Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics.

Jean Morrison *[1], Nicholas Knoblauch[1], Joseph Marcus[1], Matthew Stephens †[1,2], and Xin He ‡[1]

[1]Department of Human Genetics, University of Chicago
[2]Department of Statistics, University of Chicago

July 11, 2019

Abstract

Mendelian randomization (MR) is a valuable tool for detecting evidence of causal relationships between pairs of traits. Opportunities to apply MR are growing rapidly as the number of genome-wide association studies (GWAS) with publicly available summary statistics grows. Unfortunately, existing MR methods are prone to false positives caused by pleiotropic variants. *Correlated pleiotropy*, which arises when genetic variants affect both traits through a heritable shared factor, is a particularly challenging problem and is not addressed by most existing methods. Additionally, most MR methods only use genome-wide significant loci, which can limit power and introduce bias. We propose a new method (Causal Analysis Using Summary Effect Estimates; CAUSE) that uses genome-wide summary statistics to identify patterns that are consistent with causal effects, while accounting for pleiotropic effects, including correlated pleiotropy. We demonstrate in simulations that CAUSE is much better at controlling the false positive rate in the presence of pleiotropic effects than other methods. We apply CAUSE to study relationships between pairs of complex traits and between blood cell composition and autoimmune disorders. We find that CAUSE detects causal relationships with strong literature support, including an effect of blood pressure on heart disease risk that is not found using other methods. Our results suggest that many pairs of traits identified as causal using alternative methods may be false positives driven by pleiotropic effects.

1 Introduction

Inferring causal relationships between traits is important for understanding the etiology of disease and designing new treatments. However, inferring causal relationships is also hard. Randomized trials are considered the gold standard, but are expensive and sometimes impossible. Observational studies are cheaper and easier, but associations measured in observational studies may be biased by confounding and reverse causality.

Mendelian randomization (MR) is a potentially-powerful approach to studying causal relationships using data from observational studies. The key idea of MR is to treat genotypes as naturally occurring "randomizations" [1, 2, 3]. Suppose we are interested in the causal effect of trait M (for "Mediator") on trait Y. Under certain assumptions, summarized in Figure 1, the associations of a genetic variant G_j with traits M and Y will satisfy

$$\beta_{Y,j} = \gamma \beta_{M,j}, \tag{1}$$

where $\beta_{Y,j}$ is the association of G_j with Y, $\beta_{M,j}$ is the association of G_j with M and γ is the causal effect of M on Y. This relationship is the core of simple MR methods. Many methods based on Equation (1), including the commonly used inverse variance weighted (IVW) regression, first obtain estimates of $\beta_{Y,j}$ and $\beta_{M,j}$ for several genetic variants G_j, and then estimate γ by regressing the estimates of $\beta_{Y,j}$ on the estimates of $\beta_{M,j}$ [4].

*jeanm@uchicago.edu
†mstephens@uchicago.edu
‡xinhe@uchicago.edu

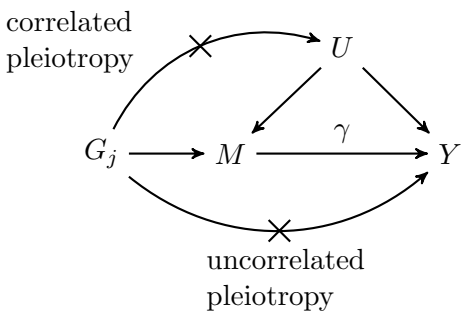


Figure 1: Causal diagram assumed by simple MR. Arrows indicate causal effects. Crosses mark causal effects that are assumed absent by simple MR methods. The causal effect of trait M on trait Y, γ, is the target of inference. Variant G_j affects trait M but is assumed to have no effects on Y that are not mediated through M. Pleiotropic effects on Y that are independent of the effect of G_j on M are referred to as uncorrelated pleiotropy. Effects that are mediated by a shared factor, U, are referred to as correlated pleiotropy.

Unfortunately, the assumptions made by simple MR methods are often violated in practice, leading to non-zero estimates of causal effects when no causal relationship exists (false positives). Specifically, simple MR relies on the assumption that no variants that affect M also affect Y through a pathway not mediated by M. Variants with such effects are often termed horizontally pleiotropic and are a major source of false positives for simple MR methods [5, 6]. Accounting for these effects has been a major focus of ongoing MR research [3]. It is helpful to distinguish between two different types of horizontal pleiotropy: *uncorrelated pleiotropy* – where the horizontal pleiotropic effects of variants on Y are uncorrelated with effects on M – and *correlated pleiotropy* where the horizontal pleiotropic effects of variants on Y are correlated with the effects on M. Uncorrelated pleiotropy occurs when a genetic variant affects Y and M through separate mechanisms, whereas correlated pleiotropy occurs when a genetic variant affects Y and M through a shared heritable factor, U (Figure 1). U may represent a shared biological process or pathway. Both types of pleiotropy may occur for any pair of traits.

