

30 9. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and
31 Primary Care, University of Cambridge, Cambridge, U.K.

32 10. Medical Research Council Population Health Research Unit, University of Oxford,
33 Oxford, UK

34 11. Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of
35 Population Health, University of Oxford, Oxford, UK

36

37

38 **# Correspondence to:**

39 Tom G Richardson | MRC Integrative Epidemiology Unit (IEU), Population Health
40 Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

41 Tom.G.Richardson@bristol.ac.uk

42

43 George Davey Smith | MRC Integrative Epidemiology Unit (IEU), Population Health
44 Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

45 KZ.Davey-Smith@bristol.ac.uk

46

47

48 **Abstract**

49

50 **Background:** Circulating blood lipids cause coronary heart disease (CHD). However, the
51 precise way in which one or more lipoprotein lipid-related entities account for this
52 relationship remains unclear. We sought to explore the causal relationships of blood
53 lipid traits with risk of CHD using multivariable Mendelian randomization.

54

55 **Methods:** We conducted GWAS of circulating blood lipid traits in UK Biobank (up to
56 $n=440,546$) for LDL cholesterol, triglycerides and apolipoprotein B to identify lipid-
57 associated SNPs. Using data from CARDIoGRAMplusC4D for CHD (consisting of 60,801
58 cases and 123,504 controls), we performed univariable and multivariable Mendelian
59 randomization (MR) analyses. Similar analyses were conducted for HDL cholesterol and
60 apolipoprotein A-I.

61

62 **Findings:** GWAS identified multiple independent SNPs associated at $P<5\times 10^{-8}$ for LDL
63 cholesterol (220), apolipoprotein B ($n=255$), triglycerides (440), HDL cholesterol (534)
64 and apolipoprotein AI (440). Between 56-93% of SNPs identified for each lipid trait had
65 not been previously reported in large-scale GWAS. Almost half (46%) of these SNPs were
66 associated at $P<5\times 10^{-8}$ with more than one lipid related trait. Assessed individually using
67 MR, each of LDL cholesterol (OR 1.66 per 1 standard deviation higher trait; 95%CI: 1.49;
68 1.86; $P=2.4\times 10^{-19}$), triglycerides (OR 1.34; 95%CI: 1.25, 1.44; $P=9.1\times 10^{-16}$) and
69 apolipoprotein B (OR 1.73; 95%CI: 1.56, 1.91; $P=1.5\times 10^{-25}$) had effect estimates

70 consistent with a higher risk of CHD. In multivariable MR, only apolipoprotein B (OR
71 1.92; 95%CI: 1.31, 2.81; $P=7.5 \times 10^{-4}$) retained a robust effect with the estimate for LDL
72 cholesterol (OR 0.85; 95%CI: 0.57; 1.27; $P=0.44$) reversing and that of triglycerides (OR
73 1.12; 95%CI: 1.02, 1.23; $P=0.01$) becoming markedly weaker.
74 Individual MR analyses showed a 1-SD higher HDL-C (OR 0.80; 95%CI: 0.75, 0.86;
75 $P=1.7 \times 10^{-10}$) and apolipoprotein A-I (OR 0.83; 95%CI: 0.77, 0.89; $P=1.0 \times 10^{-6}$) to lower the
76 risk of CHD but these effect estimates weakened to include the null on accounting for
77 apolipoprotein B.

78

79 **Conclusions:** Apolipoprotein B is of fundamental causal relevance in the aetiology of
80 CHD, and underlies the relationship of LDL cholesterol and triglycerides with CHD.

81

82 Introduction

83

84 There is incontrovertible evidence that lipids play a causal role in the aetiology of
85 coronary heart disease (CHD) [1-3]. Multiple large-scale randomized trials of lipid
86 modifying therapies have conclusively shown that lowering of cholesterol in atherogenic
87 lipoproteins leads to a reduction in risk of CHD[4, 5]. These findings have been
88 recapitulated in human genetic studies using genetic variants robustly associated with
89 low-density lipoprotein (LDL) cholesterol [6-9].

90

91 Each circulating atherogenic lipoprotein particle includes one apolipoprotein B particle
92 but the amount of cholesterol (especially in LDL particles) and the amount of
93 triglycerides (especially in very low-density lipoprotein particles) can vary extensively
94 between lipoprotein particles[10]. Thus, while the concentration of LDL cholesterol and
95 triglycerides quantifies the concentration of these lipid substances in the blood, they do
96 not precisely quantify the number of atherogenic lipoproteins; in contrast, the
97 concentration of apolipoprotein B particles is directly proportional to the number of
98 circulating atherogenic particles in the blood. Evidence from human genetics supports a
99 causal role of LDL cholesterol, triglycerides and apolipoprotein B in CHD[11-13]. While it
100 is plausible that each lipid-related entity does individually play a causal role, it is also
101 feasible that one trait predominates. Elucidating the comparative role of blood lipids in
102 the aetiology of CHD has important repercussions not only in terms of a clearer
103 understanding of the underlying pathophysiology, but also in terms of which

104 biomarker(s) should be the focus of lipid-modifying therapeutics and might have more
105 application in the clinical setting.

106

107 Disentangling the relationships of atherogenic lipoprotein lipids and risk of CHD is non-
108 trivial, given the correlated nature of these traits. One such approach is to take genetic
109 variants that associate with more than one lipid trait and scale the CHD associations for
110 a given difference in lipid, in an attempt to identify which one or more traits appears to
111 have a consistent effect on risk of CHD. However, such an approach makes use of
112 indirect extrapolations and might be liable to biases, such as differential measurement
113 error in the lipid traits. Another approach is to use Mendelian randomization (MR), a
114 genetic approach that can facilitate an assessment of causality under certain
115 assumptions.[14] Conventionally, MR involves the analysis of individual exposure to
116 outcome relationships. A recently-developed extension to MR, so-called multivariable
117 MR, permits the appraisal of multiple risk factors simultaneously. By including the
118 genetic associations for multiple exposures in the same model, multivariable MR can
119 elucidate which traits retain a causal relationship with an outcome of interest, through
120 the genetic interrogation of potential confounders[15].

121

122 In this study, we sought to use human genetics to disentangle which one or more of the
123 atherogenic lipid-related traits (apolipoprotein B, LDL cholesterol and triglycerides) is
124 the underlying causal risk factor for CHD. We first conducted *de novo* GWAS of lipid-
125 related traits using the UK Biobank to identify variants robustly associated with each

126 trait. We then conducted MR analyses, including multivariable MR, to elucidate which of
127 the atherogenic lipid traits cause CHD. Finally, we investigated whether the entity
128 underlying the causal role of atherogenic lipid-related traits in CHD also accounted for
129 the ‘cardioprotective’ association of HDL related phenotypes.

