

25 3 Institute of Cardiovascular Science, Faculty of Population Health Sciences, University
26 College London, London, WC1E 6DD, UK.

27 4 Department of Cardiology, Division Heart and Lungs, UMC Utrecht, Utrecht, 3584 CX,
28 Netherlands.

29 5 Institute of Health Informatics, University College London, London, NW1 2DA, UK.

30 6 Health Data Research UK London, University College London, NW1 2DA, UK.

31 7 UCL British Heart Foundation Research Accelerator, London, NW1 2DA, UK.

32 8 Cardiovascular Health Research Unit, Department of Medicine, University of Washington,
33 Seattle, WA, WA 98101, USA.

34 9 Medical Research Council Population Health Research Unit, University of Oxford, Oxford,
35 OX3 7LF, UK.

36 10 Clinical Trial Service and Epidemiological Studies Unit, Nuffield Department of
37 Population Health, University of Oxford, Oxford, OX3 7LF, UK.

38 11 NIHR Bristol Biomedical Research Centre, Bristol, BS8 2BN, UK.

39

40

41 **Correspondence email:** m.c.borges@bristol.ac.uk

42 **Twitter handle:** @MCarol_Borges

43 **Short title:** Mendelian randomization, fatty acids and cardiovascular diseases

44

45 **ABSTRACT**

46 Despite early interest, the evidence linking fatty acids to cardiovascular diseases remains
47 controversial. We used Mendelian randomization to explore the involvement of
48 polyunsaturated (PUFA) and monounsaturated (MUFA) fatty acids biosynthesis in the
49 aetiology of several cardiovascular disease endpoints in up to 1,153,768 European and
50 212,453 East Asian ancestry individuals. As instruments, we selected single nucleotide
51 polymorphisms (SNP) mapping to genes with well-known roles in PUFA (i.e. *FADS1/2* and
52 *ELOVL2*) and MUFA (i.e. *SCD*) biosynthesis. Our findings suggest that higher PUFA
53 biosynthesis rate (proxied by rs174576 near *FADS1/2*) is related to higher odds of multiple
54 cardiovascular diseases, particularly ischemic stroke, peripheral artery disease and venous
55 thromboembolism, whereas higher MUFA biosynthesis rate (proxied by rs603424 near *SCD*)
56 is related to lower odds of coronary artery disease among Europeans. Results were unclear
57 for East Asians as most effect estimates were imprecise. By triangulating multiple approaches
58 (i.e. uni-/multi-variable Mendelian randomization, a phenome-wide scan, genetic
59 colocalization and within-sibling analyses), our results are compatible with higher low-
60 density lipoprotein (LDL)-cholesterol (and possibly glucose) being a downstream effect of
61 higher PUFA biosynthesis rate. Our findings indicate that genetically-determined PUFA and
62 MUFA biosynthesis are involved in the aetiology of cardiovascular diseases and suggest
63 LDL-cholesterol as a potential mediating trait between PUFA biosynthesis and
64 cardiovascular diseases risk.

65 **Key words:** fatty acids, cardiovascular diseases, Mendelian randomization

66 INTRODUCTION

67

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

78 Some fatty acids can be synthesized endogenously by fatty acid synthase or taken up
79 by diet and further elongated and desaturated into longer chain fatty acids by fatty acid
80 elongases and desaturases, respectively (14). Genome-wide association studies (GWAS) have
81 reported that circulating fatty acids are strongly influenced by genetic variants near genes
82 coding fatty acid elongases and desaturases: fatty acid desaturase 1 (*FADS1* -
83 ENSG00000149485), fatty acid desaturase 2 (*FADS2* - ENSG00000134824), elongase 2
84 (*ELOVL2* - ENSG00000197977), and stearoyl-CoA desaturase (*SCD* - ENSG00000099194)
85 (15-21). The chemical reactions and pathways catalysed by the enzymes encoded by *FADS1*,
86 *FADS2*, *ELOVL2*, and *SCD* are summarised in **Figure 1**.

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

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

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

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

114 The aim of this study was to use Mendelian randomization to explore the effect of
115 fatty acids biosynthesis on a wide range of cardiovascular disease end-points in up to
116 1,153,768 European and 212,453 East Asian ancestry individuals. We extend work in
117 previous studies by using genetic variants regulating multiple rate-limiting enzymes in fatty
118 acids biosynthesis (i.e. D5D/D6D, ELOVL2 and SCD), comparing findings between
119 Europeans and East Asians and extensively exploring the key scenarios that could lead to
120 spurious findings in this and previous Mendelian randomization studies.

121 MATERIAL AND METHODS

122

123 Study design

124 We used two-sample Mendelian randomization to probe the lifelong effect of fatty
125 acids biosynthesis on the risk of multiple cardiovascular diseases in Europeans and East
126 Asian individuals. As instruments, we selected genetic variants mapping to genes with a well-
127 known role in fatty acids biosynthesis (i.e. *FADS1/2*, *ELOVL2*, and *SCD*). We used two
128 approaches to circumvent limitations in previous studies.

129 First, we used genetic variants to instrument for enzyme activity in a given fatty acids
130 biosynthesis pathway (rather than for individual fatty acids) by deriving the ratio between
131 fatty acids that are the product and the substrate of a reaction catalysed by the corresponding
132 enzyme. This allows harnessing the advantages of cis-acting variants in the vicinity of genes
133 coding for key enzymes in fatty acids biosynthesis pathways and can provide more credible
134 evidence on the likely therapeutic benefit of targeting such protein in preventing
135 cardiovascular diseases(33).

136 Second, we used a range of methods to assess the plausibility of the four key
137 scenarios that could invalidate inferences from this and previous studies: (i) horizontal
138 pleiotropy, where the genetic variant influences the outcome via a different biological
139 pathway; (ii) confounding by *linkage disequilibrium* (LD), where the selected genetic variant
140 is in LD with another genetic variant influencing the outcome independently; (iii)
141 confounding by population stratification, assortative mating or indirect genetic effects, which
142 could create a spurious association between genetic variant and outcome, and (iv) selection
143 bias, where the genetic variant (or, more likely, its downstream traits) and the outcome affect
144 selection into the sample resulting in a spurious association (**Figure 2**).

145

146 **Data sources**

147 The study included data from multiple consortia of genetic association studies (34-38)
148 and biobanks (39-45).

149

150 *Genetic associations with cardiovascular disease outcomes*

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

154 Summary data for the association between genetic variants and these cardiovascular
155 disease endpoints was obtained from UK Biobank, FinnGen, BioBank Japan and several
156 large-scale genetic consortia of cardiovascular disease outcomes. If genetic association data
157 on a cardiovascular endpoint were available from two or more independent datasets of
158 individuals from the same genetic ancestry (i.e. UK Biobank and FinnGen), genetic
159 association estimates were pooled across data sources using fixed-effect meta-analysis with
160 inverse variance weights. Characteristics of studies and criteria for case definition are
161 detailed in **Supplementary table 1** and **Supplementary methods**.

162 For individuals of European ancestry only/predominantly (i.e. UK Biobank,
163 FinnGen and large-scale genetic consortia), data were available on all outcomes of interest:
164 coronary artery disease (N cases/controls = 123,668/702,156), ischemic stroke (N
165 cases/controls = 53,395/1,030,253), haemorrhagic stroke (N cases/controls =
166 4,558/627,188), heart failure (N cases/controls = 64,696/1,089,072), atrial fibrillation (N
167 cases/controls = 77,945/1,067,430), peripheral arterial disease (N cases/controls =
168 9,836/627,950), aortic aneurysm (N cases/controls = 9,735/730,073), venous

169 thromboembolism (N cases/controls = 25,284/616,235), and aortic valve stenosis (N
170 cases/controls = 2,844/461,776).

171 For individuals of East Asian ancestry (i.e. BioBank Japan), cardiovascular
172 outcomes data were available for coronary artery disease (N cases/controls =
173 29,319/183,134), ischemic stroke (N cases/controls = 17,671/192,383), haemorrhagic stroke
174 (N cases/controls = 2,820/192,383), heart failure (N cases/controls = 9,413/203,040), atrial
175 fibrillation (N cases/controls = 8,180/28,612), and peripheral arterial disease (N
176 cases/controls = 3,593/208,860).

177

178 *Genetic associations with circulating fatty acid concentration*

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

186 For East Asian ancestry individuals, we used genetic association data on fatty acids
187 from the Singapore Chinese Health Study (SCHS) for circulating PUFA (N = 1,361) (21) and
188 from a metanalysis of the Nutrition and Health of Aging Population in China (NHAPC) and
189 the Chinese ancestry individuals of the Multi-Ethnic Study of Atherosclerosis (MESA) for
190 RBC or circulating SFA and MUFA (N = 3,521) (46, 47).

191 Characteristics of these studies are detailed in **Supplementary table 2**.

192

193 **Data analysis**

194 *Selection of genetic instruments indexing fatty acids biosynthesis*

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

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

213 For European ancestry individuals, we derived genetic association data for proxies of
214 enzyme activity by applying the GWIS (“Genome-wide Inferred Study”) method to genetic
215 association data for circulating fatty acids from the CHARGE consortium (15-17) for the
216 ratios of AA to DGLA (proxy of D5D activity), γ -linolenic acid (GLA) to LA (proxy of D6D

217 activity), docosahexaenoic acid (DHA) to docosapentaenoic acid (DPA n-3) (proxy of
218 ELOVL2 activity), and palmitoleic acid (POA) to palmitic acid (PA) (proxy of SCD activity),
219 as detailed in **Supplementary methods**. Briefly, GWIS approximates GWAS summary
220 statistics for a new variable as a function of multiple phenotypes for which GWAS summary
221 statistics, phenotypic means, and covariances are available (48).

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

226 If the selected genetic variant was a palindromic SNP, it was replaced by a non-
227 palindromic proxy variant in strong LD to avoid data harmonisation problems in subsequent
228 analyses. All SNP-trait associations were harmonised so that the allele associated with
229 increasing enzyme activity was the effect allele, indicating more active conversion. We
230 approximated the R^2 , a measure of the variance in exposure explained by the genetic variant,
231 and the F-statistics, a measure of instrument strength (49), as detailed in **Supplementary**
232 **methods**.

233

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

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

239

240 *Mendelian randomization analysis*

241 For each cardiovascular outcome, we used the Wald ratio method to estimate the odds
242 ratio of disease per standard unit increase in enzyme activity by dividing estimates for the
243 genetic association with cardiovascular outcome by estimates for the genetic association with
244 enzyme activity as detailed in the **Supplementary methods**. We used a Bonferroni
245 correction to account for the number of outcomes available in Europeans (p-value = 0.05/9
246 outcomes = 0.006) and East Asians (p-value = 0.05/6 outcomes = 0.008). We use these p-
247 value thresholds simply as a heuristic for highlighting associations worthy of follow-up.
248 Mendelian randomization analyses were performed using R software version 3.6.2 (R
249 Foundation for Statistical Computing) including the TwoSampleMR R package (50).

250

251 *Exploring bias in Mendelian randomization analyses*

252 We explored the validity of our Mendelian randomization findings by assessing
253 whether results are likely to be spurious due to horizontal pleiotropy, population
254 stratification, assortative mating, indirect genetic effects, and selection bias as follows.

255

256 Horizontal pleiotropy

257 Horizontal pleiotropy is one of the main threats to the validity of Mendelian
258 randomization studies since it is a widespread biological phenomenon (51) and it cannot be
259 empirically verified. We combined two approaches to explore the plausibility that our results
260 are explained by horizontal pleiotropy: phenome scan and multivariable Mendelian
261 randomization to adjust for the potential effect of other genes expressed in the candidate
262 genomic regions.

263

264 Phenome scan

265 We explored the potential mechanisms that might link the selected genetic variants to
266 cardiovascular diseases by testing their association with well-established cardiovascular risk
267 factors (hypothesis-driven approach) and across the phenome (hypothesis-free approach)
268 (52).

269 In the hypothesis-driven approach, we tested for the association between the selected
270 genetic variants and eight well-established risk factors for cardiovascular diseases [i.e. low-
271 density lipoprotein (LDL)-cholesterol, triglycerides, systolic, diastolic blood pressure, fasting
272 glucose, type 2 diabetes, smoking, and body mass index] using Bonferroni correction to
273 account for multiple testing (p-value = 0.05/8 risk factors). Genetic association data for
274 these risk factors were extracted from UK Biobank (only/predominantly) and BioBank
275 Japan using the IEU OpenGWAS project database (53).

276 In the hypothesis-free (phenome-wide) approach, we used an automated phenome-
277 wide scan tool from the IEU OpenGWAS project database (53) to test the association of the
278 selected genetic variants with 32,534-34,465 (non-unique) traits for European ancestry
279 individuals and 110 traits for Japanese individuals from BioBank Japan. We used a
280 Bonferroni correction to account for multiple testing considering the maximum number of
281 traits included in the phenome-wide scan in Europeans (p-value = $0.05/34,465 = 1.5 \times 10^{-6}$)
282 and East Asians (p-value = $0.05/110 = 4.5 \times 10^{-4}$).

