

Supplementary Material for: “Assessing the efficacy and safety of angiotensinogen inhibition using human genetics.”

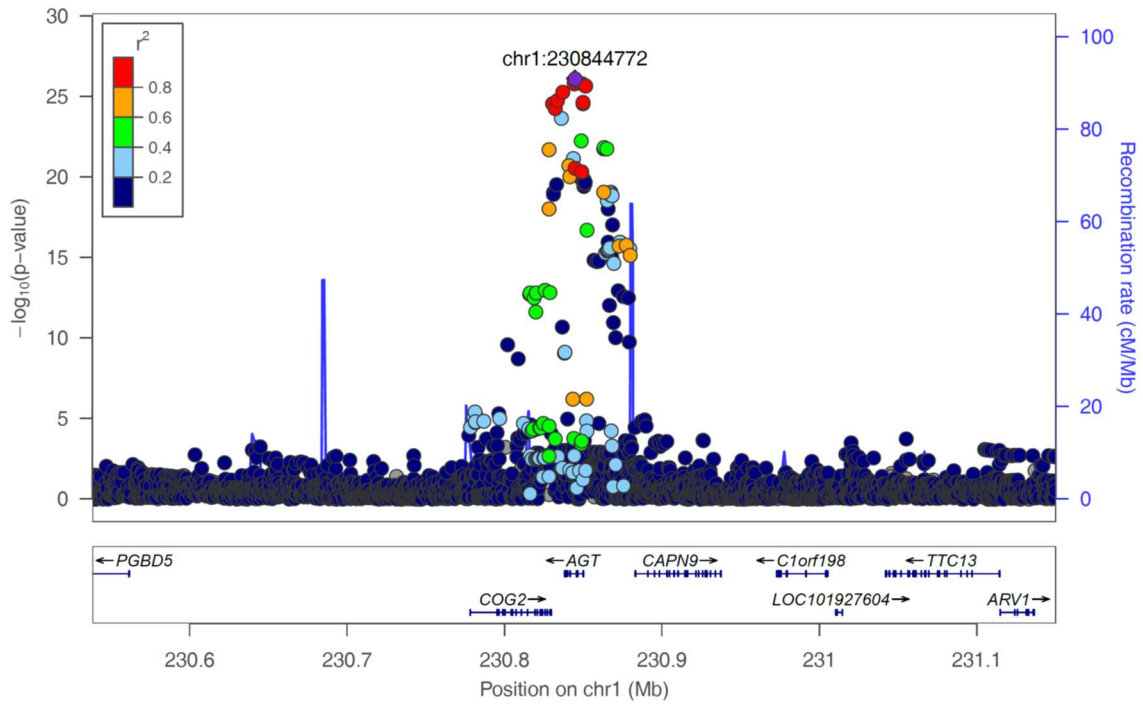


Figure S1. Regional association plot of *AGT* locus in a GWAS of SBP in UKBB. One independent genome-wide significant ($P < 5 \times 10^{-8}$) variant was detected (rs2478539; indicated as a purple point) and selected as an instrument for *AGT* inhibition. Each plotted point represents a genetic variant; colours of the plotted points indicate the linkage disequilibrium (in r^2) of each variant with the lead variant (i.e. rs2478539). The y-axis indicates the $-\log_{10}(P\text{-value})$ for association with SBP) for each plotted point, with chromosomal coordinates (and gene locations) shown on the x-axis. GWAS data were sourced from Gazal *et al.*¹

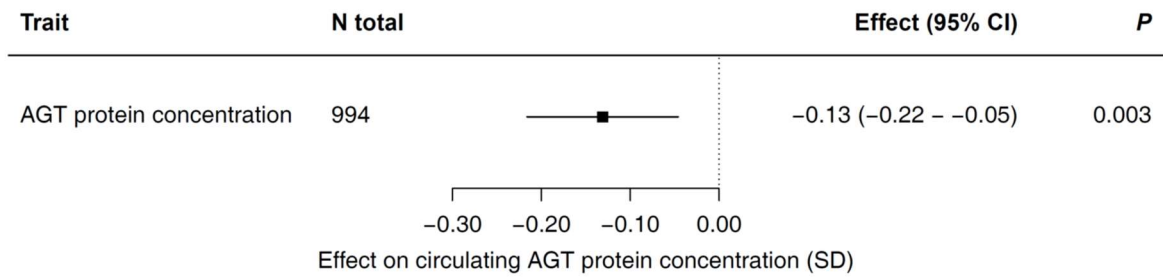


Figure S2. Per-allele association of the selected AGT instrument (rs2478539) with circulating AGT protein concentration. Data are plotted in standard deviation (SD) units and effect is per SBP-lowering allele. The box represents the point estimate of effect. Line represents 95% confidence interval (CI). Data sourced from Suhre *et al.*²