Of the two types, uncorrelated pleiotropy is easier to account for, because it only adds noise to the relationship in Equation (1) rather than inducing a systematic correlation between $\beta_{Y,j}$ and $\beta_{M,j}$. Several methods have recently been developed to deal with this, including Egger regression [7, 8], the weighted median estimator [9], and several methods that rely on outlier removal including GSMR [10] and MR-PRESSO [6]. However, because these methods do not model correlated pleiotropy, they remain vulnerable to false positives in its presence (see Figure 2a). Evidence from genetic correlation studies suggests that correlated pleiotropy is common [11]. Many traits, including pairs that are unlikely to be causally linked, share common genetic factors [11]. For example, many mental disorders are found to be genetically correlated, a pattern that is better explained by shared biological processes than by causal relationships [12]. Thus correlated pleiotropy is an important potential source of false positives in MR analyses [13, 14].

Here we present a new MR analysis method that accounts for both uncorrelated and correlated pleiotropy. Causal Analysis Using Summary Effect Estimates (CAUSE) uses genome-wide summary statistics to identify pairs of traits that are consistent with a causal effect. The intuition is that, if M indeed causally affects Y then *every* variant affecting M will have a correlated effect on Y (Figure 2b). This can be distinguished from correlated pleiotropy that leads to correlated effects for only a fraction of variants (Figure 2a). Variants with strong effects on M but little effect on Y provide evidence against a causal effect of M on Y, even if other genetic variants have strong, correlated effects on both traits. Methods like IVW regression, that assess average correlation, cannot account for this, and can be mislead by even modest levels of correlated pleiotropy.

CAUSE assesses whether observed summary statistics for a pair of traits are *consistent with a causal effect* of M on Y by determining if effects are correlated for all variants or only a subset. Extreme cases of correlated pleiotropy, where a large proportion of variants affect M and Y through a heritable shared factor or pathway, may be impossible to distinguish from causal effects. Nonetheless, we demonstrate in simulations that CAUSE makes fewer false detections in the presence of correlated pleiotropy than existing methods. In applications to GWAS data for a large number of traits, CAUSE identifies a smaller number

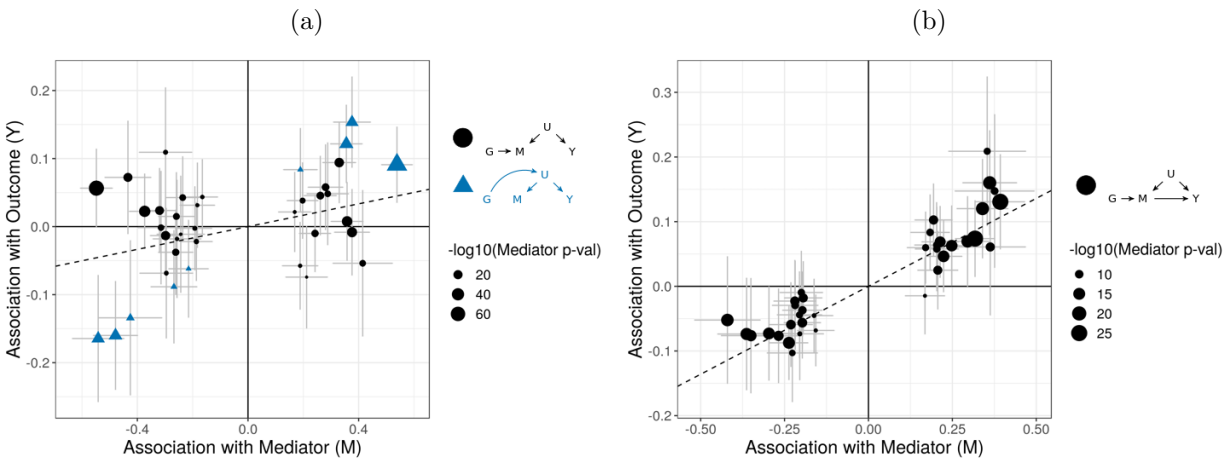


Figure 2: Simulated data illustrating (a) the pattern induced by a shared factor (correlated pleiotropy) and (b) the pattern induced by a causal effect. In both plots, effect size estimates are indicated by points. Error bars around points have length 1.96 times the standard error on each side. Only variants that are strongly associated with M ($p < 5 \cdot 10^{-8}$) are shown. (a) Summary statistics are simulated for two traits with no causal effect. 15% of the M effect variants act on M through a shared factor that affects both M and Y (triangles). Although there are many variants associated with M and not with Y, the IVW estimator (dotted line) is significantly different from zero ($p = 0.01$), driven by the effect size correlation of the 15% pleiotropic variants (triangles). (b) Summary statistics are simulated under a model where M causally affects Y. Here *every* variant associated with M is also associated with Y and the effect sizes are correlated.

of trait pairs as consistent with causal effects than methods that do not account for correlated pleiotropy. Many of the pairs that CAUSE does detect have a plausible causal connection.