130

131 **Methods**

132

133 *Data sources*

134 We used data from the UK Biobank (UKBB) under application #15825 and summary
135 estimates from CARDIoGRAMplusC4D [16]. All individual participant data used in this
136 study were obtained from the UK Biobank study who have obtained ethics approval
137 from the Research Ethics Committee (REC - approval number: 11/NW/0382). All
138 participants enrolled in UK Biobank have signed consent forms.

139

140 *Data handling*

141 Lipid-related traits in the UK Biobank were standardized/normalized using inverse rank-
142 normalization such that the mean was 0 and standard deviation was 1, allowing
143 comparison of effect estimates between traits.

144

145 *GWAS of lipid-related traits*

146 We identified SNPs associated with each of the lipid-related traits using the BOLT-LMM
147 software[17]. Analyses were adjusted for age, sex and a binary variable denoting the
148 genotyping chip individuals were allocated to in UKBB. Population stratification and
149 relatedness was accounted for using a mixed model after excluding individuals of non-
150 European descent based on k-means clustering (k=4). Further details on genotyping
151 quality control, phasing, imputation and association testing have been reported
152 previously[18, 19]. We assigned a SNP as associated with a lipid-related trait of interest

153 through use of a conventional GWAS threshold ($P < 5 \times 10^{-8}$) and SNPs were binned into
154 loci based on pair-wise LD ($r^2 < 0.001$), with the SNP with the strongest association with
155 the trait of interest (as defined by P value) being retained in each locus. We defined
156 novel SNPs as those associated with the trait of interest at $P < 5 \times 10^{-8}$ in our analyses
157 where an association had not been previously reported at $P < 5 \times 10^{-8}$, within 1MB and at
158 $r^2 < 0.001$, by the Global Lipids Genetics Consortium[20] (for LDL cholesterol,
159 triglycerides and HDL cholesterol) or by Kettunen et al[21] (for apolipoprotein B or
160 apolipoprotein A-I).

161

162 *Synthesis and characterization of genetic instruments*

163 SNPs associating with lipid related traits at conventional GWAS thresholds ($P < 5 \times 10^{-8}$)
164 were taken forward to generate genetic instruments for each phenotype. We
165 characterized the genetic instruments in several ways. First, to characterize the
166 ‘specificity’ of individual SNPs included in each genetic instrument, we elucidated how
167 many SNPs associated with traits other than the primary lipid trait of interest at
168 conventional GWAS thresholds of significance ($P < 5 \times 10^{-8}$) and used this information to
169 generate a Venn diagram. Second, we characterize instrument ‘specificity’ by taking per-
170 allele SNP estimates from our GWAS for each lipid trait and conducting inverse variance
171 weighted regressions on these summary estimates to elucidate the association of
172 genetic instruments across the various lipid related traits – these estimates are
173 presented as standardized differences per 1-standard deviation (SD) higher genetically-
174 predicted trait. While we recognize that this approach may be prone to inflation, the

175 primary motivation is to characterize the associations of lipid instruments with the lipid-
176 related traits: we do not interpret these as formal instrumental variable estimates.

177

178 *Genetic analyses to elucidate causality*

179 We first conducted univariable Mendelian randomization (MR) analyses for each lipid-
180 related trait. For this, we harmonized SNPs identified from our GWASs of blood lipids in
181 UK Biobank to those SNPs available in CARDIoGRAMplusC4D by either matching the SNP
182 directly, or by selecting proxy SNPs in high linkage disequilibrium ($r^2 > 0.8$). This led to a
183 small drop in the number of SNPs being available for MR, with a median of 93% SNPs
184 identified in GWAS available for MR (the numbers available for each trait are provided in
185 **Table 1**). We used the inverse variance weighted approach which, in brief takes the
186 form of a linear regression of the SNP-outcome association regressed on the SNP-
187 exposure association weighted by the inverse of the square of the standard error of the
188 SNP-outcome association, with the intercept constrained at the origin. We additionally
189 conducted univariable MR analyses using weighted median [22], weighted mode [23]
190 and MR-Egger [24] approaches.

191

192 We next conducted multivariable MR, which is a statistical approach that allows for the
193 association of SNPs with multiple phenotypes to be incorporated into the analysis
194 permitting an estimation of the direct effect of each phenotype on the outcome (i.e. an
195 effect which is not mediated by any other factor in the model [15]) see **Supplementary**
196 **Figure 1** for further details. In this manuscript, we use the term 'adjusted' in the context

197 of multivariable MR to mean 'direct' effects, i.e. the effect of a lipid trait on CHD while
198 accounting for either mediation or confounding by another trait included in the model.
199 For the multivariable MR analyses, we fitted a model with apolipoprotein B, LDL
200 cholesterol and triglycerides to identify which one or more traits was responsible for the
201 effect of "atherogenic" lipid-related traits on risk of CHD. We then took the atherogenic
202 trait(s) that retained an effect on CHD in the multivariable MR model forward and
203 further adjusted for apolipoprotein A-I and HDL cholesterol to assess the causal effect of
204 HDL-related phenotypes with risk of CHD. In the setting of multivariable MR, we
205 included all GWAS associated SNPs for all traits in the model. This meant that there
206 were differing numbers of SNPs in the two multivariable models tested.

207

208 We characterized instrument strengths in both the univariable and multivariable MR
209 settings, as follows: for the univariable estimates, we generated the mean F-statistic,
210 using the approximation described by Bowden et al[25]. For the multivariable estimate,
211 we generated the conditional F-statistic [26].

212

213 *Software*

214 The BOLT-LMM software was used to undertake GWAS [17]. MR analyses were
215 conducted using the 'TwoSampleMR' R package[27]. Manhattan plots were generated
216 using the 'ggplot2' package[28].The Venn diagram was generated using the online tool
217 available at <http://bioinformatics.psb.ugent.be/webtools/Venn> [accessed 13th August
218 2019].

219

220 *Analysis*

221 While we desisted from dichotomizing results of analyses purely on the basis of a P-
222 value into being ‘significant’ or not[29], as a means of grading the strength of evidence
223 against the null hypothesis, in both the univariable and multivariable Mendelian
224 randomization analyses, we used a two-sided alpha of 0.01, on the basis of testing 5
225 lipid-related traits. Such a Bonferroni adjustment to account for multiple testing can be
226 considered overly conservative given the high correlation between the lipid related
227 traits.