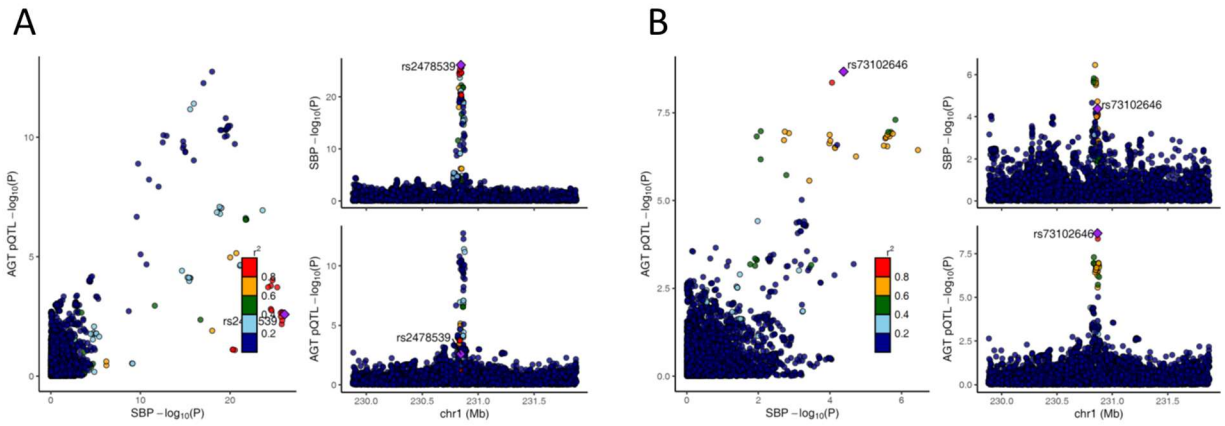


Figure S3. Colocalization of AGT protein quantitative trait loci (pQTL) data and SBP. LocusCompare plots showing the strength of association of genetic variants in the *AGT* region with AGT protein concentration and SBP. (A) Colocalization performed using unconditioned datasets. (B) Colocalization performed with data conditioned on the effect of the lead SBP-associated variant (rs2478539; pQTL data could not be conditioned - see methods for details). Each plotted point represents a genetic variant; colours of the plotted points indicate the linkage disequilibrium (in r^2) of each variant with the lead SBP-associated variant (A) or the lead pQTL (B) (indicated as a purple point). Posterior probabilities of colocalization are provided in Table S9.

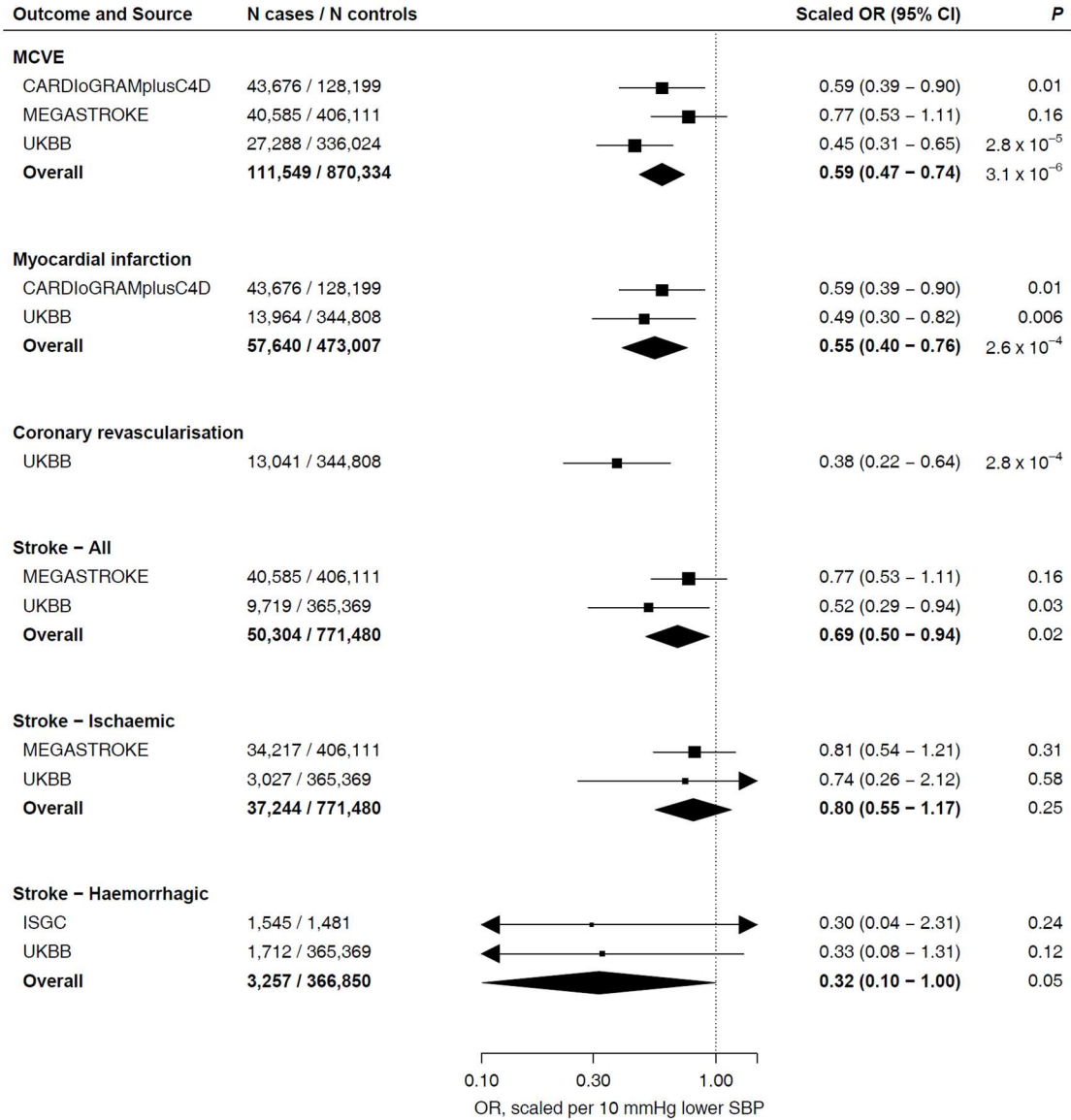


Figure S4. Scaled meta-analysis associations of the AGT instrument (rs2478539) with vascular outcomes. Estimates are scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. MCVE represents “Major Cardiovascular Events”, a composite of myocardial infarction, coronary revascularisation, and all stroke. Estimates were derived in UK Biobank (UKBB) and from various GWAS summary statistics datasets, where participants in such datasets did not overlap with UKBB. Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI). See Table S13 for the source of each GWAS dataset. CARDIoGRAMplusC4D, Coronary Artery Disease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics; ISGC, International Stroke Genetics Consortium; OR, odds ratio.

