamura, F., Micha, R., Wu, J.H., de Oliveira Otto, M.C., Otite, F.O., Abioye, A.I. and Mozaffarian, D. (2016) Effects of Saturated Fat, Polyunsaturated Fat, Monounsaturated Fat, and Carbohydrate on Glucose-Insulin Homeostasis: A Systematic Review and Meta-analysis of Randomised Controlled Feeding Trials. *PLoS Med*, **13**, e1002087.

59 Koletzko, B., Reischl, E., Tanjung, C., Gonzalez-Casanova, I., Ramakrishnan, U., Meldrum, S., Simmer, K., Heinrich, J. and Demmelmair, H. (2019) FADS1 and FADS2 Polymorphisms Modulate Fatty Acid Metabolism and Dietary Impact on Health. *Annu Rev Nutr*, **39**, 21-44.

60 Mitchell, J.A. and Warner, T.D. (2006) COX isoforms in the cardiovascular system: understanding the activities of non-steroidal anti-inflammatory drugs. *Nat Rev Drug Discov*, **5**, 75-86.

61 Ntambi, J.M., Miyazaki, M., Stoehr, J.P., Lan, H., Kendziorski, C.M., Yandell, B.S., Song, Y., Cohen, P., Friedman, J.M. and Attie, A.D. (2002) Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. *Proc Natl Acad Sci U S A*, **99**, 11482-11486.

62 Cohen, P., Miyazaki, M., Socci, N.D., Hagge-Greenberg, A., Liedtke, W., Soukas, A.A., Sharma, R., Hudgins, L.C., Ntambi, J.M. and Friedman, J.M. (2002) Role for stearoyl-CoA desaturase-1 in leptin-mediated weight loss. *Science*, **297**, 240-243.

63 Brown, J.M., Chung, S., Sawyer, J.K., Degirolamo, C., Alger, H.M., Nguyen, T., Zhu, X., Duong, M.N., Wibley, A.L., Shah, R. *et al.* (2008) Inhibition of stearoyl-coenzyme A desaturase 1 dissociates insulin resistance and obesity from atherosclerosis. *Circulation*, **118**, 1467-1475.

64 MacDonald, M.L., van Eck, M., Hildebrand, R.B., Wong, B.W., Bissada, N., Ruddle, P., Kontush, A., Hussein, H., Pouladi, M.A., Chapman, M.J. *et al.* (2009) Despite antiatherogenic metabolic characteristics, SCD1-deficient mice have increased inflammation and atherosclerosis. *Arterioscler Thromb Vasc Biol*, **29**, 341-347.

65 Jensen, P.N., Fretts, A.M., Hoofnagle, A.N., Sitlani, C.M., McKnight, B., King, I.B., Siscovick, D.S., Psaty, B.M., Heckbert, S.R., Mozaffarian, D. *et al.* (2020) Plasma Ceramides and Sphingomyelins in Relation to Atrial Fibrillation Risk: The Cardiovascular Health Study. *J Am Heart Assoc*, **9**, e012853.

66 Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C. *et al.* (2015) A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*, **47**, 1121-1130.

67 Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A.K., van der Laan, S.W., Gretarsdottir, S. *et al.* (2018) Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*, **50**, 524-537.

68 Nielsen, J.B., Thorolfsdottir, R.B., Fritsche, L.G., Zhou, W., Skov, M.W., Graham, S.E., Herron, T.J., McCarthy, S., Schmidt, E.M., Sveinbjornsson, G. *et al.* (2018) Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet*, **50**, 1234-1239.

69 Jones, G.T., Tromp, G., Kuivaniemi, H., Gretarsdottir, S., Baas, A.F., Giusti, B., Strauss, E., Van't Hof, F.N., Webb, T.R., Erdman, R. *et al.* (2017) Meta-Analysis of Genome-Wide Association Studies for Abdominal Aortic Aneurysm Identifies Four New Disease-Specific Risk Loci. *Circ Res*, **120**, 341-353.

70 Shah, S., Henry, A., Roselli, C., Lin, H., Sveinbjornsson, G., Fatemifar, G., Hedman, A.K., Wilk, J.B., Morley, M.P., Chaffin, M.D. *et al.* (2020) Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun*, **11**, 163.

71 Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M. *et al.* (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, **12**, e1001779.

72 Allen, N.E., Sudlow, C., Peakman, T., Collins, R. and Biobank, U.K. (2014) UK biobank data: come and get it. *Sci Transl Med*, **6**, 224ed224.

73 FinnGen. (2020), in press.

74 Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T. *et al.* (2017) Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol*, **27**, S2-S8.

75 Hirata, M., Kamatani, Y., Nagai, A., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Kubo, M., Muto, K., Mushiroda, T. *et al.* (2017) Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases. *J Epidemiol*, **27**, S9-S21.

76 Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba, N., Low, S.K., Okada, Y. *et al.* (2020) Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*, **52**, 669-679.

77 Low, S.K., Takahashi, A., Ebana, Y., Ozaki, K., Christophersen, I.E., Ellinor, P.T., Consortium, A.F., Ogishima, S., Yamamoto, M., Satoh, M. *et al.* (2017) Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat Genet*, **49**, 953-958.

78 Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K.E., Zheng, Z., Yengo, L., Lloyd-Jones, L.R., Sidorenko, J., Wu, Y. *et al.* (2018) Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun*, **9**, 2941.

79 Elsworth, B., Lyon, M., Alexander, T., Liu, Y., Matthews, P., Hallett, J., Bates, P., Palmer, T., Haberland, V., Smith, G.D. *et al.* (2020) The MRC IEU OpenGWAS data infrastructure. *bioRxiv*, in press., 2020.2008.2010.244293.

80 Zhu, J., Manichaikul, A., Hu, Y., Chen, Y.I., Liang, S., Steffen, L.M., Rich, S.S., Tsai, M., Siscovick, D.S., Lemaitre, R.N. *et al.* (2017) Meta-analysis of genome-wide association studies identifies three novel loci for saturated fatty acids in East Asians. *Eur J Nutr*, **56**, 1477-1484.

81 Hu, Y., Tanaka, T., Zhu, J., Guan, W., Wu, J.H.Y., Psaty, B.M., McKnight, B., King, I.B., Sun, Q., Richard, M. *et al.* (2017) Discovery and fine-mapping of loci associated with MUFAs through trans-ethnic meta-analysis in Chinese and European populations. *J Lipid Res*, **58**, 974-981.

