Distinguishing genetic correlation from causation across 52 diseases and complex traits

Luke J. O'Connor[1,2] and Alkes L. Price[1,3,4]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA
[2]Program in Bioinformatics and Integrative Genomics, Harvard Graduate School of Arts and Sciences, Cambridge, MA
[3]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
[4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA
Correspondence should be addressed to L.J.O. (loconnor@g.harvard.edu) or A.L.P. (aprice@hsph.harvard.edu)

Abstract

Mendelian randomization (MR) is widely used to identify causal relationships among heritable traits, but it can be confounded by genetic correlations reflecting shared etiology. We propose a model in which a latent causal variable mediates the genetic correlation between two traits. Under the LCV model, trait 1 is *fully genetically causal* for trait 2 if it is perfectly genetically correlated with the latent causal variable, and *partially genetically causal* for trait 2 if the latent variable has a higher genetic correlation with trait 1 than with trait 2. To quantify the degree of partial genetic causality, we define the genetic causality proportion (gcp), enabling us to describe genetically causal relationships non-dichotomously. We fit this model using mixed fourth moments $E(\alpha_1^2\alpha_1\alpha_2)$ and $E(\alpha_2^2\alpha_1\alpha_2)$ of marginal effect sizes for each trait, exploiting the fact that if trait 1 is causal for trait 2 then SNPs with large effects on trait 1 will have correlated effects on trait 2, but not vice versa. We performed simulations under a wide range of genetic architectures and determined that LCV, unlike state-of-the-art MR methods, produced well-calibrated false positive rates and reliable gcp estimates in the presence of genome-wide genetic correlations and asymmetric genetic architectures. We applied LCV to GWAS summary statistics for 52 traits (average N=326k), identifying fully or partially genetically causal effects (1% FDR) for 63 pairs of traits. Results consistent with the published literature included genetically causal effects on myocardial infarction (MI) for LDL, triglycerides and BMI. Novel findings included a genetically causal effect of LDL on bone mineral density, consistent with clinical trials of statins in osteoporosis. Our results demonstrate that it is possible to distinguish between genetic correlation and causation using genetic data.

Introduction

Mendelian Randomization (MR) is widely used to identify potential causal relationships among heritable traits, which can be valuable for designing disease interventions.[1–10] Genetic variants that are significantly associated with one trait, the "exposure," are used as genetic instruments to test for a causal effect on a second trait, the "outcome." If the exposure does have a causal effect on the outcome, then variants affecting the exposure should affect the outcome proportionally.

For example, the MR approach has been used to show that LDL[3,11] and triglycerides[4] (but not HDL[3]) have a causal effect on coronary artery disease (CAD). However, a challenge is that genetic variants can affect both traits pleiotropically, and these pleiotropic effects can induce a genetic correlation, especially when the exposure is polygenic.[2,9,10,12–14] This challenge can potentially be addressed using curated sets of genetic variants that aim to exclude pleiotropic effects, but curated sets of genetic variants are unavailable for most complex traits (curation is more likely to be successful for molecular traits with simpler genetic architectures). One potential solution has been to apply MR bidirectionally, using genome-wide significant SNPs for each trait in turn.[9,15,16] This approach relies on the assumption that if there is no genetically causal relationship, then genome-wide significant SNPs for each trait are equally likely to have correlated effects; however, this assumption can be violated due to differences in trait polygenicity or GWAS sample size.

We introduce a latent causal variable (LCV) model, under which the genetic correlation between two traits is mediated by a latent variable having a causal effect on each trait. We compare the magnitude of these effects, defining trait 1 as *partially genetically causal* for trait 2 when the effect of the latent variable on trait 1 is larger than its effect on trait 2; by comparing the size of these effects we define the *genetic causality proportion* (gcp), which is 0 when there is no partial causality and 1 when trait 1 is *fully genetically causal* for trait 2 (meaning that there is a genetic correlation of one between trait 1 and the causal variable). In simulations we confirm that LCV, unlike other methods, avoids confounding due to genetic correlations, even under asymmetric genetic architectures with unequal polygenicity or unequal power between the two traits. Applying LCV to GWAS summary statistics for 52 diseases and complex traits (average N=326k), we identify both genetically causal relationships that are consistent with the published literature and novel genetically causal relationships.

Results

Overview of methods

The latent causal variable (LCV) model assumes that the genetic correlation between trait 1 and trait 2 is mediated by a latent variable L having causal effects on trait 1 and trait 2 (Figure 1). We define trait 1 as *fully genetically causal* for trait 2 when the genetic component of trait 1 is equal to L, so that every genetic perturbation to trait 1 produces a proportional change in trait 2. We define trait 1 as *partially genetically causal* for trait 2 when the effect of the latent variable on trait 1 is stronger than its effect on trait 2. By comparing the magnitude of these effects, we define the *genetic causality proportion*, gcp, of trait 1 on trait 2, which is 0 when there is no partial causality and 1 when trait 1 is fully genetically causal for trait 2. A high value of gcp indicates that trait 1 is either strongly or perfectly genetically correlated with an underlying causal trait; it suggests that interventions targeting trait 1 are likely to have an effect on trait 2, to the extent that they mimic genetic perturbations to trait 1. (However, we caution that mechanistic hypotheses are also required before designing disease interventions, as the success of an intervention may depend on its mechanism of action and on its timing relative to disease progression.) An intermediate positive value of gcp indicates that functional insights into the genetic architecture of trait 1 may also provide insights into the etiology of trait 2. Our goals are to test for statistically significant partial genetic causality and to estimate gcp. We exploit the fact that if trait 1 is genetically causal for trait 2, then many SNPs affecting trait 1 will have proportional effects on trait 2, but

not vice versa. In particular, we compare the mixed fourth moments $E(\alpha_1^2\alpha_1\alpha_2)$ and $E(\alpha_2^2\alpha_1\alpha_2)$ of marginal effect sizes for each trait, adjusting for the genetic correlation between traits. We derive a statistical test for partial genetic causality and a posterior mean estimator of gcp using the estimated mixed fourth moments.

Under the latent causal variable (LCV) model (Figure 1) we define the genetic causality proportion (gcp) as the number x such that:

$$\frac{q_2^2}{q_1^2} = (\rho_g^2)^x, \tag{1}$$

where q_1 and q_2 denote effects of L on trait 1 and trait 2 and the genetic correlation ρ_g is equal to $q_1 q_2$. gcp is positive when trait 1 is partially genetically causal for trait 2. When gcp = 1, trait 1 is fully genetically causal for trait 2: q_1 is equal to 1, and $q_2 = \rho_g$ is the causal effect size of trait 1 on trait 2 (we note that it is possible to have gcp = 1 with a weak causal effect size). Conversely, when gcp = −1, trait 2 is fully genetically causal for trait 1. We derive a relationship between the mixed fourth moments of the marginal effect size distribution and the parameters q_1 and q_2 in the LCV model, allowing us to test for partial genetic causality and to estimate gcp: let the random variable α_k denote the marginal effect of a SNP on Y_k, including effects mediated by L and effects not mediated by L. Under the LCV model,

$$E(\alpha_1^3\alpha_2) = \kappa_\pi q_1^3 q_2 + 3\rho_g, \tag{2}$$

where π is the effect of a SNP on L and $\kappa_\pi = E(\pi^4) - 3$ is the excess kurtosis of π (see Online Methods). Our method exploits this excess kurtosis; when κ_π is zero (such as when π is normally distributed), we are unable to test for genetic causality or to estimate gcp (indeed, the model is not identifiable when π is normally distributed; see Supplementary Note). We estimate ρ_g using a modified version of cross-trait LD score regression,[14] and we use a modified version of LD score regression[17] to normalize the summary statistics. In order to estimate the gcp, we construct statistics $S(x)$ based on the difference between the estimated mixed fourth moments for each possible value of gcp = x; these estimates are corrected for possible sample overlap (see Online Methods). We estimate the variance of these statistics using a block jackknife and obtain an approximate likelihood function for the gcp. We compute a posterior mean estimate of gcp (and a posterior standard deviation) using a uniform prior on $[-1, 1]$. We test the null hypothesis of no partial genetic causality using the statistic $S(0)$. Details of the method are provided in the Online Methods section; we have released open source software implementing the method (see URLs).

Simulations with no LD: comparison with existing methods

To compare the calibration and power of LCV with existing causal inference methods, we performed simulations involving simulated summary statistics with no LD. We compared four methods: LCV, random-effect two-sample MR[5] (denoted MR), MR-Egger[7] and Bidirectional MR[9] (see Online Methods). We applied each method to simulated GWAS summary statistics (N = 100k individuals in each of two non-overlapping cohorts; M = 50k independent SNPs[18]) for two heritable traits (h^2 = 0.3), generated under the LCV model. LCV uses LD score regression[17] to normalize the summary statistics and cross-trait LD score regression[14] to estimate the genetic correlation; for simulations with no LD, we use constrained-intercept LD score regression[14] for both of these steps. In each simulation, approximately 320 SNPs on average were genome-wide significant for each

trait, explaining roughly half of h^2; MR, MR-Egger and Bidirectional MR rely exclusively on these genome-wide significant SNPs. A detailed description of these simulations is provided in the Online Methods section.

First, we performed null simulations (gcp = 0) with uncorrelated pleiotropic effects and zero genetic correlation. 1% of SNPs were causal for both traits (with independent effect sizes), 4% were causal for trait 1 but not trait 2, and 4% were causal for trait 2 but not trait 1. Results are displayed in Figure 2a (scatterplots of estimated SNP effects are displayed in Figure S1a). LCV produced conservative p-values (0.0% false positive rate at $\alpha = 0.05$); our normalization of the test statistic can lead to conservative p-values when the genetic correlation is low (see Online Methods; analyses of real phenotypes are restricted to genetically correlated traits). All three MR methods produced well-calibrated p-values. Even though the "exclusion restriction" assumption of MR– that there is no pleiotropy– is violated here, these results confirm that uncorrelated pleiotropic effects do not confound random-effect MR at large sample sizes;[19] we caution that pleiotropy is known to produce false positives if standard errors are computed using a less conservative fixed-effect approach.[20] In these simulations, all methods except LCV used the set of approximately 320 SNPs (on average) that were genome-wide significant ($p < 5 \times 10^{-8}$), either for trait 1 only (MR and MR-Egger) or for both traits (Bidirectional MR); varying the significance threshold produced similar results (Table S1).