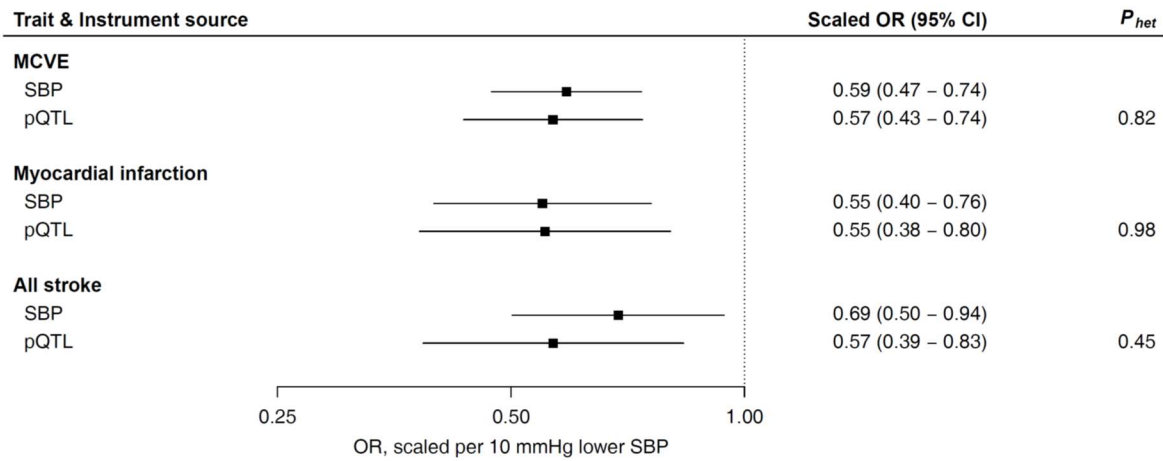


Figure S5. Associations of different AGT instruments with major outcomes of interest. Estimates are scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. “Instrument source” refers to the primary dataset used to select the instrument used in each analysis; SBP refers to the instrument selected from SBP GWAS data, and is used elsewhere in this paper (rs2478539); pQTL refers to an instrument selected from AGT protein concentration genetic association data (rs73102646). Estimates were scaled using the SBP estimate of each instrument. P_{het} refers to P -value from Cochran’s Q test for heterogeneity between the SBP and pQTL estimates. Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI).

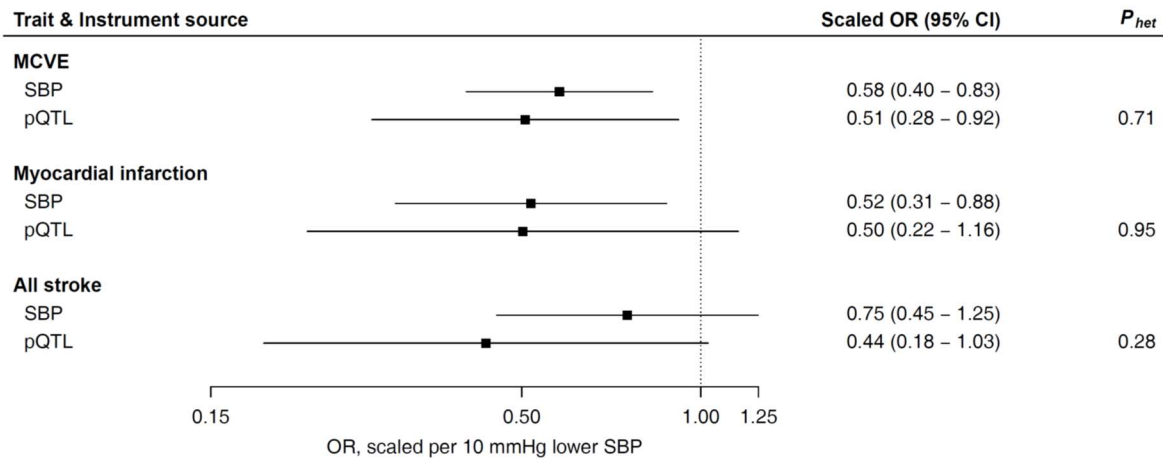


Figure S6. Conditional associations of different AGT instruments with major outcomes of interest Estimates are scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. “Instrument source” refers to the primary dataset used to select the instrument used in each analysis; SBP refers to the instrument selected from SBP GWAS data, and is used elsewhere in this paper (rs2478539); pQTL refers to an instrument selected from AGT protein concentration genetic association data (rs73102646). The association of the instrument with each outcome is conditional on the other instrument. Conditional estimates were derived using GCTA-COJO for summary-level data, and using SNPTEST for UKBB data. The estimates were scaled using the conditional SBP estimate of each instrument. P_{het} refers to P -value from Cochran’s Q test for heterogeneity between the SBP and pQTL estimates. Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI).