82 Nieuwboer, H.A., Pool, R., Dolan, C.V., Boomsma, D.I. and Nivard, M.G. (2016) GWIS: Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes. *Am J Hum Genet*, **99**, 917-927.

83 Burgess, S., Davies, N.M. and Thompson, S.G. (2016) Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*, **40**, 597-608.

84 Burgess, S., Small, D.S. and Thompson, S.G. (2015) A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res*, in press.

85 Burgess, S., Butterworth, A. and Thompson, S.G. (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*, **37**, 658-665.

86 Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R. *et al.* (2018) The MR-Base platform supports systematic causal inference across the human phenome. *Elife*, **7**.
87 Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*, **10**, e1004383.
88 Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*, **88**, 76-82.
89 Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C., Replication, D.I.G., Meta-analysis, C., Madden, P.A., Heath, A.C., Martin, N.G. *et al.* (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*, **44**, 369-375, S361-363.
90 Consortium, G.T. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318-1330.
91 Sanderson, E., Davey Smith, G., Windmeijer, F. and Bowden, J. (2018) An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol*, in press.
92 Sanderson, E., Richardson, T.G., Hemani, G. and Davey Smith, G. (2021) The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *Int J Epidemiol*, in press.
93 Hartwig, F.P., Davies, N.M. and Davey Smith, G. (2018) Bias in Mendelian randomization due to assortative mating. *Genet Epidemiol*, **42**, 608-620.
94 Davies, N.M., Howe, L.J., Brumpton, B., Havdahl, A., Evans, D.M. and Davey Smith, G. (2019) Within family Mendelian randomization studies. *Hum Mol Genet*, **28**, R170-R179.
95 Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R. and Allen, N.E. (2017) Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*, **186**, 1026-1034.

979 **TABLES**

980

981

Locus	Exposure	Outcome	PPA H1: Exposure only	PPA H2: Outcome only	PPA H3: Distinct variants	PPA H4: Shared variant	PPA H5: None
FADS1	AA:DGLA	Aortic aneurysm	1	0	0	0	0
FADS1	AA:DGLA	Atrial fibrillation	1	0	0	0	0
FADS1	AA:DGLA	Haemorrhagic stroke	1	0	0	0	0
FADS1	AA:DGLA	Ischemic stroke	0.99	0	0	0	0
FADS1	AA:DGLA	Aortic valve stenosis	1	0	0	0	0
FADS1	AA:DGLA	Coronary artery disease	0.99	0	0	0	0
FADS1	AA:DGLA	Heart failure	1	0	0	0	0
FADS1	AA:DGLA	Peripheral artery disease	0.61	0	0.12	0.27	0
FADS1	AA:DGLA	Venous thromboembolism	0	0	0.15	0.85	0
ELOVL2	DHA:DPA_n3	Aortic aneurysm	1	0	0	0	0
ELOVL2	DHA:DPA_n3	Atrial fibrillation	1	0	0	0	0
ELOVL2	DHA:DPA_n3	Haemorrhagic stroke	1	0	0	0	0
ELOVL2	DHA:DPA_n3	Ischemic stroke	1	0	0	0	0
ELOVL2	DHA:DPA_n3	Aortic valve stenosis	0.99	0	0.01	0	0
ELOVL2	DHA:DPA_n3	Coronary artery disease	0.99	0	0	0	0
ELOVL2	DHA:DPA_n3	Heart failure	1	0	0	0	0
ELOVL2	DHA:DPA_n3	Peripheral artery disease	1	0	0	0	0
ELOVL2	DHA:DPA_n3	Venous thromboembolism	1	0	0	0	0
SCD	POA:PA	Aortic aneurysm	1	0	0	0	0
SCD	POA:PA	Atrial fibrillation	1	0	0	0	0
SCD	POA:PA	Haemorrhagic stroke	1	0	0	0	0
SCD	POA:PA	Ischemic stroke	0.99	0	0.01	0	0
SCD	POA:PA	Aortic valve stenosis	1	0	0	0	0
SCD	POA:PA	Coronary artery disease	0.01	0	0	0.99	0
SCD	POA:PA	Heart failure	1	0	0	0	0
SCD	POA:PA	Peripheral artery disease	1	0	0	0	0
SCD	POA:PA	Venous thromboembolism	1	0	0	0	0

982

983 **Table 1.** Genetic colocalization results for enzyme activity and cardiovascular diseases risk
984 among European ancestry individuals

985 Results are expressed as posterior probabilities of genetic association (PPA) with fatty acid trait only
986 (H1), cardiovascular trait only (H2), both traits due to distinct causal variants (H3), both traits due to a
987 shared causal variant (H4), or no association (H5).

988 *FADS1*: fatty acid desaturase 1; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase; AA:DGLA:
989 arachidonic acid to dihomo-γ-linoleic acid ratio; DHA:DPA_n3: docosahexaenoic acid to omega-3
990 eicosapentaenoic acid ratio; POA:PA: palmitoleic acid to palmitic acid ratio.

Locus	Exposure	Outcome	PPA H1: Exposure only	PPA H2: Outcome only	PPA H3: Distinct variants	PPA H4: Shared variant	PPA H5: None
FADS1	AA:DGLA	Body mass index	0.98	0	0.02	0	0
FADS1	AA:DGLA	Diastolic blood pressure	0	0	1	0	0
FADS1	AA:DGLA	Glucose	0	0	0.36	0.64	0
FADS1	AA:DGLA	LDL-cholesterol	0	0	0.13	0.87	0
FADS1	AA:DGLA	Systolic blood pressure	0.1	0	0.9	0	0
FADS1	AA:DGLA	Smoking	1	0	0	0	0
FADS1	AA:DGLA	Type 2 diabetes	0.95	0	0.03	0.02	0
FADS1	AA:DGLA	Triglycerides	0	0	1	0	0
ELOVL2	DHA:DPA_n3	Body mass index	0.61	0	0.39	0	0
ELOVL2	DHA:DPA_n3	Diastolic blood pressure	0.97	0	0.03	0	0
ELOVL2	DHA:DPA_n3	Glucose	0.99	0	0.01	0	0
ELOVL2	DHA:DPA_n3	LDL-cholesterol	0.94	0	0.06	0	0
ELOVL2	DHA:DPA_n3	Systolic blood pressure	0.95	0	0.05	0	0
ELOVL2	DHA:DPA_n3	Smoking	1	0	0	0	0
ELOVL2	DHA:DPA_n3	Type 2 diabetes	0.99	0	0.01	0	0
ELOVL2	DHA:DPA_n3	Triglycerides	1	0	0	0	0
SCD	POA:PA	Body mass index	0	0	1	0	0
SCD	POA:PA	Diastolic blood pressure	0	0	1	0	0
SCD	POA:PA	Glucose	0.92	0	0.08	0	0
SCD	POA:PA	LDL-cholesterol	0.01	0	0.95	0.03	0
SCD	POA:PA	Systolic blood pressure	0	0	1	0	0
SCD	POA:PA	Smoking	1	0	0	0	0
SCD	POA:PA	Type 2 diabetes	0	0	1	0	0
SCD	POA:PA	Triglycerides	0.14	0	0.86	0	0

