

Circulating selenium and prostate cancer risk: a Mendelian randomization analysis

James Yarmolinsky^{1,2}, Carolina Bonilla^{1,2}, Philip C Haycock^{1,2}, Ryan JQ Langdon^{1,2}, Luca A Lotta³, Claudia Langenberg³, Caroline L Relton^{1,2}, Sarah J Lewis^{1,2}, David M Evans^{1,2,4}, the PRACTICAL consortium*, George Davey Smith^{1,2}, Richard M Martin^{1,2,5}

¹ MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

² Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

³ MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

⁴ University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia

⁵ University Hospitals Bristol NHS Foundation Trust National Institute for Health Research Bristol Nutrition Biomedical Research Unit, University of Bristol, Bristol, UK

* Members from the PRACTICAL Consortium (<http://practical.ccge.medschl.cam.ac.uk/>) are provided in the Supplementary Materials.

Corresponding author:

Professor Martin, BMBS, PhD

Population Health Sciences

Bristol Medical School

University of Bristol

Canynge Hall, 39 Whatley Road

Bristol BS8 2PS, United Kingdom

Fax: +44 117 928 7236

Tel: +44 117 928 7321

richard.martin@bristol.ac.uk

Abstract

In the Selenium and Vitamin E Cancer Prevention Trial (SELECT), selenium supplementation (causing a median 114 µg/L increase in circulating selenium) did not lower overall prostate cancer risk, but increased risk of high-grade prostate cancer and type 2 diabetes. Mendelian randomization analysis uses genetic variants to proxy modifiable risk factors and can strengthen causal inference in observational studies. We constructed a genetic risk score comprising eleven single-nucleotide polymorphisms robustly ($P < 5 \times 10^{-8}$) associated with circulating selenium in genome-wide association studies. In a Mendelian randomization analysis of 72,729 men in the PRACTICAL Consortium (44,825 cases, 27,904 controls), 114 µg/L higher genetically-elevated circulating selenium was not associated with prostate cancer (OR: 1.01; 95% CI: 0.89-1.13). Concordant with findings from SELECT, selenium was weakly associated with advanced (including high-grade) prostate cancer (OR: 1.21; 95% CI: 0.98-1.49) and type 2 diabetes (OR: 1.18; 95% CI: 0.97-1.43; in a type 2 diabetes GWAS meta-analysis with up to 49,266 cases, 249,906 controls). Mendelian randomization mirrored the outcome of selenium supplementation in SELECT and may offer an approach for the prioritization of interventions for follow-up in large-scale randomized controlled trials.

The development of interventions to prevent cancer requires robust causal knowledge, but few observational epidemiological claims are replicated in randomized controlled trials (RCTs) and some trial results are in the opposite direction to those seen observationally (i.e. causing harm) (1, 2). Such failures to translate observational data into effective cancer prevention interventions arise in part because the inherent limitations of observational research - confounding, reverse causation, and measurement error - preclude confident causal inference.

Mendelian randomization uses genetic variants as instruments (i.e. proxies) to assess whether a potential intervention target (e.g. risk factor, molecular intermediate, or gene product) has a causal effect on a disease outcome in a non-experimental (observational) setting (3). The principle of Mendelian randomization is that analysis of groups defined by common genetic variants is analogous to that of intention-to-treat analysis in an RCT. This is based on the independent assortment of genetic variants at meiosis which should allow for genotype at a population level to be largely independent of later environment and lifestyle factors (Figure 1). Using genetic variants as instruments to proxy intervention targets means that analyses should be less susceptible to confounding by environmental factors that typically distort observational associations. Furthermore, genotypes are typically measured with little error, represent life-long exposure, and are not subject to reverse causation because disease status cannot influence one's germline genotype.