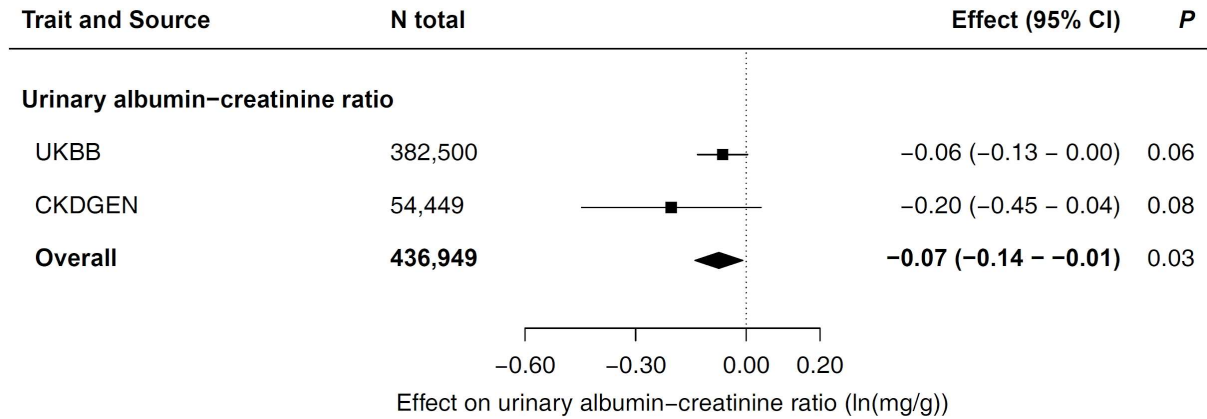


Figure S7. Scaled meta-analysis associations of the AGT instrument with urinary albumin-creatinine ratio. Estimates are in natural logarithm(mg/g) units, scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. UKBB refers to the estimate derived in UK Biobank³ and CKDGEN refers to the estimate derived in the Chronic Kidney Disease Genetics consortium.⁴ Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI).

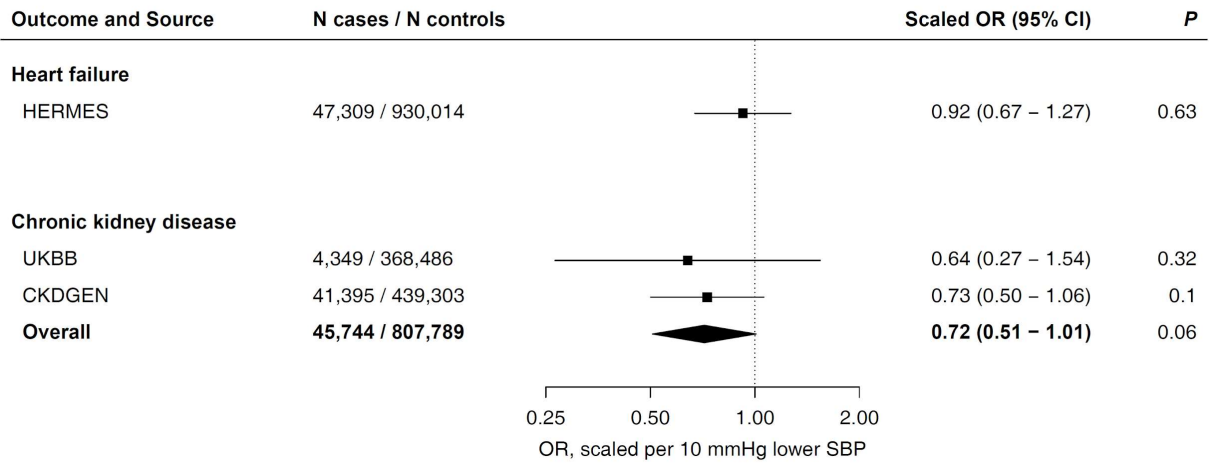


Figure S8. Scaled meta-analysis associations of the AGT instrument with heart failure and chronic kidney disease. Estimates are scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. Estimates were derived in UK Biobank (UKBB) and from GWAS summary statistics from the Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES)⁵ and Chronic Kidney Disease Genetics (CKDGEN)⁶ consortia. Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI).