Table 2. Genetic colocalization results for enzyme activity and cardiovascular risk factors among European ancestry individuals

Results are expressed as posterior probabilities of genetic association (PPA) with fatty acid trait only (H1), cardiovascular trait only (H2), both traits due to distinct causal variants (H3), both traits due to a shared causal variant (H4), or no association (H5).

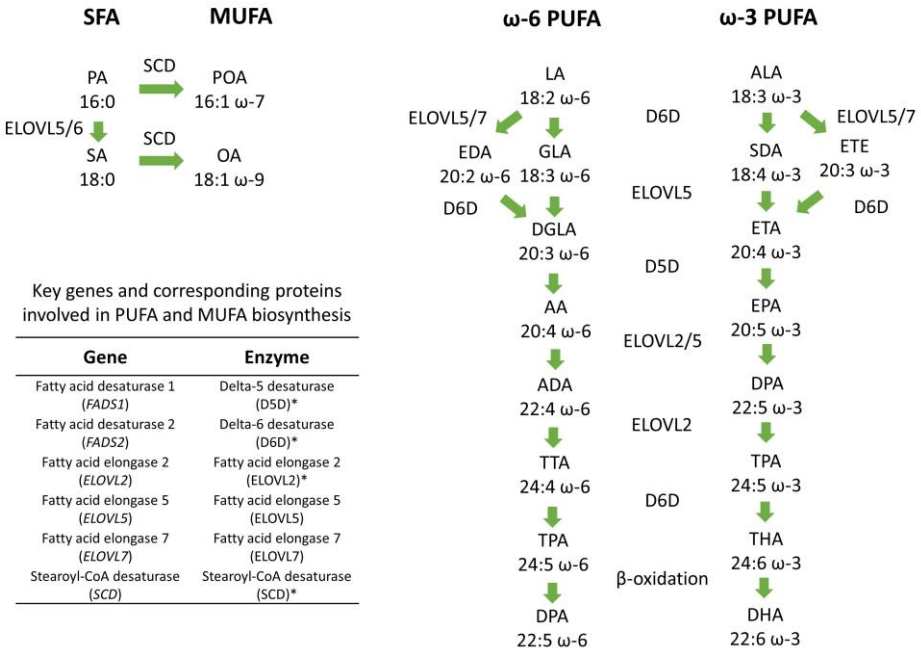
FADS1: fatty acid desaturase 1; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase; AA:DGLA: arachidonic acid to dihomo-γ-linoleic acid ratio; DHA:DPA_n3: docosahexaenoic acid to omega-3 eicosapentaenoic acid ratio; POA:PA: palmitoleic acid to palmitic acid ratio.

Locus	Chr	Enzyme	Enzyme activity proxy*	Fatty acids class	Ancestry
FADS1	11	D5D	AA:DGLA	PUFA n-6	Europeans
FADS2	11	D6D	GLA:LA DGLA:LA	PUFA n-6	Europeans East Asians
ELOVL2	6	ELOVL2	DHA:DPAn3	PUFA n-3	Europeans
SCD	10	SCD	POA:PA	MUFA/SFA	Europeans

Table 3. Genomic region, target gene and corresponding proxy of enzyme activity

* Enzyme activity was proxied based on enzyme-specific product to substrate ratio using data from circulating fatty acids. AA: arachidonic acid; DGLA: dihomo-γ-linolenic acid; DHA: docosahexaenoic acid; DPA: docosapentaenoic acid; *ELOVL2*: elongase 2; *FADS*: fatty acids desaturase; GLA: γ-linoleic acid; LA: linoleic acid; MUFA: monounsaturated fatty acids; PA: palmitic acid; POA: palmitoleic acid; PUFA: polyunsaturated fatty acids; SFA: saturated fatty acids; *SCD*: stearoyl-CoA desaturase.

FIGURE LEGENDS



Key genes and corresponding proteins involved in PUFA and MUFA biosynthesis

Gene	Enzyme
Fatty acid desaturase 1 (*FADS1*)	Delta-5 desaturase (D5D)*
Fatty acid desaturase 2 (*FADS2*)	Delta-6 desaturase (D6D)*
Fatty acid elongase 2 (*ELOVL2*)	Fatty acid elongase 2 (ELOVL2)*
Fatty acid elongase 5 (*ELOVL5*)	Fatty acid elongase 5 (ELOVL5)
Fatty acid elongase 7 (*ELOVL7*)	Fatty acid elongase 7 (ELOVL7)
Stearoyl-CoA desaturase (*SCD*)	Stearoyl-CoA desaturase (SCD)*

Figure 1. Overview of desaturation and elongation reactions involved in the conversion of MUFA from SFA and of longer-chain omega-3 and omega-6 PUFA from their shorter chain precursors.

*D5D, D6D, ELOVL2 and SCD activity were explored in the current study.

Abbreviations: *Saturated fatty acids (SFA):* PA: palmitic acid; SA: stearic acid. *Monounsaturated fatty acids (MUFA):* POA: palmitoleic acid; OA: oleic acid. *ω-6 polyunsaturated fatty acids (PUFA):* LA: linoleic acid; GLA: γ-linolenic acid; EDA: eicosadienoic acid; DGLA: dihomo-γ-linolenic acid; AA: arachidonic acid; ADA: adrenic acid; TTA: tetracosatetraenoic acid; TPA: tetracosapentaenoic acid; DPA: docosapentaenoic acid. *ω-3 polyunsaturated fatty acids (PUFA):* ALA: α-linolenic acid; SDA: stearidonic acid; ETE: eicosatrienoic acid; ETA: eicosatetraenoic acid; EPA: eicosapentaenoic acid; DPA: docosapentaenoic acid; TPA: tetracosapentaenoic acid; THA: tetracosahexaenoic acid; DHA: docosahexaenoic acid.