283

284 Gene expression and tissue-specific analyses

285 We explored the influence of higher expression of the target genes (i.e. *FADS1/2*,
286 *ELOVL2* and *SCD*), and their tissue specificity, on cardiovascular disease risk and risk factors
287 by integrating expression quantitative trait loci (eQTL) data from Genotype-Tissue
288 Expression (GTEx) version 8 with genetic association data for cardiovascular traits, as

289 detailed in **Supplementary methods**. First, we performed a cross-tissue assessment of the
290 association of the selected genetic variants with transcription of cis-genes, defined as genes
291 for which the transcription start site was 1 Mb away from the genetic variant. Second, we
292 used multivariable Mendelian randomization to jointly model the expression of a target gene
293 (i.e. *FADS1/2*, *ELOVL2*, and *SCD*) and a co-expressed cis-gene on cardiovascular outcomes
294 across tissues using the MVMR R package (54). This analysis allowed us to estimate the
295 direct contribution of changes in the expression of each target gene where the selected genetic
296 variant was related to co-expression of a non-target cis-gene.

297

298 Confounding by linkage disequilibrium

299 *Linkage disequilibrium* (LD) could lead to spurious results in Mendelian
300 randomization analysis if, by chance, the selected genetic variant influencing fatty acids
301 biosynthesis is correlated (i.e. in LD) with another genetic variant influencing the risk of
302 cardiovascular diseases independently. We used *coloc* (55), a method for pairwise genetic
303 colocalization analysis, to test whether the same genetic variant influences fatty acids
304 biosynthesis and cardiovascular diseases risk. A posterior probability of association (PPA) \geq
305 70% for association with both traits due to a single causal variant was considered as strong
306 evidence for a shared genetic variant.

307 *Coloc* assumes a single causal variant in the genomic region, and, as a result, the
308 presence of multiple conditionally independent SNPs within a region can affect the
309 performance of the method. Therefore, where *Coloc* provided evidence of a shared genetic
310 signal, we also performed approximate conditional analyses using GCTA(56, 57) (adjusting
311 for the top SNP in each genomic region) and re-ran *Coloc* using the adjusted association
312 estimates as a sensitivity analyses.

313 Colocalization analyses were restricted to European datasets as the method assumes
314 that samples are drawn from independent populations of similar genetic background (i.e.
315 allele frequencies and LD pattern are identical), which was not the case for East Asians in our
316 analyses since fatty acids and cardiovascular disease data were derived from Singaporean
317 Chinese and Japanese individuals, respectively.

318

319 Population stratification, assortative mating and indirect genetic effects

320 Mendelian randomization studies generally assume that genetic association estimates
321 from GWAS reflect the direct effect of a genetic variant on a phenotype, i.e., the downstream
322 effect of inheriting an allele. However, there is growing evidence that GWAS of unrelated
323 individuals may also capture non-direct sources of association relating to population
324 stratification, assortative mating and indirect genetic effects of parents(58). Within-family
325 GWAS/Mendelian randomization designs, such as parent-offspring trio or within-sibship
326 models, control for variation in parental genotypes, and so are not affected by these potential
327 biases (59-61).

328 There is evidence that the *FADS1/2* locus was under important selection pressure in
329 different populations and at different times, possibly as a response to dietary changes and the
330 need for adequate supply of essential long-chain PUFA from precursors (62). Despite
331 attempts to control for population stratification in genetic association data (e.g. by adjusting
332 for genomic principal components), there could still be residual population structure as has
333 recently been shown in UK Biobank (63). In addition, there is a possibility that indirect
334 genetic effects of parents bias studies among unrelated individuals given the literature
335 suggesting that maternal genotype for *FADS1/2* variants might indirectly influence offspring
336 outcomes via intrauterine effects and/or breastfeeding (64).

337 We used two approaches to explore this. First, we tested the association between the
338 selected genetic variants with two negative control outcomes, skin colour and ease of skin
339 tanning, using UK Biobank genetic association data deposited in the IEU Open GWAS
340 Project (53). Since these traits could not conceivably be affected by fatty acids biosynthesis,
341 any evidence for an association between genetic variants and these negative control outcomes
342 would be indicative of residual population stratification(65). Second, we used data from a
343 recent within-sibship GWAS, including up to 178,076 individuals (77,832 sibling pairs) from
344 23 cohorts, to evaluate if our findings are sensitive to population stratification, assortative
345 mating, and indirect genetic effects of parents. We compared the within-sibship association of
346 the selected genetic variants with cardiovascular risk factors (LDL-cholesterol, triglycerides,
347 systolic blood pressure, glycated haemoglobin, smoking, and body mass index) with
348 estimates from standard GWAS models in unrelated individuals (sample size ranging from
349 50,361 for glycated haemoglobin to 155,457 for BMI). Data on cardiovascular disease
350 endpoints and other risk factors (i.e. diastolic blood pressure, fasting glucose, and type 2
351 diabetes) were not available.

352

353 Selection bias

354 Several processes of sample selection, occurring from study design to data analyses,
355 can result in selected samples not representative of their target populations (e.g. due to
356 case/control ascertainment, participant dropout, loss to follow-up, subgroup analysis, or
357 missing data). In some instances, non-random sample selection can bias inference about the
358 causal effect of an exposure on an outcome, including when using Mendelian randomization
359 (66).

360 Although bias can result from multiple selection mechanisms, we were particularly
361 concerned about selection due to ascertainment of cardiovascular disease status given the
362 design of the studies used for cardiovascular disease endpoints. As an example, BioBank
363 Japan is a hospital-based study, in which cases for cardiovascular diseases, except atrial
364 fibrillation, were compared to a control group including a mixture of hospital-based (i.e.
365 individuals diagnosed at health centres with other diseases) and community-based (i.e.
366 individuals from population-based cohorts) controls as previously described (42). In addition,
367 UK Biobank has a response rate of 5.5% and its participants have fewer self-reported health
368 conditions and are more likely to be older, female, wealthier, leaner, non-smokers, non-
369 drinkers than the general UK population (67).

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

377

378 **RESULTS**

379

380 *Selection of genetic instruments indexing fatty acids biosynthesis*

381 In individuals of European ancestry, the selected genetic variants were rs174546
382 (*FADS1/2*, chr11q13.3), rs174576 (*FADS1/2*, chr11q13.3), rs3734398 (*ELOVL2*, chr6q15),
383 and rs603424 (*SCD*, chr10q22.1), which explained a proportion of the variance in the

384 corresponding marker of enzyme activity of 32.6% ($F = 4174$) for AA:DGLA, 6.3% ($F =$
385 580) for GLA:LA, 2.4% ($F = 218$) for DHA:n-3 DPA, and 1.1% ($F = 100$) for POA:PA,
386 respectively (**Supplementary table 3**). The *FADS1/2* SNPs (i.e. rs174546 and rs174576)
387 were in strong LD ($R^2 = 0.93$ 1000 Genomes European population), and, therefore, only the
388 SNP more strongly associated with the corresponding marker of enzyme activity was used in
389 subsequent analyses (i.e. rs174546).

390 In individuals of East Asian ancestry, the top variant in the *FADS1/2* locus was
391 palindromic and, therefore, was replaced by rs174546 (i.e. LD $R^2 = 0.93$ in 1000G East Asian
392 population), which explained 8.4% ($F = 125$) of the variance in DGLA:LA (**Supplementary**
393 **table 3**). No genetic variants were associated with markers of D5D activity (AA:DGLA). As
394 previously mentioned, genetic variants at *ELOVL2* and *SCD* loci were not associated with
395 circulating fatty acids in East Asians and, therefore, were not eligible for our analyses, which
396 may be related to the modest sample size available for East Asians ($N = 1,361-3,521$).

397

398 *Impact of genetic instruments on circulating fatty acids*

399 Overall, the effect of genetic variants on the fatty acids pool was replicable across
400 studies and ancestries. In addition, the genetic variants impact on the fatty acids pool was
401 consistent with their predicted function on fatty acids biosynthesis. The *FADS1/2* SNP
402 (rs174546) was associated with a lower concentration of shorter chain omega-3 (e.g. ALA)
403 and omega-6 (e.g. LA) fatty acids and higher concentration of longer chain omega-3 (e.g.
404 DHA) and omega-6 (e.g. AA) fatty acids. The *ELOVL2* SNP (rs3734398) was mostly
405 associated with higher concentration of DHA and lower concentration of eicosapentaenoic
406 acid (EPA) and n-3 DPA, whereas the *SCD* SNP (rs603424) was related to lower SFA,

407 particularly PA, and higher monounsaturated fatty acids (MUFA), particularly POA
408 (**Supplementary figures 1 and 2**).

409

410 *Mendelian randomization analysis*

411 In European ancestry individuals, higher *FADS1/D5D* activity (proxied by increase in
412 AA:DGLA in standard units) was related to higher odds of coronary artery disease (OR =
413 1.02; 95% CI: 1.01, 1.03; p-value = 0.006), ischemic stroke (OR = 1.03; 95% CI: 1.01, 1.05;
414 p-value = 0.004), heart failure (OR = 1.02; 95% CI: 1.01, 1.04; p-value = 0.008), atrial
415 fibrillation (OR = 1.02; 95% CI: 1.00, 1.03; p-value = 0.04), peripheral artery disease (OR =
416 1.08; 95% CI: 1.04, 1.12; p-value = 1×10^{-5}), venous thromboembolism (OR = 1.07; 95%
417 CI: 1.05, 1.09; p-value = 5×10^{-9}), and aortic valve stenosis (OR = 1.08; 95% CI: 1.01,
418 1.15; p-value = 0.02). Only results for ischemic stroke, peripheral artery disease, and venous
419 thromboembolism passed our threshold for multiple testing correction ($P < 0.006$) (**Figure 3**).
420 Overall, results were consistent across studies, except for coronary artery disease, for which
421 the estimated effect was attenuated in UK Biobank compared to other studies, and for aortic
422 aneurysm, for which the estimated effect was in different directions between UK Biobank
423 and other studies (**Supplementary figure 3**).

424 There was little evidence supporting a relationship between higher *ELOVL2/ELOVL2*
425 activity (proxied by increase in DHA:DPAn-3 in standard units) and cardiovascular
426 endpoints. However, some results were imprecisely estimated and, therefore, we cannot rule
427 out the presence of clinically meaningful effects, particularly for haemorrhagic stroke (OR =
428 0.86; 95% CI: 0.71, 1.05; p-value = 0.14) and aortic valve stenosis (OR = 1.17; 95% CI:
429 0.92, 1.48; p-value = 0.20) (**Figure 3 and Supplementary figure 3**).

430 Higher *SCD/SCD* activity (proxied by increase in POA:PA in standard units) was
431 related to lower odds of coronary artery disease (OR = 0.82; 95% CI: 0.76, 0.88; p-value =
432 1×10^{-7}) (**Figure 3**), which was consistent across studies (**Supplementary figure 3**). There
433 was little evidence supporting a relationship between higher *SCD/SCD* activity (proxied by
434 increase in POA:PA in standard units) and other cardiovascular endpoints. However, some of
435 these results were imprecisely estimated and, therefore, we cannot rule out the presence of
436 clinically meaningful effects, particularly for peripheral artery disease (OR = 0.85; 95% CI:
437 0.68, 1.06; p-value = 0.14), aortic aneurysm (OR = 1.16; 95% CI: 0.88, 1.52; p-value =
438 0.30), and aortic valve stenosis (OR = 0.85; 95% CI: 0.60, 1.22; p-value = 0.38) (**Figure 3**
439 and **Supplementary figure 3**).

440 In East Asian ancestry individuals, there was limited evidence supporting a
441 relationship between *FADS2/D6D* activity and the odds of cardiovascular endpoints.
442 However, statistical power was substantially lower for analyses in East Asian individuals and,
443 therefore, some findings could be compatible with higher *FADS2/D6D* activity (proxied by
444 increase in DGLA:LA in standard units) being related to higher odds of disease, such as for
445 atrial fibrillation (OR = 1.02; 95% CI: 0.92, 1.13; p-value = 0.68), or to lower odds of
446 diseases, such as for coronary artery disease (OR = 0.96; 95% CI: 0.91, 1.00; p-value = 0.04)
447 and haemorrhagic stroke (OR = 0.85; 95% CI: 0.75, 0.97; p-value = 0.01) (**Figure 3**).