An advantage of Mendelian randomization is that implementation does not require access to individual level data or trait measurements in all samples: it can be implemented using information on genetic variant-exposure and genetic variant-outcome associations obtained from separate samples, which greatly increases the scope of the approach ("two-sample Mendelian randomization") (4, 5). Two-sample approaches using summary data can be used in an efficient and cost-effective manner to screen hundreds of potential intervention

targets for causal relationships with cancer, without having to expensively measure these targets within individual cancer collections, and is made increasingly possible by the rapid increase in genome-wide association studies (GWAS) over the last decade; there are currently over 2000 published GWAS for over 1500 unique traits (<https://www.genome.gov/gwastudies/>).

The largest ever prostate cancer prevention trial (SELECT, N=35,533) was designed to examine whether daily supplementation with selenium, vitamin E, or both agents combined, could prevent prostate cancer (6). It was abandoned at a cost of \$114 million because of lack of efficacy compounded by possible carcinogenic (increased rates of high-grade [Gleason score ≥ 7] prostate cancer) and adverse metabolic effects (increased rates of diabetes) of the interventions (6, 7). We investigated whether Mendelian randomization could have predicted the results of the SELECT trial observed for selenium in a two-sample Mendelian randomization study of 72,729 individuals of European descent from the PRACTICAL (Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome) consortium (8).

We obtained summary GWAS statistics from analyses on 44,825 prostate cancer cases and 27,904 controls of European descent from 108 studies in PRACTICAL. Summary statistics were also obtained from analyses on 6,263 advanced prostate cancer cases (defined as Gleason ≥ 8 , prostate-specific antigen >100 ng/mL, metastatic disease (M1), or death from prostate cancer) and 27,235 controls (summary data on high-grade prostate cancer alone was not available). All studies in PRACTICAL have the relevant Institutional Review Board approval from each country, in accordance with the Declaration of Helsinki. Genotyping of PRACTICAL samples was performed using an Illumina Custom Infinium genotyping array (OncoArray), designed for the OncoArray consortium and consisting of $\sim 570,000$ single nucleotide polymorphisms (SNPs) (9). All SNPs with a poor imputation quality ($r^2 < 0.30$), a

minor allele frequency of <1%, a call rate of <98%, or evidence of violation of Hardy-Weinberg equilibrium ($P < 10^{-7}$ in controls or $P < 10^{-12}$ in cases) were removed.

To analyze the effect of selenium on type 2 diabetes we used summary GWAS data from analyses in up to 49,266 type 2 diabetes cases and 249,906 controls of European descent obtained from a meta-analysis of the DIAbetes Genetics Replication And Meta-analysis (10), EPIC-InterAct (11), and UK Biobank studies (12). Methods for this meta-analysis have been published previously (13).

A genetic risk score to proxy for circulating selenium levels was constructed by obtaining SNPs shown to robustly ($P < 5 \times 10^{-8}$) associate with selenium concentrations in a meta-analysis of blood and toenail selenium GWAS (14, 15). Of twelve selenium SNPs identified, one (rs558133) was not available in PRACTICAL and thus eleven SNPs were used as genetic instruments for overall and advanced prostate cancer analyses. As sensitivity analyses, we also constructed a restricted genetic risk score using only SNPs robustly ($P < 5 \times 10^{-8}$) associated with selenium in a GWAS of blood selenium that were replicated ($P < 0.05$) in subsequent independent studies (14-16). For these analyses, of five selenium SNPs initially identified, one (rs6859667) was not available in PRACTICAL and thus four SNPs were used as instruments for both prostate cancer analyses. All SNPs utilised for primary and sensitivity analyses and their corresponding ENSEMBL-mapped gene(s) (17) are presented in the footnote to Table 1.