228

229 Results

230

231 *GWAS of blood lipid traits*

232 The lipid traits were measured in 393,193 to 440,546 individuals with GWAS genotyping
233 (**Table 1**). On GWAS, we identified a large number of independent SNPs associated at
234 $P < 5 \times 10^{-8}$ with each lipid related trait: 220 SNPs (of which 56% had not been previously
235 reported) associated with LDL cholesterol, 440 (77% novel) for triglycerides, 255 (80%
236 novel) for apolipoprotein B, 534 (72% novel) for HDL cholesterol and 440 (93% novel) for
237 apolipoprotein A-I (**Figure 1** and **Table 1**). Full details of the SNPs associated with the
238 lipid-related traits are provided in **Supplementary Tables 1-5**.

239

240 A considerable number (477 out of a total 1044 clumped SNPs, i.e. 46%) of SNPs used in
241 each of the lipid-related genetic instruments showed associations at conventional GWAS
242 significance ($P < 5 \times 10^{-8}$) with other lipid traits (**Figure 2A**). On exploring the relationships
243 of the genetic instruments with each lipid-related trait, we identified widespread
244 associations (**Figure 2B**). For example, in addition to its association with apolipoprotein
245 B, the genetic instrument for apolipoprotein B showed strong positive associations with
246 LDL cholesterol and triglycerides and inverse associations with HDL cholesterol and
247 apolipoprotein A-I.

248

249 *Appraisal of LDL cholesterol, triglycerides and apolipoprotein B*

250 On individual assessment through conventional MR, we identified LDL cholesterol,
251 triglycerides and apolipoprotein B to have effect estimates consistent with a higher risk
252 of CHD, using data from CARDIoGRAMplusC4D (with up to 60,801 cases) (**Figure 3A**). A
253 1-SD higher LDL cholesterol had an OR of 1.66 (95%CI: 1.49; 1.86; $P=2.4 \times 10^{-19}$) for CHD
254 with the corresponding value for triglycerides being (OR 1.34; 95%CI: 1.25, 1.44;
255 $P=9.1 \times 10^{-16}$) and apolipoprotein B (OR 1.73; 95%CI: 1.56, 1.91; $P=1.5 \times 10^{-25}$). Sensitivity
256 analyses using methodological approaches that take into account potential genetic
257 pleiotropy led to no substantive change in these estimates (**Supplementary Figure 2**).

258

259 When LDL cholesterol, triglycerides and apolipoprotein B were assessed together in
260 multivariable MR, only apolipoprotein B (OR 1.92; 95%CI: 1.31, 2.81; $P=7.5 \times 10^{-4}$)
261 retained a robust causal relationship with CHD (**Figure 3A**). The point estimate for LDL
262 cholesterol reversed on mutual adjustment, to yield an adjusted OR of 0.85 (95%CI:
263 0.57; 1.27; $P=0.44$). The estimate for triglycerides was weakened substantially (adjusted
264 OR 1.12; 95%CI: 1.02, 1.23; $P=0.01$).

265

266 *Appraisal of HDL cholesterol and apolipoprotein A-I*

267 Individual analysis using conventional MR showed both HDL cholesterol and
268 apolipoprotein A-I to have effect estimates consistent with a lower risk of CHD (**Figure**
269 **3B**). The estimate for a 1-SD higher HDL cholesterol was OR 0.80 (95%CI: 0.75, 0.86;
270 $P=1.7 \times 10^{-10}$) and for apolipoprotein A-I it was OR 0.83 (95%CI: 0.77, 0.89; $P=9.96 \times 10^{-7}$).
271 The effect estimates for both HDL cholesterol and apolipoprotein A-I were diminished

272 when using methodologies that are more robust to potential pleiotropy of the genetic
273 variants used in the analysis (**Supplementary Figure 2**).

274 On appraisal in a multivariable MR analysis that included apolipoprotein B (which
275 retained a causal relationship with CHD on mutual genetic adjustment for LDL
276 cholesterol and triglycerides), the effect estimates of both HDL cholesterol and
277 apolipoprotein A-I diminished and were not distinguishable from the null. (**Figure 3B**).
278 The adjusted estimate for HDL-C was OR 0.91 (95%CI: 0.74, 1.12; P=0.36) and for
279 apolipoprotein A-I it was OR 0.94 (95%CI: 0.76, 1.17; P=0.59). When adjusted for HDL
280 cholesterol and apolipoprotein A-I, apolipoprotein B retained a robust causal effect on
281 CHD (adjusted OR 1.68; 95%CI: 1.54, 1.84; P=6.5x10⁻³²).

282

283 The F-statistics for all lipid-related genetic instruments in both the univariable and
284 multivariable MR settings were consistent with weak instruments being an unlikely
285 source of bias (**Table 1**).

286

287

288 **Discussion**

289

290 Our study provides strong evidence from human genetics that supports apolipoprotein
291 B being the underlying causal driver of the relationship of blood lipids and risk of CHD.

292 This adds further evidence to support the hypothesis that it is the number of
293 atherogenic lipoprotein particles, indexed by apolipoprotein B, rather than the amount
294 of circulating cholesterol or triglycerides, that is the important driver of CHD[30]. In
295 other words, changes in cholesterol or triglycerides that are not accompanied by
296 commensurate changes in apolipoprotein B are unlikely to lead to altered risks of CHD.

297

298

299 Our GWAS identified many hundreds of variants associated with the major lipid related
300 traits, with most SNPs identified being novel. Many SNPs identified for one lipid-related
301 trait also showed associations with other lipid traits, highlighting their pleiotropic
302 nature. Individual appraisal using univariable Mendelian randomization showed
303 widespread effects of all lipid-related traits, with LDL cholesterol, triglycerides and
304 apolipoprotein B each having effect estimates consistent with a higher risk of CHD.

305 These findings recapitulate those reported in previous studies [8, 9, 11, 12], leading to
306 the contemporary view that each atherogenic lipid trait might play a causal role in
307 vascular disease. When we estimated the direct (i.e. adjusted) effect of these traits
308 using multivariable MR (see Supplementary Figure 1 for further details), only
309 apolipoprotein B retained a robust causal effect with CHD, with the effect of LDL

310 cholesterol being reversed and that for triglycerides being largely diminished with only a
311 very weak residual effect. The apparent protective associations of HDL cholesterol and
312 apolipoprotein A-1, present on univariable MR analyses, were also markedly attenuated
313 when direct effects conditional on apolipoprotein B were estimated. Taken together,
314 these findings indicate that among the lipid related traits we investigated, it is
315 apolipoprotein B, and thus the number of atherogenic lipoprotein particles, that
316 predominates as the underlying cause of CHD.