In addition to allowing for pleiotropic effects, CAUSE has two other important differences from simple MR methods. First, it uses information at all variants, rather than only the variants most strongly associated with M. This can increase power when the GWAS for trait M has low power. Second, when there is evidence of correlated pleiotropy, CAUSE can identify which variants most likely act through a shared factor. This can be useful for elucidating the biological mechanisms driving genetic correlation of related traits.

2 Results

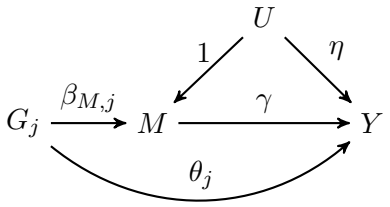
2.1 CAUSE models uncorrelated and correlated pleiotropy

We use effect estimates and standard errors from GWAS of traits M and Y to assess whether the data are consistent with a causal effect of M on Y. In most MR methods, variants with strong associations with M are pre-selected and assumed to follow the causal diagram in Figure 1. Our proposal, illustrated in Figure 3, differs in three ways. First, we use all variants genome-wide and model uncertainty about the effect of each variant on M. Second, we allow every variant to exhibit uncorrelated pleiotropy. This is modeled as an effect, θ_j, of variant G_j on trait Y that is uncorrelated with the effect of G_j on M. Third, we allow a subset of variants to exhibit correlated pleiotropy. We assume that a small proportion of variants, q, follow the causal diagram in Figure 3b, affecting M and Y through an unobserved heritable shared factor, U. The remaining variants are assumed to follow the causal diagram in Figure 3a, which is the same as the diagram in Figure 1 with the addition of a uncorrelated pleiotropic effect.

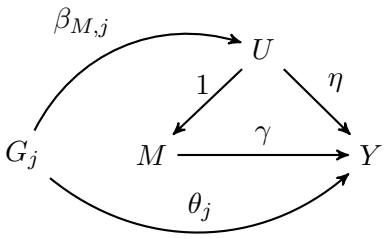
Under this model, the relationship between $\beta_{Y,j}$ and $\beta_{M,j}$ is

$$\beta_{Y,j} = \underbrace{\gamma\beta_{M,j}}_{\text{causal effect}} + \underbrace{Z_j\eta\beta_{M,j}}_{\substack{\text{correlated}\\ \text{pleiotropy}}} + \underbrace{\theta_j}_{\substack{\text{uncorrelated}\\ \text{pleiotropy}}}, \tag{2}$$

where Z_j is an indicator that is 1 if G_j affects U and 0 otherwise, and η is the effect of U on Y. This relationship is an extension of Equation (1) that includes terms allowing for both types of pleiotropic effects.



(a) Causal diagram for variants that are independent of the shared factor U.



(b) Causal diagram for variants that act on M through U.

Figure 3: CAUSE assumes that variants affect trait M through one of two mechanisms. A proportion $1-q$ of variants have the causal diagram in (a), while the remaining proportion, q, have the causal diagram in (b). (a) G_j has an effect on M of size $\beta_{M,j}$ and is independent of U. G_j can affect Y through two pathways: through the uncorrelated effect, θ_j, and through the effect of M on Y (if $\gamma \neq 0$). (b) The effect of G_j on M is mediated by U. We have assumed that U is scaled so that its effect on M is 1. This makes the effect of G_j on U equal to the effect of G_j on M, $\beta_{M,j}$. The effect of U on Y is equal to η, an unknown parameter. G_j can affect Y through three pathways: through θ_j, through the causal effect of M on Y, and through the effect of U on Y.

We assume that q is small, so that Z_j is equal to 0 for most variants (see Methods). We see in Equation 2 that, if there is no causal effect ($\gamma = 0$) and no variants affect a shared factor ($q = 0$), then $\beta_{Y,j}$ and $\beta_{M,j}$ are uncorrelated for all variants. If $\gamma = 0$ and q and η are non-zero, then $\beta_{Y,j}$ and $\beta_{M,j}$ are correlated for the small subset of variants with $Z_j = 1$. This is the pattern observed in Figure 2a. If there is a causal effect ($\gamma \neq 0$), $\beta_{Y,j}$ and $\beta_{M,j}$ are correlated for all variants (Figure 2b). We include every variant genome-wide in this model, but assume that θ_j and $\beta_{M,j}$ are equal to zero for most variants. This allows us to model uncertainty about variant effects, avoid pre-selection, and gain power from variants with weak evidence of association with M (see Methods).

We assess whether estimates of $\beta_{M,j}$ and $\beta_{Y,j}$ obtained from GWAS are consistent with a causal effect by comparing two nested models: the *sharing model* which has γ fixed at 0 allowing for only pleiotropic effects and no causal effect and the *causal model* in which γ is a free parameter. We compare the models using the expected log pointwise posterior density (ELPD; [15]), which measures how well the posterior distributions of a particular model are expected to predict a new set of summary statistics, for example, if both GWAS were repeated with new samples (see Methods). We estimate the difference in ELPD between the two models, ΔELPD, and the standard error of the estimator to generate a z-score and a corresponding one-sided p-value. A large, positive z-score provides support for the causal model, indicating that the data are consistent with a causal effect.

