Supplementary Online Content

Burgess S, Ference BA, Staley JR, et al; European Prospective Investigation Into Cancer and Nutrition–Cardiovascular Disease (EPIC-CVD) Consortium. Association of *LPA* variants with risk of coronary disease and the implications for lipoprotein(a)-lowering therapies: a mendelian randomization analysis. *JAMA Cardiol*. Published online June 20, 2018. doi:10.1001/jamacardio.2018.1470

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods 1. Details of studies and participants

Copenhagen City Heart Study (CCHS)

CCHS is a population-based prospective study initiated in 1976 with follow-up examinations from 1981 to 1983, 1991 to 1994, and 2001 to 2003. Participants were selected on the basis of the national Danish Civil Registration System to reflect the adult Danish population age 20 to 100 years. Data were obtained from a questionnaire, a physical examination, and blood samples including deoxyribonucleic acid extraction. The case definition was fatal and non-fatal myocardial infarction and other coronary events according to ICD10 codes I20-I25. Control participants are members of the CCHS cohort who were free from coronary disease at baseline and after follow-up.

An immunoturbidimetric assay (either supplied by DiaSys or Technicon Axon) was used to measure Lp(a) in CCHS. All analyses in CCHS were additionally adjusted for Lp(a) assay supplier.

Copenhagen General Population Study (CGPS)

The CGPS is a population-based prospective study initiated in 2003 with ongoing enrollment. Participants were selected on the basis of the national Danish Civil Registration System to reflect the adult Danish population age 20 to 100 years. Data were obtained from a questionnaire, a physical examination, and blood samples including deoxyribonucleic acid extraction. An immunoturbidimetric assay (either supplied by DiaSys or Denka Seiken) was used to measure Lp(a) in CGPS. All analyses in CGPS were additionally adjusted for Lp(a) assay supplier.

The Copenhagen Ischaemic Heart Disease Study (CIHDS)

CIHDS is a study comprising of 6,625 cases (4,635 men and 1,990 women) with myocardial infarction and other major acute coronary syndromes (ICD10 codes I20 to I25) and 10,368 age-and-sex matched controls from CGPS. Blood samples including DNA extraction were available. The cases were recruited between 1991 and 2009 from the Copenhagen University Hospital. In addition to a diagnosis of coronary heart disease, these cases also had stenosis or atherosclerosis on coronary angiography and/or positive results on exercise electrocardiography. For estimating associations with coronary heart disease, CGPS and CIHDS are treated together as a single case-control study. Lp(a) measurements from CIHDS were not including in this project.

Individuals in the CCHS 2001–2003 examination and the first 5592 individuals in the CGPS had lipoprotein(a) measured immediately after sampling using an assay from DiaSys (Diagnostic Systems), whereas subsequent individuals in the CGPS had lipoprotein(a) measurements done using an assay from Denka Seiken; of the latter, 12,577 were stored at -80 °C until measurement, whereas the remaining 31,211 were measured on fresh samples.

European Prospective Investigation into Cancer and Nutrition - Cardiovascular Disease Study (EPIC-CVD)

EPIC is a multi-centre prospective cohort study of 519,978 participants (366,521 women and 153,457 men, mostly aged 35–70 years) recruited between 1992 and 2000 in 23 centres located in 10 European countries. Participants were invited mainly from population-based registers (Denmark, Germany, certain Italian centres, the Netherlands, Norway, Sweden, UK). Other sampling

frameworks included: blood donors (Spain, and Turin and Ragusa in Italy); screening clinic attendees (Florence in Italy and Utrecht in the Netherlands); people in health insurance programmes (France); and health conscious individuals (Oxford, UK). About 97% of the participants were of white European ancestry. EPIC-CVD is a nested case-cohort design comprising all cardiovascular disease (CVD) cases from the large EPIC cohort and a random sample of the entire cohort ("subcohort"). The case definition was fatal and non-fatal myocardial infarction and other coronary events according to ICD10 codes I20-I25.

A particle-enhanced immunoturbidimetric assay on a Roche MODULAR ANALYTICS EVO analyser was used to measure Lp(a) in EPIC-CVD. Details of blood draw and storage procedures in each of the EPIC centres can be found in the consortium protocol paper (1).

Prospective Study of Pravastatin in the Elderly at Risk trial (PROSPER)

PROSPER was a controlled, randomised study involving 2,804 men and 3,000 women aged 70-82, with a history of, or risk factors for cardiovascular disease. Participants were randomised to either 40mg pravastatin per day or matching placebo. The case definition was fatal coronary heart disease or nonfatal myocardial infarction.

Lp(a) was not measured in PROSPER. This study was only used in testing the association of the Lp(a) genetic risk score with CHD.

West of Scotland Coronary Prevention Study (WOSCOPS)

WOSCOPS was a primary-prevention clinical trial of 6,595 men in the West of Scotland district aged 45 to 64 years with elevated cholesterol levels (moderate hypercholesterolemia). Between 1989 and 1991, 6,595 men who had no evidence of previous myocardial infarction were randomised to either receive pravastatin (40 mg once daily) or placebo. To be eligible for enrolment, participants had to have two measurements of LDL-C >155mg/dL, with at least one measurement >174mg/dL. A nested case-control design was used for this study, selecting as cases individuals who self-reported a history of coronary disease at baseline or who had a coronary event during follow-up. Controls were participants who were free of cardiovascular disease at baseline and at the end of follow-up, frequency matched to the cases for sex and age (in 5-year bands). The case definition was fatal coronary heart disease or nonfatal myocardial infarction.

An ELISA assay (supplied by Innogenetics) was used to measure Lp(a) in WOSCOPS on fasted samples (8 hours+) of blood plasma that had been stored at -70 °C for 1 to 5 years.

Details on the numbers and percentage of participants with Lp(a) levels above various threshold values are provided in Supplementary Table S2. A violin plot of the distribution of Lp(a) in each study is presented as Supplementary Figure S4. Lp(a) assays are not uniformly calibrated (2) and large differences in measurements using different assays have been demonstrated previously (3).

eMethods 2. Detailed statistical methods

LPA GENETIC SCORE

To select variants for inclusion in the *LPA* genetic score, we started with the full list of 2462 candidate variants in the *LPA* gene region (660kb window). 936 variants remained available for analysis after filtering out variants that: were monomorphic across all samples, severely deviated from Hardy—Weinberg equilibrium ($p < 1 \times 10^{-6}$ for variants with MAF ≥ 0.05 and $p < 1 \times 10^{-15}$ for variants with MAF < 0.05), or did not have a call rate > 95% in each study. Values of Lp(a) above 130 mg/dL were taken to be 130 mg/dL (winsorization) to reduce the effect of extreme outliers on the analyses. In total, 5% of Lp(a) measurements were above this threshold.

At each step of the selection algorithm, we tested the association of each variant in a linear model where the dependent variable was Lp(a) and independent variables were age, sex, study cohort, assay method, 5 principal components of ancestry, and all variants selected in a previous step of the algorithm. The variant that was associated with Lp(a) in this conditional analysis with the lowest p-value below a threshold of 5×10^{-8} was added to the set of selected variants. Once a variant was included in the analysis, any other variant that was correlated with the selected variant at $r^2 > 0.4$ was removed from the set of candidate variants. We then continued to the next step until all variants were either selected, removed due to linkage disequilibrium with a selected variant, or were not strongly associated ($p < 5 \times 10^{-8}$) with Lp(a) in the conditional analysis. A weighted genetic score was then constructed with weights being the conditional associations of each variant with Lp(a) (conditional on all the other variants in the score). The variants and their associations (marginal and conditional) with the risk factor and outcome are provided in Supplementary Table S3. Marginal associations of the variants with Lp(a) in each study are displayed in Supplementary Figure S6. Despite substantial differences between the mean and median concentrations of Lp(a) in each study, the genetic associations are similar across studies, suggesting that the assay type affects the baseline concentration of Lp(a), but not changes in Lp(a). As we adjust for study (and assay type if appropriate) in all analyses, differences in baseline concentrations of Lp(a) will not influence the analysis – only genetic associations with Lp(a) are used in our calculations.