317

318 How do these findings enhance the evidence-base relating to lipid traits and vascular
319 disease? Large-scale observational[31], interventional[4, 5] and genetic[6-9] studies
320 support LDL cholesterol as being causal in the aetiology of CHD. In recent years, genetic
321 studies have provided evidence in support of triglycerides[11, 12] also playing a causal
322 role. Both LDL cholesterol and triglycerides are carried in atherogenic lipoproteins, each
323 containing an apolipoprotein B particle. Recent narrative reviews[32] [33] point to
324 apolipoprotein B potentially being the necessary entity for atherosclerosis to occur, for
325 example, through the ‘response to retention’ hypothesis, in which apolipoprotein-B
326 containing particles become trapped in the tunica intima of the arterial wall[34] . Our
327 study builds on recent findings[30] to provide further empirical evidence that supports
328 this hypothesis, but our findings importantly do not discredit the causal roles that LDL
329 cholesterol or triglycerides play in vascular disease. This is because apolipoprotein B
330 does not occur in physiological circumstances in isolation[33], but rather is always
331 accompanied by cholesterol and triglycerides. In light of this, our findings pinpoint that

332 it is apolipoprotein B that is necessary in order for atherogenesis to occur. Indeed, our
333 findings from multivariable MR are consistent with apolipoprotein B being an essential
334 element allowing the atherogenic effects of LDL cholesterol and triglyceride to be
335 expressed.

336

337 How do these findings aid us in the context of developing drugs that modify blood lipid
338 concentrations and predicting their effects on risk of CHD? Drug-target Mendelian
339 randomization studies show that, for example, modifying triglycerides through therapies
340 such as ANGPTL3/4 inhibition may represent an emerging approach to lowering the risk
341 of CHD[35-37] – do our findings contradict these data? Not so: our findings shed light on
342 whether the concentrations of cholesterol and/or triglycerides that are carried by
343 apolipoprotein B containing lipoproteins plays a role in CHD *beyond* that of
344 apolipoprotein B. Based on these and recent data [13, 30], the primary focus of lipid
345 modifying therapies ought to be the reduction in number of atherogenic lipoproteins (as
346 measured by apolipoprotein B) rather than the reduction in cholesterol or triglycerides.
347 This is especially the case where drugs have discrepant effects across these lipid
348 traits.[10] [38] [13] Thus in predicting the vascular effects of a lipid-modifying
349 therapeutic, apolipoprotein B can, all things being equal, be used as a reliable surrogate
350 marker for the relative risk reduction in CHD – assuming, of course, that the drug under
351 investigation does not display adverse events that arise either from target-mediated
352 mechanisms, or off-target effects (notably, both can be measured directly, or
353 extrapolated, from studies in human genetics[39]).

354

355 We note that this interpretation is in keeping with two important prior investigations
356 that examined the concordance of CHD associations between SNPs associated with
357 apolipoprotein B, LDL cholesterol[13] and triglycerides[30]. Indeed, one of these prior
358 investigations conducted a form of multivariable MR analysis and obtained similar
359 findings to those we report in the present study[30]. Importantly, the analysis that we
360 conducted and report herein builds on these prior investigations by including the full
361 repertoire of GWAS-associated SNPs for each of the lipid related traits (including a de
362 novo GWAS of apolipoprotein B): such a comprehensive representation of trait-
363 associated SNPs is necessary in order to reliably interpret the MR estimates for each of
364 the entities included in the multivariable MR analysis.

365

366 The findings that we make have been made available by two recent advances. First, the
367 availability of large-scale blood lipid phenotyping and GWAS genotyping in the UK
368 Biobank, providing sufficiently large numbers to permit identification of robust genetic
369 variants (and therefore suitable genetic instruments) in order to conduct MR of each of
370 the lipid-related traits. Use of a single study with similar numbers of individuals with
371 measures available for each lipid-related trait enabled GWAS and the downstream
372 synthesis of genetic instruments for each trait in which the genetic architecture of each
373 phenotype ought to be similarly represented, allowing for a more rigorous comparative
374 assessment of the traits in both the univariable and multivariable MR setting. Second,
375 methodological developments in MR to include more than one trait (so-called

376 multivariable MR) allows for “direct” effects (i.e. the effects of an exposure on disease,
377 taking into account potential confounding and mediation by other traits) of multiple
378 exposures to be assessed simultaneously and without the risk that this introduces forms
379 of bias (such as collider bias)[15]. It is this methodological approach that allows the
380 deduction that we make: that apolipoprotein B underlies the causal effects of lipid-
381 related traits with risk of CHD. We note here an important theme that emerges: the
382 discrepancy between our findings and those derived from other MR approaches that
383 hitherto have been used in contemporary MR studies (reflected by the univariable MR
384 estimates we present in Supplementary Figure 2). While other approaches such as MR-
385 Egger and weighted median MR can provide reliable tests of causation even in the
386 presence of confounding through unbalanced horizontal pleiotropy[40] (as evidenced by
387 the diminution of the HDL cholesterol association with risk of CHD on MR-Egger,
388 Supplementary Figure 2), such approaches notably do not, with a few exceptions [41,
389 42], allow simultaneous statistical adjustment for multiple traits. The repertoire of
390 univariable MR analyses that seek to act as sensitivity analyses for potential pleiotropy
391 each makes a different series of assumptions [22]. In the context that genetic
392 confounding affects the majority of SNPs used in the genetic instruments, and when
393 such confounding is present in a dose-response manner (i.e. on average, SNPs that
394 increase the exposure of interest also increase the confounder of interest), this violates
395 the ‘inSIDE’ assumption[24] and the MR analyses yield biased estimates. This is why the
396 MR estimates for LDL cholesterol and triglycerides remain seemingly robust to MR Egger
397 and weighted median MR approaches. In this context, multivariable MR analysis can

398 help when the traits included in the analysis account fully for the unbalanced, dose-
399 related, horizontal pleiotropy. In the scenario that we investigate, apolipoprotein B does
400 just so, permitting us to conclude that it is apolipoprotein B that is ultimately
401 responsible for the underlying causal relationship of blood lipids and risk of CHD.