448

449 *Exploring bias in Mendelian randomization analyses*

450

451 Horizontal pleiotropy

452

453 Phenome scan

454 In a hypothesis-driven approach, higher D5D/D6D (*FADS1/2*) activity was related to
455 higher LDL-cholesterol, fasting glucose, and type 2 diabetes risk, but lower triglycerides and
456 diastolic blood pressure, among individuals of European (as proxied by AA:DGLA) and East
457 Asian ancestry (as proxied by DGLA:LA). Higher *ELOVL2/ELOVL2* activity (proxied by
458 DHA:DPAn-3) was not related to cardiovascular risk factors at p-value < 0.00625 and higher
459 *SCD/SCD* activity (proxied by POA:PA) was related to lower LDL-cholesterol, triglycerides,
460 systolic and diastolic blood pressure (**Figure 4**);

461 In the hypothesis-free approach, the *FADS1/2* variant (rs174546) was related not only
462 to fatty acids but also to numerous non-fatty acid traits such as lipid, glycaemic, blood cell
463 traits, physical measures (e.g. pulse, heart rate and height), immune-related disorders (e.g.
464 asthma, hypothyroidism, Crohn's disease, inflammatory bowel disease) and several
465 biomarkers (e.g. total bilirubin, insulin growth factor (IGF)-1, cystatin C, alkaline
466 phosphatase and urate) among individuals of European ancestry (**Figure 5** and
467 **Supplementary table 4**). The pleiotropic associations of the *FADS1/2* variant (rs174546)
468 were also seen in East Asians in relation to lipid, glycaemic, blood cell traits (**Figure 5** and
469 **Supplementary table 5**). The *ELOVL2* variant (rs3734398) was related to levels of an
470 unknown metabolite X-12627 and DHA and the *SCD* variant (rs603424) was related to
471 multiple SFA/MUFA, as well as to bone mineral density and blood cell-related traits (**Figure**
472 **5** and **Supplementary table 4**).

473

474 Gene expression and tissue-specific analyses

475 Among Europeans, the selected genetic instruments were strongly associated with the
476 expression of the target genes in one or more tissues (N = 208-670 individuals per tissue).
477 The *FADS1* SNP was related to *FADS1* expression in all tissues analysed, whereas the

478 *ELOVL2* and *SCD* SNPs were related to expression of *ELOVL2* in the liver and expression of
479 *SCD* in (subcutaneous and visceral) adipose tissues, respectively (**Figure 6**). This may be
480 partly related to the expression profile of these genes, since *FADS1* expression is ubiquitous,
481 whereas *ELOVL2* and *SCD* expressions are more prominent in specific tissues
482 (**Supplementary figure 4**). As expected, the alleles increasing the expression of the target
483 genes corresponded to the same alleles increasing the rate of fatty acids biosynthesis, as
484 proxied by markers of enzyme activity, across tissues, except for *FADS1* expression in whole
485 blood (**Figure 6**). Regarding expression of non-target genes, the *FADS1* SNP was related to
486 expression of eleven other genes, including *FADS2*, *FADS3*, *TMEM258*, *MYRF*, and *FEN1*,
487 across multiple tissues, whereas the *ELOVL2* SNP was related to the expression of *SYCP2L*
488 (artery, heart and whole blood tissues) and *ELOVL2-AS1* (liver) and the *SCD* SNP was
489 related to the expression of three other genes, *BLOC1S2*, *PKD2L1* and *OLMALINC*, in
490 subcutaneous adipose tissue only (**Figure 6**).

491 The fact that the selected instruments are related to the expression of nearby non-
492 target genes, which is specially the case for the *FADS1* SNP, could bias our analyses if the
493 proteins encoded by these genes directly influence cardiovascular diseases. To explore that,
494 we used multivariable Mendelian randomization, which supported a direct effect of the target
495 genes (i.e. *FADS1*, *ELOVL2* and *SCD*) on cardiovascular diseases and risk factors. In
496 Europeans, higher *FADS1* expression in multiple tissues (e.g. visceral adipose, artery, heart -
497 atrial appendage, liver and pancreas) was related to higher risk of most cardiovascular
498 diseases (**Supplementary figure 5A** and **5B**) and changes in several risk factors, including
499 higher LDL-cholesterol and fasting glucose (**Supplementary figure 6A** and **6B**). Higher
500 expression of *SCD* in subcutaneous adipose tissue was mostly associated to lower risk of
501 coronary artery disease, heart failure, venous thromboembolism and lower diastolic blood
502 pressure, whereas higher expression of *ELOVL2* was related to higher LDL-cholesterol and

503 type 2 diabetes risk (**Supplementary figure 5A-B** and **6A-B**). The conditional F statistics for
504 these analyses ranged from 18 to 433 and 5 to 50 in unadjusted and adjusted models,
505 respectively (**Supplementary table 6**).

506

507 Confounding by linkage disequilibrium

508

509 Genetic colocalization

510 We used genetic colocalization among Europeans to tease apart whether results from
511 Mendelian randomization analyses were compatible with a shared variant between proxies of
512 enzyme activity (i.e. AA:DGLA, DHA:DPAn-3, and POA:PA) and cardiovascular diseases
513 risk and risk factors, which is a necessary (though not sufficient) condition for two traits to be
514 causally related.

515 For AA:DGLA (proxying D5D activity), there was strong evidence that it colocalised
516 with risk of venous thromboembolism (PPA = 85% for a shared variant). For other
517 cardiovascular disease outcomes, PPA was 0%-27% for a shared variant accompanied by
518 PPA of 60%-100% for the variant being associated with enzyme activity only, which could
519 be a result of limited statistical power (**Supplementary table 7** and **Supplementary figure**
520 **7**). In addition, the top variant for AA:DGLA (rs174546) was in strong LD with the top
521 variants for peripheral artery disease, venous thromboembolism, ischemic stroke, and heart
522 failure ($R^2 > 0.8$ using 1000G European population), although in weak/moderate LD with top
523 variants for coronary artery disease, atrial fibrillation, and aortic valve stenosis ($R^2 = 0.19$ -
524 0.42) (**Supplementary table 8**). After conditioning the genetic association data for
525 AA:DGLA and venous thromboembolism on rs174546, there was no evidence of a distinct or
526 shared genetic signal (**Supplementary table 9**).

527 There was supportive evidence for a shared causal variant of AA:DGLA with LDL-
528 cholesterol (PPA for shared variant = 87%) and for distinct causal variants with systolic,
529 diastolic blood pressure, and triglycerides (PPA for distinct variants = 90-100%). Evidence
530 was less conclusive for colocalization of AA:DGLA with fasting glucose (PPA for shared
531 variant = 64%) (**Supplementary table 10** and **Supplementary figure 8**). The top variant for
532 AA:DGLA (rs174546) was in strong LD with the top variants for LDL-cholesterol,
533 triglycerides and glucose ($R^2 > 0.92$ using 1000G European population) (**Supplementary**
534 **table 8**).

535 For POA:PA (proxying *SCD/SCD* activity), there was strong evidence that it
536 colocalised with risk of coronary artery disease (PPA = 99% for a shared variant)
537 (**Supplementary table 7** and **supplementary figure 7**). After conditioning the genetic
538 association data for POA:PA and coronary artery disease on rs603424, there was no evidence
539 of a distinct or shared genetic signal (**Supplementary table 9**).

540 Colocalization analyses supported distinct genetic variants between POA:PA and
541 LDL-cholesterol, triglycerides, systolic, diastolic blood pressure, type2 diabetes and body
542 mass index (PPA for distinct variants = 86-100%) (**Supplementary table 10** and
543 **Supplementary figure 8**).

544

545 Confounding by population stratification, assortative mating and indirect genetic effects

546 The *FADS1/2* SNP (rs174546) was associated with both negative control outcomes
547 among Europeans: skin colour [mean change of 0.005 unit per C allele ($p\text{-value} = 5 \times 10^{-5}$)]
548 and ease of skin tanning [mean change of -0.006 unit per C allele ($p\text{-value} = 0.007$)], while
549 the *SCD* SNP was associated with ease of skin tanning [mean change of -0.006 unit increase
550 per C allele ($p\text{-value} = 0.037$)] (**Supplementary table 11**).

551 To explore the potential impact of confounding from population stratification,
552 assortative mating and indirect genetic effects, we compared within-sibship associations of
553 the selected genetic variants with cardiovascular risk factors with estimates from unrelated
554 individuals. Estimates were broadly consistent indicating that our findings are unlikely to
555 have been affected by population stratification, assortative mating or indirect genetic effects
556 (**Figure 7**). As an example, for the *FADS1/2* SNP (rs174546), each C allele was related to a
557 mean LDL-cholesterol increase of 0.041 (95% CI: 0.025; 0.057) and 0.036 (95% CI: 0.025;
558 0.046) SD units within-siblings and in unrelated individuals, respectively.

559

560 Selection bias

561

562 Positive control analyses

563 Overall, we observed the expected effect of well-established risk factors on the
564 development of cardiovascular diseases across studies (**Supplementary figure 9**). Systolic,
565 diastolic blood pressure and body mass index were related to higher odds of all
566 cardiovascular disease outcomes in studies of European ancestry individuals (i.e. UK
567 Biobank, genetic association metanalyses and FinnGen) and higher odds of most
568 cardiovascular outcomes (except for coronary and peripheral artery disease) in Biobank
569 Japan. Higher LDL-cholesterol and triglycerides, and liability to type 2 diabetes were related
570 to higher odds of coronary artery disease and peripheral artery disease across studies for both
571 ancestries, while glucose and smoking were related to higher odds of peripheral artery disease
572 in Europeans and East Asians. There were a few instances where these risk factors were
573 related to lower odds of disease, such as type 2 diabetes liability with hemorrhagic stroke

574 (Biobank Japan) and LDL-cholesterol with hemorrhagic stroke (UK Biobank and Biobank
575 Japan).

576

577 **DISCUSSION**

578

579 *Main findings*

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

591 By triangulating multiple approaches, our results are compatible with higher LDL-
592 cholesterol (and possibly glucose) being a downstream effect of higher *FADS1/D5D* activity
593 instead of explained by confounding by LD or by the co-expression of other genes in the
594 region. Given the well-established role of *FADS1/D5D* activity in PUFA biosynthesis and the
595 well-known involvement of LDL-cholesterol in the aetiology of multiple cardiovascular
596 diseases, this strengthens the evidence for a causal relationship and provides a putative
597 mediating pathway for the effect of PUFA biosynthesis on the risk of cardiovascular diseases.

598

599 *Previous literature*

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

605 Previous meta-analyses of classical observational studies indicate that higher
606 circulating long-chain omega-3 and omega-6 PUFA are either not associated or are associated
607 with lower risk of coronary artery disease and stroke (68-72), whereas higher circulating
608 MUFA and SFA are either not associated or are associated with higher risk of coronary artery
609 disease and stroke (68, 69). Recent systematic reviews of randomized controlled trials of
610 dietary advice or supplementation of omega-3 and omega-6 PUFA have suggested little to no
611 benefit in reducing the risk of cardiovascular diseases (73-75). However, most studies
612 included in these systematic reviews were at moderate to high risk of bias and there is large
613 uncertainty on the evidence linking PUFA to some cardiovascular outcomes (73-75). It is
614 important to emphasise that comparing our findings to previous classical observational and
615 randomized controlled trials deserves caution as our genetic instruments have a broad impact
616 on the fatty acids pool and, therefore, cannot be used to make inferences about individual
617 fatty acids/fatty acids classes. As an example, higher *FADS1/D5D* activity (instrumented by
618 rs174546) is related to higher longer chain omega-3 and omega-6 PUFA (e.g. AA, EPA and
619 DHA) but lower shorter chain omega-3 and omega-6 PUFA (e.g. LA and ALA).

620 Several GWAS have reported that SNPs within the *FADS1/2* locus are associated with
621 cardiovascular risk factors (e.g. LDL-cholesterol and triglycerides) (19, 76, 77) and previous

622 Mendelian randomization studies have reported that longer and shorter chain PUFA are
623 related to risk of cardiovascular diseases in contrasting directions among Europeans,
624 including coronary artery disease in CARDIoGRAMplusC4D and UK Biobank (26, 27, 78),
625 ischemic stroke in MEGASTROKE and UK Biobank (28, 29), and venous thromboembolism
626 in UK Biobank (29). Our findings expand on previous Mendelian randomization studies by
627 implicating higher *FADS1/D5D* activity in the development of a wide range of cardiovascular
628 diseases among Europeans in the largest available samples to date (up to 1,153,768
629 individuals). In addition, to our knowledge, this is the first Mendelian randomization study to
630 report a potential protective effect of higher *SCD* activity on coronary artery disease among
631 Europeans and to explore the relation between *FADS2/D6D* activity and cardiovascular
632 diseases among East Asians.

633

634 *Plausibility of Mendelian randomization assumptions*

635 A major challenge in Mendelian randomization studies is the unprovable assumption
636 that the estimated effect of the genetic instrument on the outcome is mediated by the
637 exposure, and not biased by horizontal pleiotropy, population stratification, assortative
638 mating, indirect genetic effects or selection bias (59, 66, 79). We assessed the plausibility that
639 our findings were explained by these sources of bias through a series of sensitivity analyses.

