Rare protein-disrupting variants in *CETP* and protection against vascular events

Fredrik Landfors[1,*], Mathijs de Kleer[2], John J.P. Kastelein[2,3], Elin Chorell[1]

[1] Department of Public Health and Clinical Medicine, Section of Medicine, Umea University, S-901 87 Umea, Sweden.

[2] NewAmsterdam Pharma, Naarden, the Netherlands

[3] Department of Vascular Medicine, Academic Medical Center, University of Amsterdam, the Netherlands

[*] To whom correspondence should be addressed.

Contact information for the corresponding author: Fredrik Landfors, Department of Public Health and Clinical Medicine, Umea University, S-901 87 Umea, Sweden; Email: Fredrik.Landfors@umu.se;

Abstract

Background: Sequence variants that disrupt Cholesteryl Ester Transfer Protein (CETP) provide a genetic proxy for therapeutic inhibition, but their association with atherosclerotic vascular events has not been robustly established.

Objectives: To assess whether protein-disrupting *CETP* variants conferring lifelong reduced CETP activity lower atherosclerotic vascular event rates.

Methods: We tested associations between protein-disrupting *CETP* variants and atherosclerotic vascular and coronary artery disease across the All of Us Research Program and the UK Biobank using whole-genome sequences and linked health data. Predicted protein-disrupting variants were identified with LOFTEE and AlphaMissense. Gene-based burden testing used Firth's logistic regression via REGENIE. Coronary artery disease was meta-analysed with FinnGen, Biobank Japan, and previous case-control studies. Restricted mean event-free survival to average life expectancy of 81 years was estimated in the UK Biobank.

Results: We included 1,515–4,575 carriers of *CETP* protein-disrupting variants and 848,000–1,585,000 non-carriers in combined populations with 126,000 atherosclerotic events and 246,000 coronary artery disease events, respectively. HDL cholesterol was 25% higher (0.35 mmol/L [14 mg/dL], $P=4\times10^{-138}$) and non-HDL cholesterol was 6% lower (0.23 mmol/L [9 mg/dL], $P=9\times10^{-8}$) in carriers. Atherosclerotic vascular event rates were lower in carriers (odds ratio, 0.73, 95% confidence interval, 0.62–0.87, $P=3\times10^{-4}$). Coronary artery disease risk was also lower in carriers (odds ratio, 0.79, 95% confidence interval, 0.72–0.86, $P=3\times10^{-7}$), reaching exome-wide significance. Carriers lived an average 1.0 years longer without atherosclerotic disease (95% confidence interval, 0.5-1.4, $P=6\times10^{-5}$).

Conclusion: Protein-disrupting *CETP* variants were associated with reduced atherosclerotic vascular disease risk and longer disease-free survival.

Introduction

Cholesteryl ester transfer protein (CETP) is a lipid exchange protein that facilitates the transfer of cholesteryl esters from high-density lipoprotein (HDL) to low-density (LDL) or very-low-density lipoprotein (VLDL) in exchange for triglycerides.[1,2] Because this lowers HDL cholesterol and raises atherogenic lipoproteins, CETP became an attractive therapeutic target, leading to the development of several small-molecule inhibitors. Clinical trials of these inhibitors, however, yielded mixed results. Torcetrapib, the first inhibitor, increased cardiovascular events and mortality through mineralocorticoid receptor activation, which increased aldosterone levels and blood pressure.[3,4] Subsequent inhibitors, dalcetrapib and evacetrapib, failed to reduce cardiovascular events in outcomes trials.[5,6] Anacetrapib showed promise by reducing vascular events in a phase 3 trial, but its development was halted due to its modest effect size, as well as accumulation in adipose tissue raising long-term safety concerns.[7,8] Obicetrapib is in ongoing cardiovascular outcomes trials, with no off-target effects reported to date.[9]

The failure of three of four major trials has cast doubt on the therapeutic value of CETP inhibition in atherosclerosis prevention, but whether the failures reflected poor pharmacological properties of the molecules tested or limitations intrinsic to CETP inhibition remains to be answered. Human genetics offers a way to address this question by isolating the consequences of reduced CETP activity from compound-specific off-target effects. Heterozygous carriers of *CETP* protein-disrupting variants experience approximately one-third reduced CETP activity,[10,11] effectively serving as natural experiments of lifelong, selective inhibition. Because genetic variants are assigned through random-like processes independent of environmental exposures occurring later in life, genetic studies are less susceptible to confounding and reverse causation, and can inform drug target assessment.[12–14]

Previous genome-wide association and Mendelian randomization studies have established robust associations between common variants at the *CETP* locus and both lipid levels and cardiovascular risk.[15–19] Rare protein-disrupting variants offer a complementary approach. By producing more substantial, mechanistically interpretable reductions in protein quantity or function, they may more directly model the anticipated effects of pharmacological inhibition.[20] Nomura et al. previously examined rare *CETP* protein-truncating variants and coronary heart disease.[21] However, because protein-

disrupting variant carriers are rare, prior estimates have been imprecise and have not reached exome-wide significance.

We therefore sought to confirm these findings with greater precision using larger biobanks. To do this, we ascertained *CETP* protein-disrupting variants in the UK Biobank and the All of Us Research Program to measure their association with incident and prevalent atherosclerotic vascular events, coronary artery disease, and plasma lipids. Analysing these two independent biobanks enabled replication with harmonized variant annotation and comparable outcome definitions in well phenotyped populations. We then conducted a meta-analysis combining our results with Biobank Japan, FinnGen, and the prior case-control study,[21] to improve power and test consistency across cohorts and ancestries.

Methods

Study cohorts and outcome definitions

The UK Biobank (UKB) is a population-based study with health outcomes data, genetics, and blood chemistry from 500,000 participants across the United Kingdom.[22] The study enrolled individuals aged 40-69 between 2006 and 2010, during which blood samples were collected. Ongoing follow-up ensures that health information is continuously updated through medical records, death certificates, and national registries. The data (UKB v20) were last updated on December 14, 2025. Atherosclerotic vascular disease events were defined as the first-occurring death certificate or hospital inpatient record diagnosis of coronary artery disease, ischemic stroke, peripheral atherosclerosis, precerebral atherosclerosis (carotid, vertebral, basilar), or abdominal aortic aneurysm. We also ascertained cases using procedure codes indicating revascularization for any atherosclerotic vascular disease endpoint. The international classification of disease (ICD) and OPCS Classification of Interventions and Procedures codes corresponding to each condition is found in **Supplemental Table 1**. LDL-C was measured both directly (UKB data field 30780), and calculated using the Friedewald formula from total cholesterol, HDL, and triglycerides (UKB data fields 30690, 30760, 30870). Non-HDL cholesterol was calculated by subtracting HDL from total cholesterol. Apolipoprotein A-

I, B, and Lipoprotein(a) levels were measured using immunoturbidimetry (UKB data fields 30630, 30640, 30790).