402

403 In conclusion, our findings demonstrate that apolipoprotein B underlies the causal effect
404 of lipids on CHD and that it is the trait that is responsible for the associations of LDL
405 cholesterol, triglycerides, HDL cholesterol and apolipoprotein A-I with the risk of CHD.

406

407

408

409 **Acknowledgements**

410 We are immensely grateful to study participants of the UK Biobank.

411

412

413 **Financial Disclosure Statement**

414 TGR, ES and GDS work in the Medical Research Council Integrative Epidemiology Unit at
415 the University of Bristol, which is supported by the Medical Research Council
416 (MC_UU_00011/1 and MC_UU-00011/2). TGR is a UKRI Innovation Research Fellow
417 (MR/S003886/1). MAK is supported by a Senior Research Fellowship from the National
418 Health and Medical Research Council (NHMRC) of Australia (APP1158958) and a
419 research grant from the Sigrid Juselius Foundation, Finland. The Baker Institute is

420 supported in part by the Victorian Government’s Operational Infrastructure Support
421 Program. BAF is supported by the National Institute for Health Research Cambridge
422 Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation
423 Trust. MVH works in a unit that receives funding from the UK Medical Research Council
424 and is supported by a British Heart Foundation Intermediate Clinical Research
425 Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford
426 Biomedical Research Centre. The funders had no role in study design, data collection
427 and analysis, decision to publish, or preparation of the manuscript.

428

429

430 **Competing Interests**

431 BAF reported receiving personal fees from Merck & Co., Amgen, Regeneron, Sanofi,
432 Pfizer, CiVi BioPharma, and KrKA Pharmaceuticals, and grants from Merck & Co., Amgen,
433 Novartis and Esperion Therapeutics. All other authors report no potential conflicts of
434 interest.

435

436 **References**

437

- 438 1. Cholesterol Treatment Trialists C, Fulcher J, O'Connell R, Voysey M, Emberson J,
439 Blackwell L, et al. Efficacy and safety of LDL-lowering therapy among men and women:
440 meta-analysis of individual data from 174,000 participants in 27 randomised trials.
441 Lancet. 2015;385(9976):1397-405. doi: 10.1016/S0140-6736(14)61368-4. PubMed
442 PMID: 25579834.
- 443 2. Cholesterol Treatment Trialists' Collaborators, Baigent C, Blackwell L, Emberson
444 J, Holland LE, Reith C, et al. Efficacy and safety of more intensive lowering of LDL
445 cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials.
446 Lancet. 2010;376(9753):1670-81. doi: 10.1016/S0140-6736(10)61350-5. PubMed PMID:
447 21067804; PubMed Central PMCID: PMC2988224.
- 448 3. Cholesterol Treatment Trialists' Collaborators, Mihaylova B, Emberson J,
449 Blackwell L, Keech A, Simes J, et al. The effects of lowering LDL cholesterol with statin
450 therapy in people at low risk of vascular disease: meta-analysis of individual data from
451 27 randomised trials. Lancet. 2012;380(9841):581-90. doi: 10.1016/S0140-
452 6736(12)60367-5. PubMed PMID: 22607822; PubMed Central PMCID: PMC3437972.
- 453 4. Silverman MG, Ference BA, Im K, Wiviott SD, Giugliano RP, Grundy SM, et al.
454 Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among
455 Different Therapeutic Interventions: A Systematic Review and Meta-analysis. Jama.
456 2016;316(12):1289-97. doi: 10.1001/jama.2016.13985. PubMed PMID: 27673306.

- 457 5. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al.
458 Interpretation of the evidence for the efficacy and safety of statin therapy. *The Lancet*.
459 2016;388(10059):2532-61. doi: 10.1016/S0140-6736(16)31357-5.
- 460 6. Ference BA, Robinson, J. G., Brook, R. D., Catapano AL, Chapman MJ, Neff DR,
461 Voros S, Giugliano RP, et al. Variation in PCSK9 and HMGCR and risk of cardiovascular
462 disease and diabetes. *New England Journal of Medicine*. 2016;375(22):2144-53. doi:
463 10.1056/NEJMoa1604304.
- 464 7. Ference BA, Majeed, F., Penumetcha, R., Flack JM, Brook RD. Effect of Naturally
465 Random Allocation to Lower Low-Density Lipoprotein Cholesterol on the Risk of
466 Coronary Heart Disease Mediated by Polymorphisms in NPC1L1, HMGCR, or Both.
467 *Journal of the American College of Cardiology*. 2015;65(15):1552-61. doi:
468 10.1016/j.jacc.2015.02.020. PubMed PMID: WOS:000352956500009.
- 469 8. Ference BA, Yoo, W., Alesh, I., Mahajan N, Mirowska KK, Mewada A, Kahn J,
470 Afonso L, et al. Effect of Long-Term Exposure to Lower Low-Density Lipoprotein
471 Cholesterol Beginning Early in Life on the Risk of Coronary Heart Disease A Mendelian
472 Randomization Analysis. *Journal of the American College of Cardiology*.
473 2012;60(25):2631-9. doi: 10.1016/j.jacc.2012.09.017. PubMed PMID:
474 WOS:000312527000006.
- 475 9. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al.
476 Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J*.
477 2015;36(9):539-50. doi: 10.1093/eurheartj/ehf571. PubMed PMID: 24474739.

- 478 10. Holmes MV, Ala-Korpela M. What is 'LDL cholesterol'? *Nat Rev Cardiol.*
479 2019;16(4):197-8. doi: 10.1038/s41569-019-0157-6. PubMed PMID: 30700860.
- 480 11. White J, Swerdlow DJ, Preiss D, Fairhurst-Hunter Z, Keating BJ, Asselbergs FW, et
481 al. Association of Lipid Fractions With Risks for Coronary Artery Disease and Diabetes.
482 *JAMA Cardiol.* 2016;1(6):692-9. doi: 10.1001/jamacardio.2016.1884. PubMed PMID:
483 27487401.
- 484 12. Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common
485 variants associated with plasma triglycerides and risk for coronary artery disease. *Nat*
486 *Genet.* 2013;45(11):1345-52. doi: 10.1038/ng.2795. PubMed PMID: 24097064; PubMed
487 Central PMCID: PMC3904346.
- 488 13. Ference BA, Kastelein JJP, Ginsberg HN, Chapman MJ, Nicholls SJ, Ray KK, et al.
489 Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein
490 Levels and Cardiovascular Risk. *Jama.* 2017;318(10):947-56. doi:
491 10.1001/jama.2017.11467. PubMed PMID: 28846118; PubMed Central PMCID:
492 PMCPMC5710502.
- 493 14. Davey Smith G, Ebrahim, S. Mendelian randomization: Prospects, potentials, and
494 limitations. *Int J Epidemiol.* 2004;33(1):30-42. doi: 10.1093/ije/dyh132.
- 495 15. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of
496 multivariable Mendelian randomization in the single sample and two-sample summary
497 data settings. *Int J Epidemiol.* 2019;48(3).
- 498 16. Consortium. tCD. A comprehensive 1,000 Genomes-based genome-wide
499 association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47(10):1121-30.

