

## Supplemental material for:

### Drug-target Mendelian randomization analysis supports lowering plasma ANGPTL3, ANGPTL4, and APOC3 levels as strategies for reducing cardiovascular disease risk

#### Table of contents

DATA SOURCES (STEP 1-2).....	2
<i>Selection of GWAS</i> .....	2
<i>Assessment and diagnostic criteria for diseases</i> .....	2
<i>Assessment and definitions for continuous exposures and outcomes</i> .....	2
DRUG-TARGET MENDELIAN RANDOMIZATION ANALYSES (STEP 1-2) .....	3
<i>Drug-target MR methodology</i> .....	3
<i>Generalized least squares estimator</i> .....	3
<i>Justification of Genetic IVs used in the phenome-wide MR</i> .....	4
<i>Validation of the ANGPTL4 p.E40K coding variants' association with plasma ANGPTL4 protein levels</i> .....	4
SENSITIVITY ANALYSES (STEP 1-2).....	5
<i>Colocalization analyses</i> .....	5
<i>LD matrix sensitivity</i> .....	5
<i>Sample overlap bias</i> .....	5
ANALYSES OF DELETERIOUS VARIANTS IN THE UK BIOBANK (STEP 3-4).....	6
<i>The UK Biobank resource</i> .....	6
<i>Definition of deleterious variants</i> .....	6
<i>Quality control of genetic data</i> .....	6
<i>Genetic mimicry analyses in the United Kingdom Biobank (Step 3)</i> .....	6
<i>Association of deleterious variants with coronary artery disease in the United Kingdom Biobank (Step 4)</i> .	7
<i>Association of deleterious variants with plasma triglycerides, ANGPTL3 and ANGPTL4 protein levels in the United Kingdom Biobank (Step 4)</i> .....	7
<b>SUPPLEMENTAL FIGURE 1</b> .....	<b>9</b>
<b>SUPPLEMENTAL FIGURE 2</b> .....	<b>15</b>
<b>SUPPLEMENTAL FIGURE 3</b> .....	<b>29</b>
<b>SUPPLEMENTAL FIGURE 4</b> .....	<b>30</b>
<b>SUPPLEMENTAL FIGURE 5</b> .....	<b>31</b>
<b>SUPPLEMENTAL FIGURE 6</b> .....	<b>32</b>
<b>REFERENCES</b> .....	<b>33</b>

## SUPPLEMENTAL METHODS

The supplemental methods section is presented below.

### Data sources (Step 1-2)

#### *Selection of GWAS*

The GWAS were predominantly meta-analyses of case-control and population-based cohort studies, including males and females. Genotyping was conducted using different DNA microarrays and exome-sequencing platforms. Our quality control procedures included updating the genomic reference to GRCh38 coordinates, effect and reference allele harmonization, and filtering indels and variants with an alternative allele frequency < 0.1%. We restricted the analyses to GWAS of European ancestry populations to reduce potential bias from population stratification. The exception was the Cystatin C-based estimated glomerular filtration rate (eGFR) GWAS that included a small subset (~16 %) of non-European ancestry individuals (**Table 1**). We chose studies with the largest effective sample sizes for outcomes. Here, one exception was the type 2 diabetes GWAS, where an earlier study (1), was given priority over a later study (2), due to a higher degree of SNP overlap between the exposure and outcome GWAS. For cardiometabolic risk factors, we selected GWAS based on the sample size while also considering if the exposure was measured on a clinically interpretable scale (for example, mmol/L was prioritized over standard deviations [s.d.] when possible). Supplemental Table 1 provides trait definitions, EBI GWAS catalog, and UK biobank trait IDs.

#### *Assessment and diagnostic criteria for diseases*

The primary outcomes of this study were CAD, CKD, ischemic stroke, non-alcoholic fatty liver disease (NAFLD), and T2D. Prevalent or incident CAD was defined according to the international classification of diseases (ICD-9, ICD-10) codes, or hospital registry codes, relating to the diagnosis of, complications, and procedures related to myocardial infarction and chronic coronary syndrome (primarily ICD-10 I20-I25, see reference for details) (3). CKD was defined as hospital inpatient records connected to chronic kidney disease (ICD-10 N18.\*) (4). NAFLD status was defined according to health registry or hospital record ICD-10 codes of fatty liver disease (K76.0), hepatic fibrosis (K74.0 and K74.2), non-alcoholic steatohepatitis (NASH; K75.9), or other specified liver disease (K76.9), while excluding cases with codes related to a wide range of secondary causes of liver disease, for example, alcoholism or viral hepatitis (5). Ascertainment of ischemic stroke in GIGASTROKE included analysis of health registry and hospital records corresponding to any cerebral infarction (ICD-10 codes I63.\*), according to radiographic criteria, and other methods (see reference for further details) (6). T2D definitions included health record diagnoses corresponding to T2D (for example, ICD-9 250.\*0), Hb1Ac or plasma glucose-based, WHO criteria, and physician- or patient-reported T2D (1). The UK Biobank outcomes that may be associated with abdominal lymphadenopathy were derived from hospital in-patient records. Related ICD-10 diagnoses (**Supplemental Table 5**) were grouped into outcomes using Phecode Map 1.2 (7). The FinnGen abdominal lymphadenopathy-associated phenotypes were ascertained through Finnish health registries of lymphadenitis (ICD-10: L04.\*, ICD-8, 9: 683), acute peritonitis (ICD-10: K65.\*, ICD-8, 9: 567), intestinal malabsorption (ICD-10: K90.\*, ICD-9: 579, ICD-8: 269[1-2]), and other noninfective disorders of lymphatic vessels and lymph nodes (ICD-10: I89.\*, ICD-8, 9: 457) (8).

#### *Assessment and definitions for continuous exposures and outcomes*

This study examined the following continuous exposures and outcomes: plasma protein abundance, routine clinical chemistry including plasma lipid concentrations (TGs, LDL-C, HDL-C, ApoB, ApoA1, Lipoprotein a), systolic and diastolic blood pressure (SBP/DBP), anthropometric traits (Body-mass-index [BMI], waist-to-hip ratio [WHR], body fat percentage [BFP]), kidney function (cystatin C and creatinine-based estimated glomerular filtration rate [eGFR]), magnetic resonance imaging (MRI) measurements, and high-throughput nuclear magnetic resonance (NMR) spectroscopy-determined lipid metabolite concentrations. Plasma protein abundance was measured using the SomaScan and Olink platforms (9, 10). These aptamer- and antibody-binding-based platforms report plasma protein abundance in standard deviations (s.d.). The UK biobank quality control and validation procedures of

the routine clinical chemistry tests are summarized in (11, 12). SBP/DBP was measured using manual, automatic, and ambulatory methods, and reported as untransformed millimeters of mercury (mmHg) (13). WHR was defined as the waist measurement divided by hip measurement adjusted by various factors as defined by the GIANT consortium protocol (14). BMI was defined as the body weight divided by squared height ( $\text{kg}/\text{m}^2$ ) (14). BFP was determined using bioelectrical impedance analysis (15). eGFR measurements were derived from natural log-transformed creatinine and cystatin C, using the CKD-EPI or Schwartz formula for creatinine and the Stevens or CKD-EPI formula for cystatin C (16). The Metabolomics had been performed using a commercial NMR spectroscopy-based metabolomics platform from Nightingale Health Ltd., Helsinki, Finland. Over one hundred plasma metabolite parameters are measured on this platform, such as lipoprotein size, subclass lipid contents, and composition, but also free fatty acids, and amino acids (17).

## **Drug-target Mendelian randomization analyses (Step 1-2)**

### ***Drug-target MR methodology***

MR is a method of causal inference that uses genetic variants as instrumental variables (IVs) to test the effect of an exposure on an outcome. By exploiting the random assignment of alleles at conception, MR studies can limit the influence of confounding variables, assuming certain key assumptions are satisfied (18). These assumptions include: 1) the genetic variants are reliably associated with the exposure of interest, 2) they are not associated with the factors that may confound the relationship between the exposure and outcome, and 3) the genetic variants do not influence the outcome, independently of the exposure (18).

Drug-target MR analysis is a specific type of MR used to estimate the causal effect of pharmacological silencing of a target gene (19). It requires an emphasis on assumptions different from those of traditional MR. It is less likely to be influenced by horizontal pleiotropy or reverse causality (for example, when the genetic association with the risk factor is secondary to its effect on the outcome, through feedback mechanisms, or via cross-generational effects (20)). This is because, in drug-target MR, we restrict the selection of genetic IVs to a region surrounding the target gene (2.5 kilobase pairs [Kb] in this paper). Additionally, we assume that genetic information travels from DNA to RNA and protein and not the other way around (following Cricks' central dogma of molecular biology). This means that in drug-target MR, the genetic effect on the outcome is presumed to occur distally to the causal chain of events that starts with a variation (the IV) in the target gene that leads to drug-target exposure (the exposure being changed target gene expression, increased/decreased protein abundance, or altered protein function), making bias from horizontal pleiotropy less probable. Likewise, the influence of horizontal pleiotropy is limited because both the variant and exposure (changed target gene expression, increased/decreased protein abundance, or altered protein function) are located within the causal chain given by Cricks' central dogma ([DNA – RNA – protein] – mediators – outcome), making bias from reverse causality less probable.