The All of Us Research Program is a longitudinal cohort study launched by the National Institutes of Health in 2018, having enrolled more than 873,000 participants as of 2026.[23] Participants contribute with their electronic health records, complete surveys, provide biological samples, and undergo physical measurements. Currently, over 414,000 participants are included in whole-genome sequencing data set. The data (All of Us Controlled Tier Dataset v8) were last updated on January 13, 2026. Coronary artery disease, ischemic stroke, precerebral atherosclerosis, peripheral atherosclerosis, and abdominal aortic aneurysm were ascertained using Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) codes (**Supplemental Table 2**). Total cholesterol, HDL, and triglycerides (OMOP concept IDs 3027114, 3007070, 3022192), were extracted from electronic health records, and LDL cholesterol was calculated from these values using the Friedewald formula. Non-HDL cholesterol was calculated by subtracting HDL from total cholesterol.

Prevalent coronary artery disease was previously ascertained in Biobank Japan, FinnGen, and one case-control study by Nomura et al.[21] Biobank Japan recruited 200,000 individuals from 66 hospitals between 2003-2007. The case definition was being positive for myocardial infarction (ICD-10: I21/I22/I23/I24/I25), unstable angina (ICD-10: I20.0), or stable angina (ICD-10: I20.9) diagnoses.[24,25] FinnGen consists of a network of biobanks that integrates genetics with longitudinal health data from national registries for over 520,000 Finns.[26] Coronary artery disease cases had a hospital discharge record, insurance reimbursement or death certificate positive for angina pectoris (ICD-10: I20), acute myocardial infarction or complications thereof (ICD-10: I21-I24), or chronic ischemic heart disease (ICD-10: I25), or corresponding ICD-8/9 and insurance codes (https://risteys.finngen.fi/endpoints/I9_IHD). Nomura et al. included 12 case-control studies with 67,281 participants.[21] The included studies case definitions included predominantly early-onset myocardial infarction and early-onset coronary artery disease. To avoid sample overlap, the Biobank Japan subset (N=15,598) of Nomura et al. was excluded from the present meta-analysis.

Pre-processing and quality control of genetic data

We used whole-genome sequences from approximately 460,000 UK Biobank and 390,000 All of Us participants. Whole-genome sequencing of the UK Biobank was performed between 2019 and 2022 using Illumina NovaSeq platforms, with data processed using DRAGEN v3.7.8 aligned to GRCh38 at 32.5x mean coverage.[27] Whole-genome sequencing data from the All of Us Research Program were accessed via the Researcher Workbench Controlled Tier Dataset v8, processed with DRAGEN v3.7.8 at ≥30x mean coverage aligned to GRCh38 as described previously.[28] Biobank Japan participants were genotyped using Illumina arrays and imputed using a combined 1000 Genomes and Japanese reference panel.[29] FinnGen participants were genotyped using Illumina and Affymetrix arrays and imputed using a Finnish-specific reference panel.[26] Nomura et al. participants underwent exome sequencing or genotyping.[21]

Genetic ancestry in UK Biobank was inferred using a support vector machine classifier implemented in KING,[30] trained on 1000 Genomes reference populations and applied to principal components derived from genomic data. For All of Us, ancestry assignments provided by the program (v8 ancestry predictions) were used. We restricted analyses to ancestry groups where more than 5 participants with protein-disrupting variant were expected to experience the primary outcome event, using a carrier frequency of 2‰ and an event rate of 10% for calculations. Based on this criterion, participants of European ancestry were included from UK Biobank, and participants of African, Admixed American, and European ancestry were included from All of Us.

For REGENIE Step 1, we created ancestry-stratified sets of approximately 500,000 linkage disequilibrium-pruned variants. Variants were filtered based on minor allele frequency (MAF ≥1.5% for UK Biobank; ≥1% in All of Us), minor allele count (MAC ≥100 for UK Biobank; ≥20 for All of Us), Hardy-Weinberg equilibrium exact test (HWE, $P>10^{-15}$), and genotype missingness (<10%). Linkage disequilibrium pruning was performed using a 500kb window and r^2 threshold of 0.2. Individuals with >10% missing genotypes were excluded. For burden testing of the *CETP* locus (chr16:56,961,923-56,983,845; GRCh38) in REGENIE Step 2, variants were extracted from BGEN-formatted files with more permissive filtering (MAC≥1, HWE exact test $P>10^{-5}$, genotype and sample missingness <10%) to include rare variants. Multiallelic variants were excluded in Step 1 but retained in Step 2.

For gene-based burden testing, we grouped variants based on predicted variant functional consequences. We restricted analyses to variants present in gnomAD v4.1.0 to ensure high-quality variant calls and enable consistent annotation across cohorts.[31] Predicted protein-truncating variants (PTVs) included splice acceptor/donor, stop gained, and frameshift variants. High-confidence PTVs were defined as PTVs passing Loss-Of-Function Transcript Effect Estimator (LOFTEE) filters, which identify and exclude probable false-positive loss-of-function calls based on sequence context and splice predictions.[32] Missense variant pathogenicity was predicted using AlphaMissense, a deep learning model based on AlphaFold that uses protein structure and evolutionary conservation to classify missense variants,[33] with variants classified as "likely pathogenic" (score ≥0.564) included in the missense burden mask. For the gene-based burden tests, high-confidence PTVs were combined with AlphaMissense likely pathogenic variants into a single variable, designating their grouping as protein-disrupting variants.

Statistical analysis

Association analyses were performed using REGENIE v4.1.[34] REGENIE employs a two-step whole-genome regression approach. Step 1 fits a null model using a set of linkage disequilibrium-pruned common variants to account for population structure and relatedness through leave-one-chromosome-out predictions.[34,35] Step 2 tests associations at individual variants or using gene-based burden masks while conditioning on the Step 1 predictions. Covariates included age, age^2, sex, and the first 10 genomic principal components. For continuous traits, we used linear regression (--qt), and for binary outcomes, Firth-corrected logistic regression (--bt, --firth, --firth-se). Meta-analysis across ancestries and cohorts was performed using fixed-effects inverse variance-weighting implemented in the *metafor* R package.[36] Study heterogeneity was estimated using Cochran's Q test and calculating the I^2 statistic. Exome-wide significance was defined as $P<2.5\times10^{-6}$ (0.05/20,000 genes), with uncorrected *P* values and 95% confidence intervals reported in **Figures 1-3**.

In the UK Biobank, time-to-event analyses were conducted to estimate the cumulative incidence of atherosclerotic vascular disease by carrier status. Exposure to a defective *CETP* allele was assumed to begin after birth, and participants were followed from their birth (UKB data fields 34, 54) until the occurrence of the first event, lost to

199 follow-up (UKB data field 191), death (UKB data field 40000), or the last date registries
200 were queried (Scotland, death registry 2024-11-30, hospital inpatient 2022-08-31;
201 England, death registry, 2024-08-31, hospital inpatient 2023-03-31; Wales, death registry
202 2024-08-31, hospital inpatient 2022-05-31), whichever occurred first. Kaplan-Meier
203 survival curves were constructed, and the crude hazard ratio was measured using Cox
204 regression. To estimate the lifetime benefit, restricted mean survival time was calculated
205 using the *survRM2* R package.[37] The restriction time of 81 years was set based on the U.K.
206 average life expectancy for 2023.[38]

207 Ethical Review

208 This research has been conducted using the UK Biobank Resource under Application
209 Number 148828. All of Us data were accessed under an institutional data use agreement
210 established the All of Us Research Program and the analyses were approved by the
211 Swedish Ethical Review Authority (Dnr 2025-05711-01). Summary statistics from Biobank
212 Japan, FinnGen, and Nomura et al. are publicly available,[21,24,26] and analyses of these data
213 did not require separate ethical approval. All participants provided written informed
214 consent, and the research adhered to the ethical standards set forth in the Declaration of
215 Helsinki.