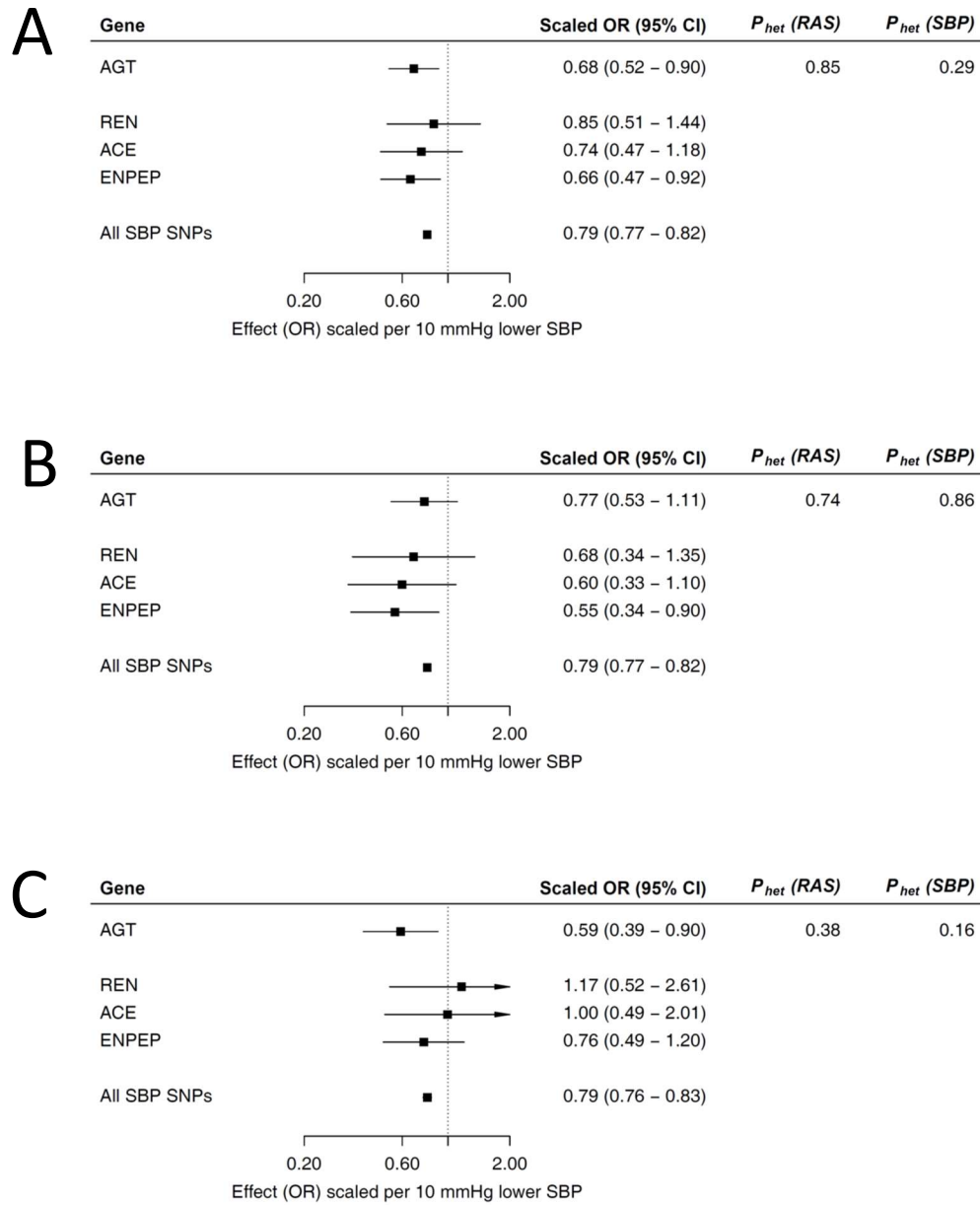


Figure S9. Comparison of scaled associations of AGT, RAS pathway genes and a genome-wide SBP-instrument, restricted to the same data sources. Scaled estimates of SBP-lowering instruments for *AGT*, *REN*, *ACE*, *ENPEP* and an “all SBP-SNPs” instrument (up to 364 SBP-associated variants and excluding any variants correlated with RAS pathway instruments) with risk of (A) major cardiovascular events (MCVE), (B) all stroke, (C) MI, derived using data from CARDIoGRAMplusC4D and MEGASTROKE consortia only. Estimates are scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. P_{het} (RAS) refers to P -value from Cochran’s Q test for heterogeneity between the RAS pathway estimates (i.e. between *AGT*, *REN*, *ACE*, and *ENPEP*). P_{het} (SBP) refers to P -value from Cochran’s Q test for heterogeneity between the *AGT* estimate and the SBP-instrument estimate. Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI). OR, odds ratio; SNPs, single-nucleotide polymorphisms.

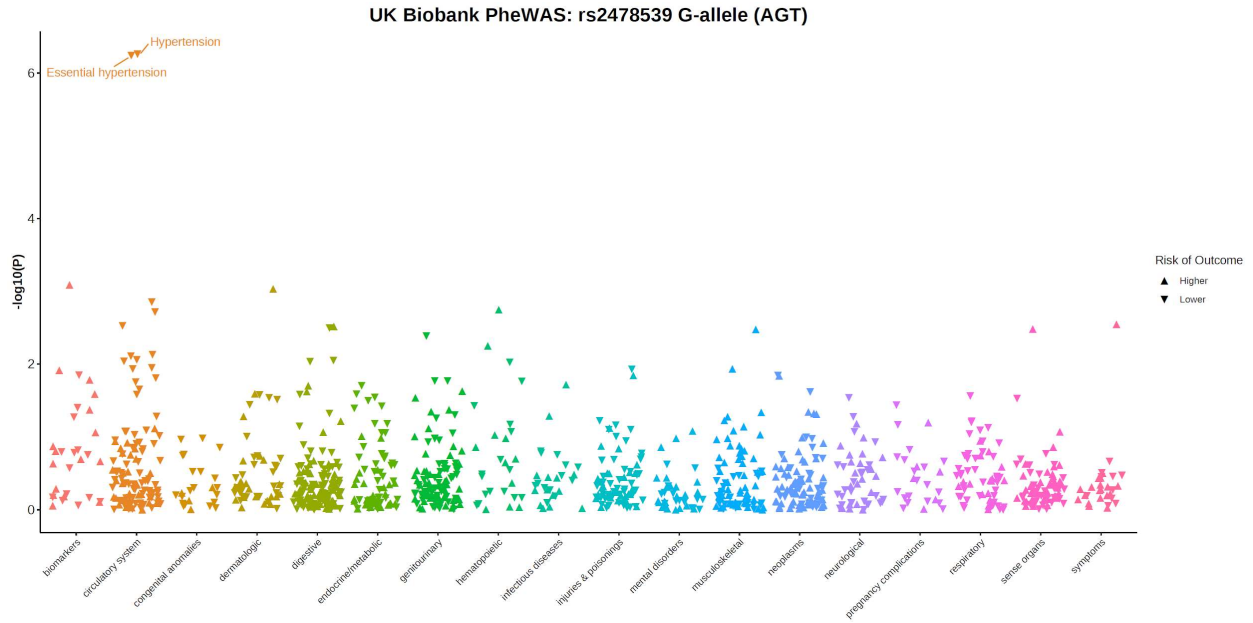


Figure S10. Phenome-wide association analysis of the AGT instrument with 1,074 binary outcomes in UK Biobank (UKBB). Each triangle on the plot refers to a binary outcome in UKBB, with outcomes grouped by phenotypic categories. Outcomes associated at a <5% false-discovery rate threshold are labelled. Direction of effect (aligned to the SBP-lowering allele) for each association is indicated by the direction in which the triangle is pointing (i.e. lower or higher risk).