### ***Generalized least squares estimator***

This study used a Generalized least squares (GLS) inverse-variance weighted (IVW) estimator to measure the association between ANGPTL3, ANGPTL4, and APOC3 protein abundance and health-related outcomes. This method can achieve a more accurate causal estimate using multiple genetic variations from a single genomic region, even if these variants are correlated via linkage disequilibrium (LD) (21). GLS adjusts the estimated effects and their standard errors to account for between-variant correlations caused by LD. To achieve this, the LD correlation matrix is used as weights. The code snippet (R programming language v.4.2.1) used for the analyses of this manuscript is provided below (modified from (21)).

```
> # run cis GLS-MR according to equations in DOI:10.1002/sim.6835
> bx = mr_harmonized$beta.exposure
> by = mr_harmonized$beta.outcome
> byse = mr_harmonized$standard_error.outcome
> rho = LD_correlation_matrix[mr_harmonized$variant_id, mr_harmonized$variant_id]
```

```

> Omega = byse %o% byse * rho
> results <- data.frame(
>   b = solve(t(bx) %**% solve(Omega) %**% bx) * t(bx) %**% solve(Omega) %**% by,
>   se = sqrt(solve(t(bx) %**% solve(Omega) %**% bx)),
>   pval = NA,
>   n.snps = nrow(mr_harmonized)
> )
> results$pval = with(results, 2 * pnorm(abs(b/se), lower.tail = FALSE))

```

Code snippet legend: 'mr\_harmonized' is the harmonized data holding the variant IDs ('variant\_id'), the association of the variants with the exposure ('beta.exposure'), the association of the variants with the outcome ('beta.outcome') and its associated standard error ('standard\_error.outcome').

### **Justification of Genetic IVs used in the phenome-wide MR**

*ANGPTL3* rs34483103-1:62604866:AGTTAATGTG>A [, c.\*52\_\*60del] is a 3 prime UTR variant associated with reduced circulating ANGPTL3 protein (effect: -0.28 s.d. per allele, 95% CI: -0.30 – -0.27, P: <  $2 \times 10^{-308}$ ) (10). *ANGPTL4* p.E40K is a missense variant (2% allele frequency in Europeans) that hinders ANGPTL4's ability to inhibit LPL by destabilizing it after secretion (22). *ANGPTL4* p.Cys80fs rs746226153-19:8364556:GC>G [p.Cys80ValfsTer12] is predicted to be a loss-of-function variant with high confidence (gnomAD Genome Aggregation Database v.4.0.0, <https://gnomad.broadinstitute.org>). *APOC3* c.55+1G>A is a high-confidence predicted loss-of-function variant (gnomAD v.4.0.0). The *APOC3* c.55+1G>A splice donor loss variant substantially impacted plasma *APOC3* levels (-2.19 s.d. protein =  $3.2 \times 10^{-142}$ , **Table 2**). *APOC3* c.55+1G>A status as a loss-of-function variant is further supported by the fact that the variant's estimated effect on plasma triglycerides was considerable (-0.86 mmol/L, P =  $3.4 \times 10^{-157}$ ; Table 2), and comparable to that of other loss-of-function variants in *APOC3* (23). In addition, the literature reports the *APOC3* c.55+1G>A splice donor loss variant as an *APOC3* loss-of-function variant (24, 25).

### **Validation of the *ANGPTL4* p.E40K coding variants' association with plasma *ANGPTL4* protein levels**

The p.E40K variant was the only variant qualifying as a *cis*-pQTL in the *ANGPTL4* region. Including missense variants in MR studies focusing on protein levels warrants careful consideration due to the possibility of epitope-binding artifacts. The association between *ANGPTL4* p.E40K coding variant carrier status and plasma *ANGPTL4* protein levels was confirmed in a separate study by ELISA using antibodies that were shown by Western blotting to similarly detect wildtype *ANGPTL4* and *ANGPTL4* containing the E40K substitution, indicating that the association was not attributable to epitope-binding artifacts.

The Western blot showing that wild-type human *ANGPTL4* protein and the *ANGPTL4* E40K protein are detected to a similar extent by the antibody used in the *ANGPTL4* Elisa (AF3485, R&D Systems) is shown in **Supplemental Figure 6**.

The association of the *ANGPTL4* p.E40K coding variant with plasma *ANGPTL4* protein levels as measured using the abovementioned *ANGPTL4* Elisa was determined in the Erasmus Rucphen Family study (26). The study included 106 *ANGPTL4* p.E40K carriers and 314 p.E40K non-carriers. The ERF study received approval from the medical ethics board of the Erasmus MC Rotterdam, the Netherlands. The investigations were conducted in compliance with the Declaration of Helsinki. Genotyping was performed using Sequenom iPLEX (MALDITOF, Sequenom Inc. San Diego, USA). To make the scale comparable to the Olink and SomaScan assays, the *ANGPTL4* concentrations measured by the ELISA were log2 transformed and then standardized to have a mean of zero and a standard deviation of one. The association was measured using linear regression, including age and sex as covariates. Having one copy of the p.E40K coding variant was associated with -0.45 s.d. plasma *ANGPTL4* protein (95% CI: -0.66 – -0.23, P =  $4.83 \times 10^{-5}$ ). The association was comparable to those measured by the Olink and SomaScan GWAS and thus confirms that plasma *ANGPTL4* levels are lower in p.E40K carriers, irrespective of the method (antibody or single-stranded DNA aptamers) used to measure the plasma *ANGPTL4* level (**Table 2**).

## Sensitivity analyses (Step 1-2)

### Colocalization analyses

Drug-target MR substantially relies on the assumption that LD (a phenomenon in which neighboring genetic variants are inherited together more frequently than anticipated by chance (27)) does not confound the association between variant and outcome. In cases where there are distinct genetic variants affecting both the exposure and the outcome, but they are connected through LD, there is a risk of making incorrect conclusions (28). To limit this issue, we performed colocalization analyses, which test whether two independent association signals in the same gene region are consistent with having a single shared causal variant (that is, testing if the association signals are ‘colocalized’) (29). The results were presented as the posterior probability, expressed in percentages, of five different hypotheses: H<sub>0</sub>) no genetic association signal for either trait, H<sub>1</sub>) a signal for the exposure but not the outcome, H<sub>2</sub>) a signal for the outcome but not the exposure, H<sub>3</sub>) a signal for both the exposure and outcome, but they do not share a single causal variant, H<sub>4</sub>) the exposure and outcome share the same single causal variant. When interpreting the results, it is important to bear in mind that the colocalization analysis we applied assumes a single causal variant. If there exist several causal variants in the same locus, the estimator will tend to favor H<sub>3</sub>. Prior probabilities were set to  $p_1 = 1 \times 10^{-4}$ ,  $p_2 = 1 \times 10^{-4}$ , and  $p_{12} = 1 \times 10^{-5}$ , meaning that we assumed that the prior probability that any SNP was causal for either trait was 1 in 1000 and that the prior probability of any SNP was causal for both traits was 1 in 10,000.

### LD matrix sensitivity

Generalized least squared (GLS)-corrected inverse-variance weighted (IVW) MR can be a powerful tool to conduct MR with correlated genetic instruments. However, given how GLS works, the method could be sensitive to which LD matrix is used. Therefore, we conducted a sensitivity analysis using an LD matrix derived from 337-thousand British ancestry individuals in the UK biobank (30). We tested for substantial heterogeneity by 1) looking at if the effect directionality was discordant, and 2) calculating a heterogeneity statistic using the formula below (v), where the resulting Z value was used to obtain a P-value from the cumulative distribution function of the normal distribution.

$$(i) \quad Z = \frac{Beta_{1000G} - Beta_{UKB LD}}{\sqrt{SE_{1000G}^2 - SE_{UKB LD}^2}}$$

The analysis indicated that the MR results were not LD matrix sensitive (see **Supplemental Table 3**)

### Sample overlap bias

There was a small-moderate sample overlap between the exposure and outcome GWAS’ (Supplemental Table 4). Assuming weak MR instruments, there will be bias towards a possibly confounded risk factor-outcome relationship if the samples overlap, and bias towards the null if they do not (31). Even though it is less likely that variants positioned close to the protein-encoding gene would violate the relevance assumption by weak instrument bias, we conducted risk of bias analyses. We assumed a worst-case scenario in which all the participants of the exposure GWAS were also included in the outcome GWAS. F-statistics were calculated to determine instrument strength for each MR model.

Formula (i) was used for the single-variant MRs, where  $Beta_{exposure}$  was the genetic association with the exposure, and  $SE_{exposure}$  was the corresponding standard error.

$$(ii) \quad F = \left( \frac{Beta_{exposure}}{SE_{exposure}} \right)^2$$

For multiple variant MRs, another formula was used (ii), where  $N_{sample\ size}$  denotes the sample size and  $N_{variants}$  denote the number of variants in the MR model.

$$(iii) \quad F = \left( \frac{N_{sample\ size} - N_{variants} - 1}{N_{variants}} \right) \left( \frac{R^2}{1 - R^2} \right)$$

The  $R^2$  in (ii) was approximated using formula (iii), where ‘MAF’ denotes the minor allele frequency.

$$(iv) \quad R^2 = 2 \text{Beta}_{exposure}^2 \text{MAF}(1 - \text{MAF})$$

Where the  $R^2$  was summed together for uncorrelated instruments in (ii).

For correlated instruments, another formula (iv) was used to approximate the  $R^2$ , where  $B_X^T$  was the transposed vector of the genetic associations with the exposure, and  $C$  was the inverse of the LD correlation matrix after Cholesky decomposition, as suggested by Burgess and Thompson (32).

$$(v) \quad R^2 = \sum 2(B_X^T C)^2 \text{MAF}(1 - \text{MAF})$$

The analyses were run using three scenarios of bias of the observational estimate (olsbias) using the code originally written by (31). These scenarios were ‘low olsbias’, ‘medium olsbias’, and ‘high olsbias’, where the effect magnitude of the biased observational estimate was 10%, 50%, and 100% of the MR estimate, respectively.