The variant selection strategy was chosen with the aim of maximizing the proportion of variance in Lp(a) explained subject to certain constraints. Mendelian randomization estimates with highly correlated genetic variants can be unstable, as the genetic correlation matrix is inverted as part of the analysis. Therefore we set a correlation threshold of $r^2 = 0.4$ to ensure that no two selected variants had a pairwise squared correlation greater than 0.4. To minimize overprediction, we strictly avoided including too many variants in the score. Hence, each variant included in the model was required to be conditionally associated with Lp(a) at a genome-wide level of significance ($p < 5 \times 10^{-8}$). The upshot of this is that less of the variance in Lp(a) is explained by the genetic score than perhaps could have been with a more liberal choice of variants; however, a sizeable proportion of variance in the risk factor is explained, and this choice of variants should lead to more robust inferences.

A regional association plot of the gene region indicating the 43 variants included in the *LPA* genetic score is provided as Supplementary Figure S1, and a schematic diagram of how the *LPA* genetic score was constructed is provided in Supplementary Figure S2. A plot showing the proportion of variance

in Lp(a) explained by the genetic score in each study and at each step of the variant selection algorithm is provided as Supplementary Figure S5.

SHAPE OF RELATIONSHIP BETWEEN LIPOPROTEIN(a) AND CORONARY HEART DISEASE RISK

The relationship between genetically predicted Lp(a) levels and CHD risk was assessed in two ways: (i) within deciles of genetically predicted Lp(a); and (ii) across the entire range of genetically predicted Lp(a) using fractional polynomials (4). These analyses are equivalent to a two-stage Mendelian randomisation analysis, where the analysis in the second stage is not restricted to be linear, and the Lp(a) genetic risk score is used as the instrument variable. The deciles of Lp(a) were created by splitting the distribution of the genetic risk score of Lp(a) across all of the studies into deciles. As the genetic risk score only depends on the number of Lp(a)-increasing variants, baseline Lp(a) level in each study should not affect the distribution of individuals in each of the decile groups. The associations of these deciles with risk of CHD were then assessed using logistic regression in each study separately, and meta-analysed using multivariate random-effects meta-analysis. To reflect the amount of information within each group, 95% confidence intervals were estimated from variances attributed to each group, including the reference group, using floating absolute risks (5). The best-fitting fractional polynomial model (of either degree 1 or 2) was also used to flexibly assess the relationship between genetically predicted Lp(a) and CHD risk using logistic regression across studies (i.e. all cohorts are fitted in a single model). A test of linearity was performed using logistic regression across studies, by testing for a non-zero quadratic term in a quadratic model. These analyses were adjusted age, sex, and the first five principal components of ancestry; the fractional polynomial and quadratic models were also adjusted for study. The mvshape R package (https://github.com/jrs95/mvshape) was used to perform these analyses.

Additionally, we estimated the relationship between log-transformed Lp(a) and log odds of CHD, as has been considered in previous epidemiological investigations (6). In contrast to the linear relationship between absolute levels of Lp(a) and log odds of CHD, the relationship between log-transformed Lp(a) and log odds of CHD was curvilinear, with an increased association between genetically predicted log-transformed Lp(a) and CHD risk at greater Lp(a) levels. The linear and curvilinear relationships are not in contradiction, but in fact are completely compatible: equal changes in log-transformed Lp(a) correspond to greater changes in absolute concentrations of Lp(a) for greater baseline values (and therefore stronger associations with CHD risk).

For untransformed Lp(a), the test of linearity using a quadratic term was not rejected ($p = 0.11$). The best-fitting fractional polynomial was the linear model. For log-transformed Lp(a), the test of linearity using a quadratic term was rejected ($p < 0.001$). The best-fitting fractional polynomial included quadratic and cubic terms.

CAUSAL EFFECT OF LIPOPROTEIN(a) ON RISK OF CORONARY HEART DISEASE

The Mendelian randomization effect of Lp(a) on CHD risk was calculated using summarized data on variants in the *LPA* genetic score. Genetic associations were derived in each study separately, and then meta-analysed across studies. This analysis assumes a linear relationship between Lp(a) and the log-odds of CHD risk. Genetic associations with Lp(a) were estimated using linear regression in participants not having a previous CHD event at baseline, adjusting for age, sex, and 5 principal components of ancestry (plus Lp(a) assay method if relevant). Genetic associations with CHD risk

were estimated using logistic regression, with the same adjustment. A matrix of genetic correlations between variants was estimated in participants not having a previous CHD event at baseline only.

The summarized genetic association estimates were combined into Mendelian randomization estimates using weighted generalized linear regression accounting for the correlation between variants. If variants were uncorrelated, this method would be equivalent to combining the variant-specific causal estimates in an inverse-variance weighted meta-analysis (or combining the variants into a single genetic score variable and calculating the Mendelian randomization ratio estimate using this score). The regression model was:

$$\beta_Y = \theta\, \beta_X + \varepsilon, \qquad \varepsilon \sim N(0, \Omega)$$

where θ is the Mendelian randomization causal estimate, β_X is a vector of the genetic associations (beta-coefficients) with the risk factor, β_Y is a vector of the genetic associations with the outcome, and the weighting matrix Ω has terms $\Omega_{j_1 j_2} = \sigma_{Y j_1}\sigma_{Y j_2}\rho_{j_1 j_2}$, where $\sigma_{Y j}$ is the standard error of the genetic association with the outcome for the *j*th variant, and $\rho_{j_1 j_2}$ is the correlation between the j_1th and j_2th variants. The causal estimate from this weighted generalized linear regression is $(\beta_X{}^T \Omega^{-1} \beta_X)^{-1} \beta_X{}^T \Omega^{-1} \beta_Y$, and the standard error is $\sigma\sqrt{(\beta_X{}^T \Omega^{-1} \beta_X)^{-1}}$, where ^{T} is a matrix transpose, and σ is the maximum of the residual standard error from the regression model and 1. This is equivalent to assuming a multiplicative random-effects model on the variant-specific causal effect estimates. By fixing σ to be no lower than 1, we ensure that the random-effects analysis is no more precise than a fixed-effect analysis would be.

The reason for using summarized data in our analysis, even though we had individual-level data available, is so that we could combine evidence first across studies, and then across genetic variants. This approach corresponds to where we would expect to see heterogeneity in our associations, and produces estimates that are less susceptible to weak instrument bias (7).

This method has been described previously (8) and was implemented using the MendelianRandomization package in R (available for download at https://cran.r-project.org/web/packages/MendelianRandomization/) (9). When run as a fixed-effect analysis, it is equivalent to the commonly-used two-stage least squares method that requires individual-level data.

CAUSAL EFFECT OF LDL-CHOLESTEROL ON RISK OF CORONARY HEART DISEASE

To assess the causal effect of LDL-C on CHD risk, we used a combined LDL score comprising 8 genetic variants in separate gene regions each of which has been specifically linked with LDL-C (it either encodes a biologically relevant compound to LDL-C, or is a proxy for an existing or proposed LDL-C lowering drug). These gene regions are: *HMGCR* (proxy for statin treatment), *PCSK9* (proxy for PCSK9 inhibition), *NPC1L1* (proxy for ezetimibe), *APOB* (encodes biologically relevant apolipoprotein B), *ABCG5/G8* (bile acid sequestrant), *SORT1* (antisense oligonucleotide RNA inhibitor targeting this pathway currently under development), *APOE* (encodes biologically relevant apolipoprotein E), and *LDLR* (encodes biologically relevant LDL receptor). The specific choice of variant in each gene region to include in the analysis was based on the lead variant from the Global Lipids Genetic Consortium's 2010 analysis. As these variants were discovered in an almost entirely non-overlapping set of studies

to the Exome+ consortium (the only overlap was up to 1772 EPIC-CVD participants from the Norfolk centre), there should be minimal bias due to winner's curse in the Mendelian randomization estimate calculated in the Exome+ consortium.

Summarized genetic associations with LDL-C were calculated in the same way as associations with Lp(a) described above, and Mendelian randomization estimates were calculated using these summarized estimates as described above.

Genetic associations with LDL-C and with CHD risk for the 8 variants in the combined LDL score are provided in Supplementary Table S1 and displayed graphically in Supplementary Figure S5. Each of the variants approximately lies on the same straight-line through the origin, indicating that the Mendelian randomization estimates based on each individual variant are similar. In particular, the Mendelian randomization estimate is relatively insensitive to which of these variants are included in the combined LDL genetic score. The similarity of these estimates gives us reasonable confidence that the Mendelian randomization assumptions are satisfied for the LDL-C score.