For computational simplicity, we use a likelihood for independent variants and prune variants for LD before estimating posterior distributions and computing test statistics. The effects of LD are discussed in more detail in Supplementary Note Section 5.4. Numerical evaluations in Section 2.2 are conducted using data simulated with a realistic LD pattern and pruned using the same method applied to biological data in Sections 2.3 and 2.5.

The causal diagram in Figure 3b is related to the model used in the latent causal variable (LCV) method proposed by O'Connor and Price [14]. The connection between CAUSE and LCV is discussed in Supplementary Note Section 5.5. LCV is conceptually different from CAUSE and other MR approaches. Rather than estimating and testing for causal effects of M on Y, LCV considers the proportion of heritability of each trait that is mediated by a heritable shared factor. This is summarized as the "genetic causality proportion" (GCP) which ranges from -1 to 1. Higher magnitude GCP indicates that the shared factor explains more genetic variation of one trait than the other, described as "partial causality" by O'Connor and Price [14]. GCP has no direct interpretation in CAUSE or other MR models and some non-causal scenarios can have large magnitude GCP. In these scenarios, LCV produces misleading results, if non-zero GCP is interpreted as evidence of a causal relationship (see Supplementary Note Section 5.5).

2.2 CAUSE can distinguish causality from correlated pleiotropy in simulations

We simulate summary statistics with realistic LD patterns to assess CAUSE in a variety of scenarios and compare performance with other MR methods. The power of trait M and Y GWAS can influence performance of all methods, so we consider low power (sample size 12,000) and a high power (sample size 40,000) settings for both traits. Both traits are simulated with a polygenic trait architecture – an average of 1,000 effect variants contribute to a total heritability of 0.25. Under the low and high power settings respectively, a median of 8 and 107 independent variants are genome-wide significant at $p < 5 \cdot 10^{-8}$. Variant effects on trait Y are generated from the relationship in Equation (2) and effect estimates are then simulated using true variant effects and the variant correlation (LD) structure using results of Zhu and Stephens [16] (see Methods). The LD structure for simulated data is estimated from the 1,000 genomes CEU samples [17].

We compare CAUSE to five MR methods: IVW regression [4], Egger regression [7], GSMR [10], MR-PRESSO [6], and the weighted median estimator [9]. Unlike the four other alternatives, the weighted median method assumes only that at least half of the variants satisfy the simple MR assumptions, and does not restrict the forms of pleiotropy allowed for the other variants. The other methods assume that no variants exhibit correlated pleiotropy. All alternative methods require pre-selection of variants with strong evidence of association with trait M. We followed the typical practice of selecting variants with $p < 5 \cdot 10^{-8}$ for association with M and pruning for LD so that no pair of variants has pairwise $r^2 > 0.1$.

We first evaluate the robustness of each method to correlated pleiotropy by simulating data with no causal effect and a proportion $q = 0 - 50\%$ of variants acting through a shared factor (Figure 4a). In all settings, methods except for CAUSE have an elevated false positive rate in the presence of correlated pleiotropy ($q > 0$). CAUSE makes more false detections when q is large and the GWAS have higher power. This is expected because the data pattern resulting from a high proportion of shared variants is similar to the pattern that results from a causal effect. Nonetheless, in all settings and for all parameter values, CAUSE has substantially lower false positive rates than any other method.

We next compare the power of each method when there is a true causal effect of M on Y and no shared factor (Figure 4b). The causal effect size is parameterized by the proportion of Y heritability explained by the causal effect of M, $\tau = \frac{\gamma^2 h_M^2}{h_Y^2}$. In our simulations $\tau = \gamma^2$ because M and Y have equal heritability. CAUSE has somewhat lower power than other methods in most settings. This is expected since CAUSE requires stronger evidence in order to conclude that the data are consistent with a causal effect. However, when the trait M GWAS has low power, CAUSE can achieve better power than other methods when causal effects are relatively large (at $N_M = 12,000$ and $N_Y = 40,000$). This is a result of using all variants genome-wide, rather than only those reaching genome-wide significance.

We note that the difference in performance of CAUSE and other MR methods cannot be explained simply by using a more stringent threshold to call causal effects in CAUSE. The CAUSE test statistic is better able to distinguish data simulated with a causal effect from data simulated with 30% correlated pleiotropy (Figure 4c). Remarkably, CAUSE is also better able to distinguish data simulated with a causal effect from data with no causal effect and no correltaed pleiotropy (Supplementary Figure S2).