640 To mitigate bias due to horizontal pleiotropy (i.e. the genetic instrument influences
641 exposure and outcome via independent pathways), we have restricted our analyses to genetic
642 variants near genes with well-established role in fatty acids biosynthesis. We have confirmed
643 that these variants have the expected impact on the circulating fatty acids pool and on the
644 expression of the target genes in key tissues (except for *FADS1* in whole blood). Previous
645 evidence confirms that the selected *FADS1/2* and *SCD* variants, or variants in high LD, are

646 related to changes in fatty acids composition across multiple sites, including adipose tissue
647 (80-82), brain (83) and liver (84). Genetic colocalization and multivariable Mendelian
648 randomization (adjusting for co-expressed genes in the region) supported a causal relation
649 between D5D activity, venous thromboembolism and LDL-cholesterol, and between SCD
650 activity and coronary artery disease. It is important to note that we were likely underpowered
651 to test colocalization between fatty acids biosynthesis and cardiovascular disease outcomes.
652 Where there was evidence that the selected genetic variant was associated with the expression
653 of non-target genes in the region in a given tissue, findings from multivariable Mendelian
654 randomization were consistent with expression of the target gene (i.e. *FADS1*) having direct
655 effects on the outcome.

656 Confounding could be introduced in Mendelian randomization studies due to
657 population stratification, assortative mating and indirect genetic effects. Of these factors,
658 population stratification is likely to be of the most concern for this study (59). Despite
659 attempts to control for population structure in genetic association data, there could still be
660 residual population structure (63). We showed that within-sibship associations of *FADS1/2*,
661 *ELOVL2*, and *SCD* variants with established cardiovascular risk factors were broadly similar
662 to estimates from unrelated individuals, suggesting that our results are unlikely to be affected
663 by population stratification, assortative mating or indirect genetic effects of parents.

664 Non-random sample selection may introduce bias in Mendelian randomization studies
665 especially if the mechanism of selection depends on the exposure and/or outcome (66, 85).
666 Using a positive control approach, we were able to identify the expected effect of well-
667 established risk factors on cardiovascular diseases across studies contributing with data on
668 cardiovascular disease endpoints, which is reassuring given our concerns that case-control
669 ascertainment could introduce bias in the analyses. Although results from the positive control
670 approach argues against selection being a major source of bias in this study, we cannot fully

671 rule out that selection might have introduced some bias in our analyses as bias due to
672 selection will depend on context-specific causal structures underlying the data under
673 consideration.

674

675 *Implications*

676 Our findings are supportive of the involvement of fatty acids biosynthesis in the
677 aetiology of cardiovascular diseases and highlight D5D and SCD as potential targets for the
678 prevention of these diseases. Further work is needed to understand the precise underlying
679 mechanism(s) are required.

680 The relation between D5D activity and cardiovascular diseases is plausibly mediated
681 by one or more fatty acids involved in the PUFA biosynthesis pathway. Given the ubiquitous
682 impact of higher D5D activity on the circulating PUFA pool, we cannot pinpoint which
683 specific fatty acids are driving these effects. For illustration, higher D5D activity decreases
684 LA and ALA (and other omega-3 and omega-6 PUFA upstream of the reaction catalysed by
685 D5D). Lower LA may relate to unfavourable metabolic changes, such as higher plasma LDL-
686 cholesterol, apolipoprotein B, and triglycerides, and haemoglobin A1c (2, 86) and, therefore,
687 is a plausible mediator of the relation between higher D5D activity and higher cardiovascular
688 diseases risk. On the other hand, higher D5D activity increases long-chain PUFA such as AA,
689 which influences key membrane/tissue functions, such as membrane fluidity, the activity of
690 membrane-bound receptors, transport proteins and signal transmission (87), and is a
691 precursor for eicosanoids (e.g. prostaglandins, leukotrienes, and thromboxane), which are
692 involved in inflammation, platelet aggregation and vascular remodeling (88).

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

695 obesity, insulin resistance, and hepatic steatosis (89-91), but show increased inflammation
696 and atherogenesis (91, 92). The putative protective effect of higher SCD activity on coronary
697 artery disease might be related to lower availability of palmitic acid and consequent lower
698 production of its toxic metabolites, such as ceramides (93).

699

700 *Conclusions*

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

705

706 **Supplemental information description**

- 707 • Supplementary methods
- 708 • Supplementary tables 1 to 10
- 709 • Supplementary figures 1 to 9

710

711 **Acknowledgements**

712

713 We want to acknowledge participants and investigators from FinnGen study, UK
714 Biobank, Biobank Japan, and the multiple large-scale GWAS consortia which made summary
715 data available, including CARDIoGRAMplusC4D, MEGASTROKE, HERMES, atrial
716 fibrillation GWAS metanalysis, aortic aneurysm GWAS metanalysis, CHARGE, and the
717 Within Family Consortium. This work was carried out using UK Biobank project 15825.
718 Data on coronary artery disease have been contributed by CARDIoGRAMplusC4D
719 investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG'. The
720 MEGASTROKE project received funding from sources specified at
721 <http://www.megastroke.org/acknowledgments.html> (a list of all MEGASTROKE authors can
722 be found in the appendix A - Supplementary methods). This work was carried out using the
723 computational facilities of the Advanced Computing Research Centre, University of Bristol -
724 <http://www.bristol.ac.uk/acrc/>. Quality Control filtering of the UK Biobank data was
725 conducted by R.Mitchell, G.Hemani, T.Dudding, L.Corbin, S.Harrison, L.Paternoster as
726 described in the published protocol (doi: 10.5523/bris.1ovaau5sxunp2cv8rcy88688v). The
727 MRC IEU UK Biobank GWAS pipeline was developed by B.Elsworth, R.Mitchell,
728 C.Raistrick, L.Paternoster, G.Hemani, T.Gaunt (doi:
729 10.5523/bris.pnoat8cxo0u52p6ynfaekeigi).

730

731 **Sources of funding**

732

733 MCB was supported by a UK Medical Research Council (MRC) Skills Development
734 Fellowship (MR/P014054/1). MCB and DAL are supported by the British Heart Foundation
735 (AA/18/7/34219) and DAL is supported by a British Heart Foundation Chair
736 (CH/F/20/90003). MCB, PH, JZ, GH, LJH, TRG, GDS, and DAL work in a Unit receives
737 funding from the University of Bristol and UK Medical Research Council (MRC)

738 (MC_UU_00011/1, MC_UU_00011/4 and MC_UU_00011/6). PCH was supported by
739 Cancer Research UK (C52724/A20138 & C18281/A29019). MVH works in a unit that
740 receives funding from the UK Medical Research Council and is supported by a British Heart
741 Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512) and the National
742 Institute for Health Research Oxford Biomedical Research Centre. The funders had no role in
743 the design, analyses, interpretation of results or writing of the paper. The views expressed in
744 this paper are those of the authors and not necessarily any of the funders.

745

746 **Declaration of interests**

747

748 DAL receives support from several national and international government and
749 charitable research funders, as well as from Medtronic Ltd and Roche Diagnostics for
750 research unrelated to that presented here. TRG receives funding from Biogen for work
751 unrelated to that presented here. MVH has consulted for Boehringer Ingelheim, and in
752 adherence to the University of Oxford's Clinical Trial Service Unit & Epidemiological
753 Studies Unit (CSTU) staff policy, did not accept personal honoraria or other payments from
754 pharmaceutical companies.

755 All other authors declare no competing interests.

756

757 **Web resources**

758 Coloc: <https://cran.r-project.org/web/packages/coloc/index.html>

759 GWIS: <https://sites.google.com/site/mgnivard/gwis>

760 IEU Open GWAS Project: <https://gwas.mrcieu.ac.uk/>

761 MVMR R package: <https://github.com/WSpiller/MVMR>

762 PWCoCco: <https://github.com/jwr-git/pwcoco>

763 TwoSampleMR R package: <https://mrcieu.github.io/TwoSampleMR>

764

765 **Data availability**

766

767 Genetic association data on fatty acids, cardiovascular diseases (CVDs) and CVDs
768 risk factors are publicly available or available on request from original investigators as
769 detailed in Supplementary tables 1 and 2.

770

771 **References**

772

773 1 Ratnayake, W.M. and Galli, C. (2009) Fat and fatty acid terminology, methods of analysis and
774 fat digestion and metabolism: a background review paper. *Ann Nutr Metab*, **55**, 8-43.

775 2 Mensink, R.P., Zock, P.L., Kester, A.D. and Katan, M.B. (2003) Effects of dietary fatty acids
776 and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and
777 apolipoproteins: a meta-analysis of 60 controlled trials. *Am J Clin Nutr*, **77**, 1146-1155.

778 3 Balk, E.M., Lichtenstein, A.H., Chung, M., Kupelnick, B., Chew, P. and Lau, J. (2006) Effects of
779 omega-3 fatty acids on serum markers of cardiovascular disease risk: a systematic review.
780 *Atherosclerosis*, **189**, 19-30.

781 4 Hooper, L., Summerbell, C.D., Thompson, R., Sills, D., Roberts, F.G., Moore, H.J. and Davey
782 Smith, G. (2012) Reduced or modified dietary fat for preventing cardiovascular disease. *Cochrane*
783 *Database Syst Rev*, in press., CD002137.

784 5 Sanders, T.A. (2009) Fat and fatty acid intake and metabolic effects in the human body. *Ann*
785 *Nutr Metab*, **55**, 162-172.

786 6 Schwingshackl, L., Strasser, B. and Hoffmann, G. (2011) Effects of monounsaturated fatty
787 acids on glycaemic control in patients with abnormal glucose metabolism: a systematic review and
788 meta-analysis. *Ann Nutr Metab*, **58**, 290-296.

789 7 Miller, P.E., Van Elswyk, M. and Alexander, D.D. (2014) Long-chain omega-3 fatty acids
790 eicosapentaenoic acid and docosahexaenoic acid and blood pressure: a meta-analysis of randomized
791 controlled trials. *Am J Hypertens*, **27**, 885-896.

792 8 Shah, M., Adams-Huet, B. and Garg, A. (2007) Effect of high-carbohydrate or high-cis-
793 monounsaturated fat diets on blood pressure: a meta-analysis of intervention trials. *Am J Clin Nutr*,
794 **85**, 1251-1256.

795 9 Mozaffarian, D. and Wu, J.H. (2011) Omega-3 fatty acids and cardiovascular disease: effects
796 on risk factors, molecular pathways, and clinical events. *J Am Coll Cardiol*, **58**, 2047-2067.

797 10 Masson, C.J. and Mensink, R.P. (2011) Exchanging saturated fatty acids for (n-6)
798 polyunsaturated fatty acids in a mixed meal may decrease postprandial lipemia and markers of
799 inflammation and endothelial activity in overweight men. *J Nutr*, **141**, 816-821.

800 11 Schwartz, E.A., Zhang, W.Y., Karnik, S.K., Borwege, S., Anand, V.R., Laine, P.S., Su, Y. and
801 Reaven, P.D. (2010) Nutrient modification of the innate immune response: a novel mechanism by
802 which saturated fatty acids greatly amplify monocyte inflammation. *Arterioscler Thromb Vasc Biol*,
803 **30**, 802-808.

804 12 Sampath, H. and Ntambi, J.M. (2005) Polyunsaturated fatty acid regulation of genes of lipid
805 metabolism. *Annu Rev Nutr*, **25**, 317-340.

806 13 Pase, M.P., Grima, N.A. and Sarris, J. (2011) Do long-chain n-3 fatty acids reduce arterial
807 stiffness? A meta-analysis of randomised controlled trials. *Br J Nutr*, **106**, 974-980.

808 14 Jakobsson, A., Westerberg, R. and Jakobsson, A. (2006) Fatty acid elongases in mammals:
809 their regulation and roles in metabolism. *Prog Lipid Res*, **45**, 237-249.

810 15 Lemaitre, R.N., Tanaka, T., Tang, W., Manichaikul, A., Foy, M., Kabagambe, E.K., Nettleton,
811 J.A., King, I.B., Weng, L.C., Bhattacharya, S. et al. (2011) Genetic loci associated with plasma
812 phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE
813 Consortium. *PLoS Genet*, **7**, e1002193.

814 16 Guan, W., Steffen, B.T., Lemaitre, R.N., Wu, J.H., Tanaka, T., Manichaikul, A., Foy, M., Rich,
815 S.S., Wang, L., Nettleton, J.A. et al. (2014) Genome-wide association study of plasma N6
816 polyunsaturated fatty acids within the cohorts for heart and aging research in genomic epidemiology
817 consortium. *Circ Cardiovasc Genet*, **7**, 321-331.