The analysis indicated that the risk of bias from sample overlap was minimal (see **Supplemental Table 4**).

## **Analyses of deleterious variants in the UK Biobank (Step 3-4)**

### ***The UK Biobank resource***

The UK Biobank is a population-based study with a vast repository of research data comprising health record information, genetic information, blood chemistry, and imaging data from approximately 500,000 participants in the United Kingdom (33). Between 2006 and 2010, participants aged 40-69 were recruited, and their blood was drawn upon recruitment. The UK Biobank follows participants, and health information is continuously updated from health records, death certificates, and registries.

### ***Definition of deleterious variants***

Variants from the population variant call files were annotated with gnomAD v.4.0.0. Deleterious variants were defined as protein-truncating variants and UK Biobank allele frequency  $<0.05$  and missense variants with REVEL pathogenicity prediction score  $\geq 0.25$  and UK Biobank allele frequency  $<0.00025$ . Protein-truncating variants included Variant Effect Predictor high impact consequences "transcript\_ablation", "splice\_acceptor\_variant", "splice\_donor\_variant", "stop\_gained", "frameshift\_variant", "stop\_lost", "start\_lost", "transcript\_amplification", "feature\_elongation", and "feature\_truncation". The resulting variants classified as deleterious are listed in the **Supplemental Table 10**.

### ***Quality control of genetic data***

We applied quality control filters on the UK Biobank exome sequencing population variant call files (UKB data field 23157). First, to avoid batch effects, we filtered variants that did not meet genotype depth of coverage (DP)  $\geq 10$  or were marked as missing for more than 90% of the population. For single nucleotide variants (SNVs), we then filtered variants with DP  $< 10$ , genotype quality (GQ)  $< 20$ , and binomial test of alternate allele departure from the heterozygous expectation of 0.5  $P > 1 \times 10^{-3}$ . For insertions and deletions (indels), we applied filters to exclude variants with DP  $< 10$  and GQ  $< 20$ .

### ***Genetic mimicry analyses in the United Kingdom Biobank (Step 3)***

We conducted genetic mimicry analyses to determine whether the metabolic effects of deleterious and common variants adjacent to *ANGPTL3*, *ANGPTL4*, and *APOC3* were metabolically concordant. For the deleterious variant models, we selected unrelated individuals of British ancestry with nuclear magnetic resonance spectroscopy (NMR) data available (UKB category 220). Ratios derived from the NMR measurements were excluded. The total number of unique metabolites amounted to 167, and the total cohort encompassed 181,672 individuals. Left-censored NMR measurements were imputed using minimum value imputation. Measurements missing due to citrate peaks, degraded sample, high ethanol, isopropyl alcohol contamination, low glutamine or high glutamate, ethanol, polysaccharide, or unknown contamination were excluded from further analyses. Each metabolite was standardized to have a mean of zero and a standard deviation of one. The effect of each deleterious variant model on the metabolites was estimated using linear regression adjusting for age, sex, age \* sex interaction, fasting time at sampling, and the five first genetic principal components (UKB data fields 21022, 22001, 74, and 22009). Each association was scaled for its 1-s.d. effect on plasma triglycerides to ease comparability with the common variant models.

For the common variant models, we included variants within 2.5 Kb of the target genes transcription start site and transcription termination sites, with an allele frequency > 0.01, significantly associated with plasma triglycerides at the genome-wide significance level ( $P < 5 \times 10^{-8}$ ). Genetic associations between these variants and the NMR metabolites (category: 'met-d-...' ) were extracted through the MRC-IEU open GWAS database (<https://gwas.mrcieu.ac.uk>). The effect of the common variants on each metabolite per 1-s.d. change of plasma triglycerides was estimated through drug-target MR of the 167 NMR metabolites using triglyceride levels as the exposure, and single metabolites as the outcome. This means the effect on each metabolite is given per 1-s.d. triglyceride change. Finally, the metabolic concordance between the deleterious and common-variant models was estimated using linear regression on the association effect estimates ('betas'). The concordance metric was reported as the coefficient of determination ( $R^2$ ).

#### ***Association of deleterious variants with coronary artery disease in the United Kingdom Biobank (Step 4)***

We conducted gene-collapsing analyses in the UK Biobank to measure the association between *ANGPTL3*, *ANGPTL4*, and *APOC3* deleterious variants and coronary artery disease (CAD). Diagnosis codes were retrieved from the death register, hospital inpatient health records, algorithmically defined myocardial infarction outcomes, and first occurrences data sets (UKB data fields 131271–131423, 40001, 40002, 41202, 41270, 42001, 42003, 42005). We selected non-related British ancestry UK Biobank participants. Cases were defined as positive for ICD-10 diagnosis codes related to ischemic heart disease (I20-I25), only excluding subjects with evidence for non-obstructive ischemic heart disease. The excluded diagnoses were: I20.1: Angina pectoris with documented spasm, I24.8: Other forms of acute ischemic heart disease (which includes I24.81: Acute coronary microvascular dysfunction), and I25.4: Coronary artery aneurysm and dissection. Controls were defined as participants without a positive diagnosis for any phenotype listed in the ICD-10 I00-I99: Diseases of the circulatory system chapter. This strategy was implemented to minimize the risk of contamination resulting from correlated or genetically related diagnoses. After applying these filters, the total UK Biobank CAD cohort amounted to 42,145 cases and 137,689 controls. Of these 179,834 individuals, there were 1169 participants with deleterious *ANGPTL3* variants, 761 with deleterious *ANGPTL4* variants, and 912 with deleterious *APOC3* variants. The association of each deleterious variant model with CAD was measured using logistic regression, adjusting for age, sex, age \* sex interaction, and the five first genetic principal components. The associations were reported as the odds ratio per 1-allele.

#### ***Association of deleterious variants with plasma triglycerides, ANGPTL3 and ANGPTL4 protein levels in the United Kingdom Biobank (Step 4)***

We measured the genetic association of deleterious variants in *ANGPTL3*, *ANGPTL4*, and *APOC3* with plasma triglycerides (UKB data field 30870). We also measured the association of deleterious *ANGPTL3* variants with *ANGPTL3* protein levels (UKB data field olink\_instance\_0.angptl3), and the

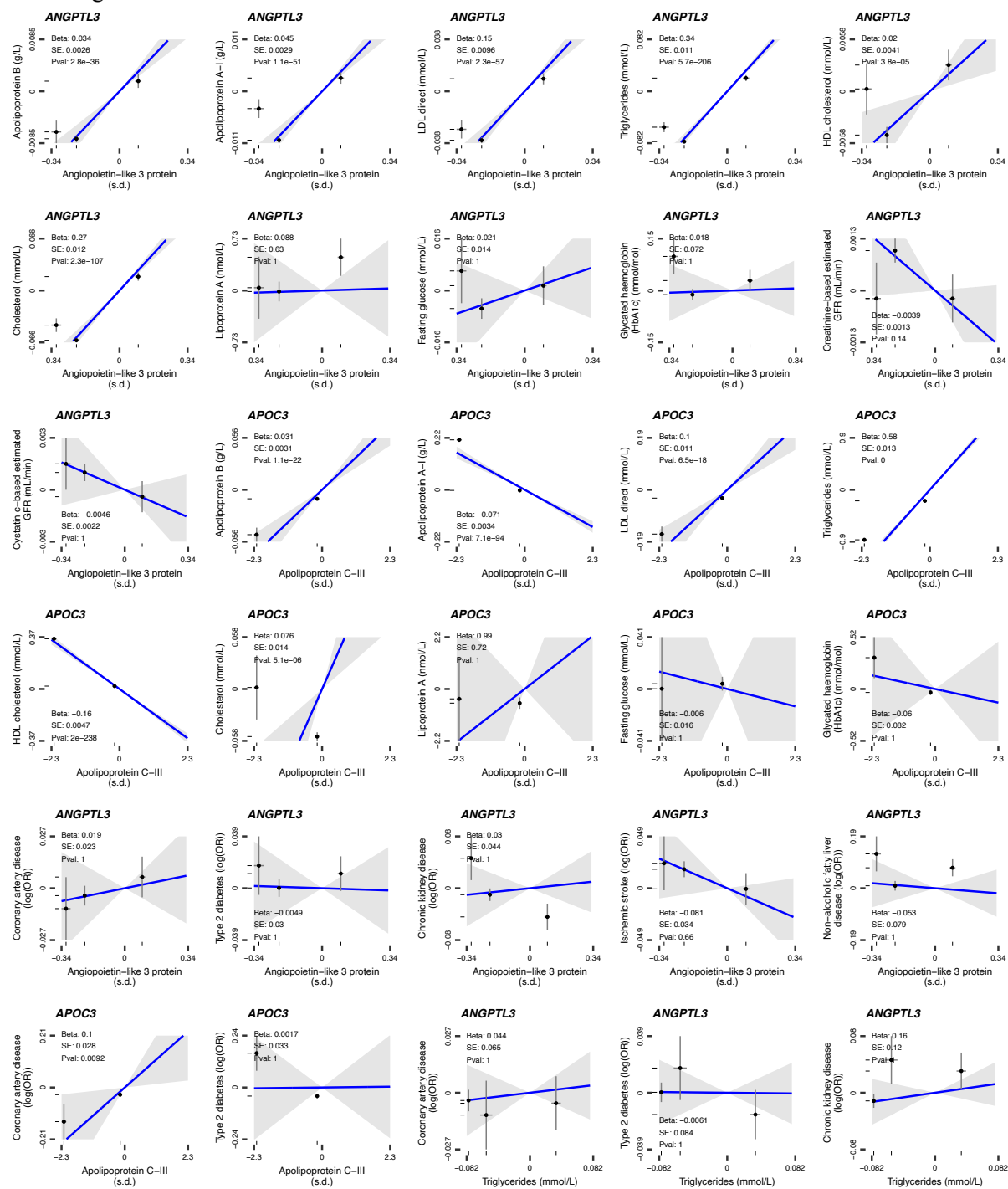
association of deleterious *ANGPTL4* variants with *ANGPTL4* protein levels (UKB data field `olink_instance_0.angptl4`). *APOC3* plasma protein levels were not measured in the UK Biobank. The association analyses were conducted in the control group of the CAD analyses described in the section above. This was for two reasons. First, it restricts bias from post-event measurements affecting the gene-triglycerides or gene-protein associations. Second, the over-recruitment of cases can affect the distribution of confounders, which might distort the gene-triglycerides or gene-protein associations.