Second, we performed null simulations with a nonzero genetic correlation: 1% of SNPs had causal effects on L, and L had effects $q_1 = q_2 = \sqrt{0.2}$ on each trait (so that $\rho_g = 0.2$); 4% of SNPs were causal for trait 2 but not trait 1, and 4% of SNPs were causal for trait 1 but not trait 2. Because the per-SNP heritability was the same on average for shared causal SNPs as for non-shared causal SNPs, these SNPs were equally likely to be genome-wide significant, and ~ 20% of significant SNPs affected both traits with correlated effect sizes. Results are displayed in Figure 2b (scatterplots in Figure S1b). Because of these correlated-effect SNPs, MR and MR-Egger both exhibited severely inflated false positive rates; in contrast, Bidirectional MR and LCV produced well-calibrated or modestly conservative p-values. Thus, correlated pleiotropic effects violate the MR exclusion restriction assumption in a manner that leads to false positives, as polygenic genetic correlations can produce correlations among genome-wide significant SNPs (Figure S1b). These simulations also violate the MR-Egger assumption that the magnitude of pleiotropic effects on trait 2 are independent of the magnitude of effects on trait 1 (the "InSIDE" assumption),[7] as SNPs with larger effects on L have larger effects on both trait 1 and trait 2 on average, consistent with known limitations.[20]

Third, we performed null simulations with a nonzero genetic correlation and differential polygenicity in the non-shared genetic architecture between the two traits: 1% of SNPs were causal for L with effects $q_1 = q_2 = \sqrt{0.2}$ on each trait, 2% were causal for trait 1 but not trait 2, and 8% were causal for trait 2 but not trait 1. Thus, the likelihood that a SNP would be genome-wide significant was higher for causal SNPs affecting trait 1 only than for causal SNPs affecting trait 2 only. We hypothesized that this ascertainment bias would cause Bidirectional MR to incorrectly infer that trait 1 was causal for trait 2. Indeed, Bidirectional MR (as well as other MR methods) exhibited inflated false positive rates, while LCV produced modestly conservative p-values (Figure 2c). We confirmed that the correlation between SNP effect sizes differs for SNPs that are significant for trait 1 and SNPs that are significant for trait 2 (Figure S1c).

Fourth, we performed null simulations with a nonzero genetic correlation and differential power for the two traits, reducing the sample size from 100k to 20k for trait 2. 0.5% of SNPs were causal

for L with effects $q_1 = q_2 = \sqrt{0.5}$ on each trait, 8% were causal for trait 1 but not trait 2, and 8% were causal for trait 2 but not trait 1. Because per-SNP heritability was higher for shared causal SNPs than for non-shared causal SNPs, shared causal SNPs but not non-shared causal SNPs were likely to reach genome-wide significance in the smaller trait 1 sample (N = 20k), while both shared and non-shared causal SNPs were likely to reach genome-wide significance in the trait 2 sample (N = 100k); thus, we hypothesized that Bidirectional MR would incorrectly infer that trait 1 was causal for trait 2. Indeed, Bidirectional MR (as well as other MR methods) exhibited inflated false positive rates, while LCV produced well-calibrated p-values (Figure 2d; scatterplots in Figure S1d).

Finally, we simulated fully genetically causal (gcp = 1) and partially genetically causal (gcp = 0.5) genetic architectures, to assess the power of each method. In the fully genetically causal case, 5% of SNPs were causal for trait 1, with proportional effects on trait 2 resulting in a genetic correlation of 0.1, and an additional 5% of SNPs were causal for trait 2 but not trait 1. In the partially genetically causal case, 5% of SNPs were causal for each trait individually, and 5% of SNPs were causal for L, explaining different amounts of heritability for each trait so that the genetic correlation was 0.1 and the gcp was 0.5. MR, Bidirectional MR and LCV (but not MR-Egger) attained very high power in the fully genetically causal case (Figure 2e; scatterplots in Figure S1e). In the partially genetically causal case, MR and LCV attained high power, followed by Bidirectional MR and MR-Egger respectively (Figure 2f; scatterplots in Figure S1f).

In summary, we determined using simulations with no LD that LCV produced well-calibrated null p-values in the presence of a nonzero genetic correlation, unlike MR and MR-Egger. LCV also avoided confounding when polygenicity or power differed between the two traits, unlike Bidirectional MR and other methods. In non-null simulations, LCV attained high power to detect a causal or partially genetically causal effect.

Simulations with no LD: LCV model violations

To investigate potential limitations of our approach, we performed simulations involving genetic architectures that violate the key assumption of the LCV model, that a single variable fully mediates the genetic correlation between two traits. Analogous to simulations reported in Figure 2, each trait had heritability 0.3 and sample size 100k (non-overlapping), with 50k SNPs and no LD. First, we performed null simulations under a model with two latent causal variables, L_1 and L_2, where L_1 had effect size 0.4 on trait 1 and 0.1 on trait 2 but L_2 had effect size 0.1 on trait 1 and 0.4 on trait 2. Thus, SNPs affecting L_1 had larger effects on trait 1 while SNPs affecting L_2 had larger effects on trait 2. These simulations can be viewed as null, because the two intermediaries collectively explained the same proportion of heritability for both traits. 2% of SNPs were causal for each latent causal variable, and an additional 4% of SNPs were causal for each trait individually. Results are displayed in Figure 3a. LCV produced conservative p-values, indicating that heterogeneity in the relative effect sizes of shared causal SNPs does not necessarily confound LCV.

Second, we repeated these simulations with differential polygenicity between the two latent causal variables: 1% of SNPs were causal for L_1, but 4% of SNPs were causal for L_2. This form of differential polygenicity is distinct from Figure 2c, which involves differential polygenicity between the non-shared genetic components of each trait. We expected that LCV would produce inflated false positive rates, as the sparse intermediary would influence the mixed fourth moments more than the polygenic intermediary. Indeed, LCV consistently produced false positives, similar to MR, MR-Egger and Bidirectional MR (Figure 3b). Thus, a limitation of our method (and existing methods) is that it can be confounded by genetic architectures involving heterogenous relative effect

sizes when the relative effects (i.e. α_1^2/α_2^2, which was higher for L_1 than for L_2) are coupled to the effect magnitudes (i.e. $\alpha_1^2\alpha_2^2$, which was also higher for L_1). This type of effect can be viewed as an asymmetric violation of the key assumption needed to derive equation (2), namely that the squared values of direct effects (γ_k^2) are uncorrelated with the squared values of mediated effects (π^2; see Online Methods). In contrast, Figure 3a involves a symmetric violation of the assumption (i.e., $corr(\gamma_1^2, \pi^2) = corr(\gamma_2^2, \pi^2) \neq 0$), leading to a violation of (2) but not false positives. Despite the fact that heterogeneity of relative effect sizes coupled with differential polygenicity can lead to false positives for LCV, genetic causality remains the most parsimonious explanation for low LCV p-values.

Third, to confirm our hypothesis that heterogeneity only confounds LCV when it is coupled with differential polygenicity, we performed null simulations in which SNP effects were drawn from a mixture of normal distributions. 4% of SNPs were causal for trait 1 only or trait 2 only, and 1% of SNPs were causal for both traits following a multivariate normal distribution with correlation 0.5, so that the relative effect sizes of shared causal SNPs were heterogenous (these SNPs explained 20% of heritability for each trait). An interpretation for this model is that shared causal SNPs act on the two traits via many different intermediaries. Results are displayed in Figure 3c. LCV produced p-values that were well-calibrated, similar to Bidirectional MR. MR and MR-Egger produced inflated p-values, similar to Figure 2b.

Fourth, we added differential polygenicity between the two traits, not coupled with the heterogeneity; 2% of SNPs were causal for trait 1 only and 8% of SNPs were causal for trait 2 only (Figure 3d). Because the differential polygenicity was not coupled with the heterogeneity, LCV produced well-calibrated p-values, while MR, MR-Egger and Bidirectional MR produced inflated p-values, similar to Figure 2c.

In summary, we determined in simulations involving LCV model violations that LCV and existing methods were confounded by complex genetic architectures involving heterogenous relative SNP effect sizes when this heterogeneity was coupled with differential polygenicity. On the other hand, heterogeneity did not confound LCV when relative SNP effects were independent of effect magnitudes, and existing methods were confounded by less complex genetic architectures in addition to complex genetic architectures.

Simulations with LD: assessing calibration and power

To further assess the calibration and power of our test for partial genetic causality and the unbiasedness and precision of our gcp estimator, we performed simulations involving real LD patterns; we note that LD can potentially impact the performance of our method, which uses a modified version of LD score regression[14,17] to normalize effect size estimates and to estimate genetic correlations. Because existing methods exhibited major limitations in simulations with no LD (Figure 2 and Figure 3), we restricted these simulations to the LCV method. We used real genotypes from the interim UK Biobank release[25] (N = 145k European-ancestry samples, M = 596k genotyped SNPs) to compute a banded LD matrix, simulated causal effect sizes for each of two traits at these SNPs, and simulated summary statistics (inclusive of LD) for each trait using the asymptotic sampling distributions.[21] We included correlations between the noise components of the summary statistics for each trait so as to mimic fully overlapping GWAS cohorts with total phenotypic correlation equal to the genetic correlation. Our initial null simulations included identical effect sizes of L on each trait ($q_1 = q_2 = \sqrt{0.2}$), 0.1% of SNPs causal for L (explaining 20% of trait h^2), 0.4% of SNPs causal for trait 1 but not trait 2 (and respectively for trait 2 but not trait 1), $h^2 = 0.3$ for each trait

and N = 100k for each cohort; we varied each of these parameters in turn. We set the proportion of causal SNPs to be lower in these simulations than in simulations without LD so as to roughly match the total number of causal SNPs and the proportion of associated SNPs (inclusive of LD) at a given p-value threshold. Further details of the simulations are provided in the Online Methods section.

First, we performed null simulations (gcp = 0) at various values of the genetic correlation ρ_g (Table 1a-c and Table S3a-e). False positive rates were approximately well-calibrated, with conservative p-values at $\rho_g = 0$ (consistent with Figure 2a) and slightly inflated p-values at higher values of ρ_g. This slight inflation was not observed in simulations with no LD because we used constrained-intercept LD score regression to estimate heritability in those simulations (variable-intercept LD score regression cannot be used when there is no LD), leading to highly precise heritability estimates; however, constrained-intercept LD score regression can produce upwardly biased heritability estimates in practice. We repeated our simulations with LD using constrained-intercept LD score regression to estimate h_g^2 (we still used variable-intercept LD score regression to estimate genetic covariance); noise in the heritability estimates was reduced (mean Z score for nonzero h_g^2 increased from $Z_h \approx 8$ to $Z_h \approx 15$), and test statistic inflation was eliminated (Table S4a-c). Thus, the slight inflation in Table 1a,c is a result of noise in the heritability estimates. To ensure that this issue would not affect our analyses of real traits, we restricted those analyses to traits with highly significant heritability ($Z_h > 7$; see below). We focus our remaining simulations on genetic architectures that include a nonzero genetic correlation, but analogous simulations with zero genetic correlation are also provided in Table S3.