We constructed a correlation matrix to estimate correlation between SNPs with reference to the HapMap 3 (release 2) dataset. The causal effect of our genetic risk score on overall prostate cancer, advanced prostate cancer, and type 2 diabetes was then examined using a maximum likelihood approach that takes into account moderate correlations between genetic variants (18). To compare the causal odds ratios from Mendelian randomization with

the hazard ratios from SELECT, we estimated the causal odds ratios per 114 μ g/L genetically increased circulating selenium, to match the measured pre- versus post-intervention blood selenium differences between supplementation and control arms in SELECT (6). All statistical analyses were performed using R version 3.0.2.

In Mendelian randomization analyses, a 114 μ g/L increase in genetically-elevated blood selenium was not associated with overall prostate cancer risk (OR: 1.01, 95% CI: 0.89-1.13, $P=0.93$) (Table 1). Genetically elevated selenium was weakly associated with advanced prostate cancer (OR: 1.21, 95% CI: 0.98-1.49, $P=0.07$) and type 2 diabetes (OR: 1.18, 95% CI: 0.97-1.43, $P=0.11$). Results for overall prostate cancer, advanced prostate cancer and type 2 diabetes were robust to sensitivity analyses employing a restricted genetic risk score (ORs were 0.95 [95% CI 0.80-1.14], 1.09 [0.81-1.47] and 1.23 [0.99-1.53]), respectively).

Limitations of our analysis include that we were only able to directly examine the effect of genetically elevated selenium on advanced prostate cancer and not high-grade prostate cancer per se (summary estimates from PRACTICAL were available only for a composite “advanced” disease classification), thus preventing direct comparison to results with SELECT. Additionally, our selenium SNPs were also associated with betaine, a putative risk factor for type 2 diabetes (19), which could introduce horizontal pleiotropy (genetic variants influencing an outcome through a different biological pathway from the exposure under investigation) into our analyses and thus invalidate instrumental variable assumptions (see footnote of Figure 1). Though we suspect that the association of these SNPs with both selenium and betaine reflects the effect of selenium on betaine in the methionine cycle (20, 21) and, consequently, that both selenium and betaine likely influence type 2 diabetes risk through the same biological pathway, we cannot rule out the possibility that at least part of a putative effect of selenium SNPs on type 2 diabetes risk is through an alternate biological pathway involving betaine.

We conclude that in contrast to findings from some (22-27) but not all prospective epidemiological studies (28, 29), our Mendelian randomization analysis using publicly available GWAS data did not find strong evidence for a causal effect of selenium on overall prostate cancer. Consistent with SELECT, we found weak evidence of a positive effect of genetically elevated selenium on advanced prostate cancer. In agreement with SELECT, we also found weak evidence of a positive effect of genetically elevated selenium on type 2 diabetes risk. The alignment of Mendelian randomization with SELECT estimates mirrors the concordance of Mendelian randomization findings with other large, phase III trials; including adverse effects of elevated LDL cholesterol on coronary heart disease (30, 31) and statin use with type 2 diabetes (32, 33); and null effects of secretory phospholipase A(2)-IIA with cardiovascular disease (34-36) and HDL cholesterol on risk of myocardial infarction (37-40). Mendelian randomization may serve as an important time-efficient and inexpensive first step in predicting both the efficacy and possible adverse effects of an intervention prior to the design of a randomized controlled trial.

Abbreviations

Randomized controlled trial (RCT), genome-wide association study (GWAS), Selenium and Vitamin E Cancer Prevention Trial (SELECT), Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL), single nucleotide polymorphisms (SNPs)

Funding

This work was supported by a Cancer Research UK programme grant (C18281/A19169), Cancer Research UK Research PhD studentships (C18281/A20988 to J.Y. and R.J.Q.L.), and a Cancer Research UK Population Research Postdoctoral Fellowship (C52724/A20138 to P.C.H.). This work was also supported by a grant awarded to S.J.L. for 3 years to identify modifiable risk factors for prostate cancer, by the World Cancer Research Fund International (grant reference number: 2015/1421). The Medical Research Council Integrative Epidemiology Unit at the University of Bristol is supported by the Medical Research Council (MC_UU_12013/1, MC_UU_12013/2, and MC_UU_12013/3) and the University of Bristol.