500 doi: 10.1038/ng.3396. PubMed PMID: 26343387; PubMed Central PMCID:
501 PMCPMC4589895.

502 17. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et
503 al. Efficient Bayesian mixed-model analysis increases association power in large cohorts.
504 Nat Genet. 2015;47(3):284-90. doi: 10.1038/ng.3190. PubMed PMID: 25642633;
505 PubMed Central PMCID: PMCPMC4342297.

506 18. Mitchell R, Elsworth BL, Raistrick CA, Paternoster L, Hemani G, Gaunt TR. MRC
507 IEU UK Biobank GWAS pipeline version 2. University of Bristol Available at:
508 <https://databrisacuk/data/dataset/pnoat8cxo0u52p6ynfaekeigi>.

509 19. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK
510 Biobank resource with deep phenotyping and genomic data. Nature.
511 2018;562(7726):203-9. doi: 10.1038/s41586-018-0579-z. PubMed PMID: 30305743.

512 20. Global Lipids Genetics Consortium, Willer CJ, Schmidt EM, Sengupta S, Peloso
513 GM, Gustafsson S, et al. Discovery and refinement of loci associated with lipid levels.
514 Nat Genet. 2013;45(11):1274-83. doi: 10.1038/ng.2797. PubMed PMID: 24097068;
515 PubMed Central PMCID: PMC3838666.

516 21. Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, et al. Genome-
517 wide study for circulating metabolites identifies 62 loci and reveals novel systemic
518 effects of LPA. Nat Commun. 2016;7:11122. doi: 10.1038/ncomms11122. PubMed
519 PMID: 27005778.

520 22. Bowden J, Davey Smith, G., Haycock, P. C., Burgess, S. Consistent Estimation in
521 Mendelian Randomization with Some Invalid Instruments Using a Weighted Median
522 Estimator. *Genetic Epidemiology*. 2016;40(4):304-14. doi: 10.1002/gepi.21965.

523 23. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data
524 Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*.
525 2017;46(6):1985-98. doi: 10.1093/ije/dyx102. PubMed PMID: 29040600; PubMed
526 Central PMCID: PMC5837715.

527 24. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid
528 instruments: effect estimation and bias detection through Egger regression. *Int J*
529 *Epidemiol*. 2015;44(2):512-25. doi: 10.1093/ije/dyv080. PubMed PMID: 26050253.

530 25. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR.
531 Assessing the suitability of summary data for two-sample Mendelian randomization
532 analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol*. 2016.
533 doi: 10.1093/ije/dyw220. PubMed PMID: 27616674.

534 26. Sanderson E, Spiller W, Bowden J. Testing and Correcting for Weak Instruments
535 in Two-sample Summary Data Multivariable Mendelian Randomisation. In preparation.
536 2019.

537 27. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-
538 Base platform supports systematic causal inference across the human phenome. *Elife*.
539 2018;7. doi: 10.7554/eLife.34408. PubMed PMID: 29846171; PubMed Central PMCID:
540 PMC5976434.

- 541 28. Wickham H. ggplot2 – Elegant Graphics for Data Analysis (2nd Edition). Journal of
542 Statistical Software. 2017;77.
- 543 29. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical
544 significance. Nature. 2019;567(7748):305-7. doi: 10.1038/d41586-019-00857-9. PubMed
545 PMID: 30894741.
- 546 30. Ference BA, Kastelein JJP, Ray KK, Ginsberg HN, Chapman MJ, Packard CJ, et al.
547 Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants
548 With Risk of Coronary Heart Disease. Jama. 2019;321(4):364-73. doi:
549 10.1001/jama.2018.20045. PubMed PMID: 30694319; PubMed Central PMCID:
550 PMCPMC6439767.
- 551 31. Emerging Risk Factors Collaboration, Di Angelantonio E, Sarwar N, Perry P,
552 Kaptoge S, Ray KK, et al. Major lipids, apolipoproteins, and risk of vascular disease. Jama.
553 2009;302(18):1993-2000. doi: 10.1001/jama.2009.1619. PubMed PMID: 19903920;
554 PubMed Central PMCID: PMC3284229.
- 555 32. Boren J, Williams KJ. The central role of arterial retention of cholesterol-rich
556 apolipoprotein-B-containing lipoproteins in the pathogenesis of atherosclerosis: a
557 triumph of simplicity. Curr Opin Lipidol. 2016;27(5):473-83. doi:
558 10.1097/MOL.0000000000000330. PubMed PMID: 27472409.
- 559 33. Ala-Korpela M. The culprit is the carrier, not the loads: cholesterol, triglycerides
560 and apolipoprotein B in atherosclerosis and coronary heart disease. Int J Epidemiol.
561 2019. doi: 10.1093/ije/dyz068. PubMed PMID: 30968109.

- 562 34. Williams KJ, Tabas I. The response-to-retention hypothesis of early
563 atherogenesis. *Arterioscler Thromb Vasc Biol.* 1995;15(5):551-61. PubMed PMID:
564 7749869; PubMed Central PMCID: PMCPMC2924812.
- 565 35. Dewey FE, Gusarova V, O'Dushlaine C, Gottesman O, Trejos J, Hunt C, et al.
566 Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J Med.*
567 2016;374(12):1123-33. doi: 10.1056/NEJMoa1510926. PubMed PMID: 26933753;
568 PubMed Central PMCID: PMCPMC4900689.
- 569 36. Graham MJ, Lee RG, Brandt TA, Tai LJ, Fu W, Peralta R, et al. Cardiovascular and
570 Metabolic Effects of ANGPTL3 Antisense Oligonucleotides. *N Engl J Med.*
571 2017;377(3):222-32. doi: 10.1056/NEJMoa1701329. PubMed PMID: 28538111.
- 572 37. Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, et al. Exome
573 sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med.*
574 2010;363(23):2220-7. doi: 10.1056/NEJMoa1002926. PubMed PMID: 20942659;
575 PubMed Central PMCID: PMCPMC3008575.
- 576 38. HPS3/TIMI55-REVEAL Collaborative Group, Bowman L, Hopewell JC, Chen F,
577 Wallendszus K, Stevens W, et al. Effects of Anacetrapib in Patients with Atherosclerotic
578 Vascular Disease. *N Engl J Med.* 2017;377(13):1217-27. doi: 10.1056/NEJMoa1706444.
579 PubMed PMID: 28847206.
- 580 39. Holmes MV. Human genetics and drug development. *N Engl J Med.*
581 2019;380:1076-9.