Second, we performed null simulations with uncorrelated pleiotropic effects, in addition to genetic correlation of 0.2. 0.2% of SNPs had direct effects on both traits with independent effect sizes, 0.2% of SNPs had direct effects on each trait (but not both), and 0.1% of SNPs had effects on L. False positive rates were approximately well-calibrated (Table 1d and Table S3f); similar to Table 1a-c, there was slight inflation as a result of noisy heritability estimates, which was eliminated when we used constrained-intercept LD score regression (Table S4d).

Third, we performed null simulations with differential polygenicity in the non-shared genetic architecture between the two traits (Table 1e and Table S3g); we note that in simulations with no LD, differences in polygenicity (in the presence of genetic correlation) confounded Bidirectional MR, but not LCV (Figure 2c). 0.2% and 0.8% of SNPs were causal for trait 1 and trait 2, respectively. False positive rates were similar to Table 1a, with slight inflation (which was eliminated when we used constrained-intercept LD score regression, Table S4e). Slightly increased inflation was observed when the difference in polygenicity was very large (0.1% and 1.6% of SNPs causal for each trait; Table S3h); we believe that this 16× difference in polygenicity represents an extreme scenario for real traits.

Fourth, we performed null simulations with differential power between the GWAS cohorts; we note that in simulations with no LD, differences in sample size (in the presence of genetic correlation) confounded bidirectional MR, but not LCV (Figure 2d). We specified a 5× difference in sample size (N_1 = 20k and N_2 = 100k). Results are displayed in Table 1f and Table S3i. Similar to Table 1a, we observed slight inflation in false positive rates, which was eliminated when we used constrained-intercept LD score regression (Table S4f). The amount of inflation was greatly increased when we further reduced N_1 to 4k (Table S3j); at this sample size, LD score regression produced unreliable heritability estimates using either variable-intercept LD score regression (average Z_h = 1.4) or constrained-intercept LD score regression (average Z_h = 2.2; Table S4g).

We generally recommend running LCV on datasets with heritability Z score $Z_h > 7$, which may preclude running LCV on small GWAS. We also performed secondary simulations under various parameter settings, including simulations involving zero genetic correlation, different environmental correlation values and different heritability values, with results that were concordant with other simulations (Table S3k-s).

Fifth, we explored the effect of population stratification in null simulations using individual-level UK Biobank genotypes from chromosome 1 (M = 43k). We added strong environmental stratification along the first principal component (explaining 1% and 2% of phenotypic variance for traits 1 and 2 respectively); this principal component approximately corresponds to latitude of origin.[22] False positive rates were severely inflated, and point estimates of gcp were severely biased (Table 1g and Table S6a-b). When residualizing summary statistics on PC1 loadings,[23] false positive rates were approximately well-calibrated (Table S6c-d). These results emphasize the importance of correcting for population stratification in order to draw valid conclusions about causal relationships between traits.

Sixth, we simulated fully causal (gcp = 1) and partially causal (gcp = 0.5) genetic architectures, to assess the power of LCV. LCV attained high power in the fully causal case and moderately high power in the partially causal case (Table 1h-i and Table S3t-u). Estimates of gcp were biased toward zero in the fully causal case (an expected consequence of our uniform prior on $[-1, 1]$), but approximately unbiased in the partially causal case. When we varied key simulation parameters in fully causal simulations, LCV attained moderate to high power across a wide range of realistic parameter values, including the sample size in both cohorts, the size of the causal effect, and the polygenicity of the causal trait (Table S3v-aa). As expected, there was no power when the genetic architecture of the causal trait was infinitesimal (Table S3bb; see Online Methods). For a putative causal trait whose genetic architecture is unknown, is difficult to predict whether LCV will be well-powered to detect a causal effect of that trait at a given sample size, since the power of LCV depends on the polygenicity of the causal trait, as well as the size of the causal effect and other unknown parameters.

Seventh, to further assess the unbiasedness of gcp posterior mean (and variance) estimates, we performed simulations in which the true value of gcp was drawn uniformly from $[-1, 1]$ and ρ_g was drawn uniformly from $[-0.5, 0.5]$ distribution. In order to be maximally realistic, these simulations also included differential polygenicity (similar to Table 1e) and differential power (similar to Table 1f); other parameters were identical to Table 1a. To mimic the process that we applied to real traits, we restricted to simulations with evidence for nonzero genetic correlation ($p < 0.05$) and evidence for partial causality ($p < 0.001$). We expected posterior-mean estimates to be unbiased in the sense that $E(\text{gcp}|\hat{\text{gcp}}) = \hat{\text{gcp}}$ (which differs from the usual definition of unbiasedness, that $E(\hat{\text{gcp}}|\text{gcp}) = \text{gcp}$).[24] Thus, we binned these simulations by $\hat{\text{gcp}}$ and plotted the mean value of gcp within each bin (Figure S2a). We determined that mean gcp within each bin was concordant with $\hat{\text{gcp}}$. Accordingly, when we regressed the true values of gcp on the estimates, the slope was close to 1 (Table S5). In addition, the root mean squared error (RMSE) was 0.15, approximately consistent with the root mean posterior variance estimate (RMPV) of 0.13 (Table S5).

In summary, in null simulations under the LCV model with real LD, we confirmed that LCV produces approximately well-calibrated null p-values under a wide range of genetic architectures with nonzero genetic correlation; these simulations included uncorrelated pleiotropic effects, differential polygenicity, high phenotypic correlations, and differential GWAS power. Some p-value inflation was observed when heritability estimates were noisy, but this is addressed in analyses of

real traits by restricting to traits with highly significant heritability ($Z_h > 7$). In non-null simulations with real LD, LCV attained high power to detect causal effects under a wide range of realistic genetic architectures, and it produced approximately unbiased posterior mean gcp estimates with well-calibrated posterior standard errors.

Application to real phenotypes

We applied our method to GWAS summary statistics for 52 diseases and complex traits, including summary statistics for 36 UK Biobank traits[25,26] computed using BOLT-LMM[27] (average N=428k) and 16 other traits (average N=54k) (see Table S7 and Online Methods). The 52 traits were selected based on the significance of their heritability estimates ($Z_h > 7$), and traits with very high genetic correlations ($|\rho_g| > 0.9$) were pruned, retaining the trait with higher Z_h. As in previous work, we excluded the MHC region from all analyses, due to its unusually large effect sizes and long-range LD patterns.[17] Of the 430 trait pairs (31%) with a nominally significant genetic correlation ($p < 0.05$), 63 trait pairs had significant evidence of full or partial genetic causality (FDR < 1%). Results for selected traits are displayed in Figure 4. 30 of these 63 trait pairs had gcp estimates less than 0.6, and many more had gcp estimates that were significantly less than 1, demonstrating that genetic causality is highly non-dichotomous. Results for the 63 significant trait pairs are reported in Table S8, and complete results are reported in Table S9. Myocardial infarction (MI) had a nominally significant genetic correlation with 31 other traits, of which six had significant evidence (FDR < 1%) for a fully or partially genetically causal effect on MI (Table 2); there was no evidence for a genetically causal effect of MI on any other trait. Consistent with previous studies, these traits included LDL,[3,11] triglycerides[4] and BMI,[28] but not HDL.[3] The effect of BMI was also consistent with prior MR studies,[28–31] although these studies did not attempt to account for pleiotropic effects (also see ref. 32, which detected no effect). There was also evidence for a genetically causal effect of high cholesterol, which was unsurprising (due to the high genetic correlation with LDL) but noteworthy because of its strong genetic correlation with MI, compared with LDL and triglycerides. There was also evidence for a genetically causal effect of fasting glucose, consistent with an MR study that reported a causal effect of type 2 diabetes (T2D) on CAD accounting for pleiotropic effects on other known CAD risk factors;[33] that study did not detect a causal effect on CAD for fasting glucose specifically, possibly due to limited power. The result for HDL and MI did not pass our significance threshold (FDR < 1%), but was nominally significant ($p = 0.02$, Table S9); we residualized HDL summary statistics on summary statistics for LDL, BMI and triglycerides, determining that residualized HDL remained genetically correlated with MI but showed no evidence of genetic causality ($p = 0.8$). On the other hand, most of the six traits with significant causal effects on MI remained significant after conditioning (Table S10). We confirmed that self-reported MI in UK Biobank was highly genetically correlated with CAD in CARDIoGRAM consortium data[35] ($\hat{\rho}_g = 1.34(0.25)$; not significantly different from 1).

We also detected evidence for a fully or partially genetically causal effect of hypothyroidism on MI (Table 2). Although hypothyroidism is not as well-established a cardiovascular risk factor as high LDL HDL, its genetic correlation with MI is comparable (Table 2), and this effect is mechanistically plausible.[36,37] While this result was robust in the conditional analysis (Table S10), and there was no strong evidence for a genetically causal effect of hypothyroidism on lipid traits (Table S9), it is possible that this effect is mediated by lipid traits. A recent MR study of thyroid hormone levels, at ~ 20× lower sample size than the present study, provided evidence for a genetically causal effect on LDL but not CAD.[38] On the other hand, clinical trials have demonstrated that treatment

of subclinical hypothyroidism using levothyroxine leads to improvement in several cardiovascular risk factors.[39–43] We also detected evidence for a genetically causal effect of hypothyroidism on T2D (Table S8), consistent with a longitudinal association between subclinical hypothyroidism and diabetes incidence.[44]

We identified four traits with evidence for a fully or partially genetically causal effect on hypertension (Table 2), which is genetically correlated with MI ($\hat{\rho}_g = 0.49(0.10)$). These included genetically causal effects of BMI, consistent with the published literature,[9,34] as well as triglycerides and HDL. The genetically causal effect of HDL indicates that there exist major metabolic pathways affecting hypertension with little or no corresponding effect on MI. The positive partially genetically causal effect of reticulocyte count, which had a low gcp estimate ($\hat{\text{gcp}} = 0.41(0.13)$), is likely related to the substantial genetic correlation of reticulocyte count with triglycerides ($\hat{\rho}_g = 0.33(0.05)$) and BMI ($\hat{\rho}_g = 0.39(0.03)$).