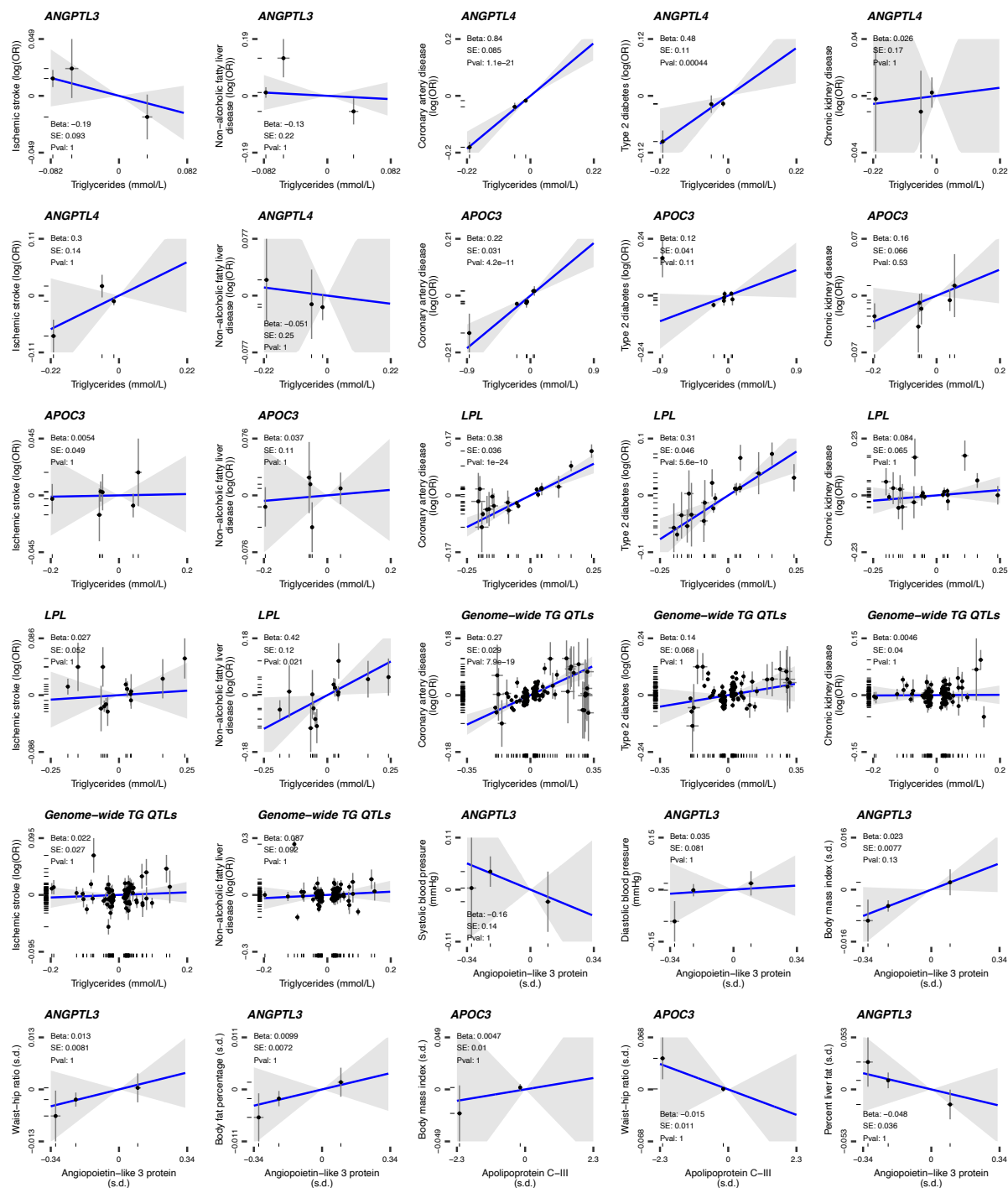
The *ANGPTL3* and *ANGPTL4* protein concentrations, measured as Olink  $\log_2$  normalized protein expression (NPX), were scaled to have a mean of zero and a standard deviation of one. In the control group, there were 131,292 individuals with available plasma triglycerides measurements, of which 886 carried deleterious *ANGPTL3* variants, 601 carried deleterious *ANGPTL4* variants, and 676 carried deleterious *APOC3* variants. There were 13,662 in the control group with available plasma *ANGPTL3* protein measurements, of which 93 had a deleterious *ANGPTL3* variant. There were 13,727 with available plasma *ANGPTL4* protein measurements in the control group, of which 77 had a deleterious *ANGPTL4* variant. The associations between the deleterious variants and plasma triglycerides, *ANGPTL3* or *ANGPTL4* protein levels were determined using linear regression adjusting for age, sex, age \* sex interaction, fasting time at sampling, and the five first genetic principal components. For triglycerides, the associations were reported as the 1-mmol/L change per 1-allele. For *ANGPTL3* and *ANGPTL4* plasma protein concentrations, the associations were reported as the 1-s.d. change per 1-allele.