216 Results

217 Cohort characteristics and protein-disrupting variants

218 Descriptive characteristics on the included 849,761 UK Biobank and All of Us participants
219 are presented in **Table 1**. The mean age at recruitment into the UK Biobank was 56.8 years
220 and 45.8% were male. The mean age of recruitment into the All of Us cohort was 58.0
221 years and 39.0% were male. We identified 775 *CETP* protein-disrupting variant carriers in
222 the UK Biobank, whereof 463 (59.7%) were LOFTEE-predicted high-confidence protein-
223 truncating variants and 312 (40.3%) were AlphaMissense-predicted pathogenic variants.
224 In All of Us, 740 protein-disrupting variant carriers were identified, whereof 437 (59.1%)
225 participants carried high-confidence protein-truncating variants, and 303 (40.9%)
226 participants carried AlphaMissense-predicted pathogenic variants.

227 To confirm that both AlphaMissense and LOFTEE-predicted protein-disrupting
228 variants altered plasma CETP activity, we measured their association with HDL

229 cholesterol and plotted groupwise HDL cholesterol distributions in the UK Biobank. We
230 tested each variant with minor allele count ≥5 individually for their association with HDL
231 cholesterol. One relatively common AlphaMissense-classified pathogenic variant
232 (GRCh38 chr16:56,973,332:G>T [p.Gly251Val], European ancestry MAF 0.19‰, HWE-
233 estimated $N_{carriers}$=188), showed no association with HDL cholesterol (fixed effects
234 estimate, +0.0069 mmol/L [0.3 mg/dL], 99.99975% CI: -0.133 to 0.15, *P*=0.82) across All
235 of Us and UK Biobank, reliably excluding the expected effect size for *CETP* protein-
236 disrupting variants (≥0.20 mmol/L [8 mg/dL]). This variant was excluded from gene-based
237 burden testing. All other predicted *CETP* protein-disrupting variants identified in UK
238 Biobank and All of Us were included in the gene-based burden tests and are listed in
239 **Supplemental tables 3-4**. To further ensure variants included in burden testing disrupted
240 CETP function, we plotted their distribution in UK Biobank. Both AlphaMissense-
241 predicted pathogenic variants and LOFTEE-predicted high-confidence protein-truncating
242 variants raised mean HDL cholesterol by 20% (+0.28 mmol/L [11 mg/dL], $P=1\times10^{-33}$) and
243 27% (+0.39 mmol/L [15 mg/dL], $P=1\times10^{-96}$) respectively, with both groups showing
244 concordant rightward-shifted distributions relative to non-carriers (**Figure 1A**). These
245 effect estimates were within the expected range of known *CETP* variants that cause
246 complete CETP deficiency in their homozygous state.[10]
247

248 ## Association with lipids and atherosclerotic vascular disease

249 The associations between *CETP* protein-disrupting variant carrier status and plasma
250 lipids are shown in **Figure 1B**. In the combined All of Us and UK Biobank analysis, *CETP*
251 protein-disrupting variant carrier status was associated with 25% higher HDL cholesterol,
252 at +0.35 mmol/L (14 mg/dL, $P=4\times10^{-138}$). Non-HDL cholesterol was 6% lower at -0.23
253 mmol/L (9 mg/dL, $P=9\times10^{-8}$), and Friedewald LDL cholesterol was 6% lower at -0.20
254 mmol/L (8 mg/dL, $P=2\times10^{-7}$). In UK Biobank, an exome-wide significant 6% reduction in
255 apolipoprotein B concentration ($P=9\times10^{-13}$) was also observed, along with an 11%
256 increase in apolipoprotein A-I ($P=5\times10^{-40}$), and a nominal 12% reduction in lipoprotein(a)
257 ($P=7\times10^{-4}$).
258 The association of *CETP* protein-disrupting variant carrier status with composite
259 atherosclerotic vascular disease in the UK Biobank and All of Us cohorts is shown in

260 **FIGURE 2**. Prevalent and incident atherosclerotic vascular disease, defined as coronary
261 artery disease, ischemic stroke, peripheral atherosclerosis, precerebral atherosclerosis,
262 or abdominal aortic aneurysm, was ascertained in 72,370 of 460,190 (16.1%) UK Biobank
263 participants and in 53,341 of 383,060 (13.9%) All of Us participants. In fixed-effects meta-
264 analysis combining estimates from UK Biobank with All of Us, carriers had 27% lower
265 odds of atherosclerotic vascular disease (odds ratio, 0.73, 95% CI: 0.62-0.87, $P=3\times10^{-4}$).
266 Time-to-event analyses in the UK Biobank showed *CETP* protein-disrupting variant
267 carriers lived an average of 1.0 years longer without atherosclerotic disease compared to
268 non-carriers (95% CI: 0.5-1.4, $P=6\times10^{-5}$), with mean survival times estimated at 77.7 years
269 and 76.7 years, respectively.
270

271 ## Coronary artery disease meta-analysis

272 The association between *CETP* protein-disrupting variant carrier status and coronary
273 artery disease in 246,000 cases and 1,344,000 controls is shown in **FIGURE 3**. Meta-
274 analysis was conducted by including Biobank Japan (32,512 cases/146,214 controls;
275 mean age 63 years at donation, 54% male),[24] FinnGen (84,088 cases/416,250 controls;
276 mean age 53 years at donation, 44% male),[26] and previous case-control studies (22,205
277 cases/45,076 controls; mean age 52 years, 57% male) published by Nomura et al.,[21]
278 together with the UK Biobank and All of Us. Since the broad atherosclerotic vascular
279 disease composite used in the previous analyses were unavailable in Biobank Japan,
280 FinnGen and the 12 case-control studies, the association between coronary artery
281 disease and *CETP* protein-disrupting variant carriers was measured in All of Us (46,459
282 cases/336,601 controls) and the UK Biobank (60,775 cases/399,415 controls) to ensure
283 harmonized endpoints.
284 In Biobank Japan, the East Asian-enriched (88-fold vs. non-Finnish Europeans)
285 16:56,982,238:G>A (GRCh38; c.1321+1G>A) splice donor LOFTEE-predicted high-
286 confidence PTV was present in approximately 2,008 participants (estimated from HWE
287 equations based on MAF 5.63‰), well-imputed (INFO=0.93), and associated with raised
288 HDL cholesterol (+0.32 mmol/L [12 mg/dL], $P=5\times10^{-122}$) and lowered LDL cholesterol (-
289 0.16 mmol/L [6 mg/dL], $P=5\times10^{-7}$). The HDL cholesterol-raising effect was well within the
290 expected range of protein-disrupting *CETP* variants in their heterozygous state, and