Acknowledgments

We are grateful to the contribution of data from EPIC-InterAct Investigators. The InterAct Study is led by Professor Nick Wareham at the MRC Epidemiology Unit, Cambridge, UK; InterAct funding was provided by the EU FP6 programme (grant number LSHM_CT_2006_037197).

Figure 1. Schematic comparison of a Randomized Controlled Trial (SELECT) to a Mendelian randomization analysis.

In an RCT, individuals are randomly allocated to an intervention or control group (In SELECT, 200 $\mu\text{g/d}$ selenium [114 $\mu\text{g/L}$ increase in blood selenium] or placebo). If the trial is adequately sized, randomization should ensure that intervention and control groups are comparable in all respects (e.g. approximately equal distribution of potential confounding factors) except for the intervention being tested. In an intention-to-treat analysis, any observed differences in outcomes between intervention and control groups can then be attributed to the trial arm to which they were allocated. In a Mendelian randomization (MR) analysis, alleles that influence levels of a trait of interest are randomly allocated at conception (In MR, the additive effects of selenium-raising alleles at eleven SNPs were scaled to mirror a 114 $\mu\text{g/L}$ increase in blood selenium). Groups defined by genotype should be comparable in all respects (e.g. distribution of both genetic and environmental confounding factors) except for their exposure to a trait of interest. Any observed differences in outcomes between groups defined by genotype can then be attributed to differences in life-long exposure to the trait of interest under study.

Mendelian randomization is an application of the technique of instrumental variable (IV) analysis. In order for a genetic variant (or a multi-allelic genetic risk score) to be used as an IV, three key assumptions must be met: 1) The instrument must be reliably associated with the exposure of interest, 2) the instrument should be independent of other factors affecting the outcome (confounders), 3) the instrument should only affect the outcome through the exposure of interest (known as the exclusion restriction criterion).

Table 1. Comparison of the effect of 114 ug/L selenium on overall prostate cancer, high-grade / advanced prostate cancer, and type 2 diabetes in SELECT and Mendelian randomization

	SELECT HR (95% CI)	Mendelian randomization OR (95% CI)
Overall prostate cancer	1.04 (0.91 to 1.19)	1.01 (0.89 to 1.13)
High-grade/advanced prostate cancer*	1.21 (0.97 to 1.52)	1.21 (0.98 to 1.49)
Type 2 diabetes	1.07 (0.87 to 1.18)	1.18 (0.97 to 1.43)

HR= Hazard Ratio; OR= Odds Ratio; CI= Confidence Interval.

*“High-grade prostate cancer” pertains to SELECT results and “Advanced prostate cancer” pertains to Mendelian randomization results.

SNPs utilised in GRS for primary analyses: ARSB (rs672413, **rs705415**), DMGDH (**rs705415**, rs3797535, rs11951068, rs921943), BHMT2 (rs10944), BHMT (rs567754, rs558133), HOMER1 (rs6859667), CBS (rs6586282, rs1789953, rs234709). SNPS utilised in GRS for sensitivity analyses: DMGDH (rs248381, rs921943), BHMT (rs7700970, rs567754), HOMER1 (rs6859667). SNPs in bold were mapped to multiple genes.

Citations

1. Wiseman M. The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *Proc Nutr Soc.* 2008;67(3):253-6.
2. Ioannidis JP. Implausible results in human nutrition research. *BMJ.* 2013;347:f6698.
3. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1-22.
4. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* 2013;178(7):1177-84.
5. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol.* 2015;30(7):543-52.
6. Lippman SM, Klein EA, Goodman PJ, Lucia MS, Thompson IM, Ford LG, et al. Effect of selenium and vitamin E on risk of prostate cancer and other cancers: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA.* 2009;301(1):39-51.
7. Klein EA, Thompson IM, Jr., Tangen CM, Crowley JJ, Lucia MS, Goodman PJ, et al. Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA.* 2011;306(14):1549-56.
8. Schumacher FR, Amin Al Olama A, Berndt SI, et al. Prostate cancer meta-analysis of more than 140,000 men identifies 63 novel prostate cancer susceptibility loci. *Nature Genetics.* 2017. In Press