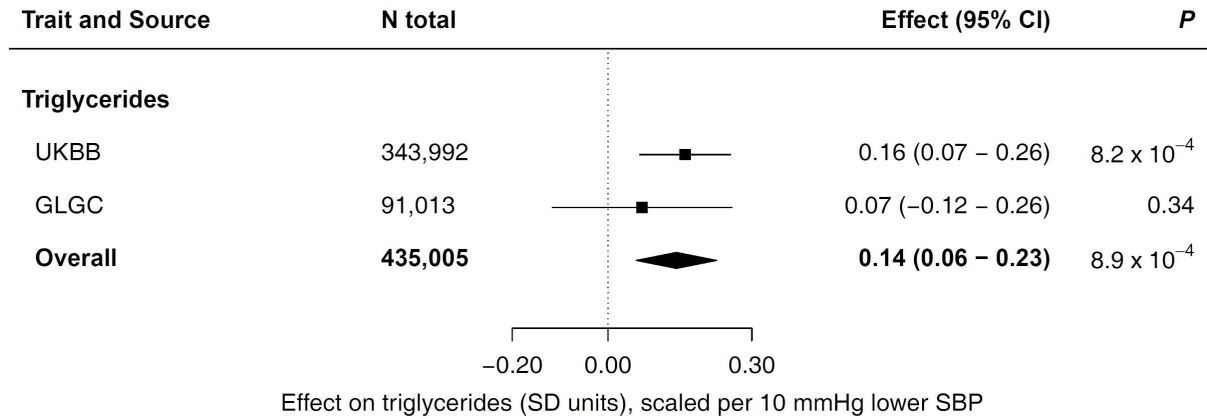
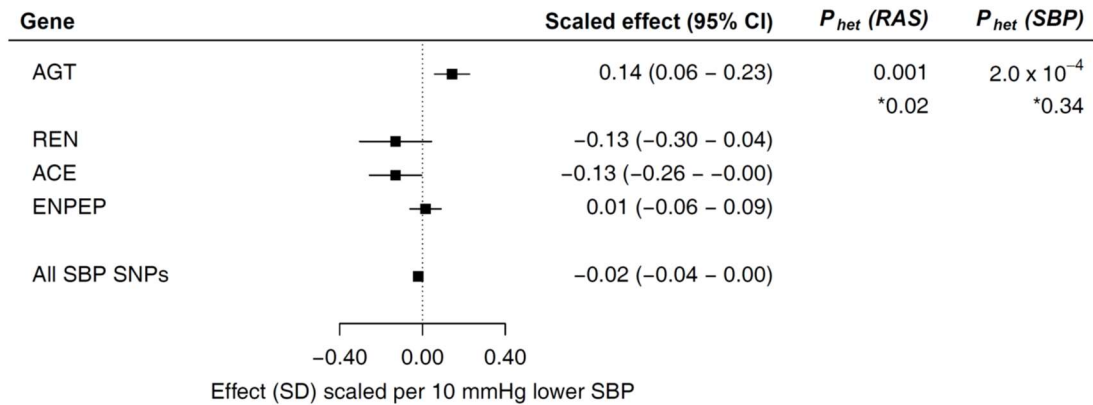


Figure S11. Scaled meta-analysis associations of the AGT instrument with triglyceride concentration. Estimates are in standard deviation (SD) units, scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. UKBB refers to the estimate derived in UK Biobank and GLGC refers to the estimate derived in the Global Lipid Genetics Consortium.⁷ Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI).

A



B

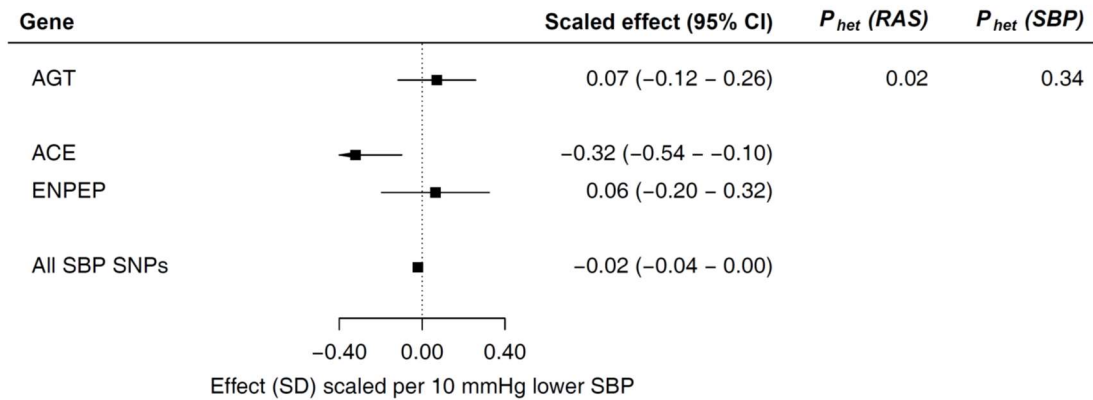


Figure S12. Comparison of scaled associations of AGT, RAS pathway genes and a genome-wide SBP-instrument, with triglyceride concentration. Scaled estimates of SBP-lowering instruments for *AGT*, *REN*, *ACE*, *ENPEP* and an “all SBP-SNPs” instrument (up to 273 SBP-associated variants and excluding any variants correlated with RAS pathway instruments) with triglyceride concentration (TG), using (A) all available data, and (B) using data from the GLGC consortium only. Estimates are scaled to a 10 mmHg lower SBP and aligned to the SBP-lowering allele. P_{het} (RAS) refers to P -value from Cochran’s Q test for heterogeneity between the RAS pathway estimates (i.e. between *AGT*, *REN*, *ACE*, and *ENPEP*). P_{het} (SBP) refers to P -value from Cochran’s Q test for heterogeneity between the *AGT* estimate and the SBP-instrument estimate. Boxes represent point estimates of effects. Lines represent 95% confidence intervals (CI). OR, odds ratio; SNPs, single-nucleotide polymorphisms.