homozygous carrier status of this variant is known to cause complete CETP deficiency with extremely elevated HDL cholesterol concentrations (mean 4.2 mmol/L [162 mg/dL] vs. 1.4 mmol/L [54 mg/dL] in unaffected family members).[10] The Finnish-enriched (48-fold vs. non-Finnish Europeans) splice donor variant 16:56,962,097:TGTAA>T (GRCh38; c.118+4_118+7del) was well-imputed (median INFO=1.00), and present in approximately 895 participants. Association analysis in the UK Biobank found this variant was nominally associated with raised HDL cholesterol (+0.22 mmol/L [9 mg/dL], *P*=0.008) and directionally consistent but non-significantly associated with lowered non-HDL cholesterol (-0.46 mmol/L [18 mg/dL], *P*=0.07). Nomura et al. found 279 splice acceptor/donor, stop gained, and frameshift variant carriers with raised HDL cholesterol (+0.58 mmol/L [22 mg/dL], $P=1\times10^{-4}$), and lowered LDL cholesterol (-0.31 mmol/L [12 mg/dL], *P*=0.043).[21] Because one Nomura et al. substudy (N=15,598) overlapped with Biobank Japan, substudies including 157 carriers were included in the present meta-analysis.

Fixed effects inverse variance-weighted meta-analysis found the risk of coronary artery disease was lower in *CETP* protein-disrupting variant carriers (odds ratio, 0.79, 95% CI: 0.72-0.86, $P=3\times10^{-7}$), meeting the gene-based exome-wide significance threshold. Effect estimates were directionally consistent across the contributing studies, with no significant heterogeneity (I^2=0%).

Discussion

In this multi-ancestry analysis of up to 1,586,000 participants including 4,575 carriers of *CETP* protein-disrupting variants, we observed a 21% reduction in coronary artery disease risk that reached exome-wide significance, along with significant reductions in composite atherosclerotic vascular disease. Carriers of *CETP* protein-disrupting variants lived an average of one additional year without atherosclerotic vascular disease compared to non-carriers in the UK Biobank. Together, these findings provide robust and convergent human genetic evidence that *CETP* loss-of-function is associated with reduced atherosclerotic disease risk.

Several analyses of common *CETP* variants have shown that genetically reduced CETP activity lowers cardiovascular risk.[15–19] Rare protein-disrupting variants offer

322 complementary evidence. They typically produce substantial reductions in CETP activity
323 through a less ambiguous loss-of-function mechanism, more directly modelling the
324 effects anticipated with pharmacological inhibition. Consistent with this, these variants
325 were associated with higher apolipoprotein A-I and HDL cholesterol as well as lower LDL
326 cholesterol, apolipoprotein B, and lipoprotein(a), mirroring the effects on lipids observed
327 with pharmacological CETP inhibition in clinical trials.[39] Extending the analysis by
328 Nomura et al.,[21] our larger study achieved gene-based exome-wide significance for the
329 association between *CETP* protein-disrupting variants and coronary artery disease
330 (P=3×10^{-7}), strengthening the evidence that *CETP* loss-of-function lower the risk of
331 atherosclerotic vascular events.

332 These results align with principles for using human genetics to evaluate drug
333 targets.[20] The genetic variants examined were predicted to disrupt CETP protein function
334 through truncation or protein-damaging missense substitutions, providing a link between
335 genotype and reduced plasma CETP activity. Protein-truncating variants typically act
336 through nonsense-mediated mRNA decay, though variants escaping mRNA decay may
337 produce truncated proteins that are unstable or non-functional,[40] while AlphaMissense-
338 predicted pathogenic missense variants may reduce protein abundance through
339 destabilization and degradation, or impair function directly by substituting catalytically or
340 structurally critical residues.[33] The similar effects of both variant classes on HDL
341 cholesterol distributions, as shown in **Figure 1A**, support their inclusion in a unified
342 burden test.

343 Several limitations warrant consideration. First, the variants we studied represent
344 constitutive germline effects that act systemically, from development through adulthood,
345 in all *CETP*-expressing tissues, whereas clinical trials administer drugs with different
346 tissue distribution and target selectivity, to adults, and generally follow participants for
347 only a few years. Lifelong genetic effects may confer benefits such as prevention of early
348 atherosclerotic plaque formation, that are not achievable with pharmacological
349 intervention later in life. Conversely, the cumulative years of exposure in genetic variant
350 carriers may exceed what is practically achievable in a clinical trial setting, which could
351 inflate estimates when translating genetic effects onto what effect would be achievable
352 in trials. Second, we examined heterozygous carriers who exhibited approximately 25%
353 higher HDL cholesterol levels (+0.35 mmol/L [9 mg/dL]), and 6% lower LDL cholesterol

354 levels (-0.20 mmol/L [8 mg/dL]). This effect magnitude is consistent with earlier
355 observations that partially CETP-deficient heterozygotes typically show 25% increases in
356 HDL cholesterol,[10,41,42] and retain 58–68% of normal CETP activity, suggesting partial
357 compensation by the wild-type allele.[11] However, this one-third inhibition only represents
358 approximately one-fifth to one-seventh of the lipid-altering potency of pharmacological
359 CETP inhibitors, which can raise HDL cholesterol by up to 140% and lower LDL
360 cholesterol by up to 44% at near-complete CETP inhibition in participants without
361 background lipid-lowering therapy.[43] Investigation of homozygous carriers, who exhibit
362 complete CETP deficiency, is challenging and was not feasible due to their extreme rarity.
363 The cardiovascular benefits of the near-complete CETP inhibition achieved with potent
364 CETP inhibitors therefore require confirmation in clinical trials. Third, FinnGen and
365 Biobank Japan genotypes were imputed from reference panels rather than directly
366 sequenced, albeit at high quality (INFO=0.93–1.00). Incorrect imputation of rare variants
367 typically results in genotypes homozygous for the reference allele, producing one-sided
368 misclassification that biases toward the null.[44] The FinnGen and Biobank Japan estimates
369 are therefore likely conservative rather than inflated by imputation error. Fourth, restricted
370 mean survival time estimates reflect the baseline risk of the study populations and may
371 differ in populations with higher or lower cardiovascular risk. Fifth, the inferential validity
372 of genetic association studies relies on the assumption that genetic variants are
373 distributed in a random-like pattern within the study population, mimicking
374 randomization in a clinical trial. Potential confounders such as population stratification
375 or assortative mating could bias estimates by creating non-causal correlations between
376 *CETP* variants and cardiovascular events.[45] We sought to minimize these biases by
377 stratifying analyses by genetically inferred ancestry groups, adjusting for genomic
378 principal components within those groups, and demonstrating consistent effect
379 estimates across multiple independent cohorts and ancestries, which together mitigate
380 concerns about confounding.