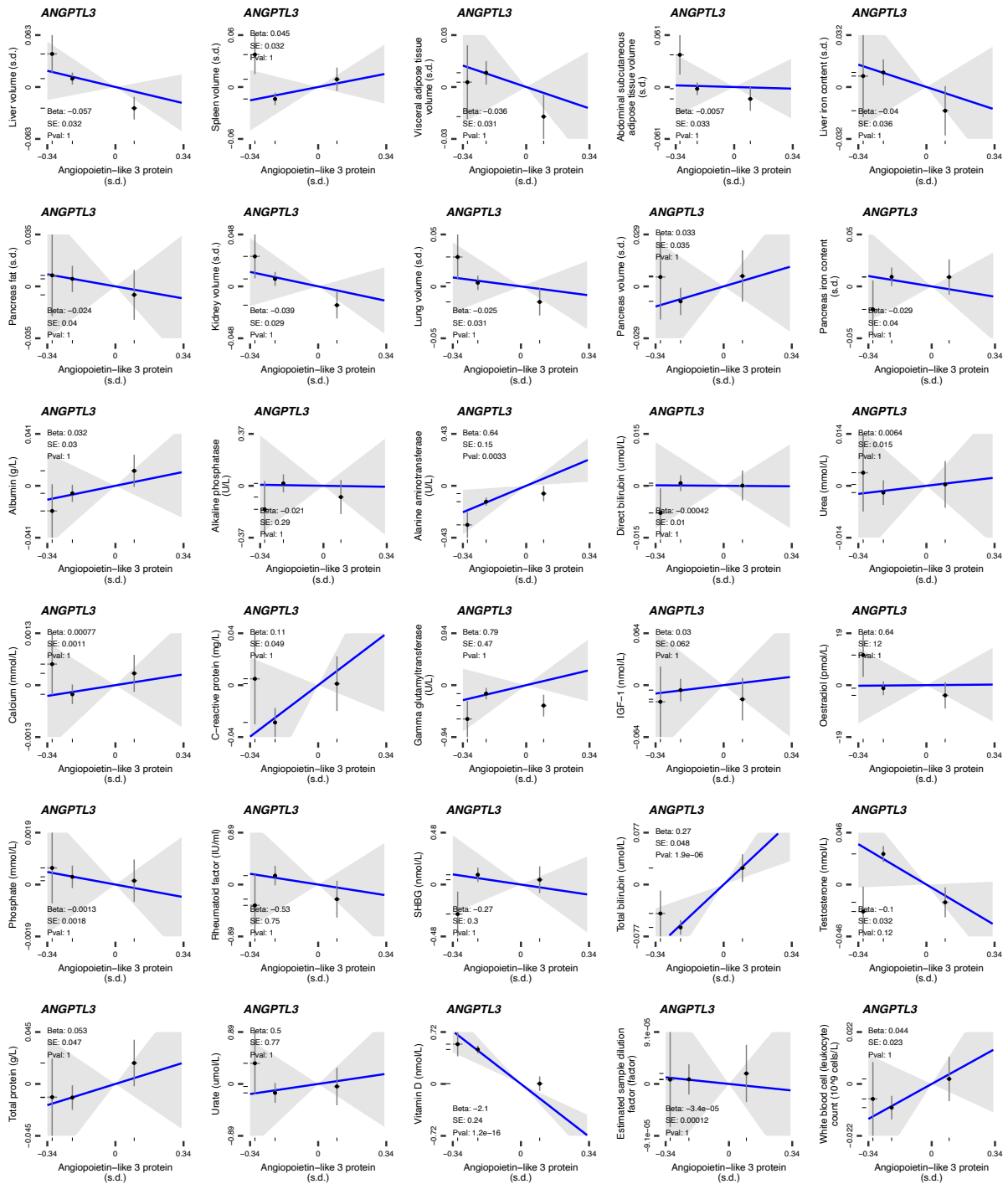


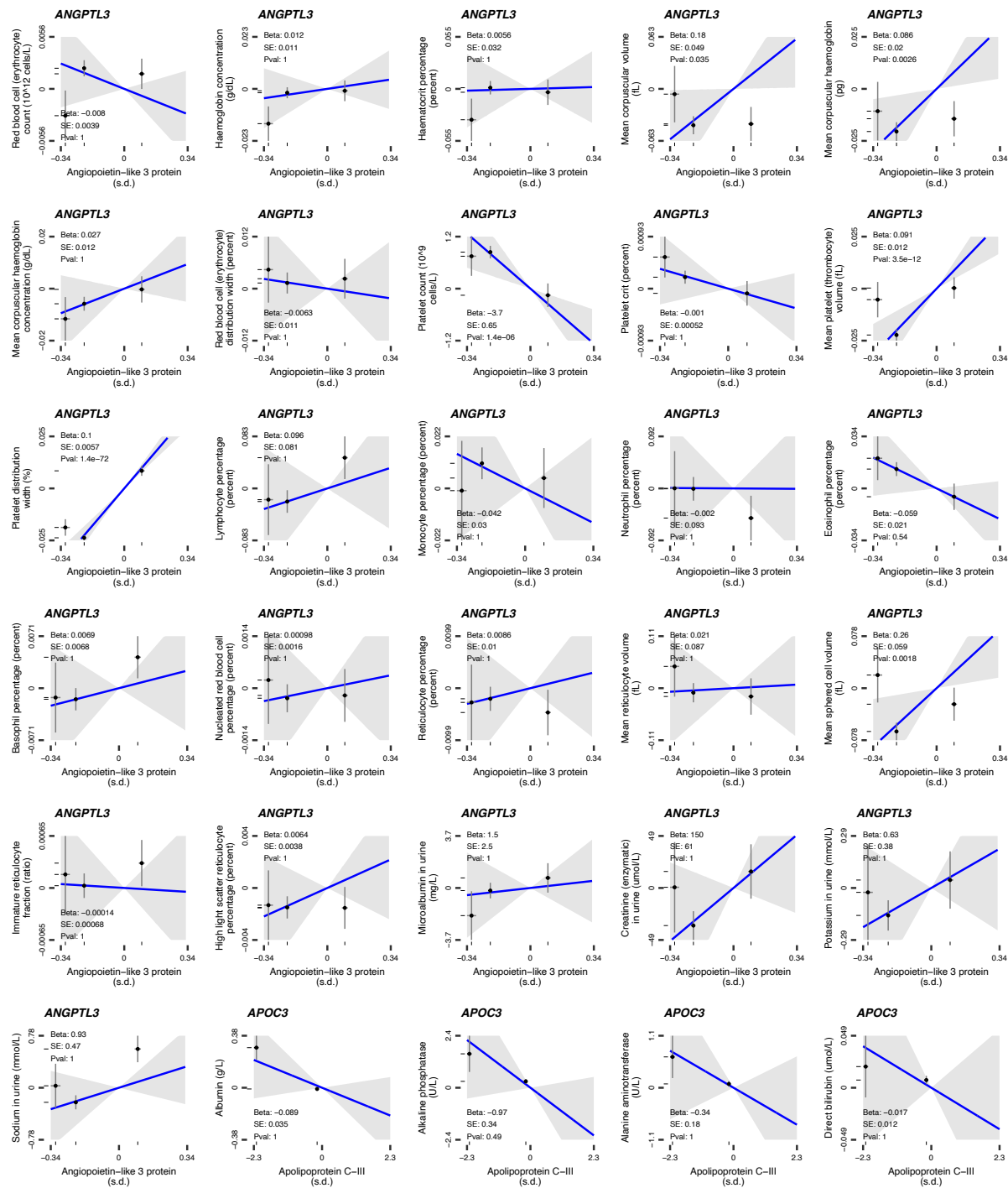
## SUPPLEMENTAL FIGURE 1

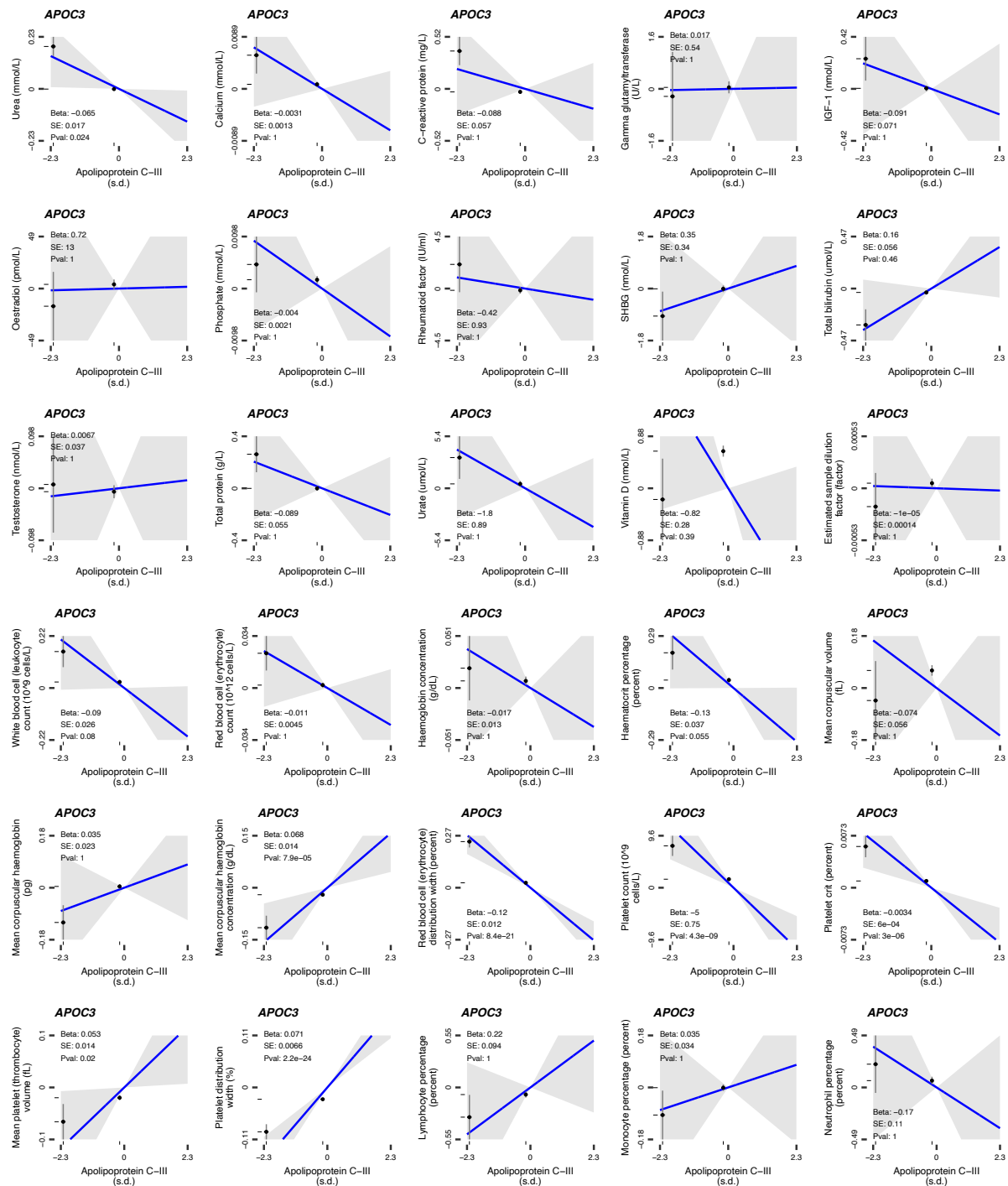
**Figure S1. Scatter plots showing the results of the drug-target MR analyses.** Each subplot represents the results of the analyses displayed in **Figures 2-4** of the main manuscript that used more than one genetic instrument.