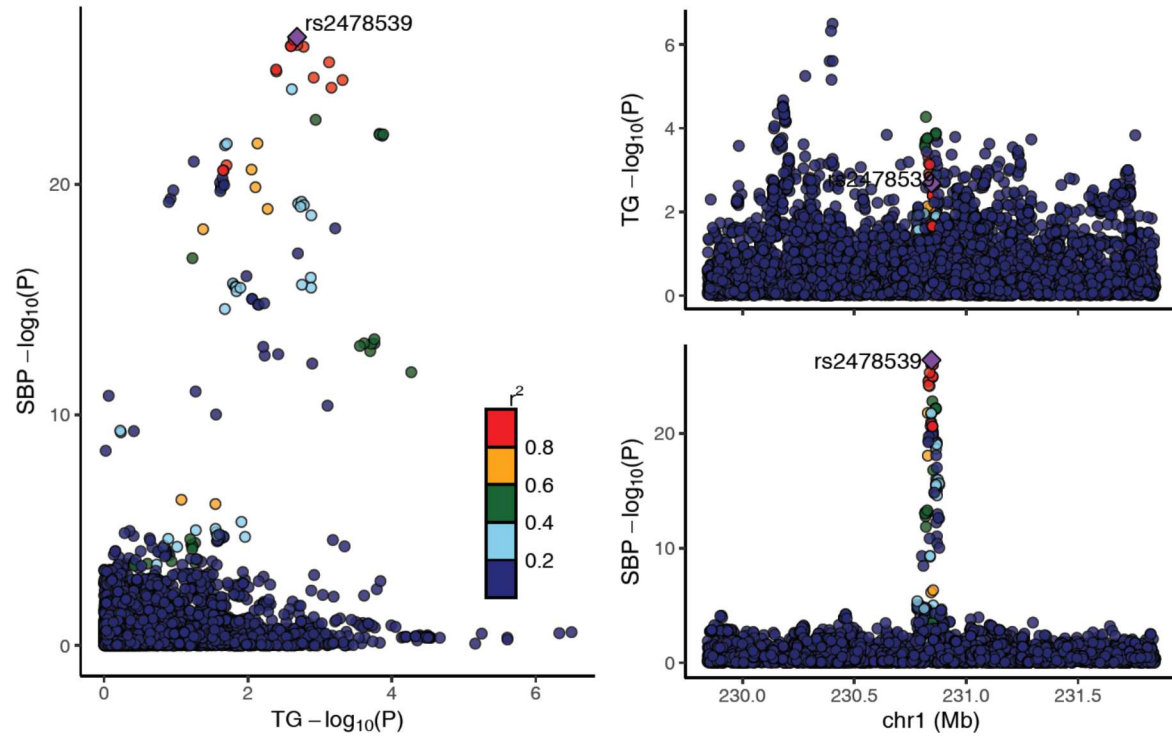


Figure S13. Colocalization of serum triglyceride concentration (TG) and SBP. LocusCompare plots showing the strength of association of genetic variants in the *AGT* region with TG and SBP. Both datasets were conditioned on nearby independent TG-associated (at $P < 5 \times 10^{-8}$) variants (rs11122450, rs149481992 and rs7529464). Each plotted point represents a genetic variant; colours of the plotted points indicate the linkage disequilibrium (in r^2) of each variant with the lead SBP-associated variant (indicated as a purple point).

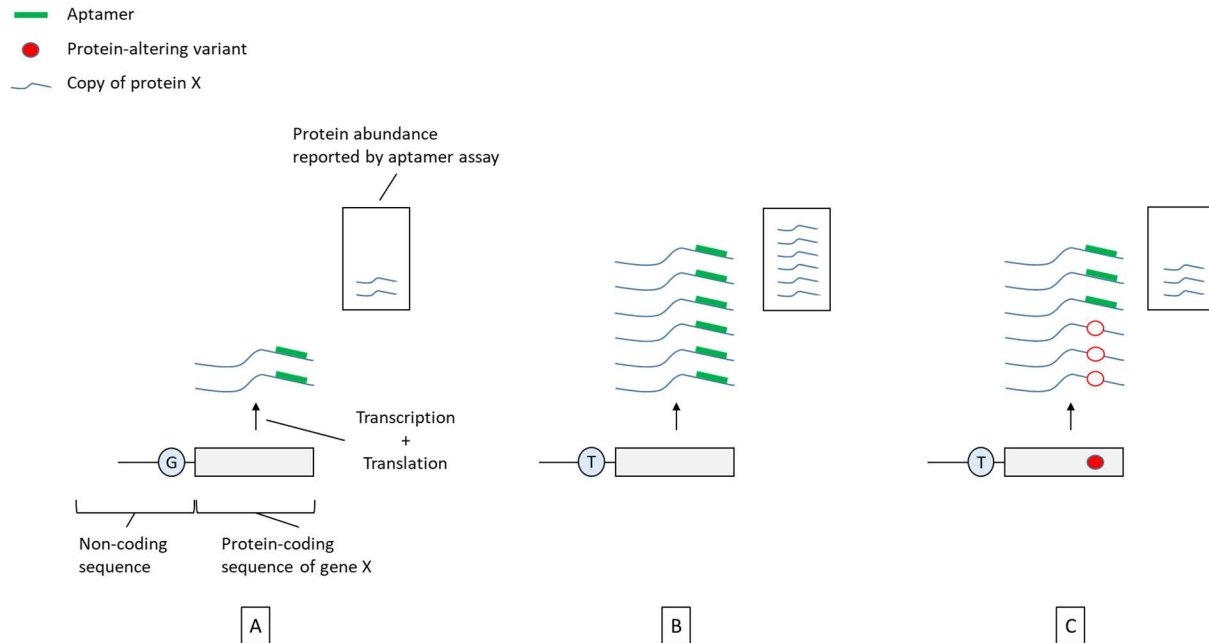


Figure S14. Protein-altering variants may bias the association of a genetic variant with concentration of a protein. In (A), the baseline scenario illustrates an allele's (G) association with abundance of protein X. (B) illustrates the alternative allele's (T) association with protein concentration. This variant is a strong pQTL given the difference in protein abundance between the two alleles. (C) illustrates a scenario where the strong pQTL from (B) is in LD with a protein-altering variant that influences aptamer binding. Copies of the protein that are affected by the protein-altering variant differentially bind to the aptamer, leading to a false-low abundance. This would lead the pQTL from (B) to have an attenuated strength of association with concentration of protein X. In the case of AGT, the (intronic) instrument selected on the basis of its association with SBP (rs2478539) is in tight LD ($r^2 = 1.0$ in European-ancestry population) with a missense variant in AGT (rs699). The presence of rs699 could therefore potentially bias the association of rs2478539 with measured AGT protein concentration, leading to a false-negative result in colocalization analysis.