818 17 Wu, J.H., Lemaitre, R.N., Manichaikul, A., Guan, W., Tanaka, T., Foy, M., Kabagambe, E.K.,
819 Djousse, L., Siscovick, D., Fretts, A.M. et al. (2013) Genome-wide association study identifies novel
820 loci associated with concentrations of four plasma phospholipid fatty acids in the de novo

- 821 lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic
822 Epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet*, **6**, 171-183.
- 823 18 Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I.,
824 Forgetta, V., Yang, T.P. *et al.* (2014) An atlas of genetic influences on human blood metabolites. *Nat*
825 *Genet*, **46**, 543-550.
- 826 19 Kettunen, J., Demirkan, A., Wurtz, P., Draisma, H.H., Haller, T., Rawal, R., Vaarhorst, A.,
827 Kangas, A.J., Lyytikainen, L.P., Pirinen, M. *et al.* (2016) Genome-wide study for circulating
828 metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun*, **7**, 11122.
- 829 20 Tintle, N.L., Pottala, J.V., Lacey, S., Ramachandran, V., Westra, J., Rogers, A., Clark, J., Olthoff,
830 B., Larson, M., Harris, W. *et al.* (2015) A genome-wide association study of saturated, mono- and
831 polyunsaturated red blood cell fatty acids in the Framingham Heart Offspring Study. *Prostaglandins*
832 *Leukot Essent Fatty Acids*, **94**, 65-72.
- 833 21 Dorajoo, R., Sun, Y., Han, Y., Ke, T., Burger, A., Chang, X., Low, H.Q., Guan, W., Lemaitre, R.N.,
834 Khor, C.C. *et al.* (2015) A genome-wide association study of n-3 and n-6 plasma fatty acids in a
835 Singaporean Chinese population. *Genes Nutr*, **10**, 53.
- 836 22 Ebrahim, S. and Davey Smith, G. (2008) Mendelian randomization: can genetic epidemiology
837 help redress the failures of observational epidemiology? *Hum Genet*, **123**, 15-33.
- 838 23 Hingorani, A. and Humphries, S. (2005) Nature's randomised trials. *Lancet*, **366**, 1906-1908.
- 839 24 Smith, G.D., Lawlor, D.A., Harbord, R., Timpson, N., Day, I. and Ebrahim, S. (2007) Clustered
840 environments and randomized genes: a fundamental distinction between conventional and genetic
841 epidemiology. *PLoS Med*, **4**, e352.
- 842 25 Holmes, M.V., Ala-Korpela, M. and Smith, G.D. (2017) Mendelian randomization in
843 cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol*, **14**, 577-590.
- 844 26 Jager, S., Cuadrat, R., Hoffmann, P., Wittenbecher, C. and Schulze, M.B. (2020) Desaturase
845 Activity and the Risk of Type 2 Diabetes and Coronary Artery Disease: A Mendelian Randomization
846 Study. *Nutrients*, **12**.
- 847 27 Liao, L.Z., Li, W.D., Liu, Y., Li, J.P., Zhuang, X.D. and Liao, X.X. (2020) Exploring the causal
848 pathway from omega-6 levels to coronary heart disease: A network Mendelian randomization study.
849 *Nutr Metab Cardiovasc Dis*, **30**, 233-240.
- 850 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
851 for circulating polyunsaturated fatty acids in ischemic stroke and modifiable factors: a Mendelian
852 randomization study. *Nutr J*, **19**, 70.
- 853 29 Yuan, S., Back, M., Bruzelius, M., Mason, A.M., Burgess, S. and Larsson, S. (2019) Plasma
854 Phospholipid Fatty Acids, FADS1 and Risk of 15 Cardiovascular Diseases: A Mendelian Randomisation
855 Study. *Nutrients*, **11**.
- 856 30 Zhang, T., Zhao, J.V. and Schooling, C.M. (2021) The associations of plasma phospholipid
857 arachidonic acid with cardiovascular diseases: A Mendelian randomization study. *EBioMedicine*, **63**,
858 103189.
- 859 31 Chen, H.Y., Cairns, B.J., Small, A.M., Burr, H.A., Ambikumar, A., Martinsson, A., Theriault, S.,
860 Munter, H.M., Steffen, B., Zhang, R. *et al.* (2020) Association of FADS1/2 Locus Variants and
861 Polyunsaturated Fatty Acids With Aortic Stenosis. *JAMA Cardiol*, **5**, 694-702.
- 862 32 Yuan, S. and Larsson, S.C. (2019) Plasma Phospholipid Fatty Acids and Risk of Atrial
863 Fibrillation: A Mendelian Randomization Study. *Nutrients*, **11**.
- 864 33 Holmes, M.V., Richardson, T.G., Ference, B.A., Davies, N.M. and Davey Smith, G. (2021)
865 Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug
866 development. *Nat Rev Cardiol*, **18**, 435-453.
- 867 34 Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou,
868 T., Nelson, C.P., Hopewell, J.C. *et al.* (2015) A comprehensive 1,000 Genomes-based genome-wide
869 association meta-analysis of coronary artery disease. *Nat Genet*, **47**, 1121-1130.
- 870 35 Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs,
871 L., Giese, A.K., van der Laan, S.W., Gretarsdottir, S. *et al.* (2018) Multiancestry genome-wide

- 872 association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes.
873 *Nat Genet*, **50**, 524-537.
- 874 36 Nielsen, J.B., Thorolfsson, R.B., Fritsche, L.G., Zhou, W., Skov, M.W., Graham, S.E., Herron,
875 T.J., McCarthy, S., Schmidt, E.M., Sveinbjornsson, G. *et al.* (2018) Biobank-driven genomic discovery
876 yields new insight into atrial fibrillation biology. *Nat Genet*, **50**, 1234-1239.
- 877 37 Jones, G.T., Tromp, G., Kuivaniemi, H., Gretarsdottir, S., Baas, A.F., Giusti, B., Strauss, E.,
878 Van't Hof, F.N., Webb, T.R., Erdman, R. *et al.* (2017) Meta-Analysis of Genome-Wide Association
879 Studies for Abdominal Aortic Aneurysm Identifies Four New Disease-Specific Risk Loci. *Circ Res*, **120**,
880 341-353.
- 881 38 Shah, S., Henry, A., Roselli, C., Lin, H., Sveinbjornsson, G., Fatemifar, G., Hedman, A.K., Wilk,
882 J.B., Morley, M.P., Chaffin, M.D. *et al.* (2020) Genome-wide association and Mendelian
883 randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun*, **11**, 163.
- 884 39 Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P.,
885 Green, J., Landray, M. *et al.* (2015) UK biobank: an open access resource for identifying the causes of
886 a wide range of complex diseases of middle and old age. *PLoS Med*, **12**, e1001779.
- 887 40 Allen, N.E., Sudlow, C., Peakman, T., Collins, R. and Biobank, U.K. (2014) UK biobank data:
888 come and get it. *Sci Transl Med*, **6**, 224ed224.
- 889 41 FinnGen. (2020), in press.
- 890 42 Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T.,
891 Tamakoshi, A., Yamagata, Z., Mushiroda, T. *et al.* (2017) Overview of the BioBank Japan Project:
892 Study design and profile. *J Epidemiol*, **27**, S2-S8.
- 893 43 Hirata, M., Kamatani, Y., Nagai, A., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z.,
894 Kubo, M., Muto, K., Mushiroda, T. *et al.* (2017) Cross-sectional analysis of BioBank Japan clinical
895 data: A large cohort of 200,000 patients with 47 common diseases. *J Epidemiol*, **27**, S9-S21.
- 896 44 Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S.,
897 Matoba, N., Low, S.K., Okada, Y. *et al.* (2020) Large-scale genome-wide association study in a
898 Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*, **52**, 669-
899 679.
- 900 45 Low, S.K., Takahashi, A., Ebana, Y., Ozaki, K., Christophersen, I.E., Ellinor, P.T., Consortium,
901 A.F., Ogishima, S., Yamamoto, M., Satoh, M. *et al.* (2017) Identification of six new genetic loci
902 associated with atrial fibrillation in the Japanese population. *Nat Genet*, **49**, 953-958.
- 903 46 Zhu, J., Manichaikul, A., Hu, Y., Chen, Y.L., Liang, S., Steffen, L.M., Rich, S.S., Tsai, M.,
904 Siscovick, D.S., Lemaitre, R.N. *et al.* (2017) Meta-analysis of genome-wide association studies
905 identifies three novel loci for saturated fatty acids in East Asians. *Eur J Nutr*, **56**, 1477-1484.
- 906 47 Hu, Y., Tanaka, T., Zhu, J., Guan, W., Wu, J.H.Y., Psaty, B.M., McKnight, B., King, I.B., Sun, Q.,
907 Richard, M. *et al.* (2017) Discovery and fine-mapping of loci associated with MUFAs through trans-
908 ethnic meta-analysis in Chinese and European populations. *J Lipid Res*, **58**, 974-981.
- 909 48 Nieuwboer, H.A., Pool, R., Dolan, C.V., Boomsma, D.I. and Nivard, M.G. (2016) GWIS:
910 Genome-Wide Inferred Statistics for Functions of Multiple Phenotypes. *Am J Hum Genet*, **99**, 917-
911 927.
- 912 49 Burgess, S., Davies, N.M. and Thompson, S.G. (2016) Bias due to participant overlap in two-
913 sample Mendelian randomization. *Genet Epidemiol*, **40**, 597-608.
- 914 50 Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess,
915 S., Bowden, J., Langdon, R. *et al.* (2018) The MR-Base platform supports systematic causal inference
916 across the human phenome. *Elife*, **7**.
- 917 51 Pickrell, J.K., Berisa, T., Liu, J.Z., Segurel, L., Tung, J.Y. and Hinds, D.A. (2016) Detection and
918 interpretation of shared genetic influences on 42 human traits. *Nat Genet*, **48**, 709-717.
- 919 52 Yang, Q., Sanderson, E., Tilling, K., Borges, M.C. and Lawlor, D.A. (2019) Exploring and
920 mitigating potential bias when genetic instrumental variables are associated with multiple non-
921 exposure traits in Mendelian randomization. *medRxiv*, in press., 19009605.

- 922 53 Elsworth, B., Lyon, M., Alexander, T., Liu, Y., Matthews, P., Hallett, J., Bates, P., Palmer, T.,
923 Haberland, V., Smith, G.D. *et al.* (2020) The MRC IEU OpenGWAS data infrastructure. *bioRxiv*, in
924 press., 2020.2008.2010.244293.
- 925 54 Sanderson, E., Davey Smith, G., Windmeijer, F. and Bowden, J. (2018) An examination of
926 multivariable Mendelian randomization in the single-sample and two-sample summary data settings.
927 *Int J Epidemiol*, in press.
- 928 55 Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and
929 Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using
930 summary statistics. *PLoS Genet*, **10**, e1004383.
- 931 56 Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide
932 complex trait analysis. *Am J Hum Genet*, **88**, 76-82.
- 933 57 Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C., Replication,
934 D.I.G., Meta-analysis, C., Madden, P.A., Heath, A.C., Martin, N.G. *et al.* (2012) Conditional and joint
935 multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex
936 traits. *Nat Genet*, **44**, 369-375, S361-363.
- 937 58 Young, A.I., Benonisdottir, S., Przeworski, M. and Kong, A. (2019) Deconstructing the sources
938 of genotype-phenotype associations in humans. *Science*, **365**, 1396-1400.
- 939 59 Morris, T.T., Davies, N.M., Hemani, G. and Smith, G.D. (2020) Population phenomena inflate
940 genetic associations of complex social traits. *Sci Adv*, **6**, eaay0328.
- 941 60 Hartwig, F.P., Davies, N.M. and Davey Smith, G. (2018) Bias in Mendelian randomization due
942 to assortative mating. *Genet Epidemiol*, **42**, 608-620.
- 943 61 Davies, N.M., Howe, L.J., Brumpton, B., Havdahl, A., Evans, D.M. and Davey Smith, G. (2019)
944 Within family Mendelian randomization studies. *Hum Mol Genet*, **28**, R170-R179.
- 945 62 Aneur, A., Enroth, S., Johansson, A., Zaboli, G., Igl, W., Johansson, A.C., Rivas, M.A., Daly,
946 M.J., Schmitz, G., Hicks, A.A. *et al.* (2012) Genetic adaptation of fatty-acid metabolism: a human-
947 specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am J*
948 *Hum Genet*, **90**, 809-820.
- 949 63 Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., Hemani, G., Davey
950 Smith, G. and Timpson, N.J. (2020) Is population structure in the genetic biobank era irrelevant, a
951 challenge, or an opportunity? *Hum Genet*, **139**, 23-41.
- 952 64 Hartwig, F.P., Davies, N.M., Horta, B.L., Ahluwalia, T.S., Bisgaard, H., Bonnelykke, K., Caspi,
953 A., Moffitt, T.E., Poulton, R., Sajjad, A. *et al.* (2019) Effect modification of FADS2 polymorphisms on
954 the association between breastfeeding and intelligence: results from a collaborative meta-analysis.
955 *Int J Epidemiol*, **48**, 45-57.
- 956 65 Sanderson, E., Richardson, T.G., Hemani, G. and Davey Smith, G. (2021) The use of negative
957 control outcomes in Mendelian randomization to detect potential population stratification. *Int J*
958 *Epidemiol*, in press.
- 959 66 Hughes, R.A., Davies, N.M., Davey Smith, G. and Tilling, K. (2019) Selection Bias When
960 Estimating Average Treatment Effects Using One-sample Instrumental Variable Analysis.
961 *Epidemiology*, **30**, 350-357.
- 962 67 Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R. and
963 Allen, N.E. (2017) Comparison of Sociodemographic and Health-Related Characteristics of UK
964 Biobank Participants With Those of the General Population. *Am J Epidemiol*, **186**, 1026-1034.
- 965 68 Borges, M.C., Schmidt, A.F., Jefferis, B., Wannamethee, S.G., Lawlor, D.A., Kivimaki, M.,
966 Kumari, M., Gaunt, T.R., Ben-Shlomo, Y., Tillin, T. *et al.* (2020) Circulating Fatty Acids and Risk of
967 Coronary Heart Disease and Stroke: Individual Participant Data Meta-Analysis in Up to 16 126
968 Participants. *J Am Heart Assoc*, **9**, e013131.
- 969 69 Chowdhury, R., Warnakula, S., Kunutsor, S., Crowe, F., Ward, H.A., Johnson, L., Franco, O.H.,
970 Butterworth, A.S., Forouhi, N.G., Thompson, S.G. *et al.* (2014) Association of dietary, circulating, and
971 supplement fatty acids with coronary risk: a systematic review and meta-analysis. *Ann Intern Med*,
972 **160**, 398-406.