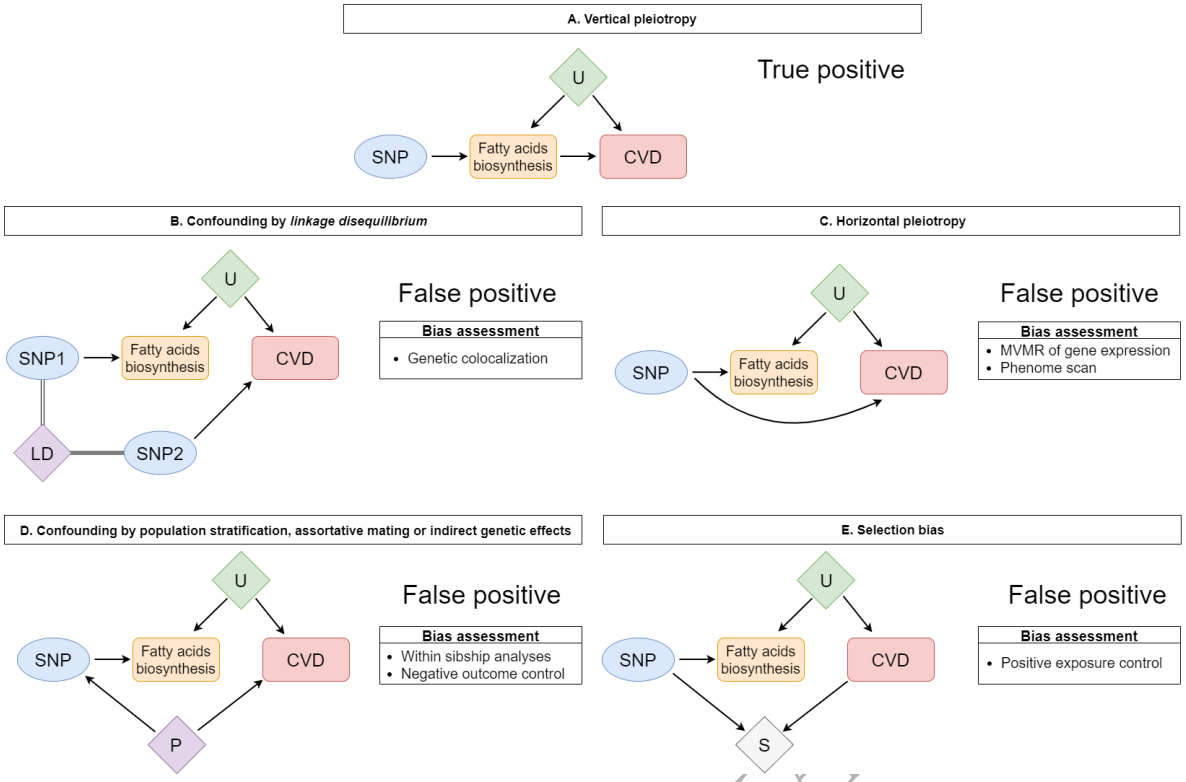


Figure 2. Schematic representation of scenarios leading to true (2A) and spurious (2B-2E) findings in Mendelian randomization analyses on the effect of fatty acids biosynthesis and cardiovascular disease (CVD) risk.

Figure 2A represents vertical pleiotropy, in which the effect of genetic instruments on CVD is mediated by fatty acids biosynthesis. Figures 2B to 2E represent alternative mechanisms that could bias Mendelian randomization findings: (B) confounding by *linkage disequilibrium* (LD), in which the selected genetic variant is in LD with another genetic variant influencing CVD independently; (C) horizontal pleiotropy, in which the genetic variant influences fatty acids biosynthesis and CVD via two different biological pathways; (D) confounding by population stratification, assortative mating or indirect genetic effects, in which different phenomena can introduce spurious association between genetic variant and CVD in samples of unrelated individuals, and (E) selection bias, in which selection into the study creates a spurious association between the genetic variant and CVD due to collider stratification bias. MVMR: multivariable Mendelian randomization; SNP: single nucleotide polymorphism; U: unobserved confounders; P: population phenomena (i.e. population stratification, assortative mating or indirect genetic effects); S: selection.