POTENTIAL CLINICAL BENEFIT OF LIPOPROTEIN(a) LOWERING THERAPIES

We here detail the approach for translating the Mendelian randomization estimate for the effect of Lp(a) on CHD risk into the effect of reducing Lp(a) in a short-term trial. We first compare the Mendelian randomization estimates for Lp(a) and for LDL-C per 10 mg/dL lowering to draw an equivalence between Lp(a) and LDL-C. The ratio between the odds ratios of 0.942 per 10 mg/dL reduction in Lp(a) and the odds ratio of 0.855 per 10 mg/dL reduction in LDL-C on the log odds ratio scale is log(0.855)/ log(0.942) = 2.63, indicating that 1 mg/dL change in LDL-C has the same Mendelian randomization estimate for CHD risk as 2.63 mg/dL change in Lp(a). A 1 mmol/L (38.67 mg/dL) reduction in LDL-C therefore has an approximately equivalent effect on CHD risk as a 38.67 × 2.63 = 101.5 mg/dL reduction in Lp(a) concentration.

This calculation is numerically equivalent to calculating the ratio between the estimates for the life-long (genetically predicted) and short-term (clinical trial) effect of LDL-C on CHD risk, and dividing the genetically predicted effect of life-long Lp(a) on CHD risk by this ratio to estimate the short-term effect of Lp(a) on CHD risk.

A 95% confidence interval for this quantity was calculated by a Monte Carlo procedure. We took a random draw from a normal distribution with mean taken as the Mendelian randomization estimate for LDL-C [log(0.855)] and standard deviation taken as the standard error of this estimate, divided by a random draw from a normal distribution with mean taken as the Mendelian randomization estimate for Lp(a) [log(0.942)] and standard deviation taken as the standard error of this estimate, and multiplied by 38.67. We repeated this procedure 100 000 times, and took the 2.5^{th} and 97.5^{th} percentiles from this distribution of values as the 95% confidence interval.

The odds ratio for the effect of Lp(a) lowering in a trial setting is calculated using this equivalence to scale the short term estimate of intervention on LDL-C obtained from clinical trials of statin treatment of 0.76 (95% CI: 0.73 – 0.78). This is the meta-analysis estimate taken from Webfigure 2 of the Cholesterol Treatment Trialists' Consortium 2010 paper with the outcome of major coronary events (10).

The odds ratio estimate for a 10 mg/dL lowering in Lp(a) can be calculated as:

exp[Change in Lp(a) × Ratio of Mendelian randomization estimates × Effect of statin (log odds ratio scale) per 38.67 mg/dL reduction in LDL-C / -38.67]

$$= \exp[-10 \times \log(0.942)/\log(0.855) \times \log(0.76)/-38.67] = 0.973. \quad (1)$$

This is a 2.7% reduction in CHD risk.

In contrast, the Mendelian randomization estimate for a 10 mg/dL lowering of Lp(a) is:

exp[Change in Lp(a) in mg/dL ×
Lp(a) Mendelian randomization estimate per 10 mg/dL reduction in Lp(a) / -10]

$$= \exp[-10 \times \log(0.942)/-10] = 0.942.$$

This is a 5.8% reduction in CHD risk. The 95% confidence interval for this quantity is 4.9—6.7%, based on the standard error of the Mendelian randomization estimate. Confidence intervals for the predicted short-term effect of Lp(a) are calculated using a bootstrap procedure as described above, except that we account for the uncertainty in three estimates (two Mendelian randomization estimates, and the effect of statins from clinical trials) by drawing from three normal distributions.

SENSITIVITY ANALYSES

We investigate the use of a different dataset for obtaining the genetic association estimates with CHD risk, and the impact of including different sets of genetic variants in the analyses. We also perform the Mendelian randomization analysis for Lp(a) separately in different study groupings according to the Lp(a) assay used in each study.

Two datasets are used to obtain genetic associations with CHD risk. These are the CHD Exome+ consortium (internal associations), and CARDIoGRAMplusC4D (2015 data release) (11). We considered using CHD associations from the CARDIoGRAM 2011 data release and the CARDIoGRAMplusC4D 2013 data release; however, few associations were available for the genetic variants included in the *LPA* genetic score. The associations of the variants with CHD in the CARDIoGRAMplusC4D consortium were referenced using PhenoScanner (12).

The CARDIoGRAMplusC4D (2015) estimates were obtained in up to 60,801 CAD cases and 123,504 controls, mostly of European descent with some South Asians and East Asians. These estimates are available for over 9 million variants. For each dataset, we used as many of the 43 variants from the *LPA* genetic score as were present in the target dataset. In total, 29 of the 43 variants in the LPA score were available in the CARDIoGRAMplusC4D dataset.

We also varied the number of genetic variants included in the analysis for the effect of Lp(a) on CHD risk. In Supplementary Table S4, we show estimates using each of the CHD association datasets listed above for all 43 variants, also for the top 30 variants (the first 30 chosen by the stepwise selection algorithm), top 20, top 10, top 5, and for the 2 variants (rs10455872 and rs3798220) that were analysed by Clarke *et al* in their previous Mendelian randomization investigation (13). Analyses using CARDIoGRAMplusC4D included all variants for which an association with CHD risk was available: estimates are based on 29 (all variants), 22 (top 30), 15 (top 20), 7 (top 10), 4 (top 5) and 2 (Clarke *et al*) variants.

We see that estimates were similar between the two datasets. Using fewer variants in the analysis led to slightly increased estimates, and more optimistic estimates for the effect of Lp(a) lowering on CHD risk than the primary analyses of this paper. This trend is somewhat expected; using more variants leads to more precise estimates, but can reduce the strength of instruments and hence increase the magnitude of weak instrument bias. In a two-sample setting, this bias is towards the null – as observed here. (Although estimates were obtained in the same studies, the use of healthy participants only for associations with the exposure means that bias operates more similarly to a two-sample setting.) However, discrepancies between estimates using different numbers of genetic variants were small.

Mendelian randomization estimates for CHD risk per 10 mg/dL lowering were similar across the three study groupings considered here (Copenhagen studies, EPIC-CVD, and WOSCOPS, Supplementary Figure S7). This is despite baseline concentrations of Lp(a) differing markedly between the studies. As previously noted with the associations of the individual variants in Supplementary Figure S6, this suggests that assay type affects the absolute concentration of Lp(a) reported, but it does not affect changes in Lp(a) concentration.

COMPARISON OF OBSERVATIONAL AND MENDELIAN RANDOMIZATION ESTIMATES

To calculate the predicted short-term effect of Lp(a) lowering on CHD risk, we first calculate the ratio of the Mendelian randomization estimates for Lp(a) and LDL-C, and then scale the short-term effect of LDL-C lowering taken from statin trials according to this ratio. To provide additional evidence that this approach is reasonable, we consider the ratio between the epidemiologic estimates for Lp(a) and LDL-C, and compare this to the ratio of the Mendelian randomization estimates. In both cases, we use published data from the Emerging Risk Factors Collaboration (ERFC) to obtain the epidemiologic estimates.