We detected evidence for a negative genetically causal effect of LDL on bone mineral density (BMD; Table 2). A meta-analysis of seven randomized clinical trials reported that statin administration increased bone mineral density, although these clinical results have generally been interpreted as evidence of a shared pathway affecting LDL and BMD.[45] Moreover, familial defective apolipoprotein B leads to high LDL cholesterol and low bone mineral density.[46] To further validate this result, we performed two-sample MR using 8 SNPs that were previously used to show that LDL affects CAD (in ref. 3; see Online Methods), finding modest evidence for a negative causal effect ($p = 0.04$). Because there is a clear mechanistic hypothesis linking each of these variants to LDL directly, this analysis provides separate evidence for a genetically causal effect (LCV does not prioritize variants that are more likely to satisfy instrumental variable assumptions). We also detected a partially genetically causal effect of height on BMD, with a lower gcp estimate (Table 2).

We detected evidence for a fully or partially genetically causal effect of triglycerides on five cell blood traits: mean cell volume, platelet distribution width, reticulocyte count, eosinophil count and monocyte count (Table 2). These results highlight the pervasive effects of metabolic pathways, which can induce genetic correlations with cardiovascular phenotypes. For example, shared metabolic pathways may explain the high genetic correlation of reticulocyte count with MI ($\hat{\rho}_g = 0.31(0.06)$) and hypertension ($\hat{\rho}_g = 0.27(0.04)$).

Finally, it has been reported that polygenic autism risk is positively genetically correlated with educational attainment[14] (and cognitive ability,[47] a highly genetically correlated trait[50]), possibly consistent with the hypothesis that common autism risk variants are maintained in the population by balancing selection.[48,49] If balancing selection involving a trait related to educational attainment explained a majority of autism risk, we would expect that most common variants affecting autism risk would affect educational attainment, leading to a partially genetically causal effect of autism on educational attainment. However, we detected no evidence of a partially genetically causal effect of autism on college education ($\hat{\text{gcp}} = 0.13(0.13)$, $\hat{\rho}_g = 0.23(0.07)$; Table S9); thus, balancing selection acting on educational attainment or a related trait is unlikely to explain the high prevalence of autism.

We discuss additional significant results (Table S8) in the Supplementary Note.

Discussion

We have introduced a latent causal variable (LCV) model to identify causal relationships among genetically correlated pairs of complex traits. We applied LCV to 52 traits, finding that many

trait pairs do exhibit partially or fully genetically causal relationships. Our results included several novel findings, including a genetically causal effect of LDL on bone mineral density (BMD) which suggests that lowering LDL may have additional benefits besides reducing the risk of cardiovascular disease.

Our method represents an advance for two main reasons. First, LCV reliably distinguishes between genetic correlation and full or partial genetic causation. Unlike existing MR methods, LCV provided well-calibrated false positive rates in null simulations with a nonzero genetic correlation, even in simulations with differential polygenicity or differential power between the two traits. Thus, positive findings using LCV are more likely to reflect true causal effects. Second, we define and estimate the genetic causality proportion (gcp) to quantify the degree of causality. This parameter, which provides information orthogonal to the genetic correlation or the causal effect size, enables a non-dichotomous description of the causal architecture. Even when both MR and LCV provide significant p-values, the p-value alone is consistent with either fully causal or partially causal genetic architectures, limiting its interpretability; our gcp estimates appropriately describe the range of likely hypotheses.

This study has several limitations. First, the LCV model includes only a single intermediary and can be confounded in the presence of multiple intermediaries, in particular when the intermediaries have differential polygenicity. Indeed, some trait pairs appear to show evidence for multiple intermediaries (Table S8). Nonetheless, causality or partial genetic causality provide a more parsimonious explanation for estimated genetically causal effects, especially when the gcp estimate is high. Second, because LCV models only two traits at a time, it cannot be used to identify conditional effects given observed confounders.[4,52] This approach was used, for example, to show that triglycerides affect coronary artery disease risk conditional on LDL.[4] However, it is less essential for LCV to model observed genetic confounders, since LCV explicitly models a latent genetic confounder. Third, LCV is not currently applicable to traits with small sample size and/or heritability, due to low power as well as incorrect calibration. However, GWAS summary statistics at large sample sizes have become publicly available for increasing numbers of diseases and traits, including UK Biobank traits.[27] Fourth, the LCV model can be confounded confounded by shared population stratification, so it is critical for association statistics to be corrected for stratification. Fifth, while many trait pairs have high gcp estimates, it is not clear whether most of these trait pairs reflect fully or partially genetically causal relationships. A gcp of 1 and a gcp of ~ 0.7 would be interpreted differently, as a gcp of ~ 0.7 suggests that only some interventions on trait 1 will modify trait 2, depending on their mechanism of action. This type of uncertainty can be reduced at higher sample size, but not eliminated entirely. Sixth, even full genetic causality must be interpreted with caution before designing disease interventions, as interventions may fail to mimic genetic perturbations. For example, factors affecting a developmental phenotype such as height might need to be modified at the correct developmental time point in order to have any effect; this limitation broadly applies to all methods for inferring causality using genetic data. Seventh, power might be increased by modeling LD explicitly, exploiting the fact that SNPs with higher LD, especially in active regulatory regions, have larger marginal effect sizes on average.[17] Nonetheless we observed high power to detect causal effects for many trait pairs. Eighth, power might also be increased by including rare and low-frequency variants; even though these SNPs explain less complex trait heritability than common SNPs,[18,53] they may contribute significantly to power if the genetic architecture among these SNPs is more sparse than among common SNPs. Ninth, we cannot infer whether inferred causal effects are linear. For example, it is plausible that BMI would

have a small effect on MI risk for low-BMI individuals and a large effect for high-BMI individuals, but this type of nonlinearity cannot be gleaned from summary statistics (unless MI summary statistics were stratified by BMI). Tenth, MR-style analyses have been applied to gene expression,[54–56] and the potential for confounding due to pleiotropy in these studies could possibly motivate the use of LCV in this setting, but LCV is not applicable to molecular traits, which may be insufficiently polygenic for the LCV random-effects model to be well-powered. Finally, we have not exhaustively benchmarked LCV against every published MR method, but have restricted our simulations to the most widely used MR methods. We note that there exist methods that aim to improve robustness by excluding or effectively down-weighting variants whose causal effect estimates appear to be outliers;[6,8,10] however, we believe that any method that relies on genome-wide significant SNPs for a single one trait is likely to be confounded by genetic correlations (Figure 2).

Despite these limitations, for most pairs of complex traits we recommend using LCV instead of MR. When the exposure is a complex trait, MR is likely to be confounded by genetic correlations, and it may be impossible to identify valid instruments. However, there are several scenarios in which MR should be used, either in addition to or instead of LCV. First, when associated variants are available that are likely to represent valid instruments because they have a mechanistically direct effect on the exposure, it is appropriate to perform MR. For example, an MR analysis identified a causal effect of vitamin D on multiple sclerosis, utilizing genetic variants near genes with well-characterized effects on vitamin D synthesis, metabolism and transport; these variants all provided consistent estimates of the causal effect.[58] As another example, cis-eQTLs can be used as genetic instruments to test for an effect of gene expression because they are unlikely to be confounded by processes mediated in trans, motivating applications of MR and related methods to gene expression[54–56] (however, these studies also have other limitations, such as the high likelihood that GWAS SNPs may approximately colocalize with an eQTL[55,57]). Second, when prior knowledge about likely pleiotropic factors is available, it is appropriate to perform MR in addition to LCV, either restricting to variants without overt pleiotropic effects or correcting for these effects in a multivariate regression model.[4,52] Third, when one of the traits has low significance for nonzero heritability, LCV may produce unreliable estimates and MR should be used either instead of or in addition to LCV. Finally, well-powered MR studies can be used to show that two traits do not have a strong, fully genetically causal relationship, as confounding due to pleiotropy is more likely to lead to false positives than false negatives. In each case, MR should be performed with multiple genetic variants, a bidirectional analysis[9,15] should be performed to reduce the potential for confounding due to genetic correlations, and consistency of causal effect estimates across variants should be assessed both manually and analytically.[10]

Acknowledgements

We are grateful to Ben Neale, Soumya Raychaudhuri, Chirag Patel, Sek Kathiresan, Bogdan Pasaniuc and Hilary Finucane for helpful discussions, and to Po-Ru Loh and Steven Gazal for producing BOLT-LMM summary statistics for UK Biobank traits. This research was conducted using the UK Biobank Resource under Application #16549 and funded by NIH grants R01 MH107649, U01 CA194393 and R01 MH101244.

URLs

Open-source software implementing our method is available at github.com/lukejoconnor/LCV.

Online Methods

Latent causal variable model

The latent causal variable (LCV) model for a pair of heritable traits Y_1 and Y_2 assumes that a single latent variable L causally affects both Y_1 and Y_2, mediating the genetic correlation between them (Figure 1). The model contains random variables γ_1, γ_2 for the marginal non-mediated effect of a SNP on each trait, a random variable π for the marginal effect of a SNP on L, and fixed scalars q_1, q_2 for the effects of L on each trait (see Methods for a full description of the LCV model). We fix $Var(\pi) = 1$ and $Var(\gamma_k) = 1 - q_k^2$, so that the variance of the effect sizes is $Var(q_k\pi + \gamma_k) = 1$. The genetic causality proportion (gcp) is defined as:

$$\text{gcp} := \frac{\log|q_2| - \log|q_1|}{\log|q_2| + \log|q_1|}, \tag{3}$$

which satisfies

$$\frac{q_2^2}{q_1^2} = (\rho_g^2)^{\text{gcp}}. \tag{4}$$

where the genetic correlation ρ_g is equal to $q_1 q_2$. gcp is positive when trait 1 is partially genetically causal for trait 2. When gcp = 1, trait 1 is fully genetically causal for trait 2: $q_1 = 1$ and the causal effect size is $q_2 = \rho_g$. Our most critical modeling assumption is that the genetic correlation is mediated by a single variable; if multiple intermediaries contribute to the genetic correlation, with different effect sizes on each trait, then the model is misspecified. The LCV model is broadly related to dimension reduction techniques such as Factor Analysis[59] and Independent Components Analysis,[60] although it differs in its modeling assumptions as well as its goal (causal inference); our inference strategy (mixed fourth moments) also differs.