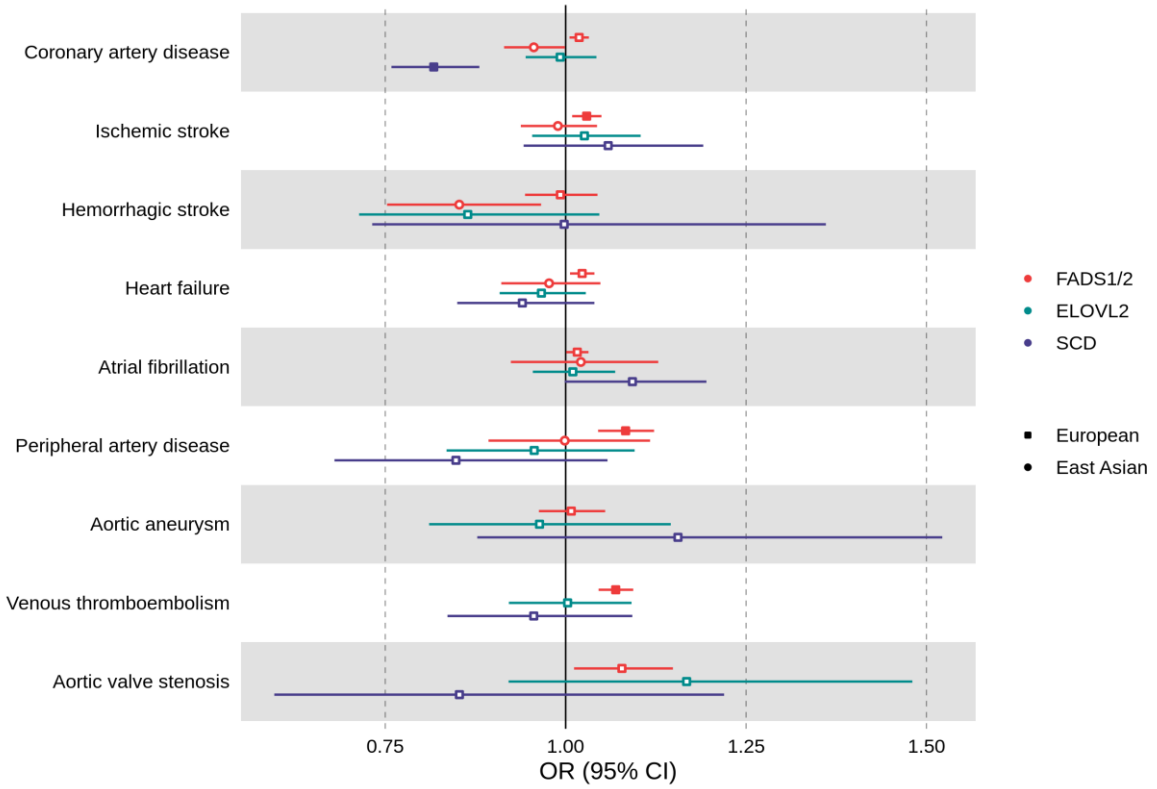


Figure 3. Mendelian randomization results for the risk of cardiovascular diseases related to increasing activity of enzymes coded by *FADS1/2* (D5D/D6D), *ELOVL2* (ELOVL2) and *SCD* (SCD) among individuals of European and East Asian ancestries.

Results are expressed as odds ratio of cardiovascular diseases per standard unit increase in the marker of enzyme activity for *FADS1/2* (i.e. AA:DGLA ratio in Europeans and DGLA:LA ratio in East Asians), *ELOVL2* (i.e. DHA:DPA ratio in Europeans) and *SCD* (i.e. POA:PA ratio in Europeans) loci. For individuals of European ancestry, SNP-cardiovascular diseases association data were metanalysed across multiple genetic association consortia, UK Biobank and FinnGen. For individuals of East Asian ancestry, SNP-cardiovascular diseases association data were extracted from BioBank Japan. Full symbols indicate associations at P-value lower than the P-value threshold accounting for multiple testing ($P < 0.00556$). AA: arachidonic acid; DGLA: dihomo-γ-linolenic acid; DHA: docosahexaenoic acid; DPA: docosapentaenoic acid; LA: linoleic acid; PA: palmitic acid; POA: palmitoleic acid; SNP: single nucleotide polymorphism; *FADS1/2*: fatty acid desaturases 1 and 2; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase.

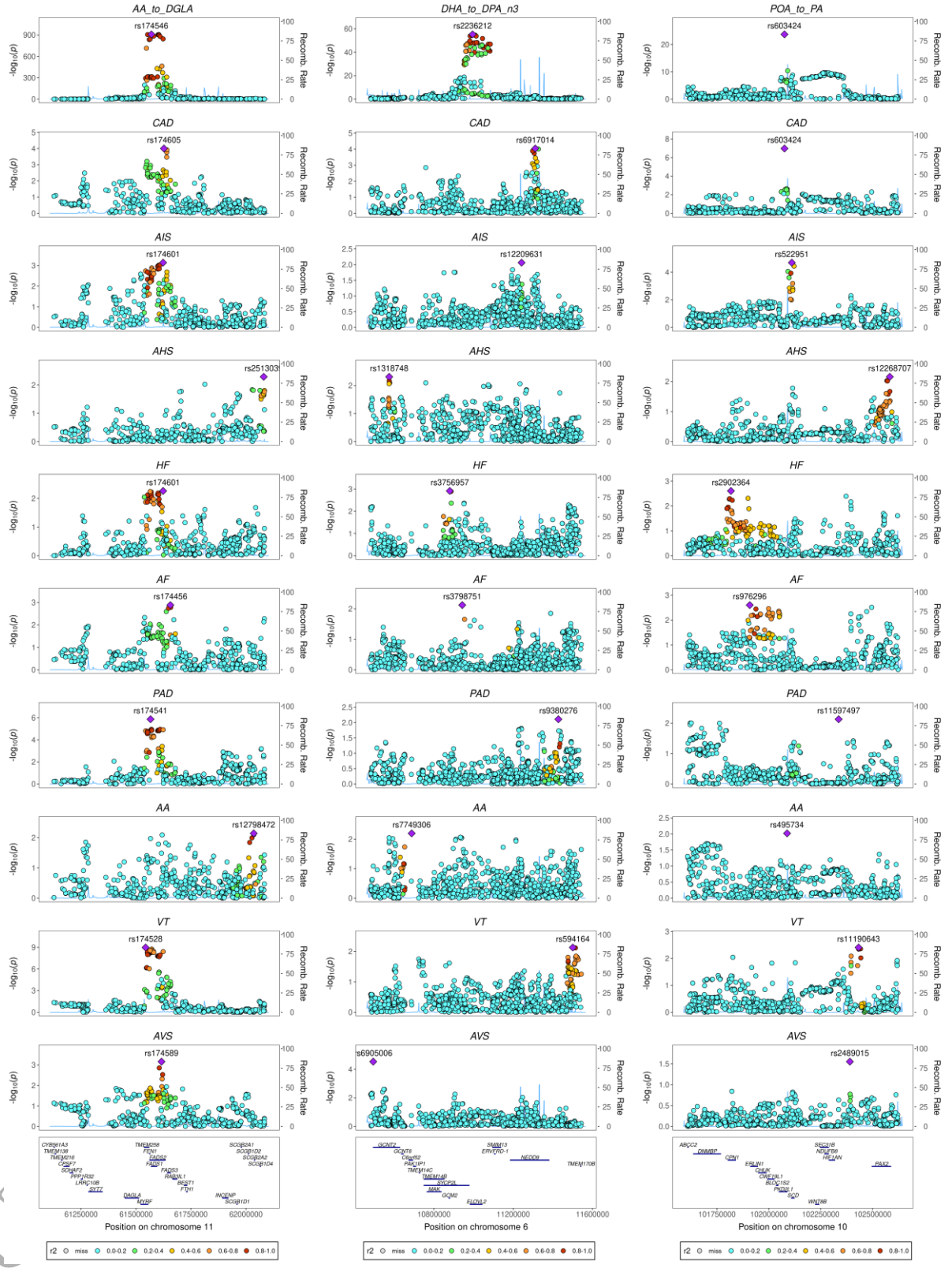


Figure 4. Genetic association plots for fatty acids enzyme activity (proxied by AA:DGLA, DHA:DPA, and POA:PA ratio) and cardiovascular diseases risk among individuals of European ancestry.