- 582 40. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A
583 framework for the investigation of pleiotropy in two-sample summary data Mendelian
584 randomization. *Stat Med*. 2017. doi: 10.1002/sim.7221. PubMed PMID: 28114746.
- 585 41. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations
586 between risk factors and common diseases inferred from GWAS summary data. *Nat*
587 *Commun*. 2018;9(1):224. doi: 10.1038/s41467-017-02317-2. PubMed PMID: 29335400;
588 PubMed Central PMCID: PMC5768719.
- 589 42. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for
590 multivariable Mendelian randomization to correct for both measured and unmeasured
591 pleiotropy. *arXiv*. 2017;2017arXiv170800272R.
- 592

Table 1. Genetic variants identified for each trait in UK Biobank.

Trait	Number with trait measured in UK Biobank with GWAS genotyping	Number of SNPs identified in GWAS ($P < 5 \times 10^{-8}$)	Number (%) of novel* SNPs	Number (%) of SNPs aligned to CARDIoGRAM-plusC4D	F-statistic overall and conditional
LDL cholesterol	440,546	220	123 (56%)	209 (95%)	164, 34
Triglycerides	441,016	440	339 (77%)	409 (93%)	116, 78
Apolipoprotein B	439,214	255	203 (80%)	234 (92%)	153, 36 [^]
HDL cholesterol	403,943	534	383 (72%)	490 (92%)	124, 67
Apolipoprotein A-I	393,193	440	407 (93%)	407 (93%)	120, 62

* We defined novel SNPs as those associated with the trait of interest at $P < 5 \times 10^{-8}$ where an association had not been previously reported at $P < 5 \times 10^{-8}$, within 1MB and at $r^2 < 0.001$ by the Global Lipids Genetics Consortium[20] (for LDL cholesterol, triglycerides and HDL cholesterol) or by Kettunen et al[21] (for apolipoprotein B or apolipoprotein A-I).

[^] The conditional F-statistic for apolipoprotein B when included in the multivariable MR model with LDL cholesterol and triglycerides was 36, and in the multivariable MR model that included HDL cholesterol and apolipoprotein A-I it was 66.

Figure 1. Manhattan plots showing findings from GWAS of lipid-related traits in UK Biobank