381 In conclusion, rare protein-disrupting variants in *CETP* conferring lifelong inhibition
382 were associated with reduced atherosclerotic vascular disease risk and longer
383 atherosclerotic vascular disease-free survival. These findings add human genetic
384 evidence for the role of CETP inhibition in atherosclerotic vascular disease prevention and

385 supports CETP as a therapeutically relevant target. Outcomes trials will determine the
386 role of small-molecule CETP inhibitors in cardiovascular prevention.
387

Acknowledgements

We gratefully acknowledge All of Us participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's All of Us Research Program for making available the participant data examined in this study. We want to acknowledge the participants and investigators of the FinnGen study. We thank the participants and hospital group of Biobank Japan. This research has been conducted using the UK Biobank Resource under Application Number 148828. This work uses data provided by patients and collected by the NHS as part of their care and support. The graphical abstract was created in BioRender (Landfors, F. (2024) https://BioRender.com/j51i966).

Funding

This work received financial support from New Amsterdam Pharma N.V.

Conflict of interest

NewAmsterdam Pharma N.V. is currently developing the CETP inhibitor Obicetrapib, which is undergoing evaluation in a phase 3 cardiovascular outcomes trial (Cardiovascular Outcome Study to Evaluate the Effect of Obicetrapib in Patients With Cardiovascular Disease [PREVAIL], ClinicalTrials.gov number, NCT05202509).[9] Fredrik Landfors received funding for this project and reports consulting fees from NewAmsterdam Pharma N.V. Mathijs de Kleer is an employee of NewAmsterdam Pharma N.V., and John J.P. Kastelein is the chief scientific officer of NewAmsterdam Pharma N.V. Authors employed by NewAmsterdam Pharma receive compensation and hold equity in the company. Elin Chorell has no conflicts of interest to declare.

Data availability

UK Biobank data are available to authorized researchers through application (https://www.ukbiobank.ac.uk). This study used data from the All of Us Research Program's Controlled Tier Dataset v8, available to authorized users on the Researcher Workbench (https://www.researchallofus.org). FinnGen summary statistics are publicly available (https://www.finngen.fi/en/access_results). Biobank Japan summary statistics are publicly available (https://pheweb.jp). The case-control data from Nomura et al. were

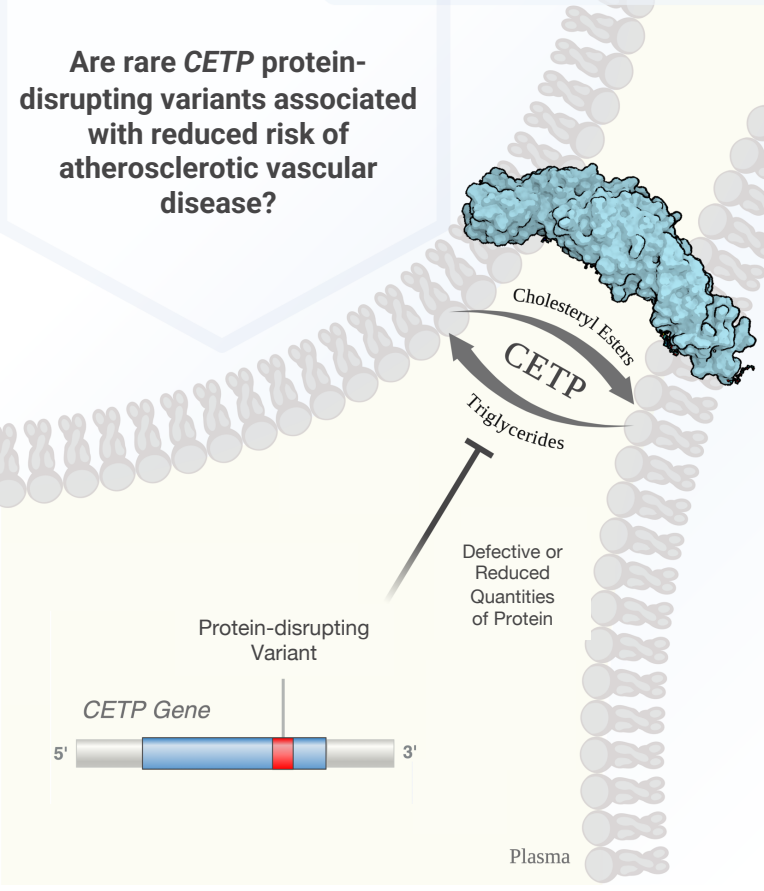
obtained from the published manuscript.[21] The UK Biobank data field identifiers and OMOP concept IDs used in the analyses are described in the **Methods** section and **Supplemental Tables 1-2**.

STUDY OF CHOLESTERYL ESTER TRANSFER PROTEIN (CETP) INHIBITION

High-density lipoprotein

Low-density lipoprotein & Very-low density lipoprotein

Are rare *CETP* protein-disrupting variants associated with reduced risk of atherosclerotic vascular disease?



In up to 1,586,000 participants (4,575 carriers), protein-disrupting variants conferring lifelong reductions CETP activity were associated with:

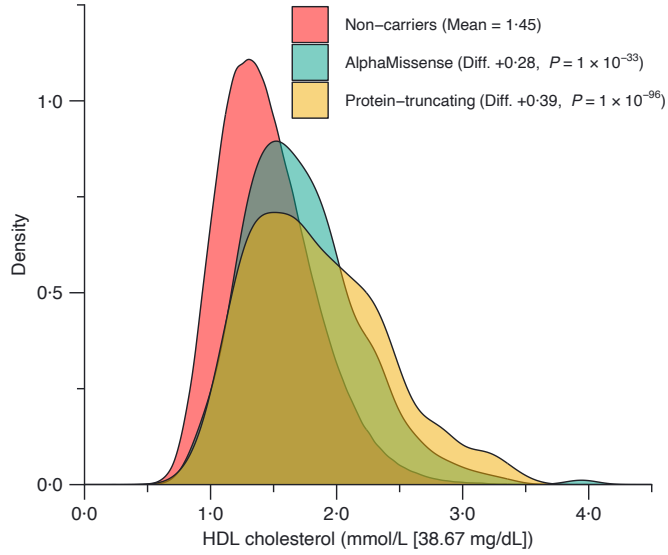
- **Raised HDL-C by 25% (+0.35 mmol/L / +14 mg/dL) and lowered non-HDL-C by 6% (0.23 mmol/L / 9 mg/dL)**
- **Reduced odds of coronary artery disease by 21% (95% CI, 14–28%)**
- **One additional year of life without vascular events (95% CI, 0.5–1.4)**

(GRAPHICAL ABSTRACT)