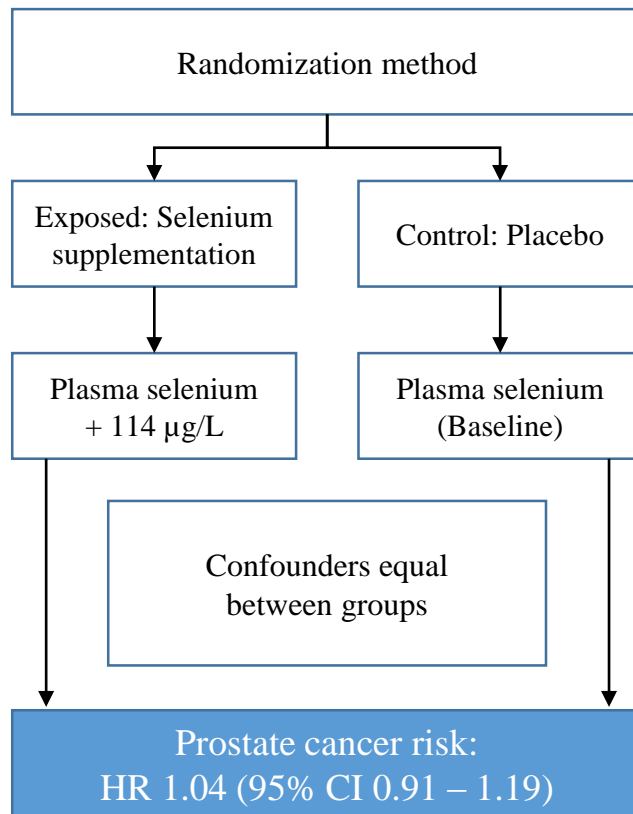
9. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev.* 2017;26(1):126-35.
10. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012;44(9):981-90.
11. InterAct C, Langenberg C, Sharp S, Forouhi NG, Franks PW, Schulze MB, et al. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia.* 2011;54(9):2272-82.
12. Collins R. What makes UK Biobank special? *Lancet.* 2012;379(9822):1173-4.
13. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, et al. Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. *PLoS Med.* 2016;13(11):e1002179.
14. Cornelis MC, Fornage M, Foy M, Xun P, Gladyshev VN, Morris S, et al. Genome-wide association study of selenium concentrations. *Hum Mol Genet.* 2015;24(5):1469-77.
15. Evans DM, Zhu G, Dy V, Heath AC, Madden PA, Kemp JP, et al. Genome-wide association study identifies loci affecting blood copper, selenium and zinc. *Hum Mol Genet.* 2013;22(19):3998-4006.
16. Mao J, Vanderlelie JJ, Perkins AV, Redman CW, Ahmadi KR, Rayman MP. Genetic polymorphisms that affect selenium status and response to selenium supplementation in United Kingdom pregnant women. *Am J Clin Nutr.* 2016;103(1):100-6.
17. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42(Database issue):D1001-6.