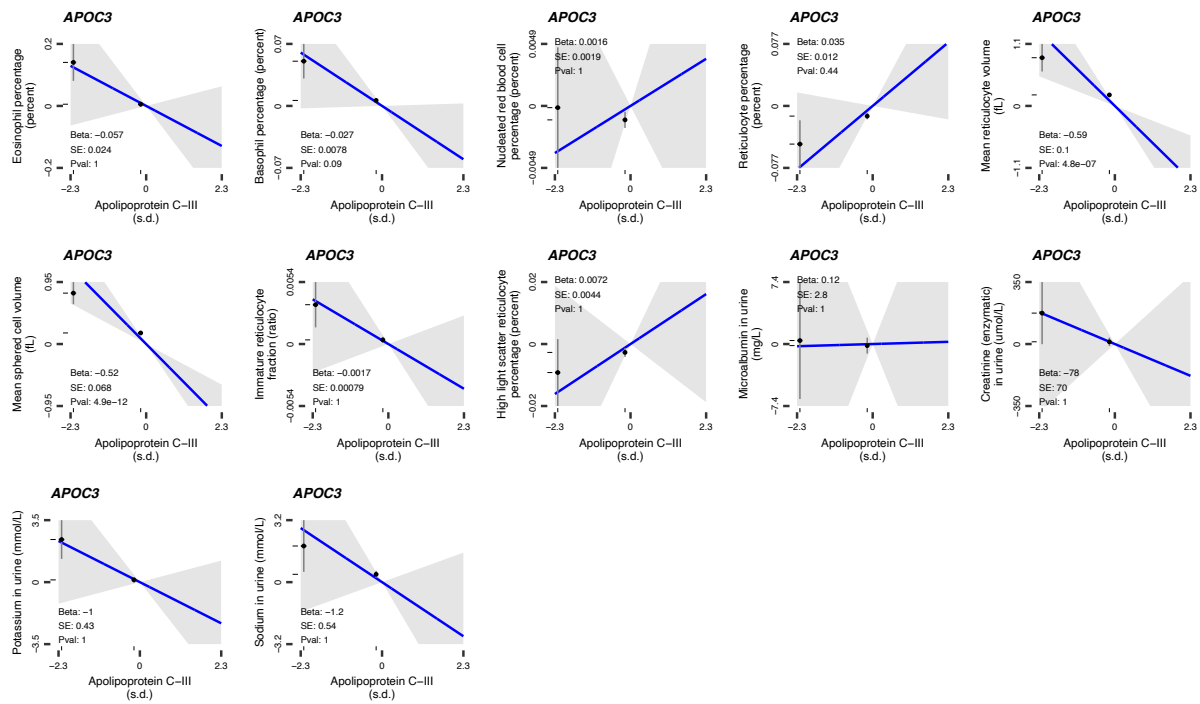








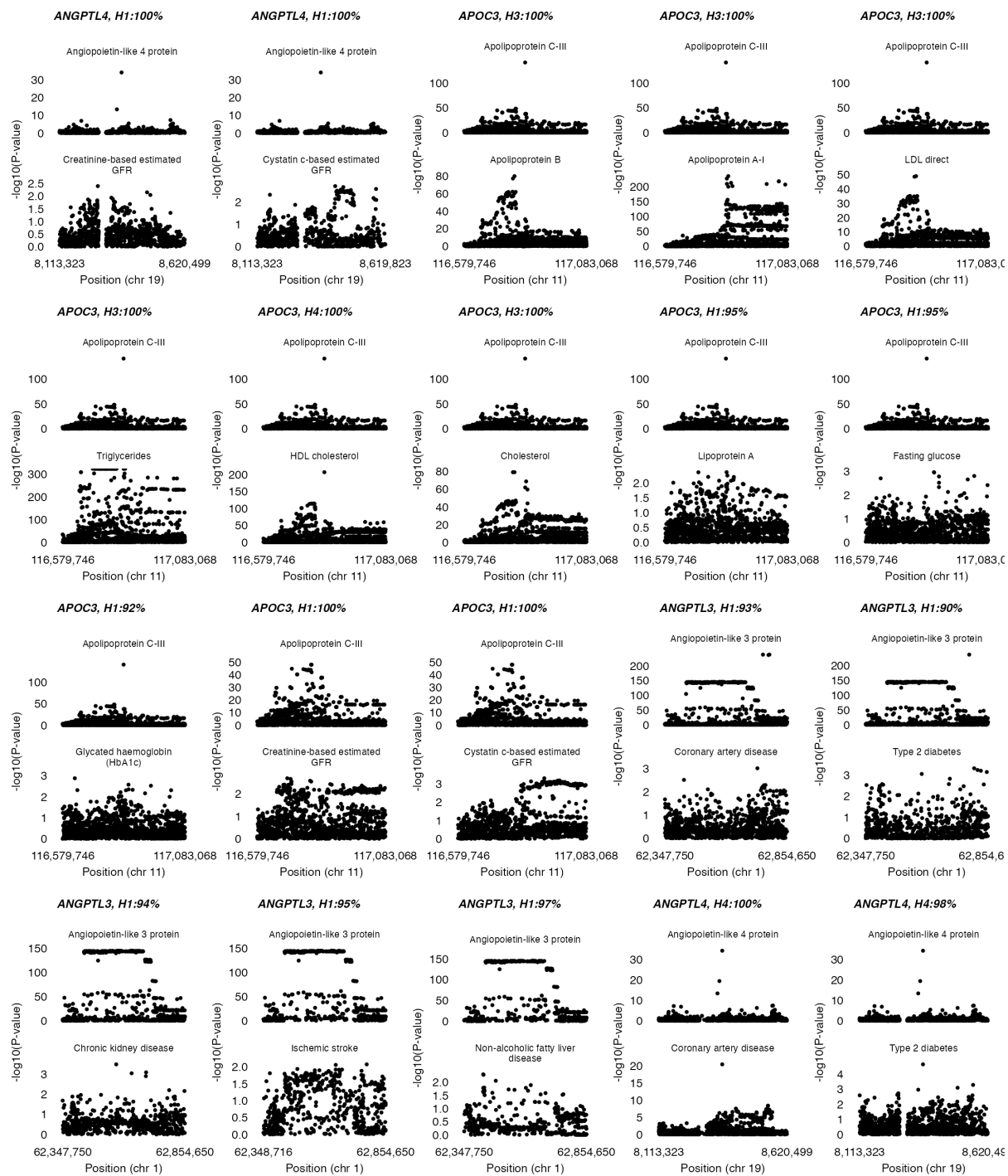




## SUPPLEMENTAL FIGURE 2

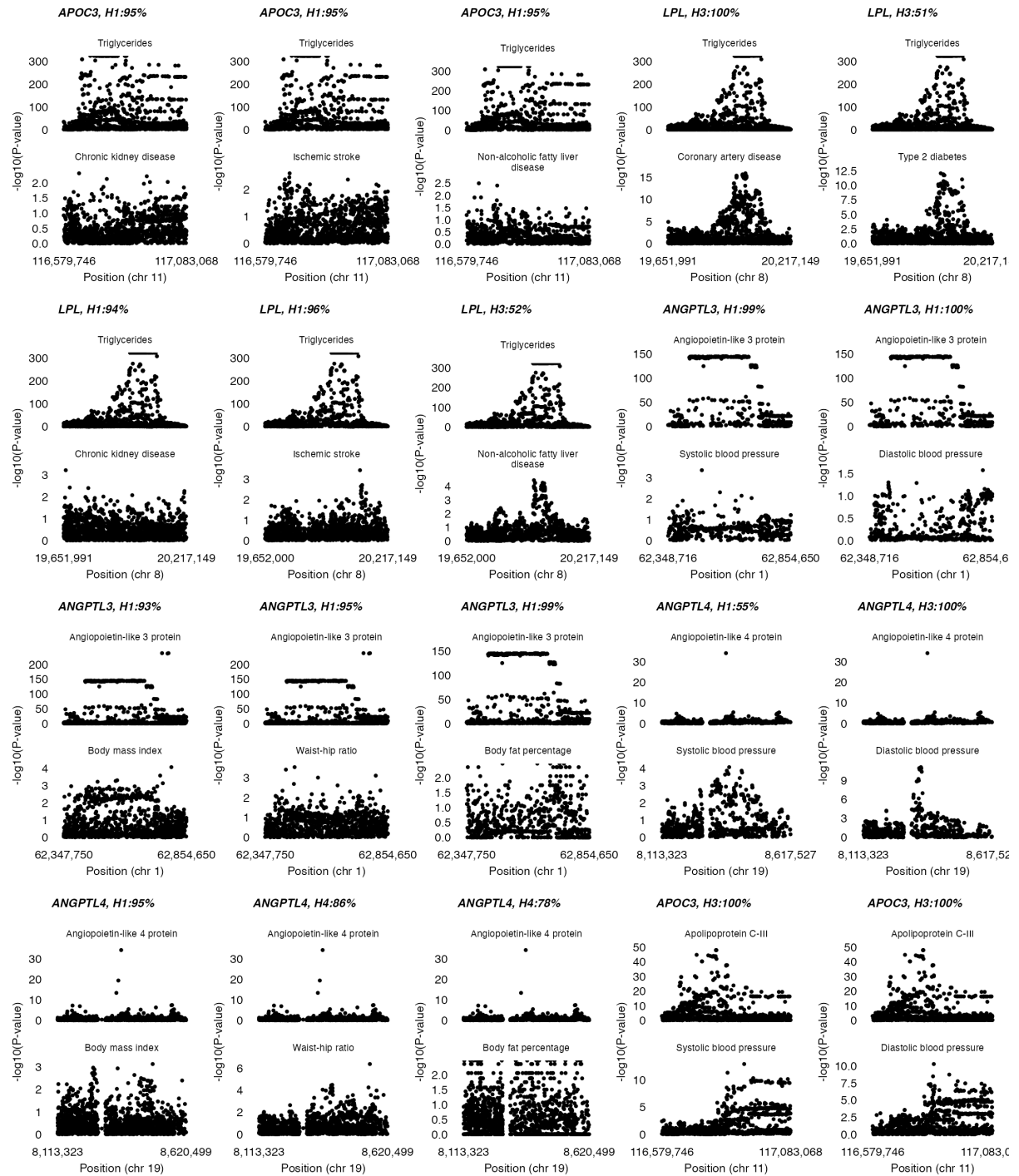
**Figure S2. Regional genetic association plots showing the results of the colocalization analyses.** Each subplot represents the results of the analyses displayed in Figure 2-4 of the main manuscript.



