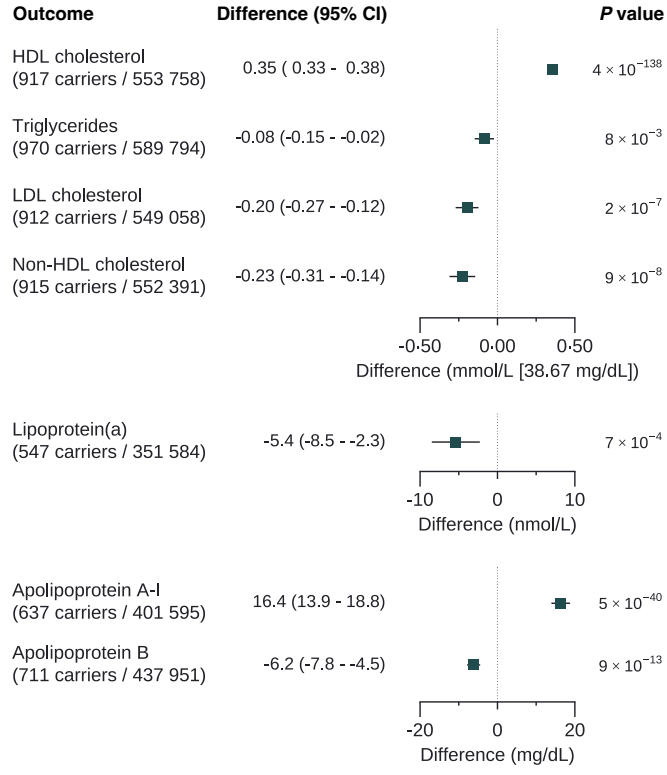
Table 1. Cohort characteristics	UK Biobank		All of Us Research Program	
	Mean (SD) / No. (%)			
	Non-carriers (n = 459 415)	*CETP* Protein-disrupting Variant Carriers (n = 775)	Non-carriers (n = 388 831)	*CETP* Protein-disrupting Variant Carriers (n = 740)
Age at recruitment, years	56.8 (8.0)	56.5 (7.8)	58.0 (17.0)	55.6 (16.7)
Male	210 330 (45.8%)	355 (45.8%)	151 815 (39.0%)	263 (35.5%)
Blood pressure, mmHg				
Systolic	138.0 (18.6)	138.6 (19.3)	125.6 (15.3)	125.1 (16.0)
Diastolic	82.2 (10.1)	82.8 (10.7)	76.6 (9.8)	76.7 (10.2)
BMI, kg/m^2	27.4 (4.8)	27.5 (4.9)	29.8 (7.4)	30.1 (7.7)
Self-reported diabetes status	22 193 (4.8%)	42 (5.4%)	60 732 (15.6%)	112 (15.1%)
Self-reported smoking status	48 069 (10.5%)	79 (10.2%)	59 392 (15.3%)	119 (16.1%)
Taking lipid-lowering medication	30 409 (6.6%)	40 (5.2%)	128 456 (33.0%)	206 (27.8%)
Taking antihypertensive medication	42 971 (9.4%)	66 (8.5%)	114 773 (29.5%)	176 (23.8%)
Prevalent or incident hypertension	151 788 (33.0%)	251 (32.4%)	131 586 (33.8%)	203 (27.4%)
Lipids, mmol/L				
HDL cholesterol	1.45 (0.38)	1.80 (0.51)	1.37 (0.42)	1.69 (0.57)
LDL cholesterol, direct	3.57 (0.87)	3.39 (0.83)	-	-
LDL cholesterol, Friedewald	3.46 (0.99)	3.25 (0.97)	2.63 (0.82)	2.51 (0.79)
Non-HDL cholesterol	4.26 (1.08)	4.03 (1.04)	3.24 (0.90)	3.10 (0.89)
Triglycerides	1.75 (1.03)	1.70 (1.00)	1.34 (0.84)	1.25 (0.70)
Lipoproteins				
Apolipoprotein A-I, mg/dL	154 (27.1)	170 (30.5)	-	-
Apolipoprotein B, mg/dL	103 (23.8)	97 (22.9)	-	-
Lipoprotein(a), nmol/L	44.0 (49.3)	39.2 (46.8)	-	-
Variant types				
AlphaMissense pathogenic	0 (0%)	312 (40.3%)	0 (0%)	303 (40.9%)
Protein-truncating	0 (0%)	463 (59.7%)	0 (0%)	437 (59.1%)
Genetically inferred ancestry				
European	459 415 (100%)	775 (100%)	229 034 (58.9%)	369 (49.9%)
African	0 (0%)	0 (0%)	82 062 (21.1%)	164 (22.2%)
Admixed American	0 (0%)	0 (0%)	77 735 (20.0%)	207 (28.0%)
Prevalent or incident disease combined				
Any atherosclerotic vascular disease	72 279 (15.7%)	91 (11.7%)	53 933 (13.9%)	71 (9.6%)
Coronary artery disease	60 699 (13.2%)	76 (9.8%)	46 974 (12.1%)	63 (8.5%)

(TABLE 1)

A. Association of *CETP* Protein-disrupting Variant Types with HDL Cholesterol in UK Biobank

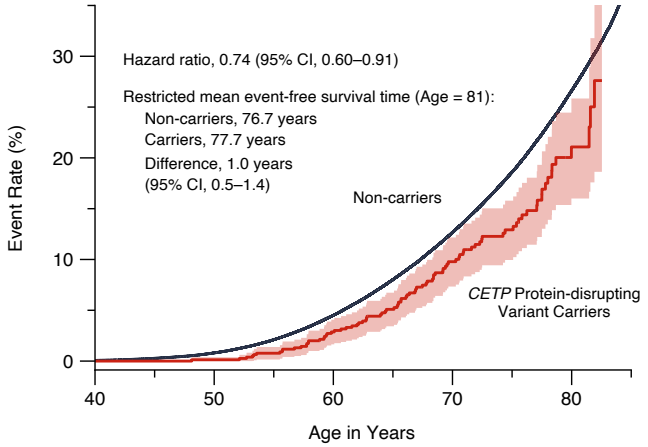


B. Association of *CETP* Protein-disrupting Variants with Plasma Lipids and Blood Pressure



(FIGURE 1)

A. Survival Analysis of *CETP* Protein-disrupting Variants and Any First Atherosclerotic Vascular Disease Event in UK Biobank



No. at Risk

Non–carrier	459111	455108	385915	234796	46759
Carrier	775	774	662	395	76

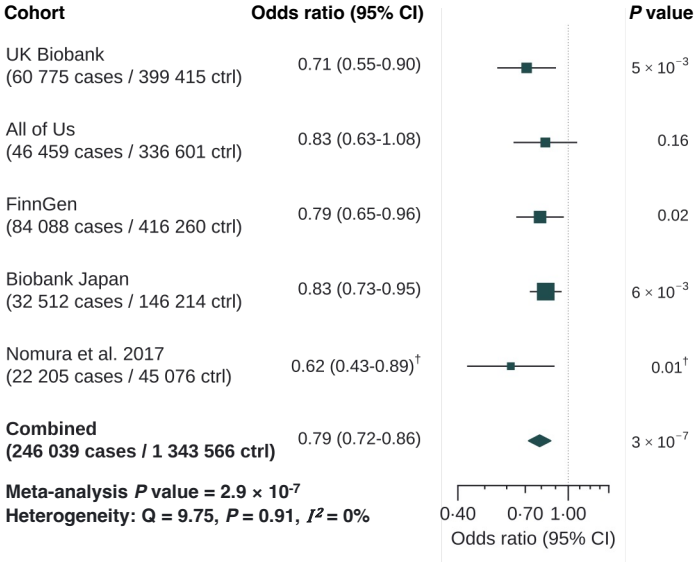
B. Association of *CETP* Protein-disrupting Variants with Any Atherosclerotic Vascular Disease