Results for each trait are expressed as $\log_{10}$ P-values for the *FADS*, *ELOVL2*, and *SCD* locus (columns 1, 2, and 3, respectively). AA: arachidonic acid; DGLA: dihomo-γ-linoleic acid; DHA: docosahexaenoic acid; DPA: docosapentaenoic acid; LA: linoleic acid; PA: palmitic acid; POA: palmitoleic acid; *FADS*: fatty acids desaturase; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase; CAD: coronary artery disease; AIS: any ischemic stroke; AHS: any haemorrhagic stroke; HF: heart failure; AF: atrial fibrillation; PAD: peripheral artery disease; AA: aortic aneurysm; VT: venous thromboembolism; AVS: aortic valve stenosis.

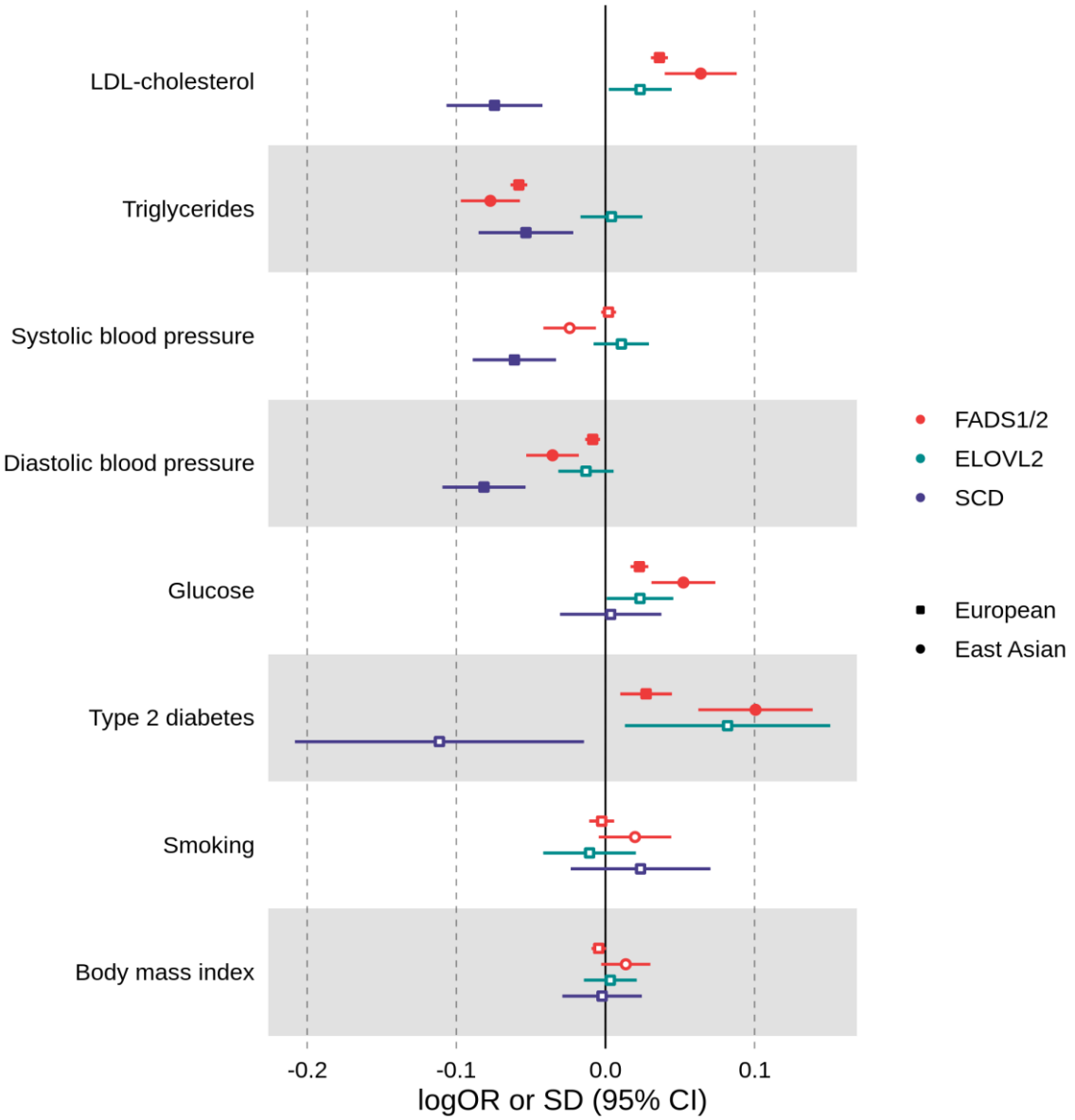


Figure 5. Mendelian randomization results for cardiovascular risk factors related to increasing activity of enzymes coded by *FADS1/2* (D5D/D6D), *ELOVL2* (ELOVL2) and *SCD* (SCD) among individuals of European and East Asian ancestries.

Results are expressed as change in standard units (SD) or log odds ratio (logOR) of cardiovascular disease risk factors per standard unit increase in the marker of enzyme activity for *FADS1/2* (i.e. AA:DGLA ratio in Europeans and DGLA:LA ratio in East Asians), *ELOVL2* (i.e. DHA:DPA ratio in Europeans) and *SCD* (i.e. POA:PA ratio in Europeans) loci. For individuals of European ancestry, data was extracted from UK Biobank or genetic association studies. For individuals of East Asian ancestry, data was extracted from BioBank Japan. Full symbols indicate associations at P-value lower than the P-value threshold accounting for multiple testing ($P < 0.00625$). Smoking is represent by pack years of smoking and number of cigarettes per day in European and East Asian ancestry individuals, respectively. AA: arachidonic acid; DGLA: dihomo-γ-linolenic acid; DHA: docosahexaenoic acid; DPA: docosapentaenoic acid; LA: linoleic acid; PA: palmitic acid; POA: palmitoleic acid; *FADS1/2*: fatty acid desaturases 1 and 2; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase; LDL-cholesterol: low-density lipoprotein-cholesterol.