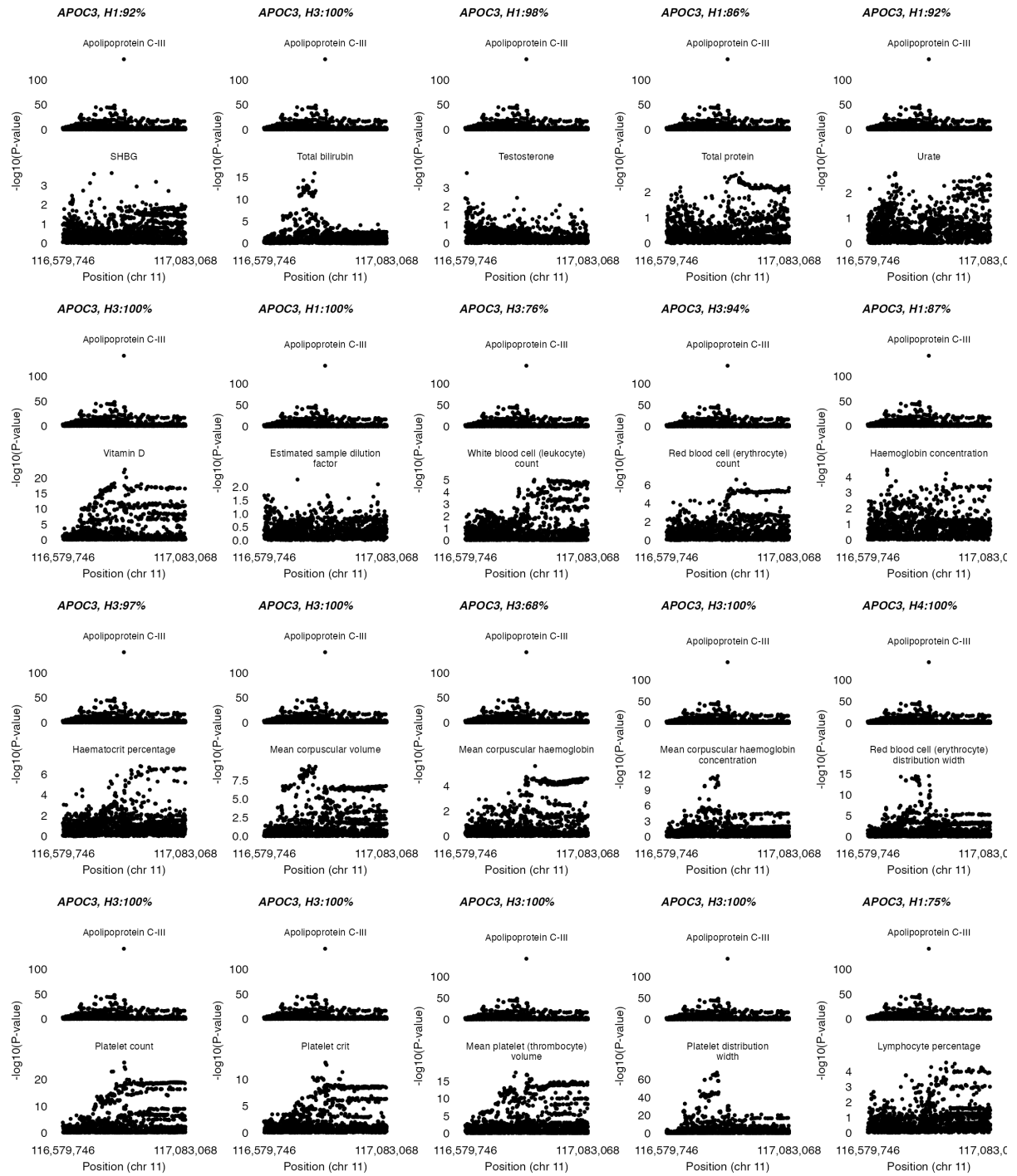


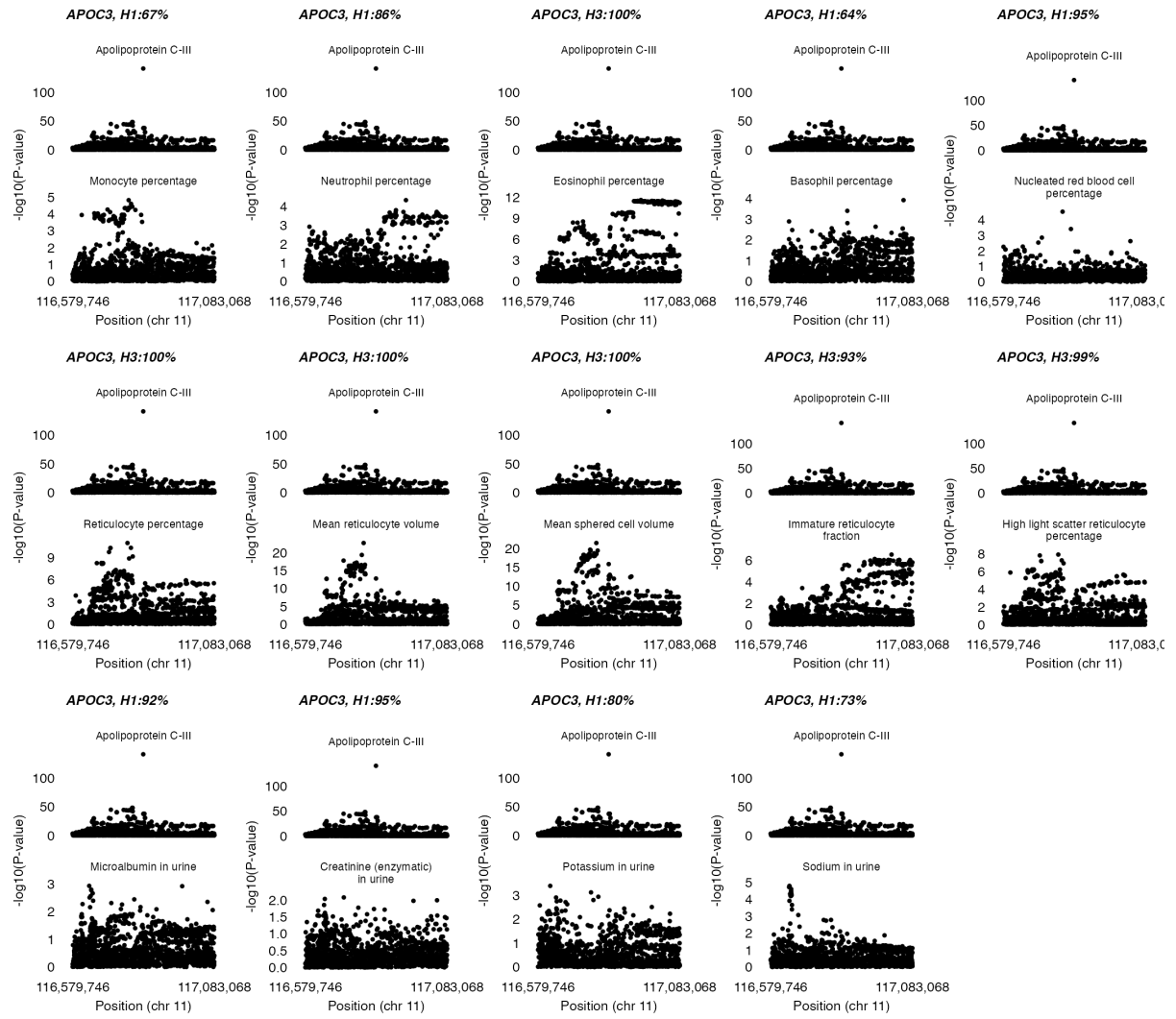






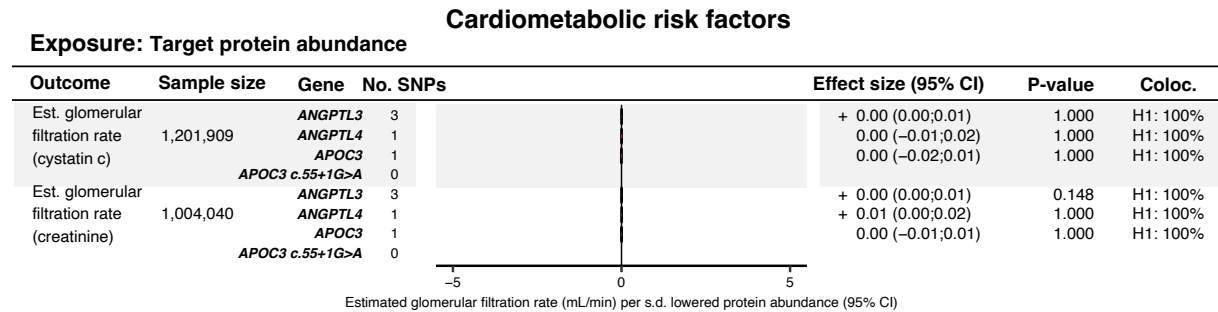






### SUPPLEMENTAL FIGURE 3

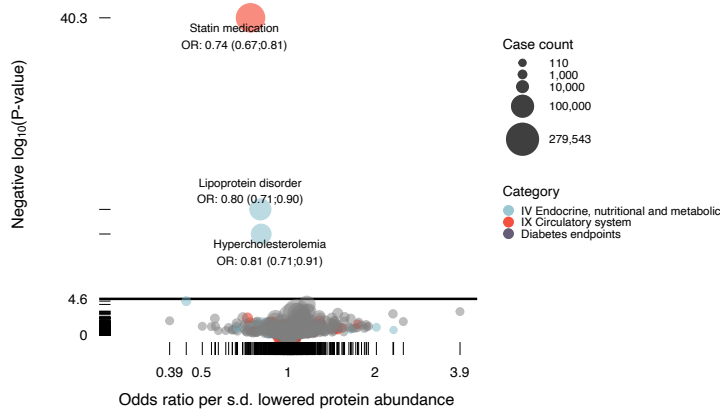
**Figure S3. Results of cis-pQTL MR analyses of the estimated glomerular filtration rate by (eGFR) by Cystatin C and plasma Creatinine.** The findings are displayed in bar graphs, illustrating the level of the effect per s.d. decrease in protein abundance. The red lines represent the 95% CI.