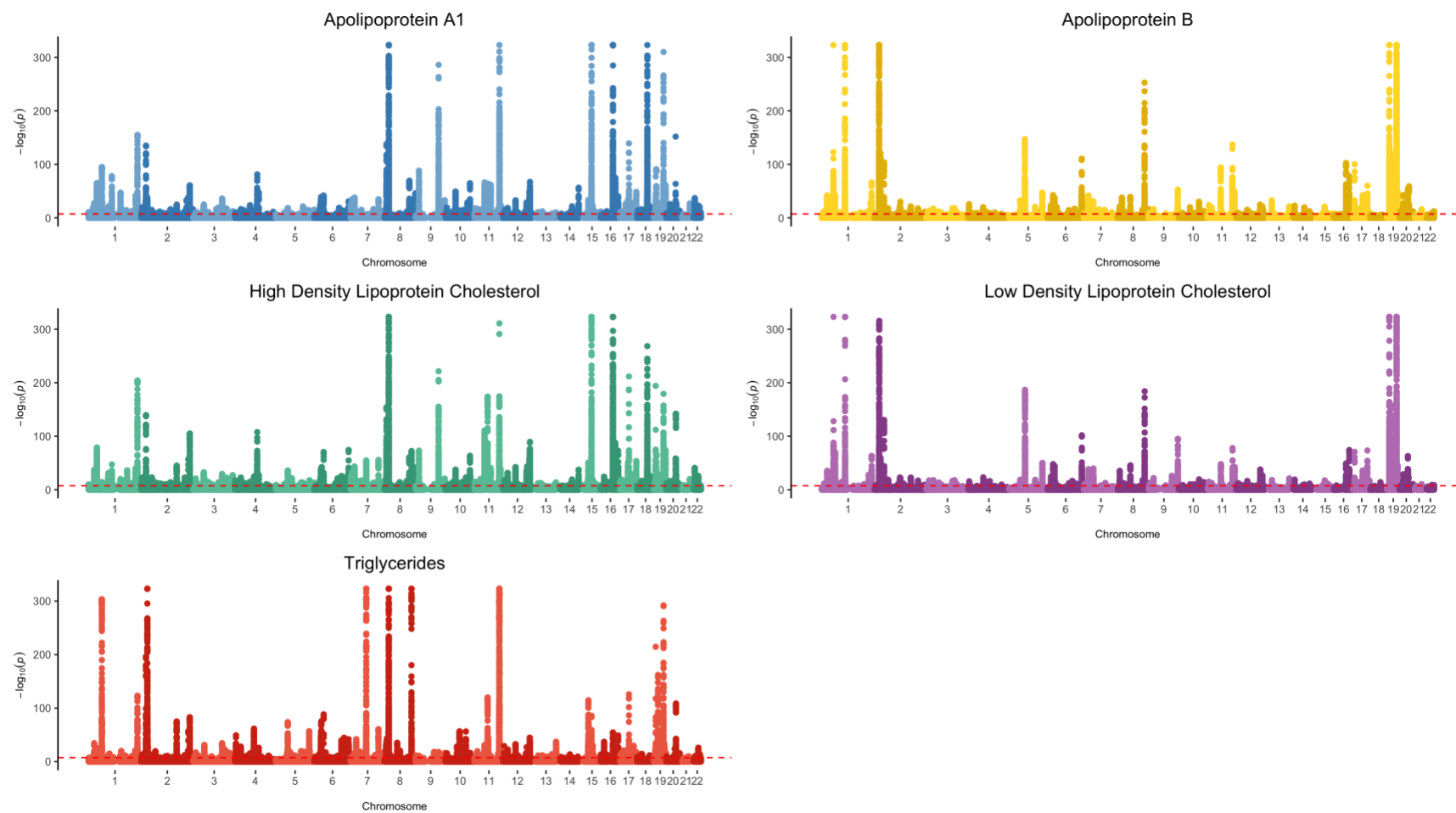
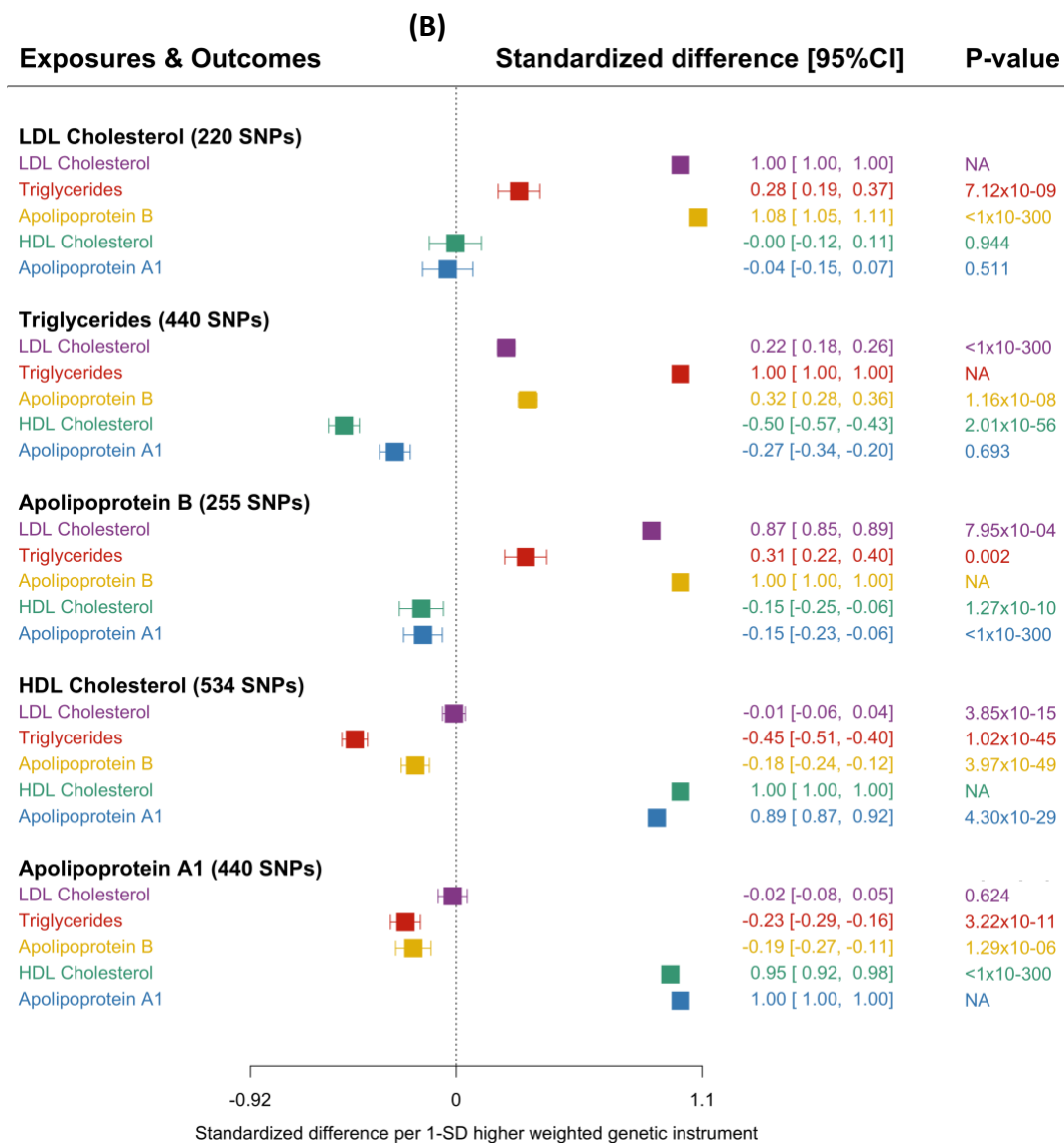
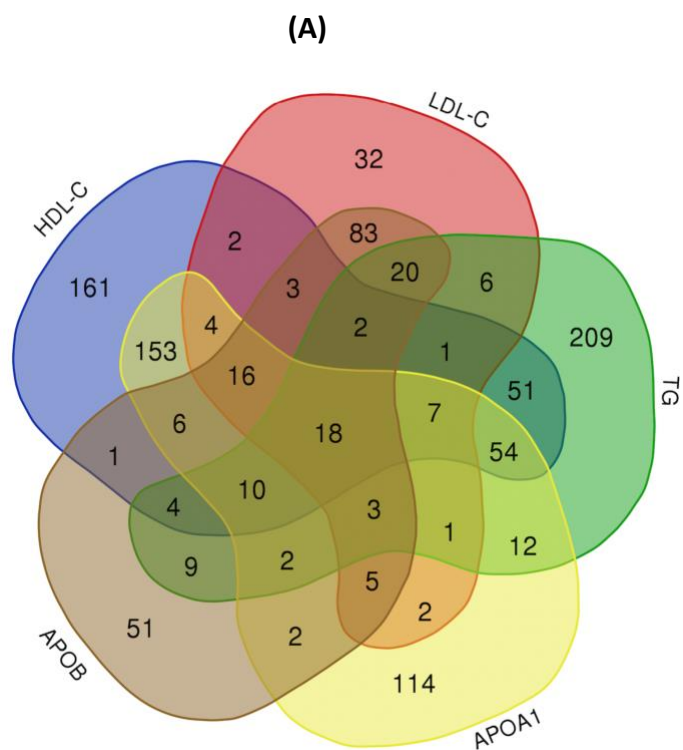


Figure 2. Characteristics of genetic instruments developed for lipid-related traits: (A) overlap of SNPs and (B) associations with lipids and apolipoproteins



Legend: In Panel A, SNPs are grouped according to whether they associate with only the primary lipid-related trait of interest, or whether they associate with other traits, based on $P < 5 \times 10^{-8}$. Panel B displays the associations of genetic instruments with lipid related traits, using the inverse variance weighting approach. While we note the potential for overfitting of estimates displayed in Panel B, we present these data for illustrative purposes; the Mendelian randomization estimates presented in Figure 3 use a two-sample approach (with no overlapping data).

Legend: In both (A) and (B), univariable Mendelian randomization (MR) estimates were derived using the inverse variance weighted approach. For a more comprehensive repertoire of estimates derived from univariable MR approaches, please see Supplementary Figure 2.