Cohort	Odds ratio (95% CI)	*P* value
UK Biobank (72 370 cases / 387 820 ctrl)	0.68 (0.54-0.86)	1×10^{-3}
All of Us (53 341 cases / 329 719 ctrl)	0.80 (0.62-1.02)	0.07
Pooled estimate (125 711 cases / 717 539 ctrl)	**0.73 (0.62-0.87)**	3×10^{-4}

Meta-analysis *P* value = 3.1×10^{-4}
Heterogeneity: Q = 1.26, *P* = 0.74, I^2 = 0%

(FIGURE 2)

Association of *CETP* Protein-disrupting Variants with Coronary Artery Disease



(FIGURE 3)

Legends

Table 1

Cohort characteristics of the included UK Biobank and All of Us participants.

Figure 1

Association of *CETP* protein-disrupting variants with plasma lipids.

(A) Distribution of HDL cholesterol concentrations in UK Biobank participants stratified by *CETP* variant status: non-carriers (red, mean 1.45 mmol/L [56 mg/dL]), AlphaMissense-predicted pathogenic variant carriers (blue; difference +0.28 mmol/L [11 mg/dL], $P=1\times10^{-33}$), and protein-truncating variant carriers (yellow; difference +0.39 mmol/L [15 mg/dL], $P=1\times10^{-96}$).

(B) Forest plots showing the association between *CETP* protein-disrupting variant carrier status and plasma lipid concentrations in the combined UK Biobank and All of Us analysis adjusting for age, age^2, sex, and the first 10 genomic principal components. Apolipoprotein A-I and apolipoprotein B were available in UK Biobank only. Error bars represent 95% confidence intervals.

Figure 2

Association of *CETP* protein-disrupting variants with atherosclerotic vascular disease.

(A) Cumulative incidence of first atherosclerotic vascular disease event by age among *CETP* protein-disrupting variant carriers and non-carriers in the UK Biobank. Shaded areas represent 95% confidence intervals. The crude unadjusted hazard ratio is shown.

(B) Forest plot showing odds ratios for composite atherosclerotic vascular disease (coronary artery disease, ischemic stroke, peripheral atherosclerosis, precerebral atherosclerosis (carotid, vertebral, basilar), or abdominal aortic aneurysm) in *CETP* protein-disrupting variant carriers versus non-carriers adjusting for age, age^2, sex, and the first 10 genomic principal components, stratified by cohort and genetically inferred ancestry. For All of Us, the pooled multi-ancestry estimate is presented to comply with data protection policies prohibiting reporting of results derived from cell counts <20. Heterogeneity statistics were calculated including UK Biobank European ancestry and All

450 of Us Admixed American, African, and European ancestry substudies individually in the
451 fixed-effects inverse variance-weighted estimator.

452 ## Figure 3

453 **Meta-analysis of *CETP* protein-disrupting variants and coronary artery disease**

454 Forest plot showing the association between *CETP* protein-disrupting variant carrier
455 status and coronary artery disease across UK Biobank, All of Us, FinnGen, Biobank Japan,
456 and case-control studies from Nomura et al. (2017).[21] The pooled multi-ancestry estimate
457 is displayed for All of Us to comply with data protection policies that prohibit reporting
458 summary statistics derived from cell counts <20. For consistency, Nomura et al. (2017) is
459 also displayed as a combined estimate. Heterogeneity statistics were calculated using
460 individual substudies from All of Us (n=3) and Nomura et al. (n=12).

461 [†] The Nomura et al. (2017) combined estimate excludes the Biobank Japan substudy
462 (N=15,598) to prevent sample overlap, and therefore differs from the originally reported
463 estimate.[21]

464 # Appendices

465 ## Supplemental tables

466 *Supplemental table 1*

467 **UK Biobank disease definitions**

468 Disease phenotype definitions used for outcome ascertainment in UK Biobank.

469 *Supplemental table 2*

470 **All of Us disease definitions**

471 Disease phenotype definitions used for outcome ascertainment in the All of Us Research
472 Program.

473 *Supplemental table 3*

474 **UK Biobank *CETP* protein-disrupting variants.**

475 *CETP* variants classified as protein-disrupting identified in UK Biobank whole-genome
476 sequencing data.

477 *Supplemental table 4*

478 **All of Us *CETP* protein-disrupting variants.**

479 *CETP* variants classified as protein-disrupting identified in All of Us Research Program.

480

References

1. Morton RE, Zilversmit DB. Inter-relationship of lipids transferred by the lipid-transfer protein isolated from human lipoprotein-deficient plasma. *Journal of Biological Chemistry*. 1983;258(19):11751-11757. doi:10.1016/S0021-9258(17)44293-1

2. Drayna D, Jarnagin AS, McLean J, et al. Cloning and sequencing of human cholesteryl ester transfer protein cDNA. *Nature*. 1987;327(6123):632-634. doi:10.1038/327632a0

3. Barter PJ, Caulfield M, Eriksson M, et al. Effects of Torcetrapib in Patients at High Risk for Coronary Events. *N Engl J Med*. 2007;357(21):2109-2122. doi:10.1056/NEJMoa0706628

4. Johns DG, Duffy J, Fisher T, Hubbard BK, Forrest MJ. On- and Off-Target Pharmacology of Torcetrapib: Current Understanding and Implications for the Structure Activity Relationships (SAR), Discovery and Development of Cholesteryl Ester-Transfer Protein (CETP) Inhibitors. *Drugs*. 2012;72(4):491-507. doi:10.2165/11599310-000000000-00000

5. Schwartz GG, Olsson AG, Abt M, et al. Effects of Dalcetrapib in Patients with a Recent Acute Coronary Syndrome. *N Engl J Med*. 2012;367(22):2089-2099. doi:10.1056/NEJMoa1206797

6. Lincoff AM, Nicholls SJ, Riesmeyer JS, et al. Evacetrapib and Cardiovascular Outcomes in High-Risk Vascular Disease. *N Engl J Med*. 2017;376(20):1933-1942. doi:10.1056/NEJMoa1609581

7. The HPS3/TIMI55–REVEAL Collaborative Group. Effects of Anacetrapib in Patients with Atherosclerotic Vascular Disease. *N Engl J Med*. 2017;377(13):1217-1227. doi:10.1056/NEJMoa1706444

8. Johns DG, Wang SP, Rosa R, et al. Impact of drug distribution into adipose on tissue function: The cholesteryl ester transfer protein (CETP) inhibitor anacetrapib as a test case. *Pharmacol Res Perspect*. 2019;7(6):e00543. doi:10.1002/prp2.543