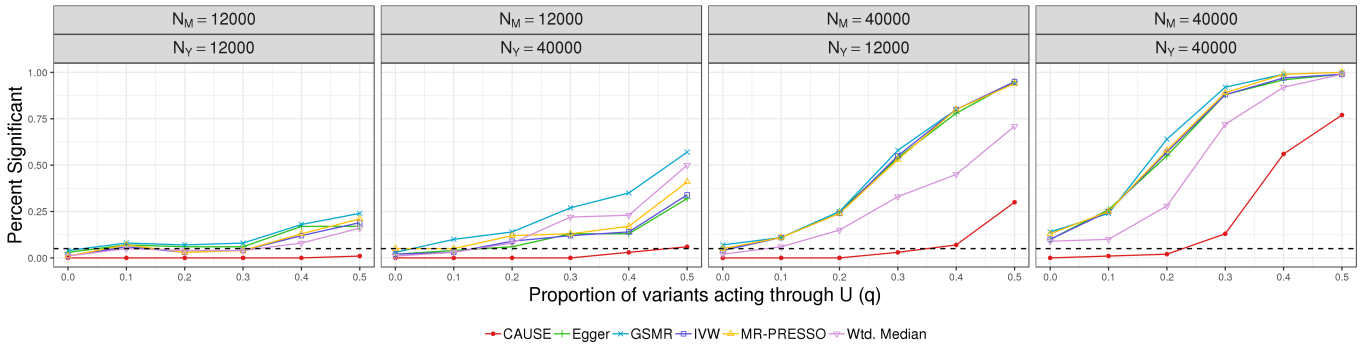
2.3 Identifying causal relationships between pairs of complex traits

We use CAUSE to analyze pairwise relationships for 20 traits with summary statistics available from published GWAS, (Supplementary Table S1). These include anthropometric traits, cardio-metabolic biomarkers (e.g. blood pressure, lipids) and complex diseases such as coronary artery disease (CAD) and type 2 diabetes (T2D). For each pair of traits we use CAUSE to compare the sharing model with the causal model. To account for multiple testing, we use the Benjamini-Hochberg procedure [18] to convert CAUSE p-values to q-values [19].

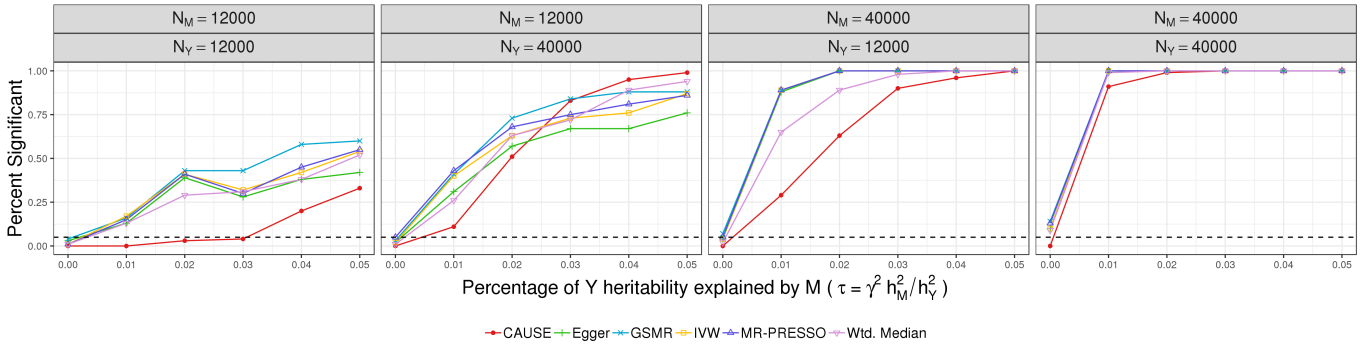
We analyze the same data using IVW regression, [4] Egger regression [7] and MR-PRESSO [6] to test for a causal effect of M on Y for each pair of traits. As with CAUSE, p-values for each method are converted to q-values. For all three methods, we include variants with $p < 5 \cdot 10^{-8}$ and mutual $r^2 < 0.1$ in the GWAS of trait M (see Methods). The results of these methods are qualitatively similar (Supplementary Figure S4) so we limit our discussion to comparing CAUSE and IVW regression.

Using CAUSE, 49 pairs of traits are significant at $q < 0.05$, while IVW regression identifies 104 pairs as significant at the same threshold (Figure 5). Seven pairs detected by CAUSE are not detected by IVW.