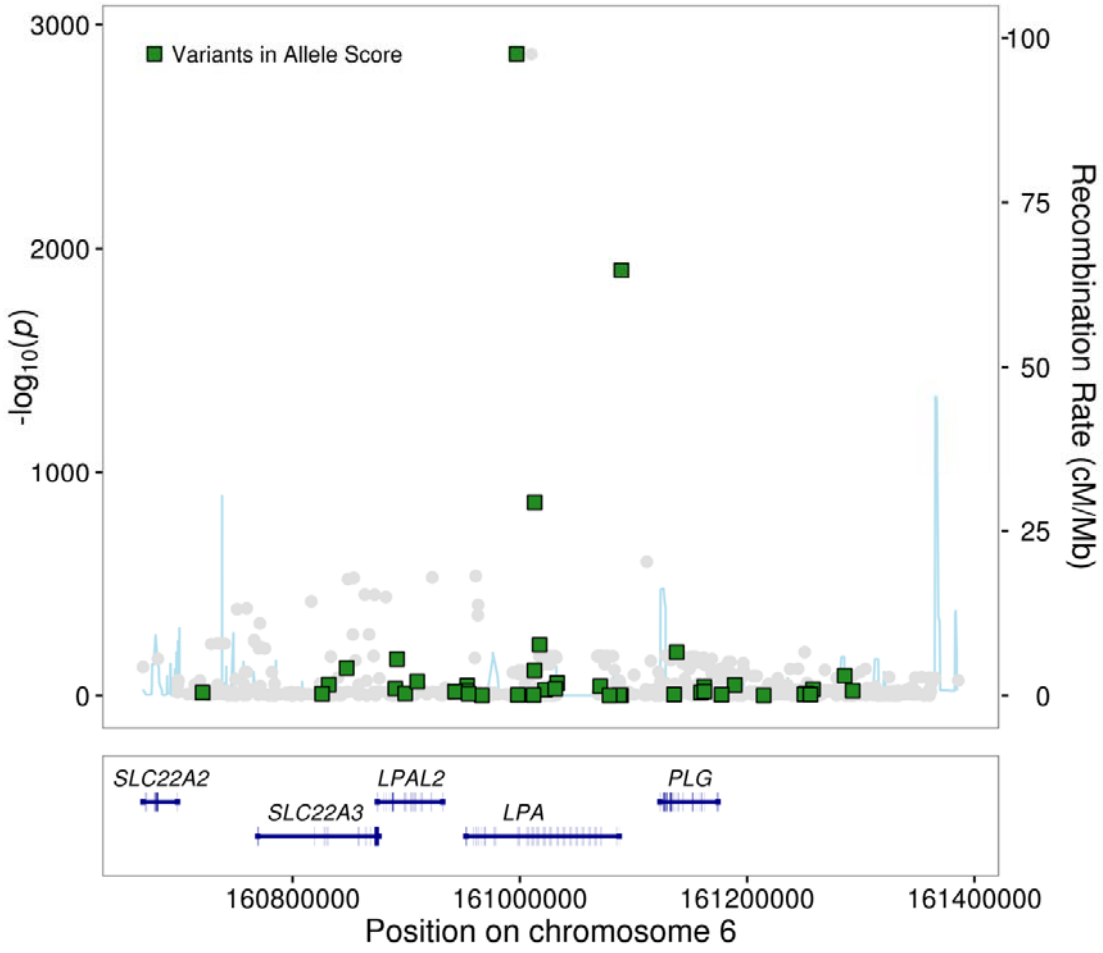
For Lp(a), the ERFC published estimates of the association with CHD risk per 1 standard deviation increase (approximately 3.5-fold increase) in log-transformed Lp(a), and for the upper tertile versus lower tertile of the Lp(a) distribution (3). Neither of these estimates is particularly useful in calculating the association with CHD for absolute changes in Lp(a). We therefore took the values of Lp(a) and CHD risk in the deciles of the distribution of Lp(a) from Figure 2A (left panel) of the ERFC publication, and plotted these on a graph changing the horizontal axis from a log scale to an absolute scale (Supplementary Figure S9). Associations with disease are for the outcome “non-fatal myocardial infarction and coronary death”, and the associations are adjusted for age and sex only. As can be seen from the graph (Supplementary Figure S9), the epidemiologic association is compatible with a linear relationship between log-transformed CHD risk and absolute changes in Lp(a). We then regressed the log-risk of CHD in each decile against the level of Lp(a) with the intercept fixed at zero, and obtained an estimate of the log odds ratio of CHD risk per 10 mg/dL change in Lp(a). To account for regression dilution bias, we divided by 0.87 (the regression dilution ratio for Lp[a] reported in the same paper). We compared this to the estimate provided by the ERFC for usual levels of non-HDL-cholesterol adjusted for age and sex only, and with non-fatal myocardial infarction and coronary death (14). This estimate was already corrected for regression dilution bias.

It is important to note that although the ERFC reported the association between non-HDL-C concentration and risk of CHD in the primary analysis, all studies included in this meta-analysis measured total cholesterol, high-density lipoprotein cholesterol (HDL-C) and triglycerides; and

reported the calculated LDL-C concentration as estimated by the Friedewald equation. The authors point out that any regression model that includes terms for non-HDL-C, HDL-C and triglycerides is a simple mathematical rearrangement of a model that includes terms for calculated LDL-C, HDL-C and triglycerides. Therefore, in the ERFC analysis, the effect of LDL-C is exactly equal to the effect of non-HDL-C on the risk of CHD by definition in the analysis. The authors confirmed this fact by demonstrating that in a sub-sample of 8 studies involving 44,234 individuals, the effect of directly measured LDL-C on the risk of CHD was nearly identical to the effect of non-HDL-C (and calculated LDL-C) per mmol/L.

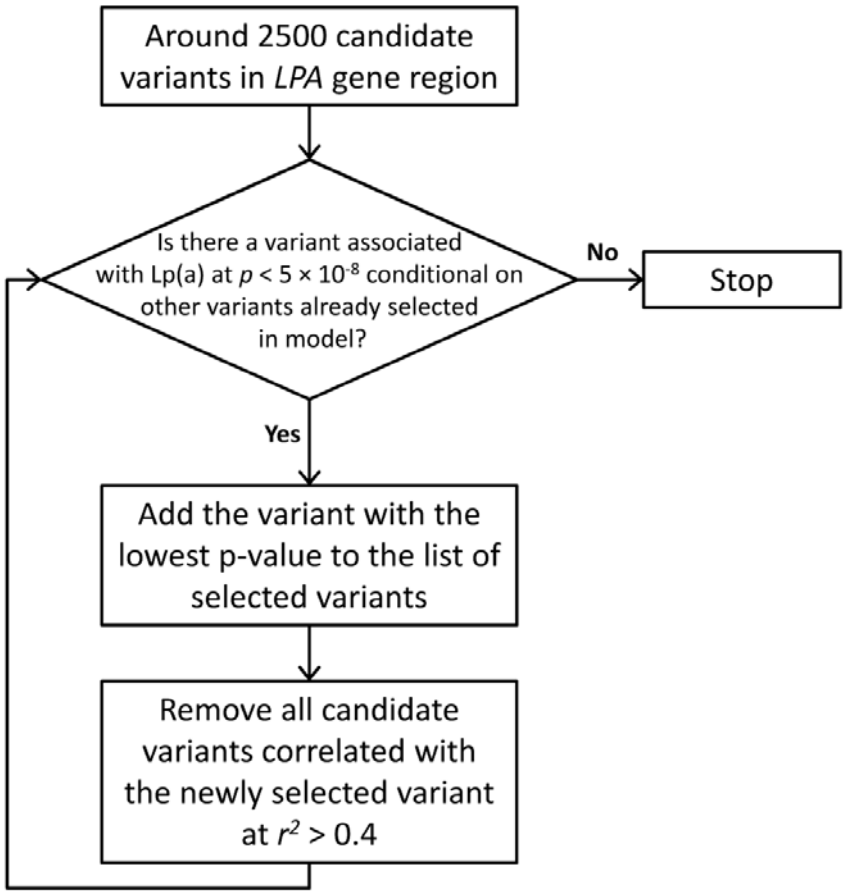
The observational association with CHD risk per 10 mg/dL increase in Lp(a) was a log odds ratio of 0.0383. The observational association with CHD risk per 43 mg/dL increase in non-HDL-C (a 1 standard deviation change) was an odds ratio of 1.63. This translates to a log odds ratio of 0.1136 per 10 mg/dL increase in LDL-C, meaning that the ratio between the epidemiologic observational associations is 2.96. In contrast, the ratio of the Mendelian randomization estimates was 2.63 (=0.1567/0.0597). The similarity of these two ratios (12% difference) gives additional confidence in the assumptions relating to the estimate of short-term lowering of Lp(a).