Fix q_1 and q_2. For each SNP, marginal effect sizes $(\pi, \gamma_1, \gamma_2)$ are drawn from some distribution D (because we consider marginal effect sizes, it is not expected that SNPs will be independent). The effect size of a SNP on trait k is $\alpha_p = q_k\pi + \gamma_k$, and we observe GWAS estimates of α for M SNPs. The asymptotic sampling distribution of estimated effect sizes for a SNP on each trait is bivariate normal, centered at the true effect sizes, with a covariance matrix that we can estimate using LD score regression.[14,17]

Assume that $(\pi, \gamma_1, \gamma_2)$ are independent, mean-zero random variables, with $E(\pi^2) = 1$ and $E(\gamma_k^2) = 1 - q_k^2$. Let $\alpha_k = q_k\pi + \gamma_k$ (note that $E(\alpha_k^2) = 1$). We derive equation (2) as follows:

$$\begin{aligned}
E(\alpha_1^3\alpha_2) =& E((\gamma_1 + q_1\pi)^3(\gamma_2 + q_2\pi)) \\
=& q_1^3 q_2 E(\pi^4) + 3q_1q_2E(\pi^2\gamma_1^2) \\
=& q_1^3 q_2 E(\pi^4) + 3q_1q_2E(\pi^2)E(\gamma_1^2) \\
=& q_1^3 q_2 E(\pi^4) + 3q_1q_2(1)(1 - q_1^2) \\
=& q_1^3 q_2 (E(\pi^4) - 3) + 3q_1q_2.
\end{aligned} \tag{5}$$

In the second line, we used the independence assumption to discard cross-terms of the form $\gamma_p \pi^3$, $\gamma_1 \gamma_2^3$, and $\gamma_1^3 \pi$. In the third and fourth lines, we used that $E(\gamma_1^2 \pi^2) = E(\gamma_1^2)E(\pi^2) = E(\gamma_1^2) = 1 - q_1^2$. The factor $E(\pi^4) - 3$ is the excess kurtosis of π, which is zero when π follows a Gaussian distribution; in order for equation (2) to be useful for inference, $E(\pi^4) - 3$ must be nonzero, and in order for the model to be identifiable, π must be non-Gaussian (see Supplementary Note).

Independence of $(\pi, \gamma_1, \gamma_2)$ was a stronger assumption than we needed. More specifically, we need:

1. $E(\gamma_1 \gamma_2) = 0$, so that U fully explains the genetic correlation between the two traits;
2. $E(\gamma_1 \gamma_2^3) = E(\gamma_2 \gamma_1^3) = 0$, so that the non-correlation between γ_1, γ_2 extends to SNPs with large non-mediated effects on each trait;
3. $E(\pi^2 \gamma_1 \gamma_2) = 0$, so that non-mediated effects do not have a tendency to either cancel out or augment mediated effects;
4. $E(\pi^3 \gamma_p) = E(\pi \gamma_p^3) = 0$;
5. And most importantly, $E(\pi^2 \gamma_p^2) = E(\pi^2)E(\gamma_p^2)$, so that SNPs with a large mediated effect do not tend to also have an additional non-mediated effect.

We do not need to assume that $corr(\gamma_1^2, \gamma_2^2) = 0$; we allow for unsigned pleiotropy between non-mediated effects (see Table S3f,n). Assumption (1) is an essential feature of the model definition, as otherwise there is no interpretation for L. Assumptions (2-4) are highly plausible, as they involve odd-numbered exponents; we are not aware of any biological or evolutionary interpretation for violations of these assumptions. Assumption (5) is the most likely to be violated in practice. First, it could be violated if some regions of the genome harbor many SNPs affecting different traits, while others do not. This phenomenon would most likely lead to *symmetric* violations of assumption (5); estimates of gcp would be biased toward zero, and power to detect a partially causal effect would be reduced. Second, if there are multiple intermediaries affecting both traits, it could lead to either symmetric or asymmetric violations of assumption (5). SNPs apparently affecting L will appear to have an additional non-mediated effect, as the compromise values of q that are fit by the model will differ from the true values of q for both intermediaries. This type of model misspecification can lead to bias and false positives (see Figure 3).

Estimation

Let $a_1 = \alpha_1 + \epsilon_1$, $a_2 = \alpha_2 + \epsilon_2$ be estimated effect sizes for the two traits. These effect estimates are normalized so that $var(\alpha_p) = 1$; we perform this normalization using a slightly modified version of LD score regression,[17] with LD scores computed from UK10K data.[51] In particular, we run LD score regression using a slightly different weighting scheme, matching the weighting scheme in our mixed fourth moment estimators; the weight of SNP i was:

$$w_i := \max(1, 1/\ell_i^{\text{HapMap}}), \tag{6}$$

where ℓ_i^{HapMap} was the estimated LD score between SNP i and other HapMap3 SNPs (this is approximately the set of SNPs that were used in the regression). This weighting scheme is motivated by the fact that SNPs with high LD to other regression SNPs will be over-counted in the regression

(see ref. 17). Similar to ref. 14, we improve power by excluding large-effect variants when computing the LD score intercept; for this study, we chose to exclude variants with χ^2 statistic 30× the mean (but these variants are used when computing $\bar{\chi}^2$). Then, we divide the summary statistics by $s = \sqrt{\bar{\chi}^2 - \hat{\sigma}_\epsilon^2}$, where $\bar{\chi}^2$ is the weighted mean χ^2 statistic and $\hat{\sigma}_\epsilon^2$ is the LD score intercept. We also divide the LD score intercept by s^2 for use in subsequent calculations. We assess the significance of the heritability by performing a block jackknife on s, defining the significance Z_h as s divided by its estimated standard error. We estimate the mixed fourth moments using:

$$\begin{aligned} E(a_1 a_2^3 | \alpha_1, \alpha_2) =& \alpha_1 \alpha_2^3 + E(\epsilon_1 \epsilon_2^3) + 3E(\alpha_1 \alpha_2 \epsilon_2^2) + E(\alpha_2^2 \epsilon_1 \epsilon_2) \\ =& \alpha_1 \alpha_2^3 + 3E(\epsilon_1 \epsilon_2) E(\epsilon_2^2) + 3\alpha_1 \alpha_2 E(\epsilon_2^2) + \alpha_2^2 E(\epsilon_1 \epsilon_2) \\ =& \alpha_1 \alpha_2^3 + 3\hat{\sigma}_{\epsilon_1 \epsilon_2} \hat{\sigma}_{\epsilon_2}^2 + 3\alpha_1 \alpha_2 \hat{\sigma}_{\epsilon_2}^2 + \alpha_2^2 \hat{\sigma}_{\epsilon_1 \epsilon_2}. \end{aligned} \tag{7}$$

We estimate $\hat{\sigma}_{\epsilon_1 \epsilon_2}$ using a modified version of cross-trait LD score regression.[14] Similar to our implementation of LD score regression, we perform cross-trait LD score regression using the weights defined in equation (6), and the intercept is computed while excluding variants with a large effect ($\chi^2 > 30\bar{\chi}^2$) on either trait. (For simulations with no LD, we use $E(\epsilon_k^2) = 1/sN_k$ and $E(\epsilon_1 \epsilon_2) = 0$ instead of estimating these values.) Then, we estimate $E(\alpha_1 \alpha_2^3) - 3\rho_g$ as:

$$\hat{k}_1 := -3\hat{\rho}_g + \frac{\sum_{i=1}^M w_i [a_{i1} a_{i2}^3 - 3a_{i1} a_{i2} \hat{\sigma}_{\epsilon_2}^2 - 3(a_{i2}^2 - \hat{\sigma}_{\epsilon_2}^2) \hat{\sigma}_{\epsilon_1 \epsilon_2}]}{\sum_{i=1}^M w_i}. \tag{8}$$

To obtain posterior mean and variance estimates for gcp, we define a collection of statistics $S(x)$ for $x \in X = \{-1, -.01, -.02, ..., 1\}$:

$$S(x) := \frac{A(x) - B(x)}{\max(1/|\hat{\rho}_g|, \sqrt{A(x)^2 + B(x)^2})} \quad A(X) = |\rho_g|^{-x} \hat{k}_1, \quad B(x) = |\rho_g|^x \hat{k}_2, \tag{9}$$

The motivation for utilizing the normalization by $\sqrt{A(x)^2 + B(x)^2}$ is that the magnitude of $A(x)$ and $B(x)$ tend to be highly correlated, leading to increased standard errors if we only use the numerator of S. However, the denominator tends to zero when the genetic correlation is zero, leading to instability in the test statistic and false positives. The use of the threshold leads to conservative, rather than inflated, when the genetic correlation is zero or nearly zero. In practice, we only analyze trait pairs with a significant genetic correlation, and this threshold usually has no effect on the results.

We estimate the variance of $S(x)$ using a block jackknife with $k = 100$ blocks, resulting in minimal non-independence between blocks. We compute an approximate likelihood, $L(S|\text{gcp} = x)$, by assuming (1) that $L(S|\text{gcp} = x) = L(S(x)|\text{gcp} = x)$ and (2) that if gcp $= x$ then $S(x)/\hat{\sigma}_{S(x)}$ follows a T distribution with 98 degrees of freedom. Imposing a uniform prior on gcp, the posterior mean estimate of gcp is:

$$\hat{\text{gcp}} := \frac{1}{|X|} \sum_{x \in X} x L(x) \tag{10}$$

The estimated standard error is:

$$\hat{\text{se}} := \sqrt{\frac{1}{|X|} \sum_{x \in X} (x - \hat{\text{gcp}})^2 L(x)}. \tag{11}$$

In order to compute p-values, we apply a T-test to the statistic $S(0)$.

Existing Mendelian randomization methods

Two-sample MR. We ascertained significant SNPs ($p < 5 \times 10^{-8}$, χ^2 test) for the exposure and performed an unweighted regression, with intercept fixed at zero, of the estimated effect sizes on the outcome with the estimated effect sizes on the exposure (in practice, a MAF-weighted and LD-adjusted regression is often used; in our simulations, all SNPs had equal MAF, and there was no LD). To assess the significance of the regression coefficient, we estimated the standard error as $\mathrm{se} = \sqrt{\frac{\frac{1}{K}\sum_{k=1}^{K}\bar{\beta}_{k2}^2}{\sum_{k=1}^{K}\hat{\beta}_{k1}^2}}$, where $\bar{\beta}_{k2}$ is the kth residual, N_2 is the sample size in the outcome cohort, and K is the number of significant SNPs. This estimate of the standard error allows the residuals to be overdispersed compared with the error that is expected from the GWAS sample size. To obtain p values, we applied a two-tailed t-test to the regression coefficient divided by its standard error, with $K - 1$ degrees of freedom.

MR-Egger. We ascertained significant SNPs for the exposure and coded them so that the alternative allele had a positive estimated effect on the exposure. We performed an unweighted regression with a fitted intercept of the estimated effect sizes on the outcome on the estimated effect sizes on the exposure. We assessed the significance of the regression using the same procedure as for two-sample MR, except that the t-test used $K - 2$ rather than $K - 1$ degrees of freedom.