(a) False Positive Rate



(b) Power



(c) False Positive Rate vs Power

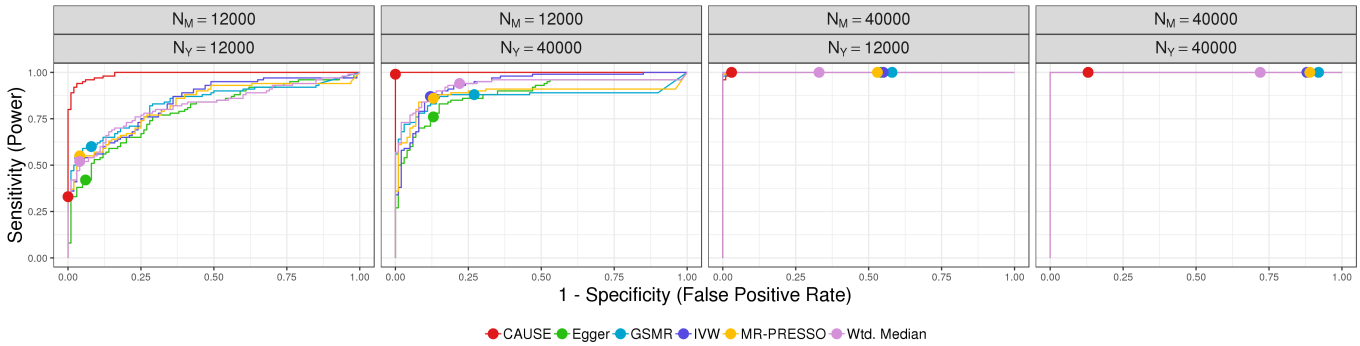


Figure 4: Performance of CAUSE and other MR methods in simulated data. (a) False positive rate averaged over 100 simulated data sets in settings with no causal effect and $\eta = \sqrt{0.05}$. The proportion of correlated pleiotropic variants (q) ranges from 0 to 50%. (b) Power averaged over 100 simulated data sets in settings with a causal effect and no shared factor. Causal effect size is parametrized by τ, the proportion of trait Y heritability explained by the causal effect of M. (c) We compare the power when $\gamma = \sqrt{0.05}$ to the false positive rate when there is no causal effect, but a proportion $q = 0.3$ of variants act through a shared factor with effect $\eta = \sqrt{0.05}$ on Y. There are 100 simulations each in the causal and non-causal scenarios. Curves are created by varying the significance threshold from very stringent (lower left) to very permissive (upper right). Points indicate the power and false positive rate achieved at a threshold of $p \leq 0.05$.

Four of these are effects of blood pressure phenotypes on CAD risk. The others are an effect of systolic blood pressure on stroke risk, an effect of birth length on adult height, and an effect of fasting glucose on type 2 diabetes risk. Effects of blood pressure on CAD are plausibly real causal effects based on evidence from clinical trials [20]. The failure of IVW regression to detect these effects is likely due to low power in the GWAS of the blood pressure phenotypes. The four phenoytpes have between five and nine independent variants reaching genome-wide significance. However, there are many sub-threshold variants that contribute to the CAUSE test statistic to bolster evidence for a causal effect (Figure 6a).

Many pairs of traits detected by both methods are plausible effects such as BMI and LDL cholesterol on CAD risk (Figure 6b) and blood pressure on stroke risk. Others are very closely related phenotypes such as body fat percentage and BMI. In many of these cases, CAUSE indicates that the data are consistent with causal effects in both directions. However, we interprete of these results to indicate that almost all effect variants are shared rather than that the traits are mutually causal (see Discussion).

IVW regression detects a large number of trait pairs that are not detected by CAUSE. Many of these may be false positives driven by variants affecting shared factors. For example, IVW detects significant causal effects of CAD on a number of traits, some of which are highly implausible, e.g. CAD effect on birth weight (Figure 5). Other examples include effects of blood pressure on BMI, and total cholesterol and triglycerides on adult height.

The analysis of CAD and LDL cholesterol provides an illustrative example of how correlated pleiotropy can induce errors using IVW regression and how these are avoided by CAUSE. Both IVW and CAUSE detect an effect of LDL cholesterol on CAD risk (IVW q-value $3.4 \cdot 10^{-30}$; CAUSE q-value $3.2 \cdot 10^{-5}$; Figure 6b), which is consistent with the current predominant view of the relationship between LDL and CAD risk [21]. However, IVW also detects a causal effect in the opposite direction, of CAD on LDL cholesterol (Figure 6c), which is most plausably a false detection. This occurs because some of the variants associated with CAD are acting through LDL and therefore have correlated effects on both traits, leading to the false finding by IVW. CAUSE avoids this problem because there are many variants that affect CAD and are independent of LDL. These variants provide evidence against a causal effect of CAD on LDL and make large contributions to the CAUSE test statistic (Figure 6c).

CAUSE does not entirely avoid likely false detections. For example, recent research suggests that the signal of an effect of HDL cholesterol on CAD is driven by variants that affect both LDL and HDL cholesterol [22, 23]. This effect is detected by both IVW regression and CAUSE. This is likely driven by a very high proportion of shared variants, which is hard to differentiate from a causal effect. Under the sharing model, CAUSE estimates that 54% of HDL variants are shared with CAD (Supplementary Figure S3).

The CAUSE results suggest that many pairs of traits share variants that act through a common pathway. Under the sharing model, the posterior median of q exceeds 10% for 114 trait pairs (the prior median is 7%; Supplementary Figure S3). These include 78 (75%) of the trait pairs detected as significant by IVW regression. This overlap is expected because a positive IVW result indicates that top trait M variants have, on average, correlated effects on Y. High levels of sharing are also suggested by large correlation estimates made by cross-trait LD-score regression [24] (Supplementary Figure S5). However, in many cases, CAUSE results suggest that these patterns can be explained by a shared factor rather than a causal effect. Together, these results suggest that shared factors may be major contributors to false positives using traditional MR methods.