eFigure 1. Regional association plot



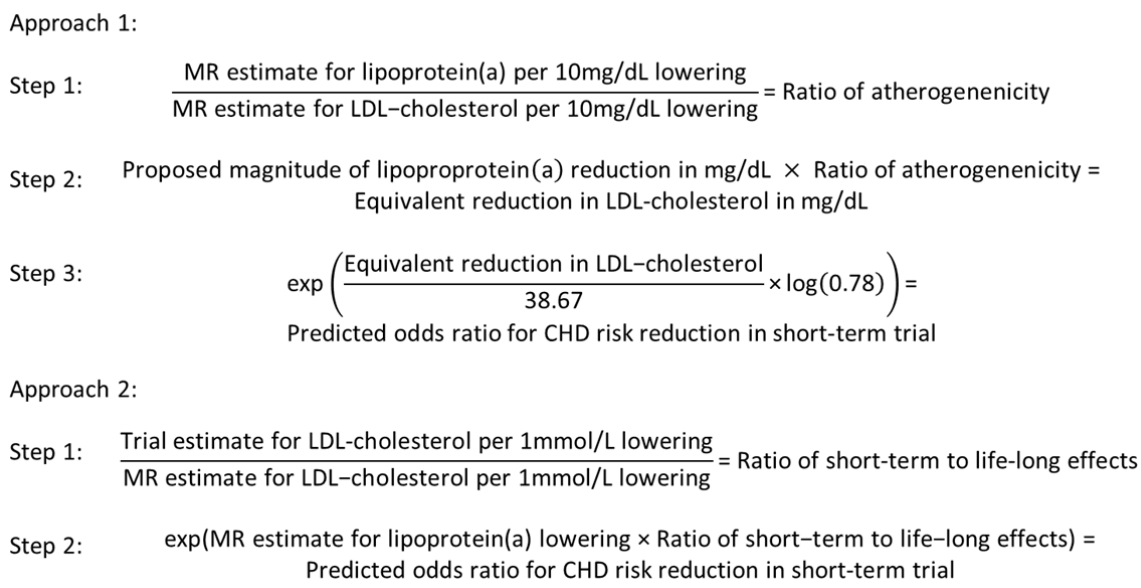
Regional association plot of *LPA* gene region showing 43 genetic variants included in *LPA* genetic score. Horizontal axis indicates position on the chromosome (GRCh37/hg19), vertical axis for points is $\log_{10}$-transformed p-value for univariable (marginal) association of the variant with plasma Lp(a) concentration, vertical axis for the line is the recombination rate (measured in centimorgans per megabase). Green boxes indicate the 43 variants included in the score, which all have pairwise correlations of $r^2 < 0.4$. Grey circles indicate variants in the *LPA* gene region not included in the score.

eFigure 2. Schematic diagram of stepwise selection algorithm



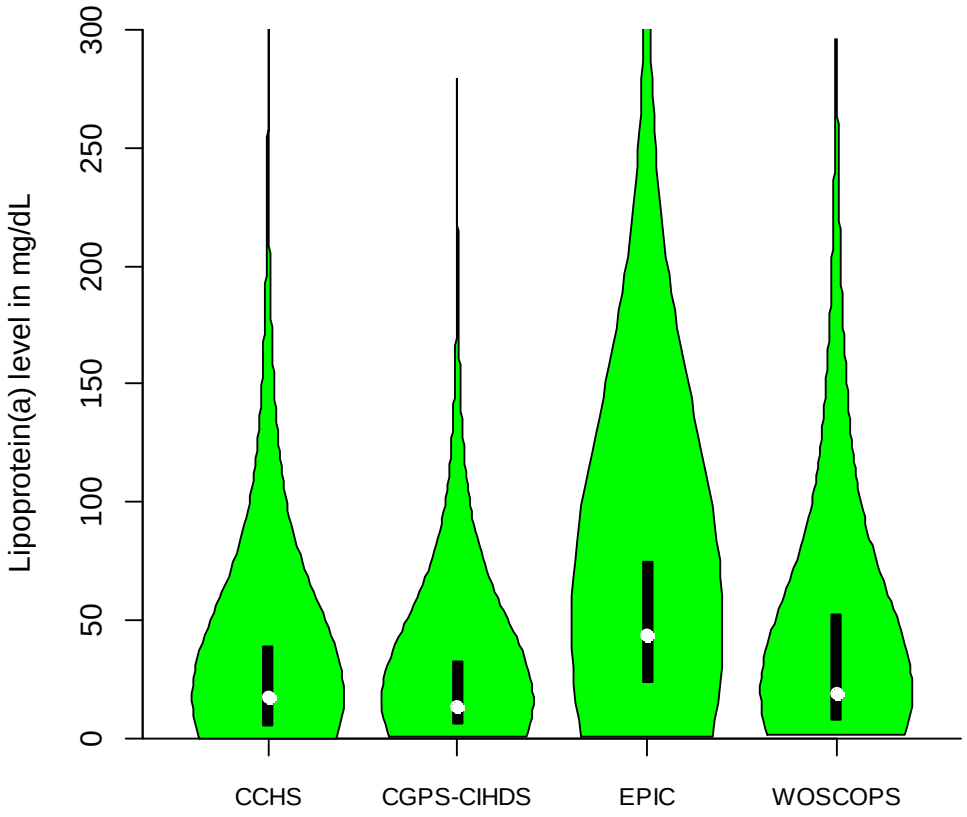
Schematic diagram of forward selection algorithm to select genetic variants to include in *LPA* genetic score

eFigure 3. Summary of analyses for estimating short-term effect of Lp(a)



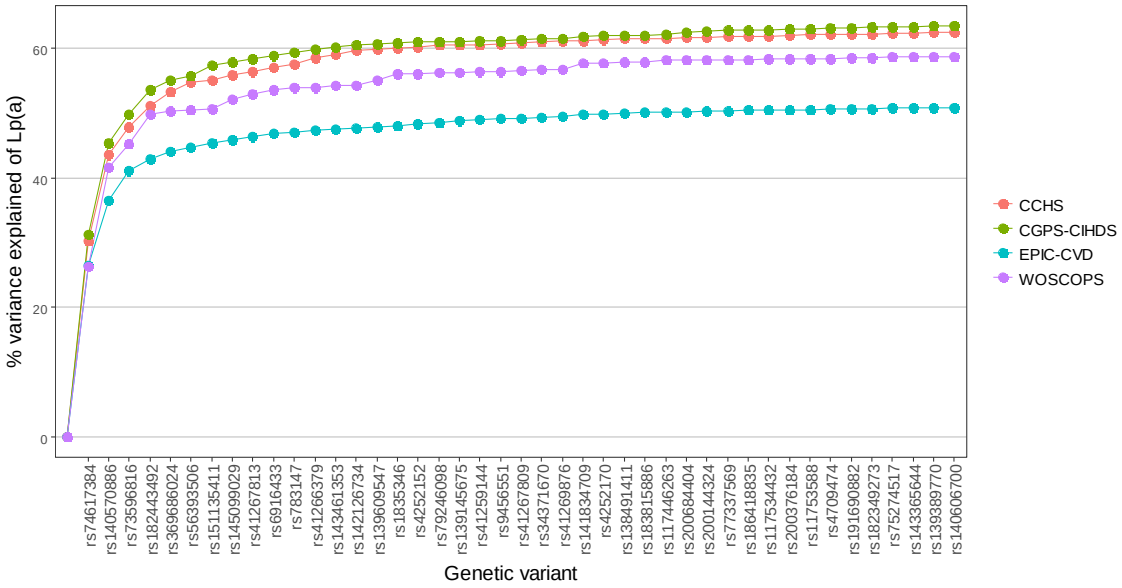
Summary of analyses for estimating effect of lowering Lp(a) in a proposed clinical trial. Both approach 1 and approach 2 give the same overall answer. Note: for LDL-cholesterol, 38.67 mg/dL = 1 mmol/L. The short-term effect of reduction in LDL-cholesterol is an odds ratio of 0.78 per 1 mmol/L (estimate from the Cholesterol Treatment Trialists' Collaboration). All estimates are assumed to be log odds ratios.

eFigure 4. Violin plot of distribution of Lp(a) in each study



Violin plot of distribution of lipoprotein(a) in each study. A violin plot is a box plot with a kernel density plot superimposed on both sides. The plot is truncated at 300 mg/dL; 49 individuals in EPIC-CVD had Lp(a) levels above 300 mg/dL (maximum value: 747.6 mg/dL). The center of the box plot is the median estimate, and the ends of the box are the lower and upper quartiles.