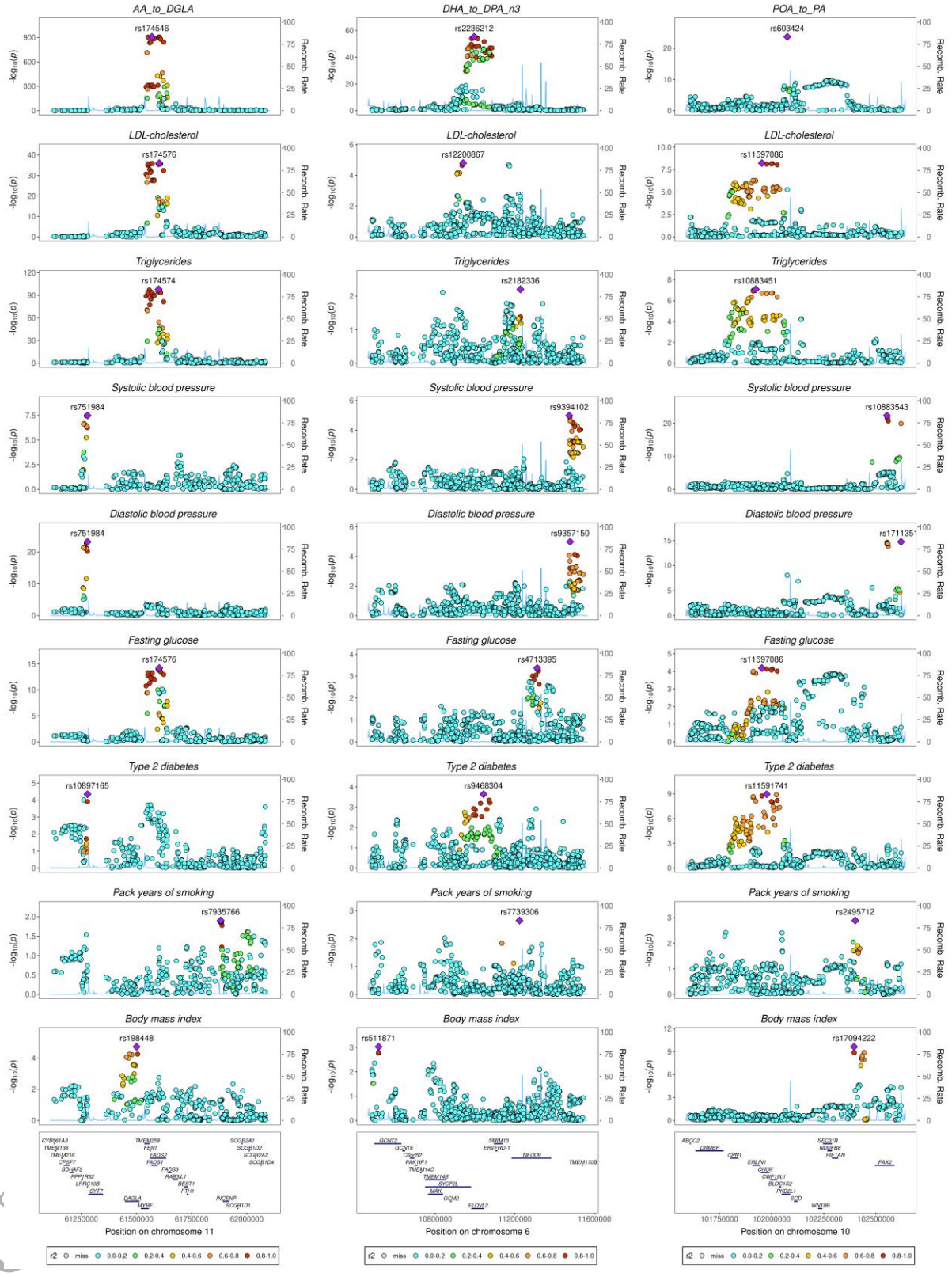


Figure 6. Genetic association plots for fatty acids enzyme activity (proxied by AA:DGLA, DHA:DPA, and POA:PA ratio) and cardiovascular disease risk factors among individuals of European ancestry.

Results for each trait are expressed as $\log_{10}$ P-values for the *FADS*, *ELOVL2*, and *SCD* locus (columns 1, 2, and 3, respectively). AA: arachidonic acid; DGLA: dihomo-γ-linoleic acid; DHA: docosahexaenoic acid; DPA: docosapentaenoic acid; LA: linoleic acid; PA: palmitic acid; POA: palmitoleic acid; *FADS*: fatty acids desaturase; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase; LDL: low-density lipoprotein.