- 973 70 Chowdhury, R., Stevens, S., Gorman, D., Pan, A., Warnakula, S., Chowdhury, S., Ward, H.,
974 Johnson, L., Crowe, F., Hu, F.B. *et al.* (2012) Association between fish consumption, long chain
975 omega 3 fatty acids, and risk of cerebrovascular disease: systematic review and meta-analysis. *BMJ*,
976 **345**, e6698.
- 977 71 Iso, H., Sato, S., Umemura, U., Kudo, M., Koike, K., Kitamura, A., Imano, H., Okamura, T.,
978 Naito, Y. and Shimamoto, T. (2002) Linoleic acid, other fatty acids, and the risk of stroke. *Stroke*, **33**,
979 2086-2093.
- 980 72 Marklund, M., Wu, J.H.Y., Imamura, F., Del Gobbo, L.C., Fretts, A., de Goede, J., Shi, P., Tintle,
981 N., Wennberg, M., Aslibekyan, S. *et al.* (2019) Biomarkers of Dietary Omega-6 Fatty Acids and
982 Incident Cardiovascular Disease and Mortality. *Circulation*, **139**, 2422-2436.
- 983 73 Abdelhamid, A.S., Martin, N., Bridges, C., Brainard, J.S., Wang, X., Brown, T.J., Hanson, S.,
984 Jimoh, O.F., Ajabnoor, S.M., Deane, K.H. *et al.* (2018) Polyunsaturated fatty acids for the primary and
985 secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*, **7**, CD012345.
- 986 74 Hooper, L., Al-Khudairy, L., Abdelhamid, A.S., Rees, K., Brainard, J.S., Brown, T.J., Ajabnoor,
987 S.M., O'Brien, A.T., Winstanley, L.E., Donaldson, D.H. *et al.* (2018) Omega-6 fats for the primary and
988 secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*, **7**, CD011094.
- 989 75 Abdelhamid, A.S., Brown, T.J., Brainard, J.S., Biswas, P., Thorpe, G.C., Moore, H.J., Deane,
990 K.H., Summerbell, C.D., Worthington, H.V., Song, F. *et al.* (2020) Omega-3 fatty acids for the primary
991 and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*, **3**, CD003177.
- 992 76 Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx,
993 B.W., Janssens, A.C., Wilson, J.F., Spector, T. *et al.* (2009) Loci influencing lipid levels and coronary
994 heart disease risk in 16 European population cohorts. *Nat Genet*, **41**, 47-55.
- 995 77 Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L.,
996 Bennett, D., Li, Y., Tanaka, T. *et al.* (2009) Common variants at 30 loci contribute to polygenic
997 dyslipidemia. *Nat Genet*, **41**, 56-65.
- 998 78 Park, S., Lee, S., Kim, Y., Lee, Y., Kang, M., Kim, K., Kim, Y., Han, S., Lee, H., Lee, J. *et al.* (2021)
999 Causal Effects of Serum Levels of n-3 or n-6 Polyunsaturated Fatty Acids on Coronary Artery Disease:
1000 Mendelian Randomization Study. *Nutrients*, **13**.
- 1001 79 Hemani, G., Bowden, J. and Davey Smith, G. (2018) Evaluating the potential role of
1002 pleiotropy in Mendelian randomization studies. *Hum Mol Genet*, **27**, R195-R208.
- 1003 80 Baylin, A., Ruiz-Narvaez, E., Kraft, P. and Campos, H. (2007) alpha-Linolenic acid, Delta6-
1004 desaturase gene polymorphism, and the risk of nonfatal myocardial infarction. *Am J Clin Nutr*, **85**,
1005 554-560.
- 1006 81 Vaittinen, M., Walle, P., Kuosmanen, E., Mannisto, V., Kakela, P., Agren, J., Schwab, U. and
1007 Pihlajamaki, J. (2016) FADS2 genotype regulates delta-6 desaturase activity and inflammation in
1008 human adipose tissue. *J Lipid Res*, **57**, 56-65.
- 1009 82 Marklund, M., Morris, A.P., Mahajan, A., Ingelsson, E., Lindgren, C.M., Lind, L. and Riserus, U.
1010 (2018) Genome-Wide Association Studies of Estimated Fatty Acid Desaturase Activity in Serum and
1011 Adipose Tissue in Elderly Individuals: Associations with Insulin Sensitivity. *Nutrients*, **10**.
- 1012 83 Freemantle, E., Lalovic, A., Mechawar, N. and Turecki, G. (2012) Age and haplotype
1013 variations within FADS1 interact and associate with alterations in fatty acid composition in human
1014 male cortical brain tissue. *PLoS One*, **7**, e42696.
- 1015 84 Wang, L., Athinarayanan, S., Jiang, G., Chalasani, N., Zhang, M. and Liu, W. (2015) Fatty acid
1016 desaturase 1 gene polymorphisms control human hepatic lipid composition. *Hepatology*, **61**, 119-
1017 128.
- 1018 85 Gkatzionis, A. and Burgess, S. (2019) Contextualizing selection bias in Mendelian
1019 randomization: how bad is it likely to be? *Int J Epidemiol*, **48**, 691-701.
- 1020 86 Imamura, F., Micha, R., Wu, J.H., de Oliveira Otto, M.C., Otite, F.O., Abioye, A.I. and
1021 Mozaffarian, D. (2016) Effects of Saturated Fat, Polyunsaturated Fat, Monounsaturated Fat, and
1022 Carbohydrate on Glucose-Insulin Homeostasis: A Systematic Review and Meta-analysis of
1023 Randomised Controlled Feeding Trials. *PLoS Med*, **13**, e1002087.

- 1024 87 Koletzko, B., Reischl, E., Tanjung, C., Gonzalez-Casanova, I., Ramakrishnan, U., Meldrum, S.,
1025 Simmer, K., Heinrich, J. and Demmelmair, H. (2019) FADS1 and FADS2 Polymorphisms Modulate
1026 Fatty Acid Metabolism and Dietary Impact on Health. *Annu Rev Nutr*, **39**, 21-44.
- 1027 88 Mitchell, J.A. and Warner, T.D. (2006) COX isoforms in the cardiovascular system:
1028 understanding the activities of non-steroidal anti-inflammatory drugs. *Nat Rev Drug Discov*, **5**, 75-86.
- 1029 89 Ntambi, J.M., Miyazaki, M., Stoehr, J.P., Lan, H., Kendzioriski, C.M., Yandell, B.S., Song, Y.,
1030 Cohen, P., Friedman, J.M. and Attie, A.D. (2002) Loss of stearoyl-CoA desaturase-1 function protects
1031 mice against adiposity. *Proc Natl Acad Sci U S A*, **99**, 11482-11486.
- 1032 90 Cohen, P., Miyazaki, M., Socci, N.D., Hagge-Greenberg, A., Liedtke, W., Soukas, A.A., Sharma,
1033 R., Hudgins, L.C., Ntambi, J.M. and Friedman, J.M. (2002) Role for stearoyl-CoA desaturase-1 in
1034 leptin-mediated weight loss. *Science*, **297**, 240-243.
- 1035 91 Brown, J.M., Chung, S., Sawyer, J.K., Degirolamo, C., Alger, H.M., Nguyen, T., Zhu, X., Duong,
1036 M.N., Wibley, A.L., Shah, R. *et al.* (2008) Inhibition of stearoyl-coenzyme A desaturase 1 dissociates
1037 insulin resistance and obesity from atherosclerosis. *Circulation*, **118**, 1467-1475.
- 1038 92 MacDonald, M.L., van Eck, M., Hildebrand, R.B., Wong, B.W., Bissada, N., Ruddle, P.,
1039 Kontush, A., Hussein, H., Pouladi, M.A., Chapman, M.J. *et al.* (2009) Despite antiatherogenic
1040 metabolic characteristics, SCD1-deficient mice have increased inflammation and atherosclerosis.
1041 *Arterioscler Thromb Vasc Biol*, **29**, 341-347.
- 1042 93 Jensen, P.N., Fretts, A.M., Hoofnagle, A.N., Sitlani, C.M., McKnight, B., King, I.B., Siscovick,
1043 D.S., Psaty, B.M., Heckbert, S.R., Mozaffarian, D. *et al.* (2020) Plasma Ceramides and Sphingomyelins
1044 in Relation to Atrial Fibrillation Risk: The Cardiovascular Health Study. *J Am Heart Assoc*, **9**, e012853.
- 1045
- 1046

1047 **FIGURE TITLES AND LEGENDS**

1048

1049 **Figure 1.** Overview of desaturation and elongation reactions involved in the conversion of
1050 MUFA from SFA and of longer-chain omega-3 and omega-6 PUFA from their shorter chain
1051 precursors.

1052 Red boxes highlight enzymes encoded by genes mapped to genetic variants strongly associated with circulating
1053 fatty acids in genome-wide association studies: delta-5 desaturase (D5D), delta-6 desaturase (D6D), fatty acid
1054 elongase 2 (ELOVL2), and stearoyl-CoA desaturase (SCD) encoded by *FADS1* (ENSG00000149485), *FADS2*
1055 (ENSG00000134824), *ELOVL2* (ENSG00000197977), *SCD* (ENSG00000099194), respectively.

1056 *Saturated fatty acids (SFA)*: PA: palmitic acid; SA: stearic acid. *Monounsaturated fatty acids (MUFA)*: POA:
1057 palmitoleic acid; OA: oleic acid. *ω-6 polyunsaturated fatty acids (PUFA)*: LA: linoleic acid; GLA: γ-linolenic
1058 acid; EDA: eicosadienoic acid; DGLA: dihomo-γ-linolenic acid; AA: arachidonic acid; ADA: adrenic acid;
1059 TTA: tetracosatetraenoic acid; TPA: tetracosapentaenoic acid; DPA: docosapentaenoic acid. *ω-3*
1060 *polyunsaturated fatty acids (PUFA)*: ALA: α-linolenic acid; SDA: stearidonic acid; ETE: eicosatrienoic acid;
1061 ETA: eicosatetraenoic acid; EPA: eicosapentaenoic acid; DPA: docosapentaenoic acid; TPA:
1062 tetracosapentaenoic acid; THA: tetracosahexaenoic acid; DHA: docosahexaenoic acid.

1063

1064 **Figure 2.** Schematic representation of multiple putative mechanisms underlying the
1065 association of genetic variants with fatty acids biosynthesis and cardiovascular disease
1066 (CVD) risk.

1067 Figure 2A represents vertical pleiotropy, in which the effect of genetic instruments on CVD is mediated by fatty
1068 acids biosynthesis. Figures 2B to 2E represent alternative mechanisms that could bias Mendelian randomization
1069 findings: horizontal pleiotropy (2B), in which the genetic variant influences fatty acids biosynthesis and CVD
1070 via two different biological pathways, (2C) confounding by *linkage disequilibrium* (LD), in which the selected
1071 genetic variant is in LD with another genetic variant influencing CVD independently, (2D) confounding, in
1072 which different phenomena can introduce spurious association between genetic variant and CVD in samples of
1073 unrelated individuals, and (2E) selection bias, in which selection into the study creates a spurious association
1074 between genetic variant and CVD due to collider stratification bias. MVMR: multivariable Mendelian
1075 randomization.

1076

1077 **Figure 3.** Mendelian randomization results for the risk of cardiovascular diseases related to
1078 increasing activity of enzymes coded by *FADS1/2* (D5D/D6D), *ELOVL2* (ELOVL2) and
1079 *SCD* (SCD) among individuals of European and East Asian ancestries.