eFigure 5. Variance explained by *LPA* score in each study

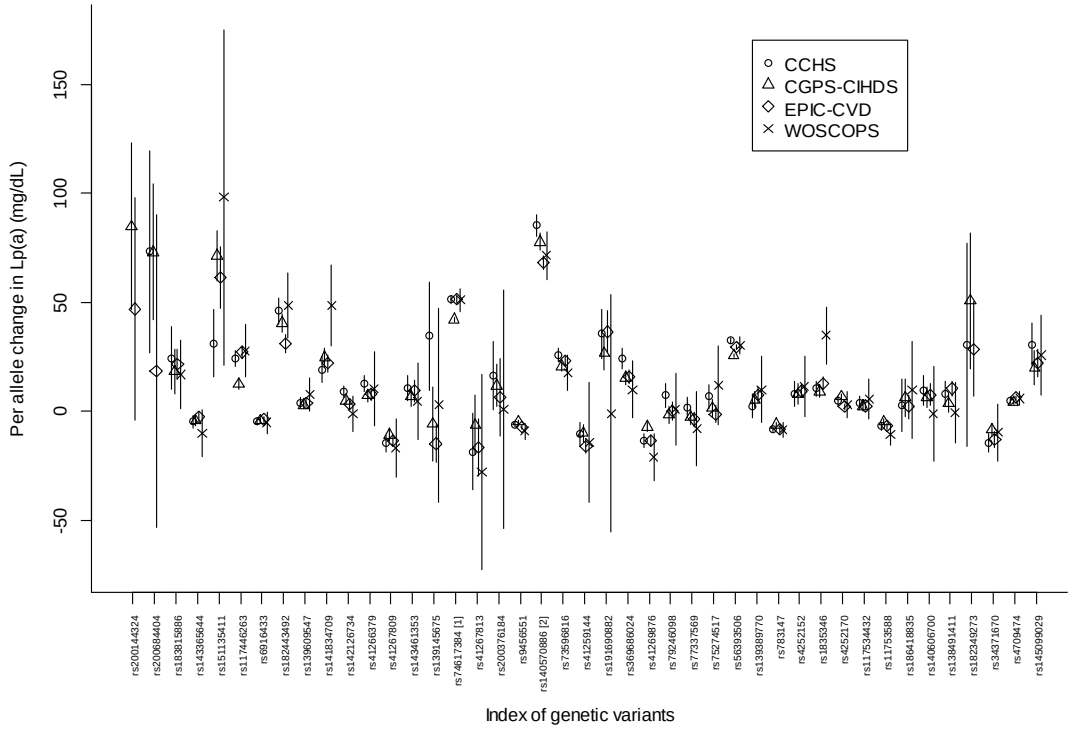


Variance explained by *LPA* genetic score in each study at each step of the variant selection algorithm.

rs74617384 is the closest proxy of rs10455872 (used previously in Clarke et al), with r2 = 0.99 in 1000 Genomes, European populations

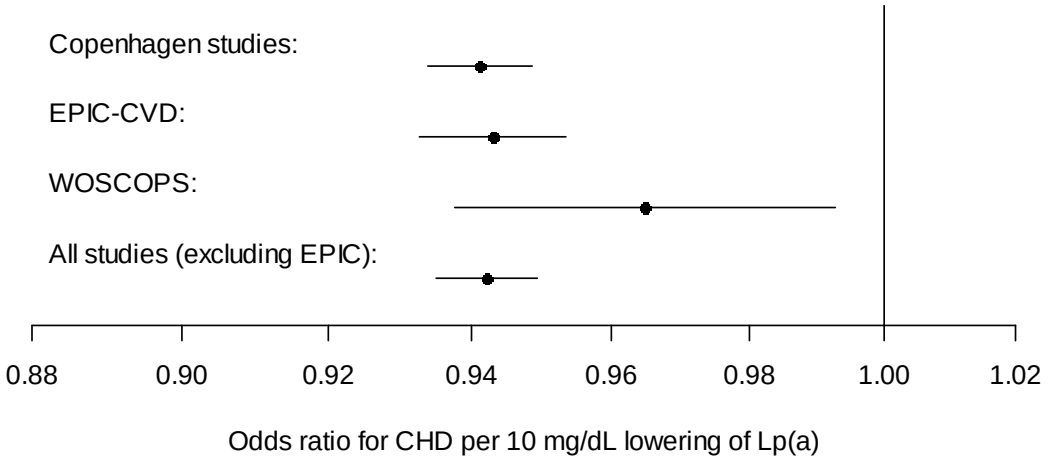
rs140570886 is the closest proxy of rs3798220 (used previously in Clarke et al), with r2 = 0.81 in 1000 Genomes, European populations

eFigure 6. Genetic associations with Lp(a) in each study



Marginal genetic associations of the 43 variants in the *LPA* score with Lp(a) concentration estimated in each study separately.

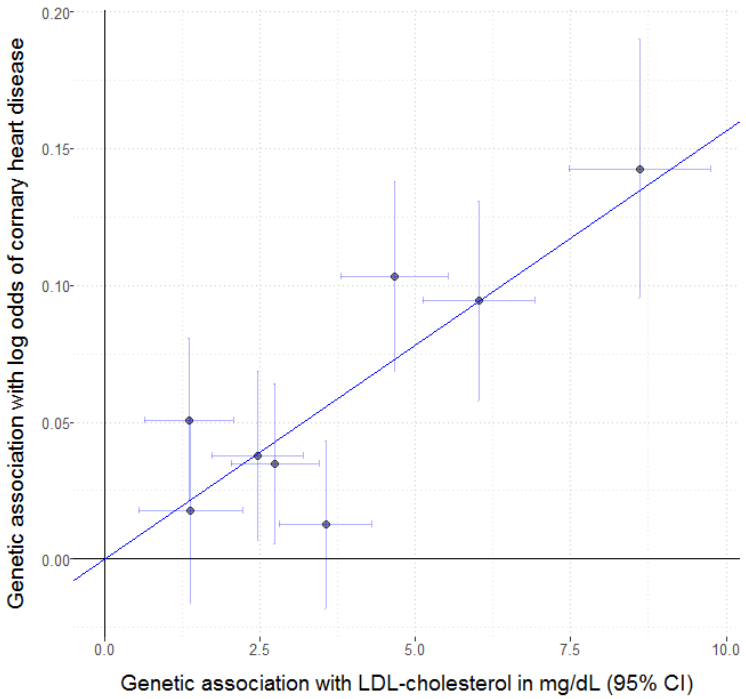
eFigure 7. Mendelian randomization estimates for Lp(a) in study groupings



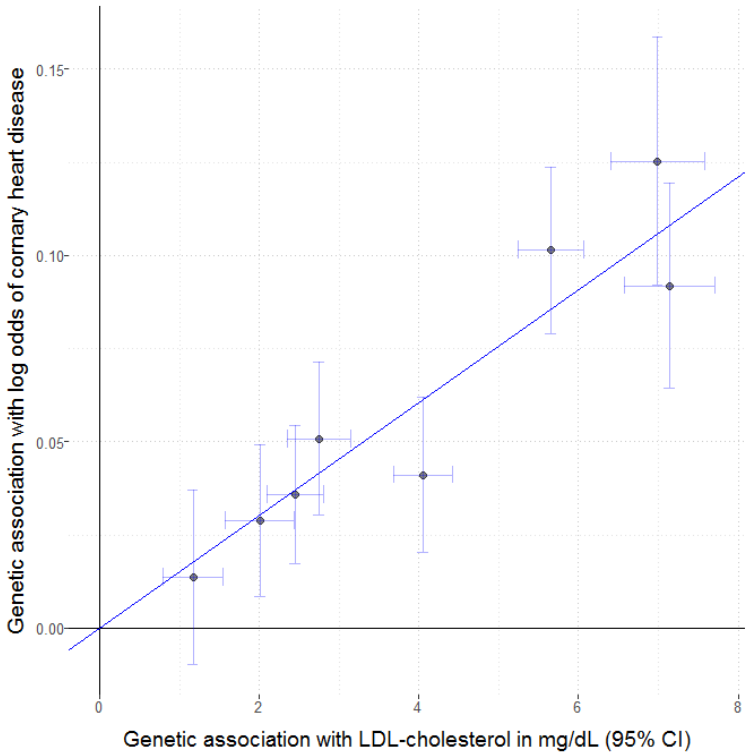
Mendelian randomization estimates per 10 mg/dL lowering of genetically-predicted Lp(a) for study groupings according to Lp(a) assay type.

eFigure 8. Genetic associations of variants in LDL-C genetic score with LDL-C and with CHD risk

A.



B.