9. Placebo Controlled, Double Blind, Randomized Cardiovascular Outcome Study to Evaluate the Effect of 10 mg Obicetrapib in Participants With ASCVD Not Adequately Controlled Despite Maximally Tolerated Lipid Modifying Therapies. Published online November 22, 2021. Accessed February 7, 2022. https://clinicaltrials.gov/study/NCT05202509

10. Inazu A, Brown ML, Hesler CB, et al. Increased High-Density Lipoprotein Levels Caused by a Common Cholesteryl-Ester Transfer Protein Gene Mutation. *N Engl J Med*. 1990;323(18):1234-1238. doi:10.1056/NEJM199011013231803

11. Boekholdt SM, Thompson JF. Natural genetic variation as a tool in understanding the role of CETP in lipid levels and disease. *Journal of Lipid Research*. 2003;44(6):1080-1093. doi:10.1194/jlr.R200018-JLR200

12. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856-860. doi:10.1038/ng.3314

13. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. Marchini J, ed. *PLoS Genet*. 2019;15(12):e1008489. doi:10.1371/journal.pgen.1008489

14. Trajanoska K, Bhérer C, Taliun D, Zhou S, Richards JB, Mooser V. From target discovery to clinical drug development with human genetics. *Nature*. 2023;620(7975):737-745. doi:10.1038/s41586-023-06388-8

15. Johannsen TH, Frikke-Schmidt R, Schou J, Nordestgaard BG, Tybjærg-Hansen A. Genetic Inhibition of CETP, Ischemic Vascular Disease and Mortality, and Possible Adverse Effects. *Journal of the American College of Cardiology*. 2012;60(20):2041-2048. doi:10.1016/j.jacc.2012.07.045

16. Ference BA, Kastelein JJP, Ginsberg HN, et al. Association of Genetic Variants Related to CETP Inhibitors and Statins With Lipoprotein Levels and Cardiovascular Risk. *JAMA*. 2017;318(10):947. doi:10.1001/jama.2017.11467

17. Nordestgaard LT, Christoffersen M, Lauridsen BK, et al. Long-term Benefits and Harms Associated With Genetic Cholesteryl Ester Transfer Protein Deficiency in the General Population. *JAMA Cardiol*. 2022;7(1):55. doi:10.1001/jamacardio.2021.3728

18. Cupido AJ, Reeskamp LF, Hingorani AD, et al. Joint Genetic Inhibition of PCSK9 and CETP and the Association With Coronary Artery Disease: A Factorial Mendelian Randomization Study. *JAMA Cardiol*. 2022;7(9):955. doi:10.1001/jamacardio.2022.2333

19. Dunca D, Chopade S, Gordillo-Marañón M, et al. Comparing the effects of CETP in East Asian and European ancestries: a Mendelian randomization study. *Nature Communications*. 2024;15(1):5302. doi:10.1038/s41467-024-49109-z

20. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*. 2013;12(8):581-594. doi:10.1038/nrd4051

21. Nomura A, Won HH, Khera AV, et al. Protein-Truncating Variants at the Cholesteryl Ester Transfer Protein Gene and Risk for Coronary Heart Disease. *Circulation Research*. 2017;121(1):81-88. doi:10.1161/CIRCRESAHA.117.311145

22. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779

23. The All of Us Research Program Investigators. The “All of Us” Research Program. *N Engl J Med*. 2019;381(7):668-676. doi:10.1056/NEJMsr1809937

24. Kanai M, Ulirsch JC, Karjalainen J, et al. Insights from complex trait fine-mapping across diverse populations. *Genetic and Genomic Medicine*. Preprint posted online September 5, 2021. doi:10.1101/2021.09.03.21262975

25. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415-1424. doi:10.1038/s41588-021-00931-x

26. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508-518. doi:10.1038/s41586-022-05473-8

27. The UK Biobank Whole-Genome Sequencing Consortium, Manuscript Writing Group, Carss K, et al. Whole-genome sequencing of 490,640 UK Biobank participants. *Nature*. 2025;645(8081):692-701. doi:10.1038/s41586-025-09272-9

28. The All of Us Research Program Genomics Investigators, Manuscript Writing Group, Bick AG, et al. Genomic data in the All of Us Research Program. *Nature*. 2024;627(8003):340-346. doi:10.1038/s41586-023-06957-x

29. Ishigaki K, Akiyama M, Kanai M, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*. 2020;52(7):669-679. doi:10.1038/s41588-020-0640-3

30. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867-2873. doi:10.1093/bioinformatics/btq559

31. Minikel EV, Karczewski KJ, Martin HC, et al. Evaluating drug targets through human loss-of-function genetic variation. *Nature*. 2020;581(7809):459-464. doi:10.1038/s41586-020-2267-z

32. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

33. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381(6664):eadg7492. doi:10.1126/science.adg7492

34. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097-1103. doi:10.1038/s41588-021-00870-7

35. Gurinovich A, Li M, Leshchyk A, et al. Evaluation of GENESIS, SAIGE, REGENIE and fastGWA-GLMM for genome-wide association studies of binary traits in correlated data. *Front Genet*. 2022;13:897210. doi:10.3389/fgene.2022.897210

36. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Soft*. 2010;36(3):1-48. doi:10.18637/jss.v036.i03

37. Uno H, Claggett B, Tian L, et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. *JCO*. 2014;32(22):2380-2385. doi:10.1200/JCO.2014.55.2208

38. United Nations D of E and SA Population Division. World Population Prospects 2024, Online Edition. Published online 2024. https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=Most%20used

39. Nicholls SJ, Nelson AJ, Ray KK, et al. Impact of Obicetrapib on Major Adverse Cardiovascular Events in High-Risk Patients. *JACC*. 2025;86(14):1046-1056. doi:10.1016/j.jacc.2025.07.056

40. Iha H, Kikutake C, Suyama M. Integrative and accurate annotations enhance current nonsense-mediated mRNA decay rules. *Nucleic Acids Research*. 2025;53(22):gkaf1306. doi:10.1093/nar/gkaf1306

41. Brown ML, Inazu A, Hesler CB, et al. Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins. *Nature*. 1989;342(6248):448-451. doi:10.1038/342448a0

42. Inazu A, Jiang XC, Haraki T, et al. Genetic cholesteryl ester transfer protein deficiency caused by two prevalent mutations as a major determinant of increased levels of high density lipoprotein cholesterol. *Journal of Clinical Investigation*. 1994;94(5):1872-1882. doi:10.1172/jci117537

43. Ford J, Lawson M, Fowler D, et al. Tolerability, pharmacokinetics and pharmacodynamics of TA -8995, a selective cholesteryl ester transfer protein (CETP) inhibitor, in healthy subjects. *Brit J Clinical Pharma*. 2014;78(3):498-508. doi:10.1111/bcp.12380

44. Stahl K, Gola D, König IR. Assessment of Imputation Quality: Comparison of Phasing and Imputation Algorithms in Real Data. *Front Genet*. 2021;12:724037. doi:10.3389/fgene.2021.724037

45. Morris TT, Davies NM, Hemani G, Smith GD. Population phenomena inflate genetic associations of complex social traits. *Sci Adv*. 2020;6(16):eaay0328. doi:10.1126/sciadv.aay0328