18. 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.
19. Svingen GF, Schartum-Hansen H, Pedersen ER, Ueland PM, Tell GS, Mellgren G, et al. Prospective Associations of Systemic and Urinary Choline Metabolites with Incident Type 2 Diabetes. *Clin Chem*. 2016;62(5):755-65.
20. Davis CD, Uthus EO. Dietary folate and selenium affect dimethylhydrazine-induced aberrant crypt formation, global DNA methylation and one-carbon metabolism in rats. *J Nutr*. 2003;133(9):2907-14.
21. Uthus EO, Ross SA. Dietary selenium affects homocysteine metabolism differently in Fisher-344 rats and CD-1 mice. *J Nutr*. 2007;137(5):1132-6.
22. Allen NE, Travis RC, Appleby PN, Albanes D, Barnett MJ, Black A, et al. Selenium and Prostate Cancer: Analysis of Individual Participant Data From Fifteen Prospective Studies. *J Natl Cancer Inst*. 2016;108(11).
23. Brooks JD, Metter EJ, Chan DW, Sokoll LJ, Landis P, Nelson WG, et al. Plasma selenium level before diagnosis and the risk of prostate cancer development. *J Urol*. 2001;166(6):2034-8.
24. Li H, Stampfer MJ, Giovannucci EL, Morris JS, Willett WC, Gaziano JM, et al. A prospective study of plasma selenium levels and prostate cancer risk. *J Natl Cancer Inst*. 2004;96(9):696-703.
25. Nomura AM, Lee J, Stemmermann GN, Combs GF, Jr. Serum selenium and subsequent risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev*. 2000;9(9):883-7.
26. van den Brandt PA, Zeegers MP, Bode P, Goldbohm RA. Toenail selenium levels and the subsequent risk of prostate cancer: a prospective cohort study. *Cancer Epidemiol Biomarkers Prev*. 2003;12(9):866-71.

27. Yoshizawa K, Willett WC, Morris SJ, Stampfer MJ, Spiegelman D, Rimm EB, et al. Study of prediagnostic selenium level in toenails and the risk of advanced prostate cancer. *J Natl Cancer Inst.* 1998;90(16):1219-24.
28. Allen NE, Appleby PN, Roddam AW, Tjønneland A, Johnsen NF, Overvad K, et al. Plasma selenium concentration and prostate cancer risk: results from the European Prospective Investigation into Cancer and Nutrition (EPIC). *Am J Clin Nutr.* 2008;88(6):1567-75.
29. Helzlsouer KJ, Huang HY, Alberg AJ, Hoffman S, Burke A, Norkus EP, et al. Association between alpha-tocopherol, gamma-tocopherol, selenium, and subsequent prostate cancer. *J Natl Cancer Inst.* 2000;92(24):2018-23.
30. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. *J Am Coll Cardiol.* 2012;60(25):2631-9.
31. Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet.* 2005;366(9493):1267-78.
32. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JE, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *Lancet.* 2015;385(9965):351-61.
33. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJ, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet.* 2010;375(9716):735-42.

34. Holmes MV, Simon T, Exeter HJ, Folkersen L, Asselbergs FW, Guardiola M, et al. Secretory phospholipase A(2)-IIA and cardiovascular disease: a mendelian randomization study. *J Am Coll Cardiol*. 2013;62(21):1966-76.
35. Millwood IY, Bennett DA, Walters RG, Clarke R, Waterworth D, Johnson T, et al. A phenome-wide association study of a lipoprotein-associated phospholipase A2 loss-of-function variant in 90 000 Chinese adults. *Int J Epidemiol*. 2016;45(5):1588-99.
36. Clinical Study Report for Study AN-CVD2233. Evaluation of the Safety and Efficacy of Short-Term A-002 Treatment in Subjects with Acute Coronary Syndromes.; 2012.
37. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380(9841):572-80.
38. Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med*. 2007;357(21):2109-22.
39. Group HTC, Landray MJ, Haynes R, Hopewell JC, Parish S, Aung T, et al. Effects of extended-release niacin with laropiprant in high-risk patients. *N Engl J Med*. 2014;371(3):203-12.
40. Group AS, Ginsberg HN, Elam MB, Lovato LC, Crouse JR, 3rd, Leiter LA, et al. Effects of combination lipid therapy in type 2 diabetes mellitus. *N Engl J Med*. 2010;362(17):1563-74.

Randomized Controlled Trial (SELECT)



Mendelian Randomization