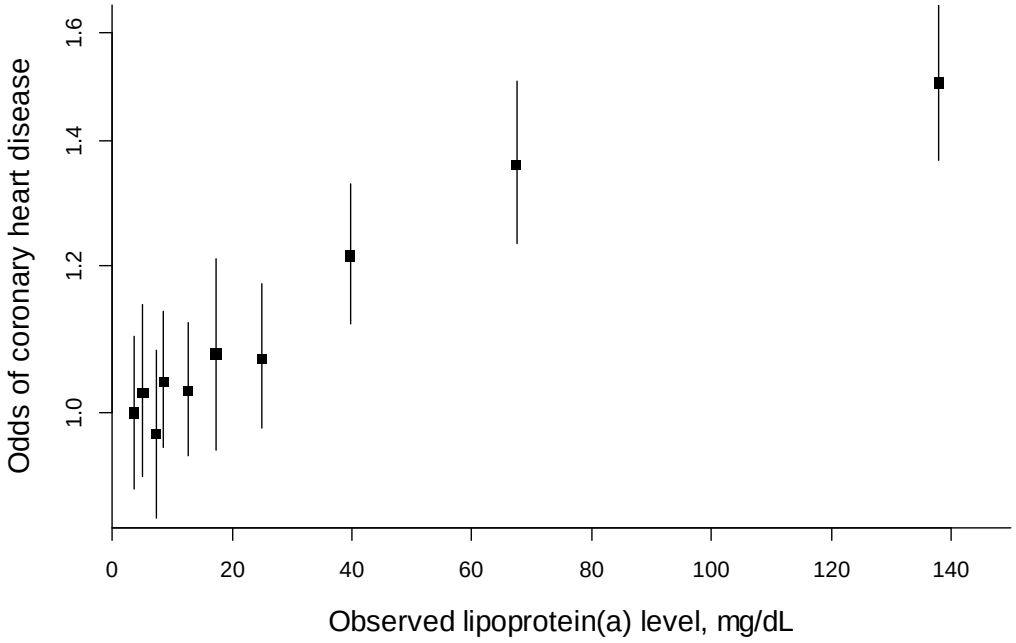


Genetic associations with low-density lipoprotein cholesterol (LDL-C) in mg/dL units and with log odds of CHD for 8 genetic variants in the combined LDL-C genetic score.

A. Genetic associations estimated in the CHD Exome+ consortium.

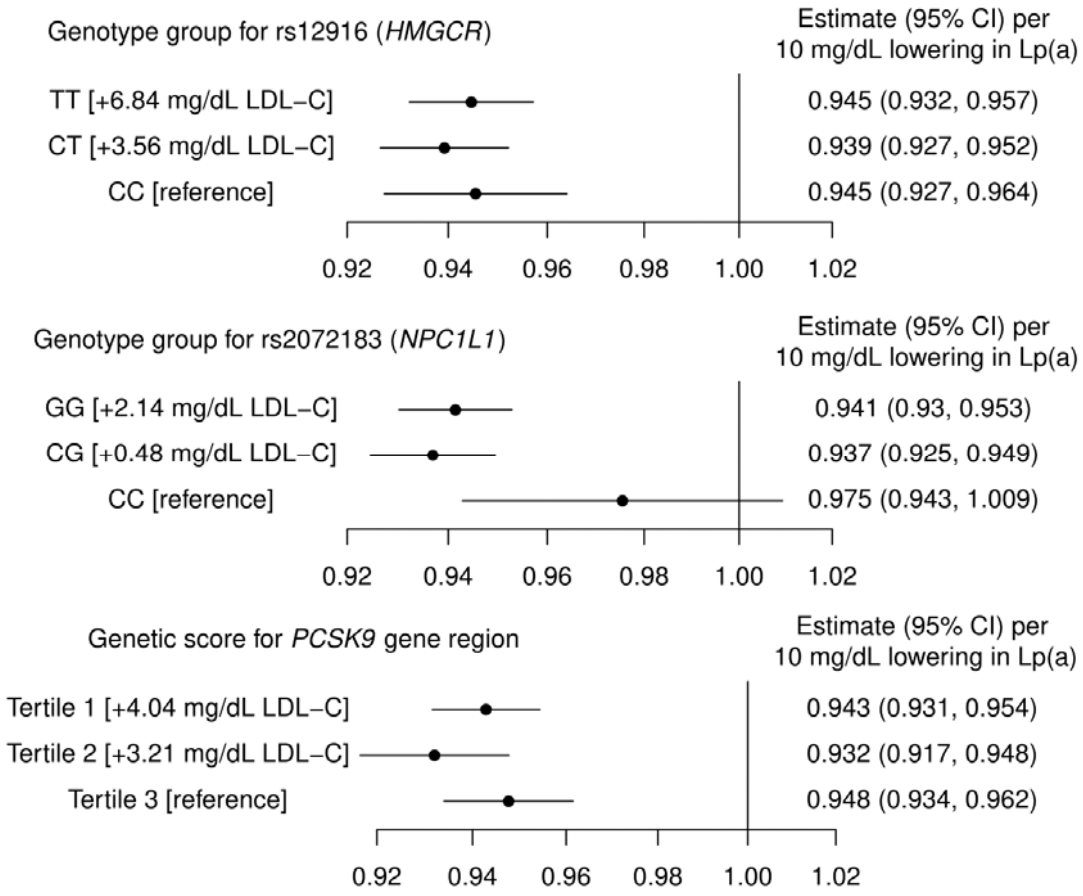
B. Genetic associations estimated in the Global Lipids Genetics Consortium (GLGC) and in the CARDIoGRAMplusC4D consortium (2015 data release). Genetic associations are orientated to the LDL-C increasing allele.

eFigure 9. Observational associations with Lp(a) and CHD risk by deciles



Observational associations with Lp(a) and CHD risk for deciles of the population in the Emerging Risk Factors Collaboration. This figure is adapted from Figure 2A (left panel) of the original publication (3), with the horizontal axis converted to the absolute scale.

eFigure 10. Effect estimates for Lp(a) lowering stratified by LDL-C–related genotypes



Effect estimates for Lp(a) lowering stratified by LDL-C-related genotypes. Odds ratios (95% confidence interval, CI) per 10 mg/dL lowering of Lp(a) in 3 groups defined based on the number of LDL-C lowering alleles each participant inherited for variants located in the *HMGCR*, *PCSK9,* and *NPC1L1* gene regions.

eTable 1. Genetic variants in LDL-C genetic score

Variant	Gene region	Effect allele	Association with LDL-C Estimate (SE)	Association with CHD risk Estimate (SE)
rs12916	*HMGCR*	C	2.729 (0.360)	0.035 (0.015)
rs2479409	*PCSK9*	G	1.354 (0.366)	0.051 (0.015)
rs2072183	*NPC1L1*	C	1.377 (0.425)	0.018 (0.017)
rs1367117	*APOB*	A	3.544 (0.374)	0.013 (0.016)
rs4299376	*ABCG5/G8*	G	2.458 (0.375)	0.038 (0.016)
rs629301	*SORT1*	T	4.654 (0.437)	0.103 (0.018)
rs4420638	*APOE*	G	6.003 (0.461)	0.094 (0.019)
rs6511720	*LDLR*	G	8.583 (0.578)	0.143 (0.024)

Per-allele genetic associations (beta-coefficients and standard errors [SE]) with low-density lipoprotein cholesterol (LDL-C) from linear regression (mg/dL change) and with coronary heart disease (CHD) risk from logistic regression (log odds ratio) for 8 genetic variants in combined LDL-C genetic score estimated in the CHD Exome+ consortium.

eTable 2. Distribution of Lp(a) in each study with respect to potential treatment cut-offs

	Number (%) of individuals in study with Lp(a) level exceeding:				
Study	50 mg/dL	100 mg/dL	120 mg/dL	150 mg/dL	200 mg/dL
CCHS	1434 (19.4%)	525 (7.1%)	326 (4.4%)	143 (1.9%)	30 (0.4%)
CGPS-CIHDS	1614 (16.2%)	366 (3.7%)	191 (1.9%)	66 (0.7%)	7 (0.1%)
EPIC-CVD	7281 (45.8%)	2518 (15.8%)	1645 (10.3%)	871 (5.5%)	306 (1.9%)
WOSCOPS	260 (25.6%)	124 (12.2%)	85 (8.4%)	44 (4.3%)	17 (1.7%)

Distribution of Lp(a) in each study with respect to potential treatment cut-offs (in controls / participants without CHD at baseline only).