Bidirectional MR. We implemented bidirectional mendelian randomization in a manner similar to Pickrell et al.[9] Significant SNPs were ascertained for each trait. If the same SNP was significant for both traits, then it was assigned only to the trait where it ranked higher (if a SNP ranked equally high for both traits, it was excluded from both SNP sets). The Spearman correlations r_1, r_2 between the z scores for each trait was computed on each set of SNPs, and we applied a χ_1^2 test to

$$\chi^2 = \frac{1}{\frac{1}{K_1-3} + \frac{1}{K_2-3}}(\operatorname{atanh}(r_1) - \operatorname{atanh}(r_2))^2, \tag{12}$$

where K_j is the number of significant SNPs for trait j. In Pickerell et al.,[9] the statistics $\operatorname{atanh}(r_j)$ are also used, but a relative likelihood comparing several different models is reported instead of a p-value. We chose to report p-values for Bidirectional MR in order to allow a direct comparison with other methods.

Application of MR to LDL and BMD. We applied two-sample MR (see above) to 8 curated SNPs that were previously used to show that LDL has a causal effect on CAD in ref. 3. 10 SNPs were used in ref. 3, of which summary statistics were available for 8 SNPs: rs646776, rs6511720, rs11206510, rs562338, rs6544713, rs7953249, rs10402271 and rs3846663.

Simulations with no LD

In order to simulate summary statistics with no LD, first, we chose causal effect sizes for each SNP on each trait according to the LCV model. The causal effect size vector for trait k was

$$\beta_k = \frac{h_k^2}{M}(q_k\pi + \gamma_k), \tag{13}$$

where in all simulations except for Table S2, q_k was a scalar, and π and γ_k were $1 \times M$ vectors. In Table S2, q_k was a 1×2 vector and π was a $2 \times M$ vector. Entries of π were drawn from i.i.d. point-normal distribution with mean zero, variance 1, and expected proportion of causal SNPs equal to p_π. Entries of γ_k were drawn from i.i.d. point-normal distributions with expected proportion of causal SNPs equal to p_{γ_k}; we modeled colocalization between non-mediated effects by fixing some expected proportion of SNPs $p_{\gamma_{1,2}} < \min(p_{\gamma_1}, p_{\gamma_2})$ as having nonzero values of both γ_1 and γ_2. Then, we centered and re-scaled the nonzero entries of π and γ_k, so that they had mean 0 and variance 1 and $1 - q_k^2$, respectively. For simulations in Figure 3a-b, q_k was a 1×2 vector and π was a $2 \times M$ matrix. For these simulations, entries of π were drawn from independent point-normal distributions with proportion of causal SNPs equal to p_{π_1} for the first row of π and p_{π_2} for the second row. Entries of γ_k were drawn from a point-normal distribution with expected proportion of causal SNPs equal to p_{γ_k} and variance $1 - \|q_k\|^2$. For simulations in Figure 3c-d, effect sizes were drawn from a mixture of Normal distributions: there was a point mass at (0,0); a component with $\sigma_1^2 = 0, \sigma_2^2 \neq 0$; a component with $\sigma_1^2 \neq 0, \sigma_2^2 = 0$; and a component with $\sigma_1^2 \neq 0, \sigma_2^2 \neq 0, \sigma_{12} = \sqrt{\sigma_1^2 \sigma_2^2}$. Values of $M, N_k, N_{\text{shared}}, \rho_{\text{total}}, p_{\gamma_k}, p_{\gamma_{1,2}}, h_k^2, p_\pi, q_k$ for each simulation can be found in Table S11.

Second, we simulated summary statistics as

$$\hat{\beta}_k \sim N(\beta_k, \frac{1}{N_k} I), \tag{14}$$

where β_k is the vector of true causal effect sizes for trait k and N_k is the sample size for trait k. When we ran LCV on these summary statistics, we used constrained-intercept LD score regression rather than variable-intercept LD score regression both to normalize the effect estimates[17] and to estimate the genetic correlation,[14] with LD scores equal to one for every SNP.

Simulations with LD

In simulations with LD, we first simulated causal effect sizes for each trait in the same manner as simulations with no LD. Then, we obtained summary statistics in one of two ways, either using real genotypes or using real LD only.

For simulations with real genotypes modeling population stratification (Table 1g and Table S6), we chose effect sizes for each SNP and each trait from the LCV model with various parameters and multiplied these effect size vectors by real genotype vectors from UK Biobank,[25] adding noise to obtain simulated phenotypes. For computational efficiency, we restricted these genotypes to chromosome 1 (M = 43k). We added stratification directly to the phenotype values along PC1 (computed on 43k SNPs and $N_1 + N_2$ individuals), with effect sizes $\sqrt{0.01}$ and $\sqrt{0.02}$ for trait 1 and trait 2, respectively. We then re-normalized phenotypes to have variance 1; afterwards, ~ 1% and ~ 2% of variance were explained by PC1 for each trait respectively. We estimated SNP effect sizes for each trait by correlating each SNP with the phenotypic values in N_k individuals. In corrected simulations (Table S6b,d,f), we residualized the PC1 SNP loadings (computed on all $N_1 + N_2$ individuals) from the SNP effect estimates, a procedure which is effectively equivalent to correction of the individual-level data.[23]

For other simulations, we simulated summary statistics without first simulating phenotypic values, using the fact that the sampling distribution of Z-scores is approximately:[21]

$$Z \sim N(\sqrt{N} R\beta, R), \tag{15}$$

where R is the LD matrix and β is the vector of true effect sizes. We estimated R from the $N = 145k$ UK Biobank cohort using plink with an LD window size of 2Mb ($M = 596k$), which we converted into a block diagonal matrix with 1001 blocks. The number 1001 was chosen instead of the number 1000 so that the boundaries of these blocks would not align with the boundaries of our 100 jackknife blocks; the use of blocks allowed us to avoid diagonalizing a matrix of size 596k, while not significantly changing overall LD patterns (there are $\sim 50,000$ independent SNPs in the genome, and $1001 << 50,000$). Because the use of a 2Mb window causes the estimated LD matrix to be non-positive semidefinite (even after converting it into a block diagonal matrix), each block was converted into a positive semidefinite matrix by diagonalizing it and removing its negative eigenvalues: that is, we replaced each block $A = V\Sigma V^T$ with the matrix B, where $B = V \max(0, \Sigma) V^T$. Then, because the removal of negative eigenvalues causes B' to have entries slightly different from one, we re-normalized each block: $C = D^{-1/2} B D^{-1/2}$, where D is the diagonal matrix corresponding to the diagonal of B. Even though the diagonal elements of B are close to 1 (mostly between 0.99 and 1.01), this step is important to obtain reliable heritability estimates using LD score regression because otherwise the diagonal elements of the LD matrix will be strongly correlated with the LD scores ($r^2 \approx 0.5$) and the heritability estimates will be upwardly biased, especially at low sample sizes.

We concatenated the blocks $C_1, ..., C_{1001}$ to obtain a positive semi-definite block-diagonal matrix R'. We also computed and concatenated the matrix square root of each block. In order to obtain samples from a Normal distribution with mean $R'\beta$ and variance $\frac{1}{N}R'$, we multiplied a vector having independent standard normal entries by the matrix square root of R' and added this noise vector to the vector of true marginal effect sizes, $R'\beta$. We computed LD scores directly from R. For simulations with sample overlap, the summary statistics were correlated between the two GWAS: the correlation between the noise term in the estimated effect of SNP i on trait 1 and the estimated effect of SNP j on trait 2 was $R'_{ij}\rho_{\text{total}}N_{\text{shared}}/\sqrt{N_1 N_2}$, which is the amount of correlation that would be expected if the total (genetic plus environmental) correlation between the traits is ρ_{total}.[14]

References

[1] Davey Smith, George, and Shah Ebrahim. "Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease?" International journal of epidemiology 32.1 (2003): 1-22.

[2] Davey Smith, George, and Gibran Hemani. "Mendelian randomization: genetic anchors for causal inference in epidemiological studies." Human molecular genetics 23.R1 (2014): R89-R98.

[3] Voight, Benjamin F., et al. "Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study." The Lancet 380.9841 (2012): 572-580.

[4] Do, Ron, et al. "Common variants associated with plasma triglycerides and risk for coronary artery disease." Nature genetics 45.11 (2013): 1345-1352.

[5] Burgess, Stephen, Adam Butterworth, and Simon G. Thompson. "Mendelian randomization analysis with multiple genetic variants using summarized data." Genetic epidemiology 37.7 (2013): 658-635.

[6] Kang, Hyunseung, et al. "Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization." Journal of the American Statistical Association 111.513 (2016): 132-144.

[7] Bowden, Jack, George Davey Smith, and Stephen Burgess. "Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression." International journal of epidemiology 44.2 (2015): 512-525.

[8] Bowden, Jack, et al. "Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator." Genetic epidemiology 40.4 (2016): 304-314.

[9] Pickrell, Joseph K., et al. "Detection and interpretation of shared genetic influences on 42 human traits." Nature genetics 48.7 (2016): 709.

[10] Verbanck, Marie, et al. "Widespread pleiotropy confounds causal relationships between complex traits and diseases inferred from Mendelian randomization." bioRxiv (2017): 157552.

[11] Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. New England Journal of Medicine 354 (2006): 1264?72.

[12] Paaby, Annalise B., and Matthew V. Rockman. "The many faces of pleiotropy." Trends in Genetics 29.2 (2013): 63-73.

[13] VanderWeele, Tyler J., et al. "Methodological challenges in Mendelian randomization." Epidemiology (Cambridge, Mass.) 25.3 (2014): 427.

[14] Bulik-Sullivan, Brendan, et al. "An atlas of genetic correlations across human diseases and traits." Nature genetics 47.11 (2015): 1236-1241.

[15] Welsh, Paul, et al. "Unraveling the directional link between adiposity and inflammation: a bidirectional Mendelian randomization approach." The Journal of Clinical Endocrinology & Metabolism 95.1 (2010): 93-99.

[16] Vimaleswaran, Karani S., et al. "Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts." PLoS Med 10.2 (2013): e1001383.

[17] Bulik-Sullivan, Brendan K., et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies." Nature genetics 47.3 (2015): 291-295.

[18] Yang, Jian, et al. "Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index." Nature genetics 47.10 (2015): 1114.

[19] Kolesar, Michal, et al. "Identification and inference with many invalid instruments." Journal of Business & Economic Statistics 33.4 (2015): 474-484.

[20] Burgess, Stephen, and Simon G. Thompson. "Interpreting findings from Mendelian randomization using the MR-Egger method." European Journal of Epidemiology (2017): 1-13.