1080 Results are expressed as odds ratio of cardiovascular diseases per standard unit increase in the marker of enzyme
1081 activity for *FADS1/2* (i.e. AA:DGLA ratio in Europeans and DGLA:LA ratio in East Asians), *ELOVL2* (i.e.
1082 DHA:DPA ratio in Europeans) and *SCD* (i.e. POA:PA ratio in Europeans) loci. For individuals of European
1083 ancestry, SNP-cardiovascular diseases association data were meta-analysed across multiple genetic association
1084 consortia, UK Biobank and FinnGen. For individuals of East Asian ancestry, SNP-cardiovascular diseases

1085 association data were extracted from BioBank Japan. Full symbols indicate associations at P-value lower than
1086 the P-value threshold accounting for multiple testing ($P < 0.006$ for Europeans and $P < 0.008$ for East Asians).
1087 AA: arachidonic acid; DGLA: dihomo- γ -linolenic acid; DHA: docosahexaenoic acid; DPA: docosapentaenoic
1088 acid; LA: linoleic acid; PA: palmitic acid; POA: palmitoleic acid; SNP: single nucleotide polymorphism;
1089 *FADS1/2*: fatty acids desaturase 1/2; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase.

1090

1091 **Figure 4.** Mendelian randomization results for cardiovascular risk factors related to
1092 increasing activity of enzymes coded by *FADS1/2* (D5D/D6D), *ELOVL2* (*ELOVL2*) and
1093 *SCD* (*SCD*) among individuals of European and East Asian ancestries.

1094 Results are expressed as change in standard units or log odds ratio of cardiovascular disease risk factors per
1095 standard unit increase in the marker of enzyme activity for *FADS1/2* (i.e. AA:DGLA ratio in Europeans and
1096 DGLA:LA ratio in East Asians), *ELOVL2* (i.e. DHA:DPA ratio in Europeans) and *SCD* (i.e. POA:PA ratio in
1097 Europeans) loci. For individuals of European ancestry, data was extracted from multiple consortia of genetic
1098 association studies. For individuals of East Asian ancestry, data was extracted from BioBank Japan. Full
1099 symbols indicate associations at P-value lower than the P-value threshold accounting for multiple testing ($P <$
1100 0.00625). Smoking is represent by pack years of smoking and number of cigarettes per day in European and
1101 East Asian ancestry individuals, respectively. *FADS1/2*: fatty acids desaturase 1/2; *ELOVL2*: elongase 2; *SCD*:
1102 stearoyl-CoA desaturase; LDL-cholesterol: low-density lipoprotein-cholesterol.

1103

1104 **Figure 5.** Phenome wide association scan of *FADS1/2* (rs174546), *ELOVL2* (rs3734398) and
1105 *SCD* (rs603424) genetic variants in European and East Asian ancestry individuals.

1106 Results are expressed as the Z-statistic for the variant-trait association for the allele increasing enzyme
1107 expression/activity. Red circles denote P-value $< 1.5 \times 10^{-6}$ in Europeans P-value $< 4.5 \times 10^{-4}$ in East Asians
1108 . *FADS*: fatty acids desaturase; *ELOVL2*: elongase 2; *SCD*: stearoyl-CoA desaturase.

1109

1110 **Figure 6.** Association of *FADS1/2* (rs174546), *ELOVL2* (rs3734398) and *SCD* (rs603424)
1111 genetic variants with expression of cis-genes across multiple tissues among Europeans.

1112 Results are expressed as z-statistics (effect estimate / standard error) for the SNP-gene expression association
1113 per allele increasing enzyme activity. Only cis-genes (i.e. SNP within 1Mb up- or down-stream of the gene's
1114 transcription start site) were included in the analysis. Genes are labelled in the plot if P-value $< 5\%$ FDR per
1115 tissue.

1116

1117 **Figure 7.** Association of *FADS1/2* (rs174546), *ELOVL2* (rs3734398) and *SCD* (rs603424)
1118 genetic variants with cardiovascular risk factors among unrelated individuals and within
1119 siblings of European ancestry.

1120 Results are expressed as change in standard deviation (SD) units (or risk difference), and 95% confidence
1121 intervals (95%CI), of cardiovascular risk factors per allele increasing enzyme activity.

Figure 1

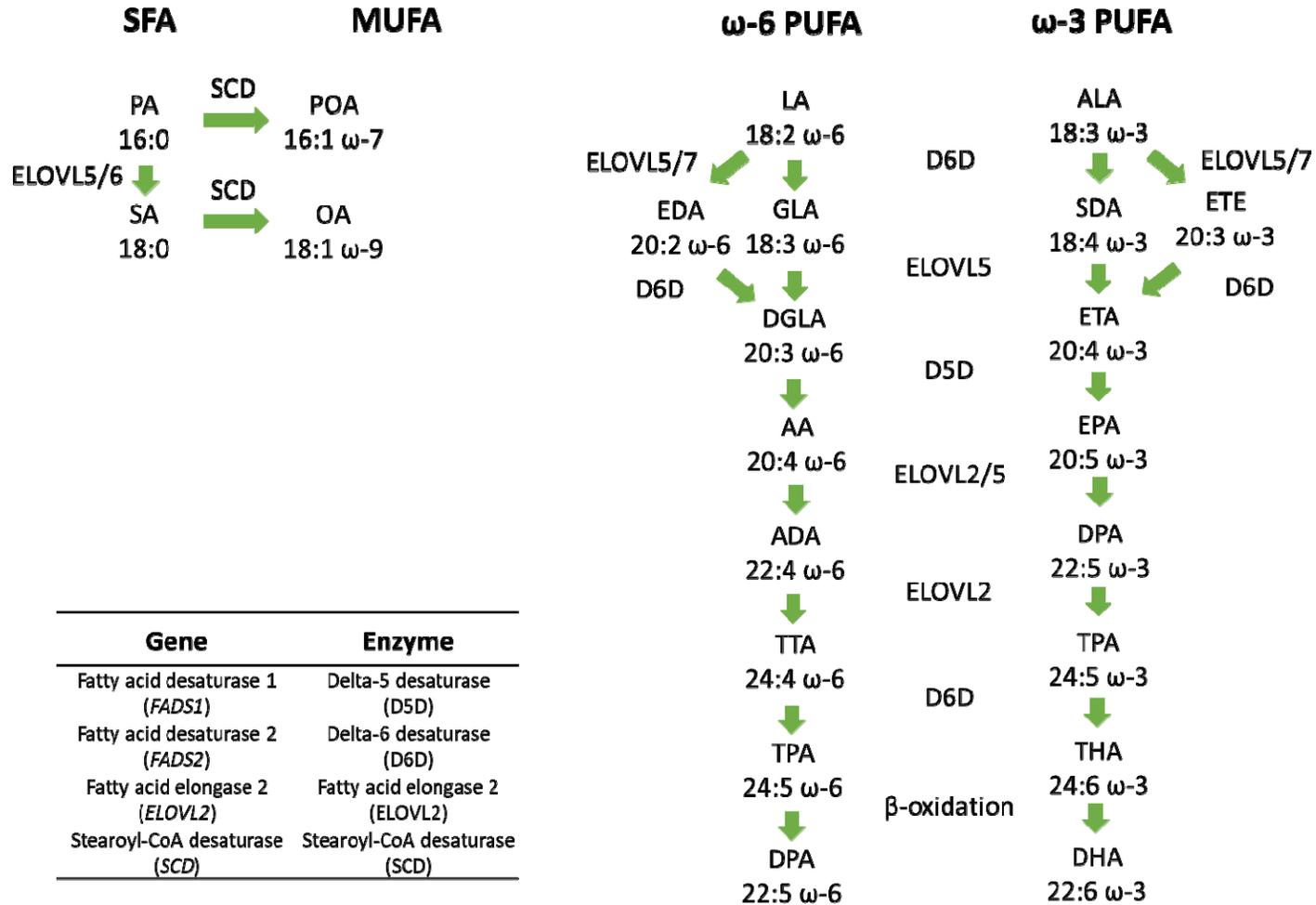
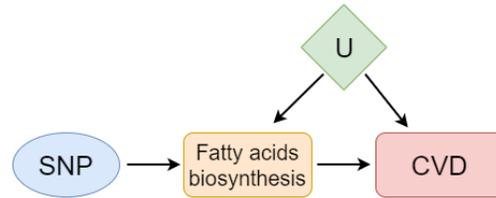