eTable 3. List of genetic variants included in the *LPA* score

Chromosome: Position (GRCh37/hg19)	rsID	Effect allele	Other allele	Minor allele frequency	Conditional association with Lp(a), mg/dL Beta (SE)	Marginal association with Lp(a), mg/dL Beta (SE)	Marginal association with CHD risk, log odds ratio Beta (SE)
6:160997118	rs74617384[1]	T	A	0.072	42.4 (0.5)	46.2 (0.4)	0.292 (0.027)
6:161013013	rs140570886[2]	C	T	0.011	80.2 (0.8)	71.6 (1.1)	0.552 (0.067)
6:161017363	rs73596816	A	G	0.034	19.2 (0.6)	22.1 (0.7)	0.087 (0.040)
6:160891897	rs182443492	A	C	0.009	36.8 (1.0)	36.3 (1.4)	0.218 (0.079)
6:161032800	rs369686024	A	G	0.014	19.2 (0.8)	16.8 (1.1)	0.039 (0.061)
6:161089307	rs56393506	T	C	0.169	12.4 (0.4)	27.9 (0.3)	0.144 (0.019)
6:160831796	rs151135411	A	G	0.001	69.5 (2.7)	57.8 (3.9)	0.389 (0.223)
6:161292838	rs145099029	C	A	0.003	17.8 (1.8)	21.8 (2.2)	0.089 (0.122)
6:160998199	rs41267813	A	G	0.001	-58.8 (2.9)	-12.8 (4.1)	0.238 (0.230)
6:160890350	rs6916433	T	A	0.140	-4.7 (0.3)	-4.1 (0.4)	-0.039 (0.021)
6:161137990	rs783147	A	G	0.450	-2.0 (0.3)	-7.3 (0.2)	-0.045 (0.015)
6:160953137	rs41266379	C	T	0.020	7.1 (0.7)	8.5 (0.9)	0.026 (0.054)
6:160954800	rs143461353	T	C	0.008	13.1 (1.0)	8.3 (1.4)	0.028 (0.083)
6:160942926	rs142126734	A	G	0.049	7.5 (0.5)	4.9 (0.6)	-0.046 (0.034)
6:160899049	rs139609547	-	A	0.054	4.4 (0.4)	3.4 (0.5)	-0.028 (0.032)
6:161162290	rs1835346	G	A	0.022	5.2 (0.7)	10.4 (0.8)	0.028 (0.050)
6:161159366	rs4252152	G	T	0.014	9.1 (0.9)	8.5 (1.1)	0.04 (0.0670)
6:161078894	rs79246098	C	T	0.010	6.2 (0.9)	0.9 (1.3)	-0.123 (0.074)
6:160966559	rs139145675	A	G	0.001	-22.5 (2.4)	-8.6 (3.7)	-0.259 (0.233)
6:161022107	rs41259144	T	C	0.011	-9.6 (0.8)	-12.8 (1.2)	-0.026 (0.073)
6:161012805	rs9456551	C	T	0.350	3.6 (0.2)	-5.9 (0.3)	-0.039 (0.015)
6:160953642	rs41267809	G	A	0.022	-6.6 (0.6)	-12.4 (0.9)	-0.165 (0.053)
6:161257953	rs34371670	T	C	0.016	-8.4 (0.7)	-11.0 (1.0)	-0.010 (0.058)
6:161070653	rs41269876	A	C	0.028	-8.2 (0.6)	-10.5 (0.7)	-0.114 (0.044)
6:160909667	rs141834709	A	T	0.009	8.7 (1.0)	22.0 (1.3)	0.109 (0.076)
6:161162406	rs4252170	C	T	0.082	3.2 (0.4)	4.2 (0.5)	0.023 (0.027)
6:161251940	rs138491411	G	A	0.012	5.0 (0.8)	7.2 (1.2)	0.031 (0.069)
6:160720804	rs183815886	C	G	0.003	14.8 (1.8)	19.0 (2.5)	0.091 (0.141)
6:160847571	rs117446263	A	G	0.022	-5.2 (0.6)	19.6 (0.8)	0.097 (0.049)
6:160543317	rs200684404	T	C	0.000	67.7 (9.2)	66.7 (12.4)	-0.171 (0.956)
6:160493099	rs200144324	T	C	0.000	81.5 (11.3)	71.0 (15.6)	0.568 (1.416)
6:161087652	rs77337569	G	T	0.013	5.2 (0.8)	-2.2 (1.1)	-0.172 (0.066)
6:161214526	rs186418835	A	G	0.004	-9.7 (1.5)	3.0 (2.2)	-0.104 (0.123)
6:161177443	rs117534432	T	C	0.036	3.3 (0.5)	2.6 (0.7)	0.041 (0.039)
6:161011999	rs200376184	C	G	0.001	17.5 (2.7)	11.6 (3.8)	0.163 (0.237)
6:161189071	rs11753588	A	G	0.109	-2.4 (0.3)	-5.9 (0.4)	-0.011 (0.024)
6:161285760	rs4709474	G	A	0.490	1.7 (0.2)	4.9 (0.2)	0.015 (0.015)
6:161031132	rs191690882	A	G	0.002	-13.2 (1.9)	31.3 (2.7)	0.415 (0.153)
6:161255668	rs182349273	G	A	0.000	34.4 (5.7)	35.0 (8.5)	-0.896 (0.607)
6:161088956	rs75274517	A	G	0.010	-6.5 (1.0)	1.4 (1.2)	-0.13 (0.077)
6:160825930	rs143365644	T	A	0.035	3.7 (0.5)	-3.8 (0.7)	-0.059 (0.04)
6:161135746	rs139389770	G	T	0.011	-5.2 (0.9)	5.5 (1.2)	0.042 (0.071)
6:161250301	rs140606700	G	A	0.007	6.4 (1.2)	7.3 (1.5)	-0.061 (0.095)
6:160961137	rs3798220	T	C	0.014		-51.2 (1.0)	-0.376 (0.060)
6:161010118	rs10455872	G	A	0.072		46.2 (0.4)	0.292 (0.027)

Per-allele association estimates for 43 genetic variants included in the *LPA* score, and 2 variants previously used in Mendelian randomization investigation (Clarke *et al*). Variants are listed in their order of inclusion in the genetic score from the stepwise selection algorithm.

[1] This variant is the closest proxy of rs10455872, with r2 = 0.99 in 1000 Genomes, European populations

[2] This variant is the closest proxy of rs3798220, with r^2 = 0.81 in 1000 Genomes, European populations

eTable 4. Sensitivity analysis for effect of Lp(a) on CHD risk using different numbers of genetic variants

Number of variants	CHD association dataset	
	Exome+	CARDIoGRAMplusC4D (2015)
All variants (43)	0.942 (0.933, 0.951)	0.948 (0.941, 0.955)
Top 30 [1]	0.941 (0.933, 0.949)	0.947 (0.940, 0.954)
Top 20	0.942 (0.933, 0.950)	0.947 (0.939, 0.955)
Top 10	0.940 (0.931, 0.948)	0.947 (0.936, 0.959)
Top 5	0.938 (0.929, 0.948)	0.943 (0.934, 0.952)
2 variants (Clarke) [2]	0.937 (0.927, 0.946)	0.938 (0.930, 0.946)

Mendelian randomization odds ratio estimate (95% confidence interval) for coronary heart disease (CHD) risk per 10 mg/dL lowering in lipoprotein(a) [Lp(a)] using six different choices of genetic score and two different datasets for obtaining genetic associations with CHD risk.

[1] First 30 variants included in the genetic score from the stepwise selection algorithm.

[2] These are the 2 variants (rs10455872 and rs3798220) that were analysed by Clarke *et al* in their previous Mendelian randomization investigation (13).

eReferences.

1. Danesh J, Saracci R, Berglund G, Feskens E, Overvad K, Panico S, et al. EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. Eur J Epidemiol. 2007;22(2):129–141.

2. Marcovina SM, Albers JJ, Scanu AM, Kennedy H, Giaculli F, Berg K, et al. Use of a reference material proposed by the International Federation of Clinical Chemistry and Laboratory Medicine to evaluate analytical methods for the determination of plasma lipoprotein(a). Clin Chem. 2000 Dec;46(12):1956–67.

3. Emerging Risk Factors Collaboration, Erqou S, Kaptoge S, Perry PL, Di Angelantonio E, Thompson A, et al. Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvascular mortality. JAMA. 2009 Jul 22;302(4):412–23.

4. Staley JR, Burgess S. Semiparametric methods for estimation of a non-linear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. Genet Epidemiol. 2017;

5. Easton DF, Peto J, Babiker AGAG. Floating absolute risk: An alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. Stat Med. 1991 Jul;10(7):1025–35.

6. Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, Thompson A, et al. Major lipids, apolipoproteins, and risk of vascular disease. J Am Med Assoc. 2009;302(18):1993–2000.

7. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol. 2011;40(3):755–764.

8. 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.

9. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;

10. Cholesterol Treatment Trialists' (CTT) Collaboration. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170 000 participants in 26 randomised trials. The Lancet. 2010 Nov;376(9753):1670–81.

11. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121–30.

12. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics. 2016;doi:10.1093/bioinformatics/btw373.

13. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med. 2009;361(26):2518–2528.

14. Emerging Risk Factors Collaboration, Di Angelantonio E, Sarwar N, Perry P, Kaptoge S, Ray KK, et al. Major lipids, apolipoproteins, and risk of vascular disease. JAMA. 2009 Nov 11;302(18):1993–2000.