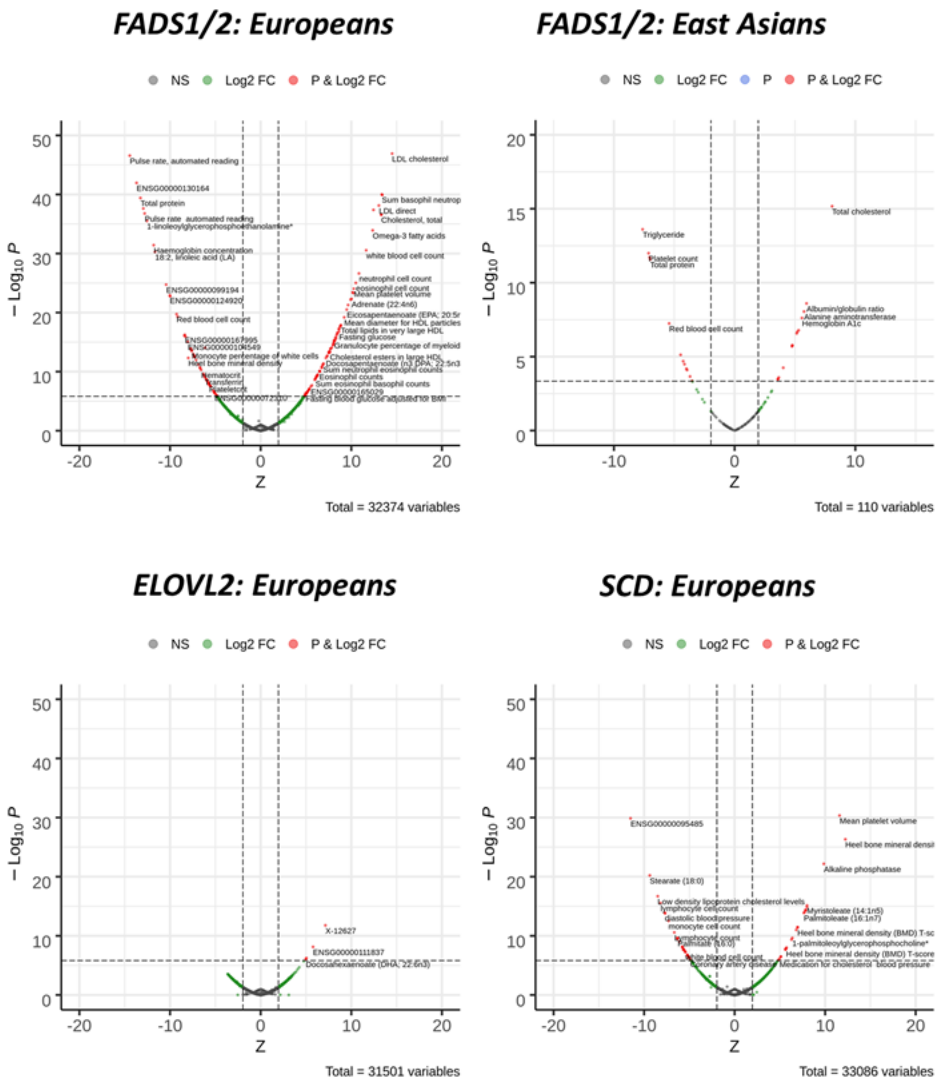


Figure 7. Phenome wide association scan of *FADS1/2* (rs174546), *ELOVL2* (rs3734398) and *SCD* (rs603424) genetic variants in European and East Asian ancestry individuals.

Results are expressed as the Z-statistic for the variant-trait association for the allele increasing enzyme expression/activity. Red circles denote P-value < 1.5×10^{-6} in Europeans P-value < 4.5×10^{-4} in East Asians . *FADS1/2*: fatty acid desaturases 1 and 2; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase.

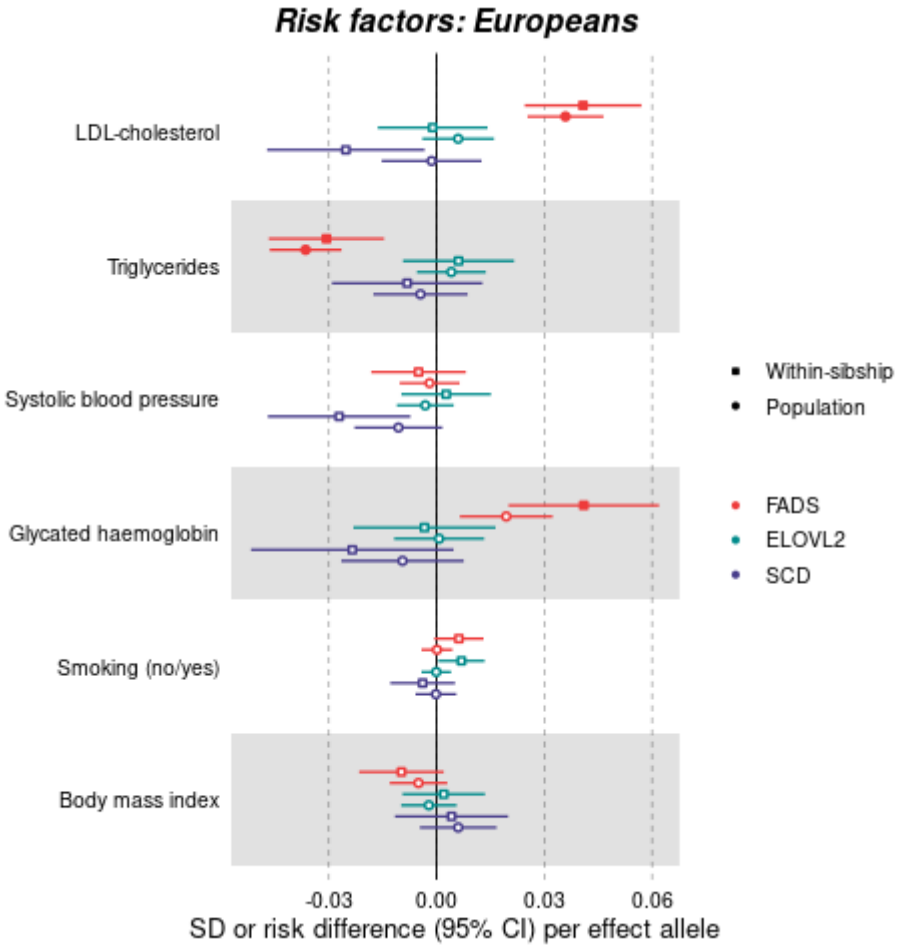


Figure 8. Association of *FADS1/2* (rs174546), *ELOVL2* (rs3734398) and *SCD* (rs603424) genetic variants with cardiovascular risk factors among unrelated individuals and within siblings of European ancestry.

Results are expressed as change in standard deviation (SD) units (or risk difference), and 95% confidence intervals (95% CI), of cardiovascular risk factors per allele increasing enzyme activity. *FADS1/2*: fatty acid desaturases 1 and 2; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase

ABBREVIATIONS

AA: arachidonic acid

ALA: α-linolenic acid

CHARGE: The Cohorts for Heart and Aging Research in Genomic Epidemiology

D5D: delta-5 desaturase

D6D: delta-6 desaturase

DGLA: dihomo-γ-linolenic acid

DHA: docosahexaenoic acid

DPA: docosapentaenoic acid

ELOVL2: elongase 2

eQTL: expression quantitative trait loci

FADS1: fatty acid desaturase 1

FADS2: fatty acid desaturase 2

GLA: γ-linolenic acid

GTEx: Genotype-Tissue Expression

GWAS: genome-wide association studies

GWIS: Genome-wide Inferred Study

LA: linoleic acid

LA: linoleic acid

LD: linkage disequilibrium

LDL-cholesterol: low-density lipoprotein-cholesterol

MESA: Multi-Ethnic Study of Atherosclerosis

MUFA: monounsaturated fatty acids

NHAPC: Nutrition and Health of Aging Population in China

PA: palmitic acid

POA: palmitoleic acid

PPA: posterior probability of association

PUFA polyunsaturated fatty acids

SCD: stearoyl-CoA desaturase

SCHS: Singapore Chinese Health Study

SFA: saturated fatty acids

SNP: single nucleotide polymorphism