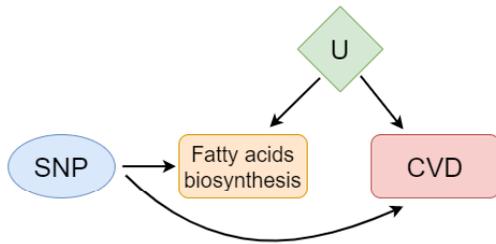
Figure 2

A. Vertical pleiotropy



True positive

B. Horizontal pleiotropy

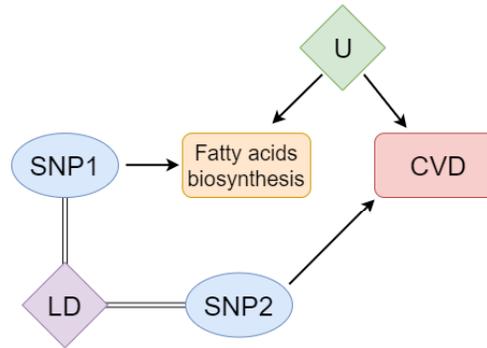


False positive

Bias assessment

- MVMR of gene expression
- Phenome scan

C. Confounding by linkage disequilibrium

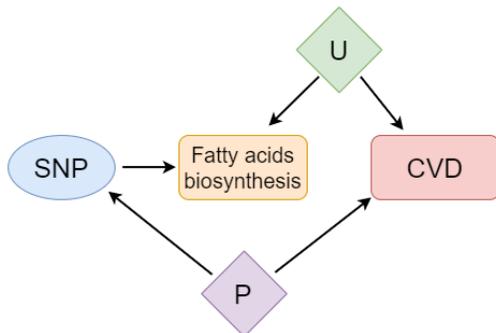


False positive

Bias assessment

- Genetic colocalization

D. Confounding by population stratification, assortative mating or indirect genetic effects

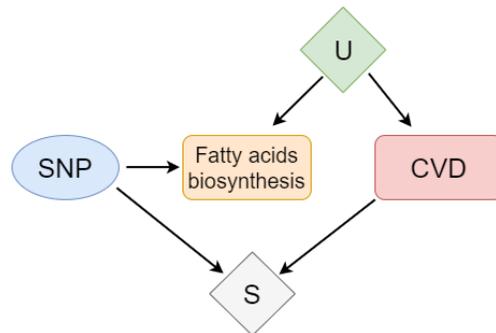


False positive

Bias assessment

- Within sibship analyses
- Negative outcome control

E. Selection bias



False positive

Bias assessment

- Positive exposure control

Figure 3

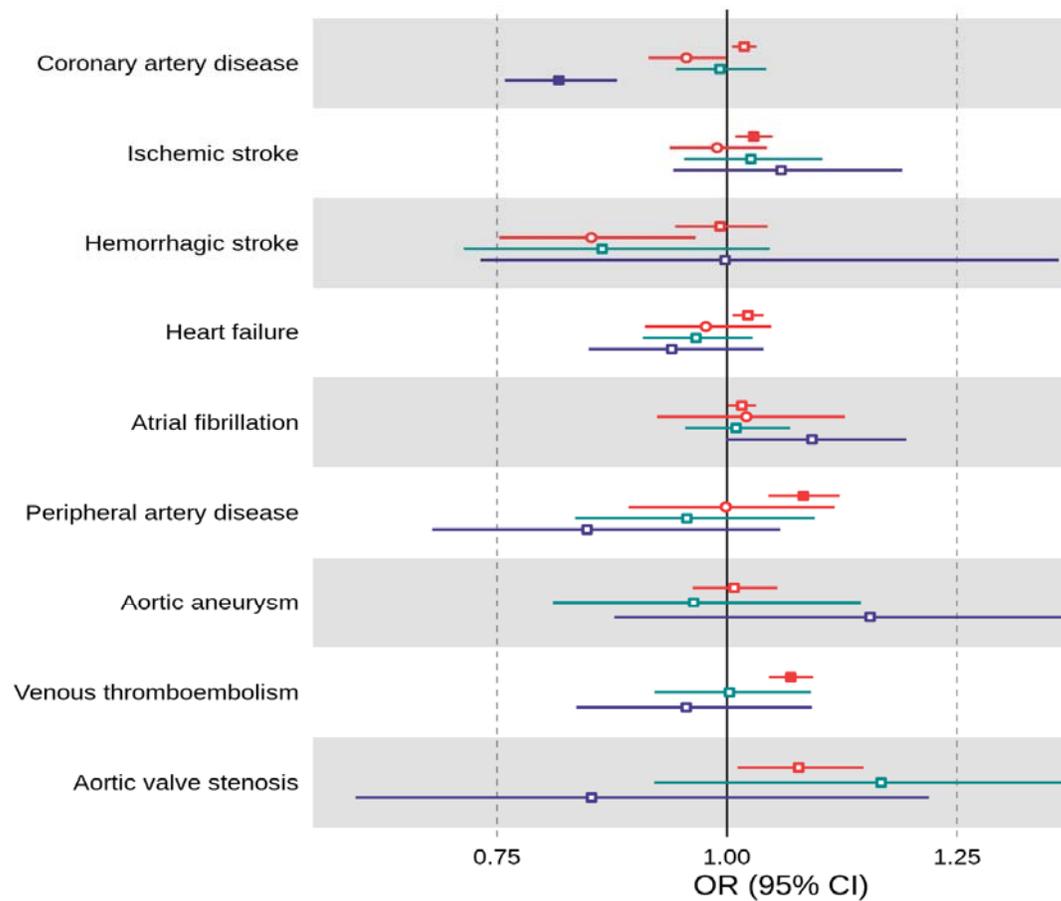


Figure 4

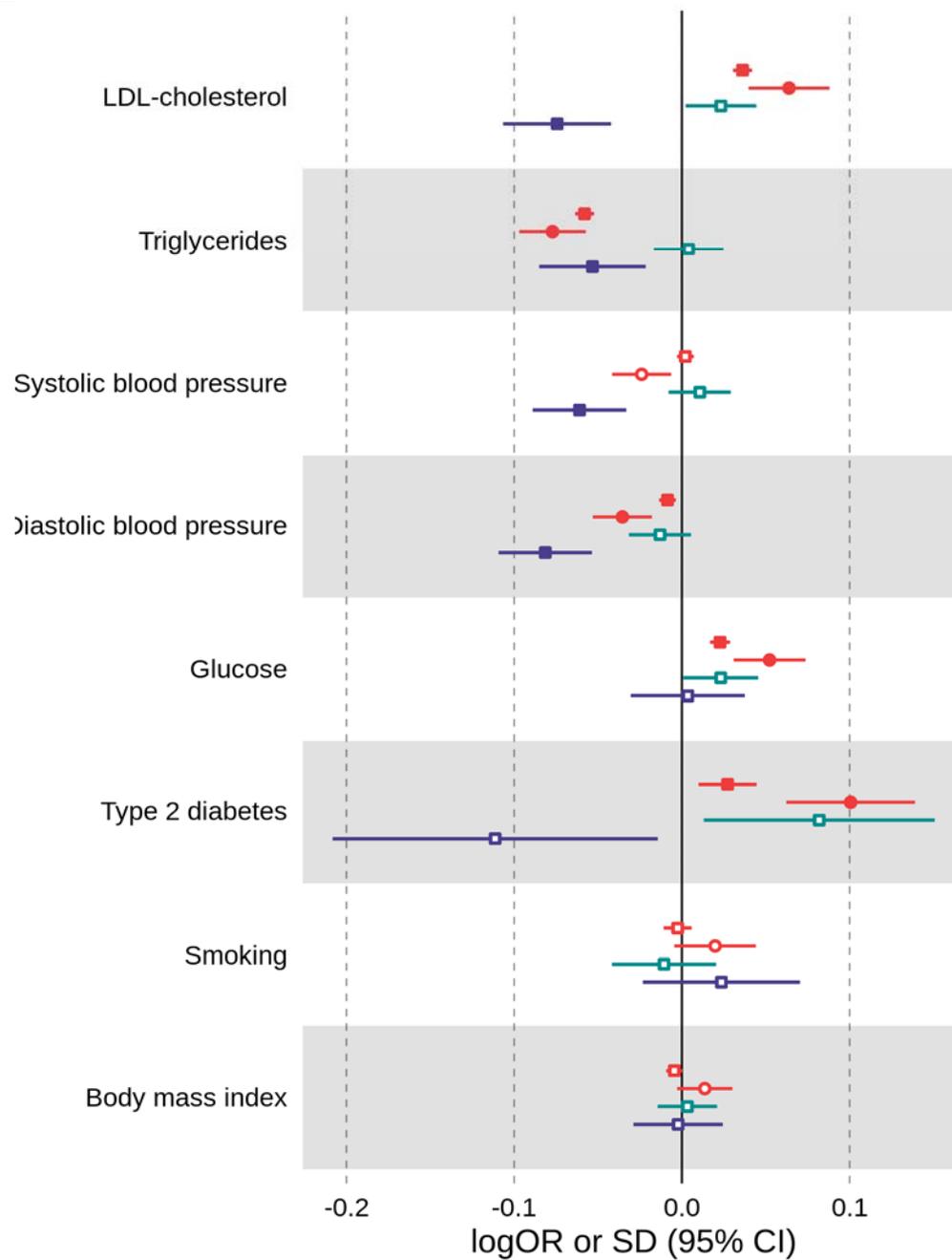
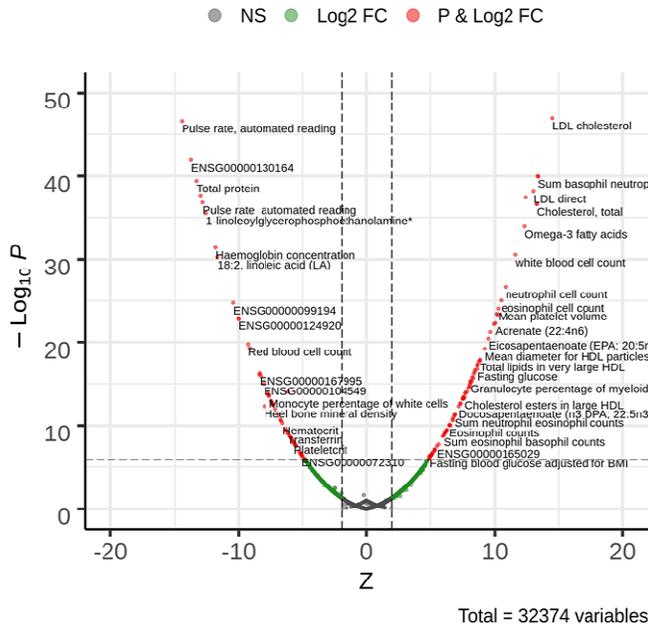
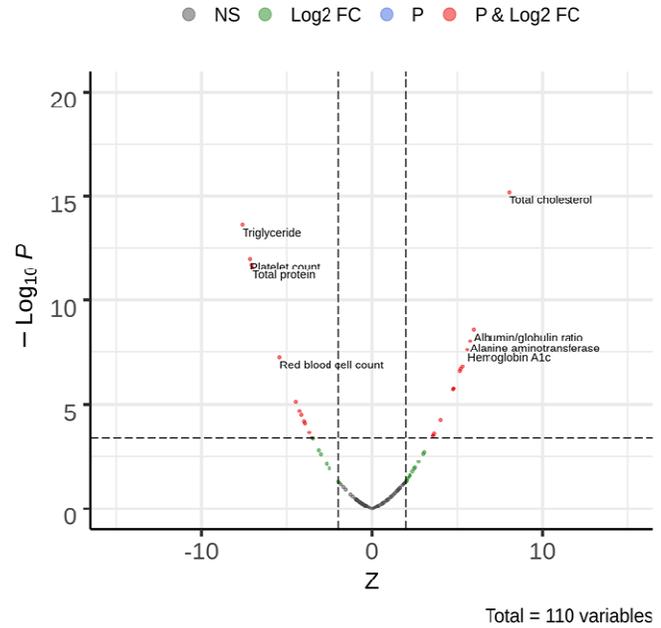


Figure 5

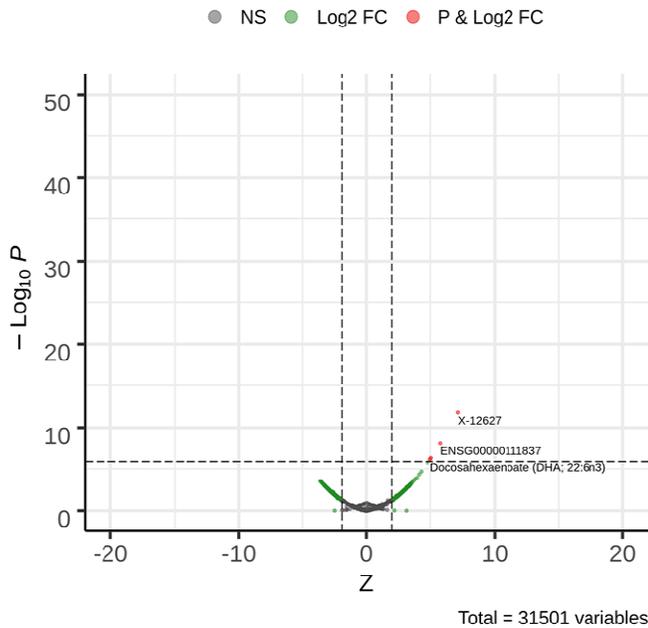
FADS1/2: Europeans



FADS1/2: East Asians



ELOVL2: Europeans



SCD: Europeans

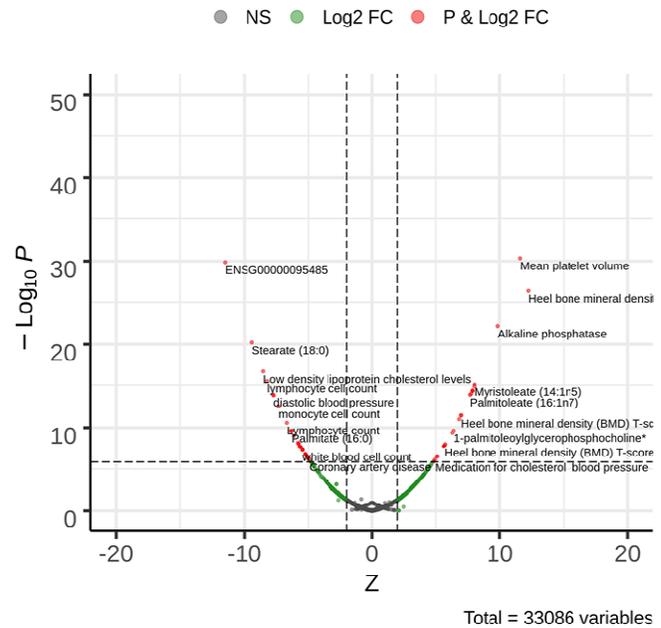


Figure 6

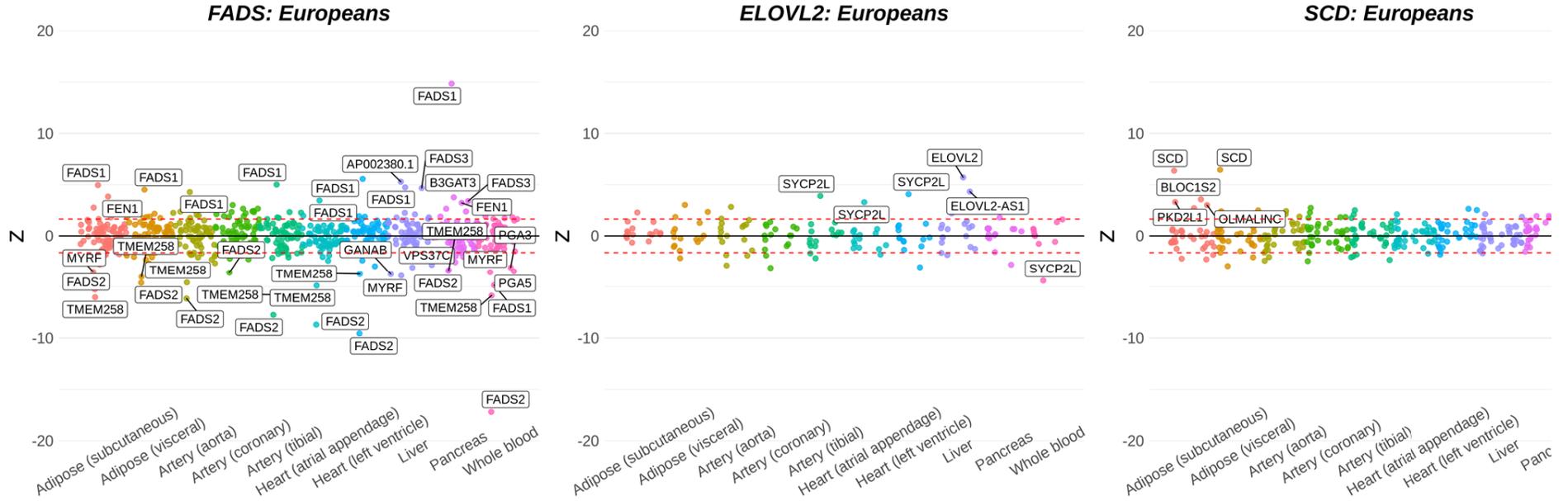


Figure 7