[21] Conneely, Karen N., and Michael Boehnke. "So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests." The American Journal of Human Genetics 81.6 (2007): 1158-1168.

[22] Galinsky, Kevin J., et al. "Population structure of UK Biobank and ancient Eurasians reveals adaptation at genes influencing blood pressure." The American Journal of Human Genetics 99.5 (2016): 1130-1139.

[23] Bhatia, Gaurav, et al. "Correcting subtle stratification in summary association statistics." bioRxiv (2016): 076133.

[24] Goddard, Michael E., et al. "Estimating effects and making predictions from genome-wide marker data." Statistical Science 24.4 (2009): 517-529.

[25] Sudlow, Cathie, 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 medicine 12.3 (2015): e1001779.

[26] Bycroft, Clare, et al. "Genome-wide genetic data on 500,000 UK Biobank participants." bioRxiv (2017): 163298.

[27] Loh, Po-Ru, et al. "Mixed model association for biobank-scale data sets." bioRxiv (2017): 194944.

[28] Holmes, Michael V., Mika Ala-Korpela, and George Davey Smith. "Mendelian randomization in cardiometabolic disease: challenges in evaluating causality." Nature Reviews Cardiology (2017): 577-590.

[29] Smith, George Davey, et al. "The association between BMI and mortality using offspring BMI as an indicator of own BMI: large intergenerational mortality study." Bmj 339 (2009): b5043.

[30] Nordestgaard, Brge G., et al. "The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach." PLoS Med 9.5 (2012): e1001212.

[31] Hgg, Sara, et al. "Adiposity as a cause of cardiovascular disease: a Mendelian randomization study." International journal of epidemiology 44.2 (2015): 578-586.

[32] Holmes, Michael V., et al. "Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis." The American Journal of Human Genetics 94.2 (2014): 198-208.

[33] Ross, Stephanie, et al. "Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease." European heart journal 36.23 (2015): 1454-1462.

[34] Lyall, Donald M., et al. "Association of body mass index with cardiometabolic disease in the UK Biobank: a Mendelian randomization study." JAMA cardiology 2.8 (2017): 882-889.

[35] Schunkert, Heribert, et al. "Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease." Nature genetics 43.4 (2011): 333-338.

[36] Klein, Irwin, and Kaie Ojamaa. "Thyroid hormone and the cardiovascular system." New England Journal of Medicine 344.7 (2001): 501-509.

[37] Grais, Ira Martin, and James R. Sowers. "Thyroid and the heart." The American journal of medicine 127.8 (2014): 691-698.

[38] Zhao, Jie V., and C. Mary Schooling. "Thyroid function and ischemic heart disease: a Mendelian randomization study." Scientific reports 7:8515 (2017): 8515.

[39] Monzani, F. et al. "Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: a double blind, placebo-controlled study." J. Clin. Endocrinol. Metab. 86 (2001): 1110-1115.

[40] Meier, C. et al. "TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism: a double blind, placebo-controlled trial (Basel Thyroid Study)." J. Clin. Endocrinol. Metab. 86 (2001): 4430-4863.

[41] Monzani, F. et al. "Effect of levothyroxine replacement on lipid profile and intima-media thickness in subclinical hypothyroidism: a double-blind, placebo- controlled study." J. Clin. Endocrinol. Metab. 89 (2004): 2099-2106 .

[42] Razvi, S. et al. "The beneficial effect of L-thyroxine on cardiovascular risk factors, endothelial function, and quality of life in subclinical hypothyroidism: randomized, crossover trial." J. Clin. Endocrinol. Metab. 92 (2007): 1715-1723.

[43] Nagasaki, T. et al. "Decrease of brachial-ankle pulse wave velocity in female subclinical hypothyroid patients during normalization of thyroid function: a double-blind, placebo-controlled study." Eur. J. Endocrinol. 160 (2009): 409-415.

[44] Chaker, Layal, et al. "Thyroid function and risk of type 2 diabetes: a population-based prospective cohort study." BMC medicine 14.1 (2016): 150.

[45] Wang, Zongze, et al. "Effects of Statins on Bone Mineral Density and Fracture Risk: A PRISMA-compliant Systematic Review and Meta-Analysis." Medicine 95.22 (2016): e3042.

[46] Yerges, Laura M., et al. "Decreased bone mineral density in subjects carrying familial defective apolipoprotein B-100." The Journal of Clinical Endocrinology & Metabolism 98.12 (2013): E1999-E2005.

[47] Clarke, T. K., et al. "Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population." Molecular psychiatry 21.3 (2016): 419-425.

[48] Keller, Matthew C., and Geoffrey Miller. "Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best?" Behavioral and Brain Sciences 29.4 (2006): 385-404.

[49] Mullins, Niamh, et al. "Reproductive fitness and genetic risk of psychiatric disorders in the general population." Nature communications 8 (2017): 15833.

[50] Davies, Gail, et al. "Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112,151)." Molecular psychiatry 21.6 (2016): 758.

[51] UK10K Consortium. "The UK10K project identifies rare variants in health and disease." Nature 526.7571 (2015): 82.

[52] Burgess, Stephen, et al. "Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways." International journal of epidemiology 44.2 (2014): 484-495.

[53] Schoech, Armin, et al. "Quantification of frequency-dependent genetic architectures and action of negative selection in 25 UK Biobank traits." bioRxiv (2017): 188086.

[54] Gamazon, Eric R., et al. "A gene-based association method for mapping traits using reference transcriptome data." Nature genetics 47.9 (2015): 1091-1098.

[55] Gusev, Alexander, et al. "Integrative approaches for large-scale transcriptome-wide association studies." Nature genetics 48 (2016): 245-252.

[56] Zhu, Zhihong, et al. "Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets." Nature genetics 48 (2016):481:487.

[57] The GTEx consortium, et al. "Genetic effects on gene expression across human tissues." Nature 550.7675 (2017): 204.

[58] Mokry, Lauren E., et al. "Vitamin D and risk of multiple sclerosis: a Mendelian randomization study." PLoS medicine 12.8 (2015): e1001866.

[59] Child, Dennis. "The essentials of factor analysis." A&C Black (2006).

[60] Comon, Pierre. "Independent component analysis, a new concept?" Signal processing 36.3 (1994): 287-314.

[61] Thyagarajan, Bharat, et al. "Longitudinal association of body mass index with lung function: the CARDIA study." Respiratory research 9.1 (2008): 31.

[62] Ellis, Justine A., Margaret Stebbing, and Stephen B. Harrap. "Polymorphism of the androgen receptor gene is associated with male pattern baldness." Journal of investigative dermatology 116.3 (2001): 452-455.

[63] Tyrrell, Jessica, et al. "Height, body mass index, and socioeconomic status: mendelian randomization study in UK Biobank." bmj 352 (2016): i582.

[64] Skaaby, Tea, et al. "Estimating the causal effect of body mass index on hay fever, asthma, and lung function using Mendelian randomization." Allergy (2017).

[65] Haase, Christiane L., et al. "High-density lipoprotein cholesterol and risk of type 2 diabetes: a Mendelian randomization study." Diabetes (2015): db141603.

[66] Cole, Stephen R., et al. "Illustrating bias due to conditioning on a collider." International journal of epidemiology 39.2 (2009): 417-420.

Tables

		ρ_g	$p_{\text{LCV}} < .05$	$p_{\text{LCV}} < .001$	Mean $\hat{\text{gcp}}$
a	Default parameter values	0.2	0.058	0.003	0.00
b	Zero genetic correlation	0	0	0	0.00
c	Very high genetic correlation	0.8	0.058	0.002	-0.00
d	Uncorrelated pleiotropic effects	0.2	0.054	0.001	0.00
e	Differential polygenicity	0.2	0.062	0.002	0.01
f	Differential power	0.2	0.063	0.004	-0.01
g	Population stratification	0.25	0.34	0.126	-0.21
h	Causal	0.2	0.97	0.94	-0.76
i	Partially causal	0.2	0.71	0.35	-0.56

Table 1: Simulations with LD. We report the positive rate (α = 0.05 and α = 0.001) for a causal (or partially causal) effect for LCV, as well as the mean $\hat{\text{gcp}}$ ($\hat{\text{gcp}}$ standard error is less than 0.01 in each row). (a) Default parameter values (see text). (b) Zero genetic correlation (ρ_g = 0). (c) Very high genetic correlation (ρ_g = 0.75). (d) Uncorrelated pleiotropic effects. (e) Differential polygenicity (0.2% and 0.8% of SNPs were causal for trait 1 and trait 2, respectively). (f) Differential power (N_1 = 20k and N_2 = 500k). (g) Population stratification. (h) Full genetic causality (gcp = 1). (i) Partial genetic causality (gcp = 0.5). Results for each panel are based on 5,000 simulations.

trait 1	trait 2	p_{LCV}	$\hat{\rho}_g$ (se)	$\hat{\mathrm{gcp}}$ (pse)	MR ref
BMI	Myocardial infarction	5×10^{-9}	0.34 (0.09)	0.94 (0.11)	30, 34
Triglycerides	Myocardial infarction	2×10^{-31}	0.30 (0.06)	0.90 (0.08)	4
LDL	Myocardial infarction	4×10^{-31}	0.17 (0.08)	0.73 (0.13)	3, 11
Hypothyroidism	Myocardial infarction	1×10^{-11}	0.26 (0.05)	0.72 (0.16)	
High cholesterol	Myocardial infarction	2×10^{-4}	0.52 (0.12)	0.71 (0.19)	
Fasting glucose	Myocardial infarction	4×10^{-4}	0.19 (0.07)	0.62 (0.23)	33
Triglycerides	Hypertension	1×10^{-38}	0.25 (0.04)	0.95 (0.04)	
HDL	Hypertension	1×10^{-21}	-0.29 (0.06)	0.87 (0.09)	
BMI	Hypertension	3×10^{-16}	0.38 (0.03)	0.83 (0.11)	9, 34
Reticulocyte count	Hypertension	4×10^{-4}	0.27 (0.04)	0.41 (0.13)	
LDL	Bone mineral density - heel	7×10^{-34}	-0.12 (0.05)	0.80 (0.12)	
Height	Bone mineral density - heel	5×10^{-14}	-0.09 (0.04)	0.50 (0.14)	
Triglycerides	Mean cell volume	2×10^{-18}	-0.20 (0.04)	0.86 (0.11)	
Triglycerides	Platelet distribution width	1×10^{-16}	0.19 (0.04)	0.81 (0.13)	
Triglycerides	Reticulocyte count	5×10^{-10}	0.33 (0.05)	0.79 (0.14)	
Triglycerides	Eosinophil count	6×10^{-17}	0.14 (0.05)	0.75 (0.16)	
Triglycerides	Monocyte count	1×10^{-4}	0.14 (0.04)	0.67 (0.21)	

Table 2: Causal or partially genetically causal risk factors for selected trait pairs. We report all traits with a significant genetic correlation ($p < .05$) and significant evidence of partial genetic causality (1% FDR) on MI, hypertension and bone mineral density, as well as all significant associations (ρ_g $p < 0.05$ and LCV FDR < 1%) between triglycerides and blood cell traits. p_{LCV} is the p-value for the null hypothesis of no partial genetic causality; $\hat{\rho}_g$ is the estimated genetic correlation, with standard error; $\hat{\mathrm{gcp}}$ is the posterior mean estimated genetic causality proportion, with posterior standard error. We also provide references for all published evidence of causal relationships between these traits that we are currently aware of.