2.4 CAUSE provides insights on biological mechanisms driving genetic correlation of trait pairs

CAUSE posterior estimates under the sharing model can yield insights into the relationship between M and Y when correlation is driven by a shared factor rather than a causal effect. CAUSE estimates a posterior probability of acting through U for each variant, under the sharing model. Identifying variants that act through a shared factor may help identify biological processes that influence both traits.

To illustrate, we consider the relationship between birth weight and T2D risk, treating birth weight as the mediator. IVW regression identifies an average negative correlation of effect sizes between the two traits (IVW q-value 7×10^{-4}), however, CAUSE does not find the data to be consistent with a causal effect (CAUSE q-value 0.32; Figure 6d). To identify candidate genes and pathways that mediate sharing between these traits, we first assign variants to genes using cis-eQTLs from the GTEx project. Variants that are

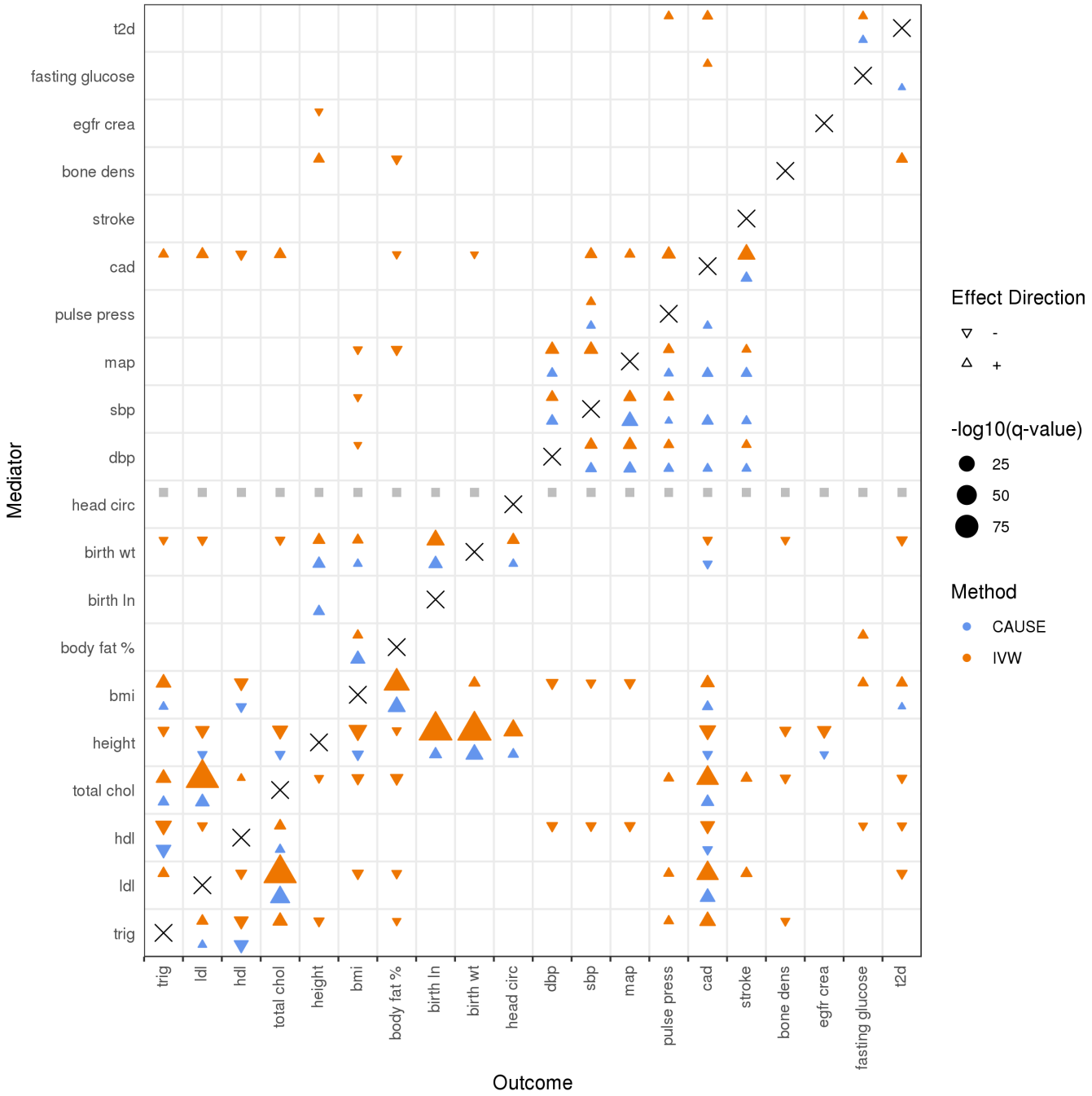


Figure 5: Each cell summarizes the IVW (upper, orange) and CAUSE (lower, blue) results for a pair of traits. Points are shown only for pairs significant at an FDR threshold of 0.05. Grey squares indicate missing values and occur for IVW regression when there are fewer than 2 genome-wide significant loci.