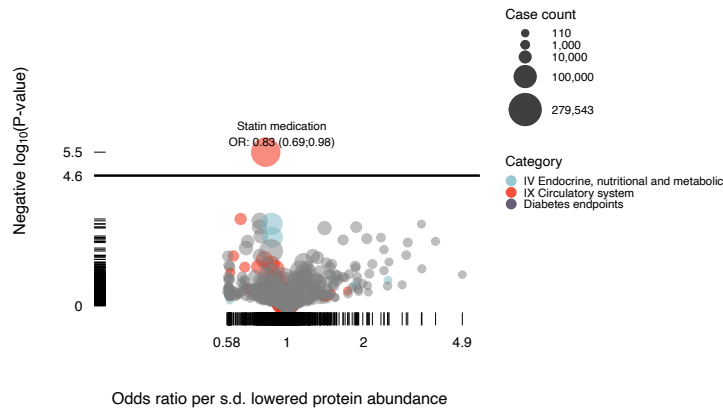
## SUPPLEMENTAL FIGURE 4

**Figure S4. Results of phenome-wide MR analysis on FinnGen outcomes using *ANGPTL3* *rs34483103-1:62604866:AGTTAATGTG>A* [3 prime UTR, c.\*52\_\*60del] and *APOC3* *rs138326449-11:116830638:G>A* [donor loss, c.55+1G>A] variants. **A:** Volcano plot displaying the results of *ANGPTL3* cis-pQTL phenome-wide MR. **B:** *APOC3* cis-pQTL phenome-wide MR volcano plot. The y-axis solid straight lines indicate the phenome-wide significance threshold. ‘OR’ indicates the odds ratio with 95% confidence intervals with Bonferroni correction for the 694 FinnGen outcomes.**

### A. *ANGPTL3* *rs34483103-1:62604866:AGTTAATGTG>A* [3 prime UTR, c.\*52\_\*60del]

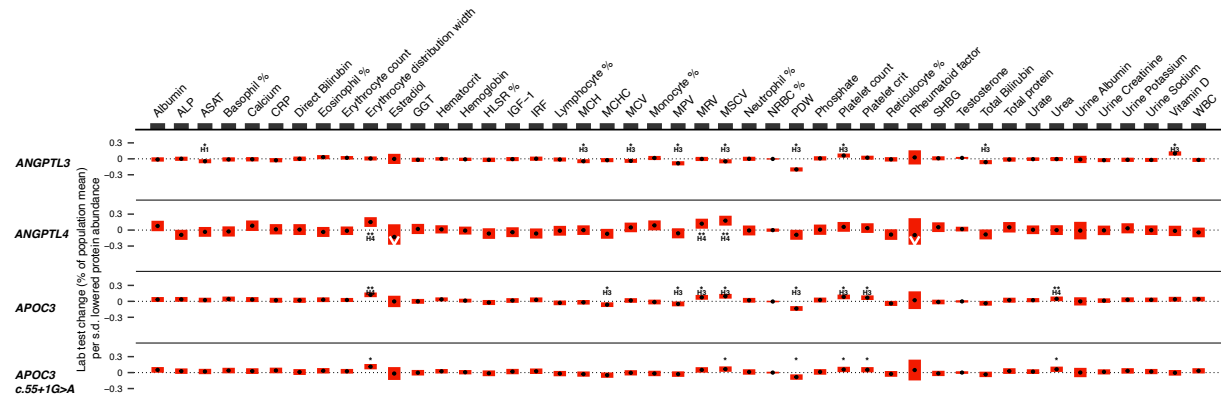


### B. *APOC3* *rs138326449-11:116830638:G>A* [donor loss, c.55+1G>A]



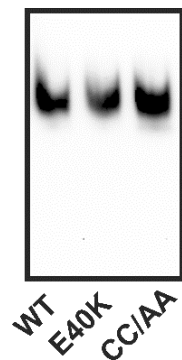
## SUPPLEMENTAL FIGURE 5

**Figure S5. Cis-pQTL MR of the clinical laboratory outcomes, showing the results on a 1-s.d. scale.** The red bars indicate the 95 % CI. The black dots indicate the effect point estimate. ‘\*’ indicate  $P < 0.05$ . ‘\*\*’ indicates  $P < 0.05$  with a shared causal variant ( $H_4$ ). A list explaining the abbreviations is provided in the supplemental material (**Supplemental Table 1**).



**SUPPLEMENTAL FIGURE 6**

**Figure S6. Wildtype human ANGPTL4 protein and the ANGPTL4 E40K variant protein are detected to a similar extent by the antibody used in the human ANGPTL4 ELISA.** Western blot of the precipitated medium of HEK293 cells transfected with expression vectors for different mutant forms of human ANGPTL4 fused to a V5-tag (34). CC/AA is an oligomerization defective ANGPTL4 variant. SDS-PAGE was performed using a loading buffer without DTT or other disulfide-reducing agent. Equal amounts of medium were loaded. Membranes were blotted with an antibody from R&D Systems that is used for the human ANGPTL4 Elisa (AF3485, 1:2500). Secondary antibody was goat anti-rabbit IgG conjugated to HRP (1:5000).





## REFERENCES

1. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;50(11):1505-13.
2. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet.* 2022;54(5):560-72.
3. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet.* 2022;54(12):1803-15.
4. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51(6):957-72.
5. Ghodsian N, Abner E, Emdin CA, Gobeil É, Taba N, Haas ME, et al. Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non-alcoholic fatty liver disease. *Cell Rep Med.* 2021;2(11):100437.
6. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature.* 2022;611(7934):115-23.
7. Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform.* 2019;7(4):e14325.
8. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature.* 2023;613(7944):508-18.
9. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrismiddottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* 2021;53(12):1712-21.
10. Dhindsa RS, Burren OS, Sun BB, Prins BP, Matelska D, Wheeler E, et al. Rare variant associations with plasma protein levels in the UK Biobank. *Nature.* 2023;622(7982):339-47.
11. Elliott P, Peakman TC, Biobank U. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol.* 2008;37(2):234-44.
12. Gaye A, Peakman T, Tobin MD, Burton PR. Understanding the impact of pre-analytic variation in haematological and clinical chemistry analytes on the power of association studies. *Int J Epidemiol.* 2014;43(5):1633-44.
13. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet.* 2018;50(10):1412-25.
14. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet.* 2019;28(1):166-74.

15. Franssen FM, Rutten EP, Groenen MT, Vanfleteren LE, Wouters EF, Spruit MA. New reference values for body composition by bioelectrical impedance analysis in the general population: results from the UK Biobank. *J Am Med Dir Assoc*. 2014;15(6):448.e1-6.
16. Stanzick KJ, Li Y, Schlosser P, Gorski M, Wuttke M, Thomas LF, et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat Commun*. 2021;12(1):4350.
17. Ritchie SC, Surendran P, Karthikeyan S, Lambert SA, Bolton T, Pennells L, et al. Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. *medRxiv*. 2021:2021.09.24.21264079.
18. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.
19. Schmidt AF, Finan C, Gordillo-Marañón M, Asselbergs FW, Freitag DF, Patel RS, et al. Genetic drug target validation using Mendelian randomisation. *Nat Commun*. 2020;11(1):3255.
20. Burgess S, Swanson SA, Labrecque JA. Are Mendelian randomization investigations immune from bias due to reverse causation? *Eur J Epidemiol*. 2021;36(3):253-7.
21. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med*. 2016;35(11):1880-906.
22. Yin W, Romeo S, Chang S, Grishin NV, Hobbs HH, Cohen JC. Genetic variation in ANGPTL4 provides insights into protein processing and function. *J Biol Chem*. 2009;284(19):13213-22.
23. Saleheen D, Natarajan P, Armean IM, Zhao W, Rasheed A, Khetarpal SA, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature*. 2017;544(7649):235-9.
24. Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med*. 2014;371(1):32-41.
25. Tg, Hdl Working Group of the Exome Sequencing Project NHL, Blood I, Crosby J, Peloso GM, Auer PL, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371(1):22-31.
26. Henneman P, Aulchenko YS, Frants RR, van Dijk KW, Oostra BA, van Duijn CM. Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: the Erasmus Rucphen Family study. *J Med Genet*. 2008;45(9):572-7.
27. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9(6):477-85.
28. Zuber V, Grinberg NF, Gill D, Manipur I, Slob EAW, Patel A, et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am J Hum Genet*. 2022;109(5):767-82.
29. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383.
30. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat Genet*. 2020;52(12):1355-63.

31. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* 2016;40(7):597-608.
32. Burgess S, & Thompson, S.G. *Mendelian Randomization: Methods for Causal Inference Using Genetic Variants* (2nd ed.): Chapman and Hall/CRC; 2021.
33. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.
34. Jonker JT, Smit JW, Hammer S, Snel M, van der Meer RW, Lamb HJ, et al. Dietary modulation of plasma angiotensin-like protein 4 concentrations in healthy volunteers and in patients with type 2 diabetes. *Am J Clin Nutr.* 2013;97(2):255-60.