Figures

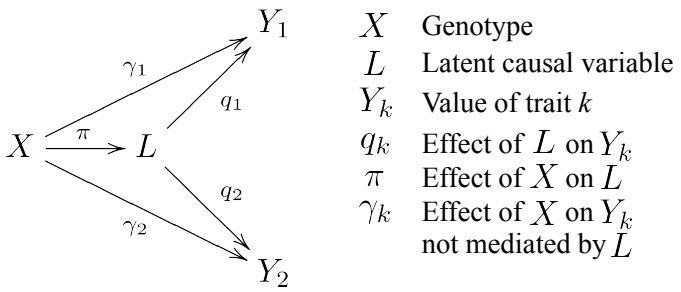

Figure 1: Latent causal variable model. We display the relationship between genotypes X, latent causal variable L and trait values Y_1 and Y_2.

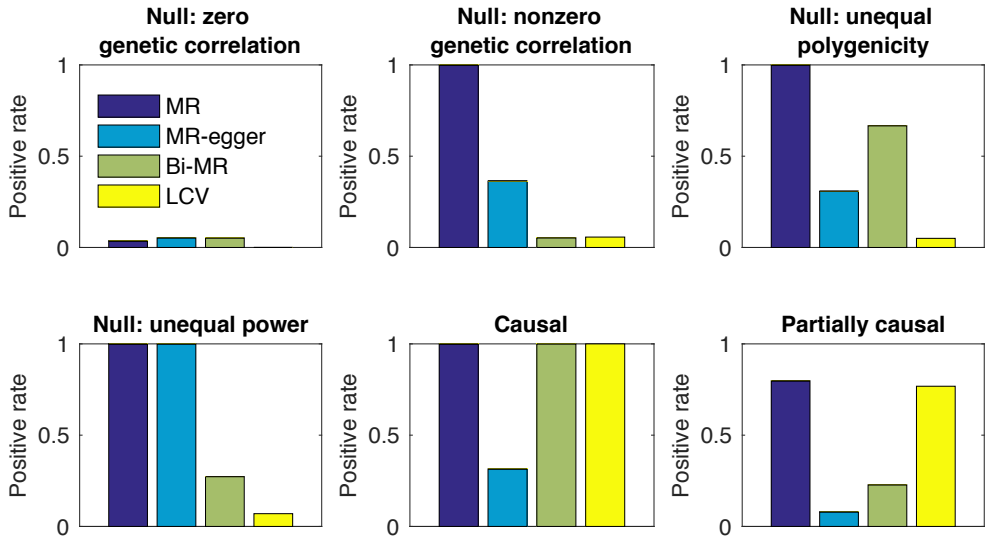

Figure 2: Simulations with no LD. We compared LCV to three MR methods (two-sample MR, MR-Egger and Bidirectional MR). We report the positive rate (α = 0.05) for a causal (or partially causal) effect. MR methods utilized ~ 320 genome-wide significant SNPs. (a) Null simulation (gcp = 0) with uncorrelated pleiotropic effects and zero genetic correlation. (b) Null simulation with nonzero genetic correlation. (c) Null simulation with nonzero genetic correlation and differential polygenicity between the two traits. (d) Null simulation with nonzero genetic correlation and different sample size for the two traits, in addition to different per-SNP heritability for shared and non-shared genetic effects. (e) Non-null simulation with full genetic causality (gcp = 1). (f) Non-null simulation with partial genetic causality (gcp = 0.5). Results for each panel are based on 2,000 simulations. Numerical results are reported in Table S1.

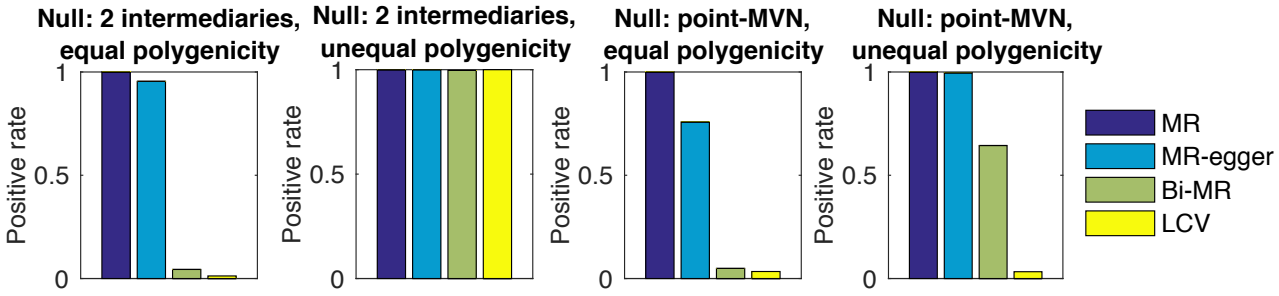

Figure 3: Null simulations with no LD and LCV model violations. We report the positive rate ($\alpha = 0.05$) for a causal (or partially causal) effect for LCV, two-sample MR, MR-Egger and Bidirectional MR. (a) Null simulation with two intermediaries with different effects on each trait; the intermediaries together explain 25% of heritability for each trait. (b) Null simulation with two intermediaries with differential polygenicity. (c) Null simulation with SNP effects drawn from a mixture of multi-variate normal distributions; one mixture component has correlated effects on each trait. (d) Null simulation with SNP effects drawn from a mixture of multi-variate normal distributions, and differential polygenicity between the two traits. Results for each panel are based on 2,000 simulations. Numerical results are reported in Table S2.

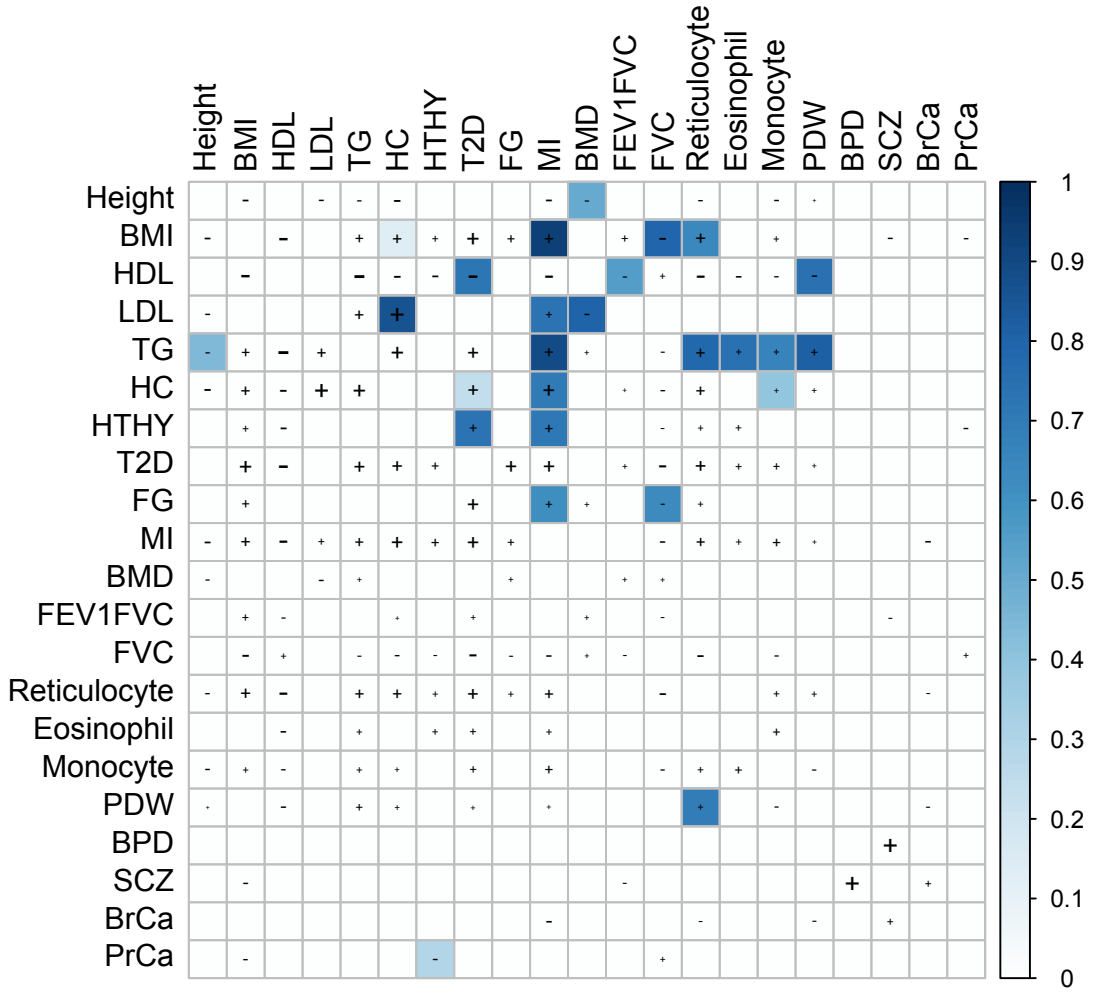

Figure 4: Genetically causal and partially genetically causal relationships between selected complex traits. Color scale indicates posterior mean $\hat{gcp}$ for the effect of the row trait on the column trait. Shaded squares indicate significant evidence for a causal or partially causal effect of the row trait on the column trait, at 1% FDR for genetically correlated trait pairs. "+" or "-" signs indicate trait pairs with a nominally significant (positive or negative) genetic correlation ($p < .05$), and the size of the "+" or "-" size is proportional to the genetic correlation. Entries without a significant genetic correlation are not shaded. Complete results are reported in Table S9. HTHY: hypothyroidism. FG: fasting glucose. PDW: platelet distribution width. BPD: bipolar disorder. SCZ: schizophrenia. BrCa: breast cancer: PrCa: prostate cancer.