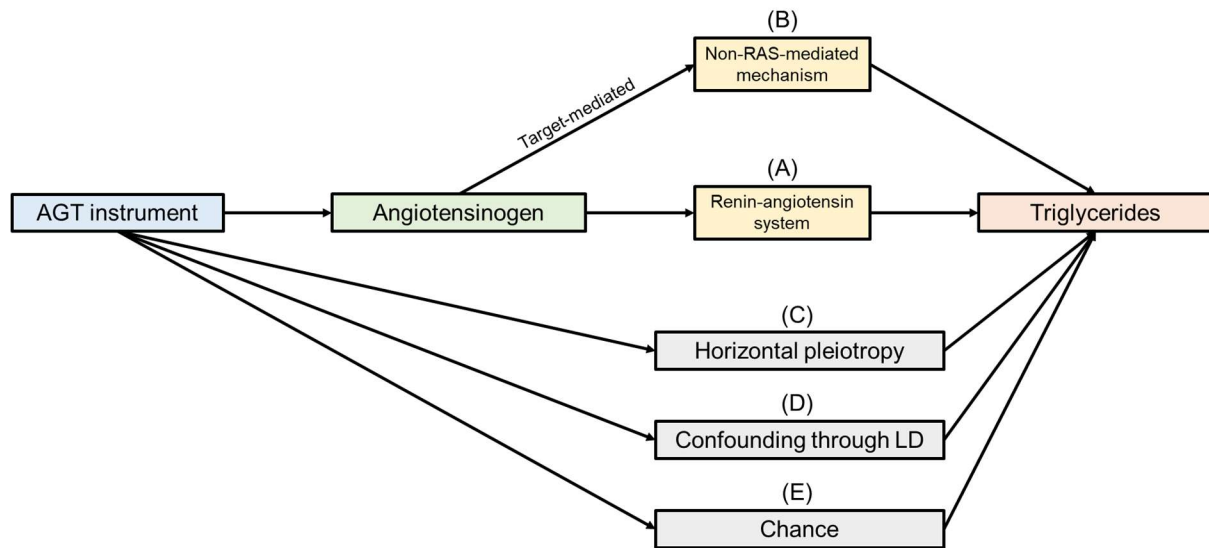


Figure S15. Potential explanations for association of the AGT instrument with triglyceride concentration (TG). This diagram illustrates some of the potential explanations for the observed association between the AGT genetic instrument and TG. In (A), the scenario where the instrument’s association with TG is mediated via the RAS pathway. The heterogeneity observed between the AGT TG estimate and the estimates relating to other RAS pathway genes, suggests that this scenario is unlikely. Since AGT lies upstream from the other RAS components, it remains possible that the instrument’s association with TG is mediated by AGT but via a non-RAS-related mechanism (B). The genetic instrument’s association with TG may however be biased in various ways. The instrument may have an influence on biological pathways other than those mediated via *AGT* (e.g. via regulation of expression of other genes; termed horizontal or pre-translational pleiotropy [C]), or it may be in linkage disequilibrium (LD) with another variant that is causally related to TG (D). Lastly, it is possible that the association of the instrument with TG is a chance occurrence (i.e. a false-positive; [E]).

Supplementary Tables

Table S1. Definitions of hypertension in UKBB.

Table S2. Definitions of vascular and renal outcomes in UKBB.

Table S3. Genome-wide association study datasets included in analyses.

Table S4. SBP-associated variants selected for two-sample MR analysis of SBP and all stroke.

Table S5. SBP-associated variants selected for two-sample MR analysis of SBP and myocardial infarction.

Table S6. SBP-associated variants selected for two-sample MR analysis of SBP and triglyceride concentration.

Table S7. Colocalization analyses of SBP and *AGT* gene expression across various tissues, using unconditioned data.

Table S8. Colocalization analyses of SBP and *AGT* gene expression across various tissues, using conditioned data.

Table S9. Colocalization analyses of SBP and circulating AGT protein concentration using unconditioned and conditioned data.

Table S10. Selected instruments for renin-angiotensin system genes.

Table S11. Two-sample Mendelian randomization analysis of SBP and all stroke.

Table S12. Two-sample Mendelian randomization analysis of SBP and myocardial infarction.

Table S13. Phenome-wide association analysis of rs2478539 in FinnGen.

Table S14 Phenome-wide association analysis of serum biomarkers in UKBB.

Table S15. Two-sample Mendelian randomization analysis of SBP and triglyceride concentration.

Table S16. Colocalization analyses of SBP and triglycerides.

References

- 1 Gazal S, Loh P-R, Finucane HK, *et al.* Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat Genet* 2018; **50**: 1600–7.
- 2 Suhre K, Arnold M, Bhagwat AM, *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017; **8**: 14357.
- 3 Haas ME, Aragam KG, Emdin CA, *et al.* Genetic Association of Albuminuria with Cardiometabolic Disease and Blood Pressure. *Am J Hum Genet* 2018; **103**: 461–73.
- 4 Teumer A, Tin A, Sorice R, *et al.* Genome-wide Association Studies Identify Genetic Loci Associated With Albuminuria in Diabetes. *Diabetes* 2016; **65**: 803–17.
- 5 Shah S, Henry A, Roselli C, *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun* 2020; **11**: 163.
- 6 Wuttke M, Li Y, Li M, *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* 2019; **51**: 957–72.
- 7 Willer CJ, Schmidt EM, Sengupta S, *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013; **45**: 1274–83.