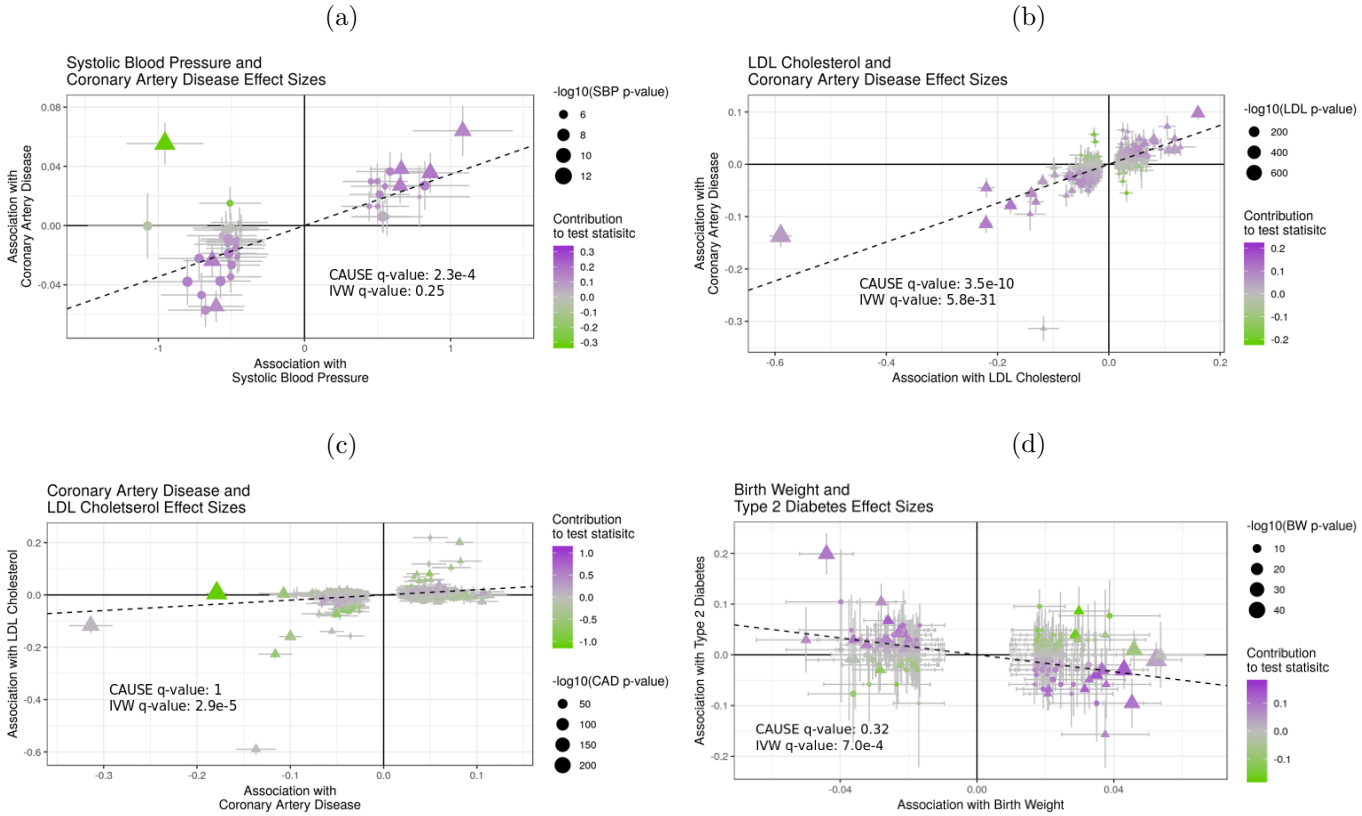


Figure 6: Effect size estimates and variant level contribution to CAUSE test statistics for four trait pairs. Effect estimates for trait M (horizontal axis) are plotted against estimates for trait Y (vertical axis). Point size is proportional to the p-value for trait M. Error bars have length 1.96 times the standard error of the estimate. Triangles indicate variants reaching genome-wide significance for trait M ($p < 5 \cdot 10^{-8}$). Only points with trait M p-value $< 5 \cdot 10^{-5}$ are shown. Color indicates the contribution of each variant to the CAUSE test statistic with positive values (purple) in favor of the causal model and negative values (green) in favor of the sharing model. Dotted lines show the IVW estimate obtained using only genome-wide significant variants. a) Systolic blood pressure (M) and CAD (Y). Only seven variants reach genome-wide significance so IVW regression is under-powered. Sub-threshold variants contribute to evidence in favor of the causal model using CAUSE. b) LDL cholesterol (M) and CAD (Y). Both methods detect a strong signal consistent with a causal effect. c) CAD (M) and LDL cholesterol (Y). Although the average correlation in effect estimates significantly different from zero, variants with large effects on CAD and no effect on LDL prevent CAUSE from making a false discovery. d) Birth weight (M) and type 2 diabetes (Y). IVW regression results suggest a causal effect while CAUSE results indicate that the data may be explained by a shared factor, because many variants (green) contradict the ovreall negative correlation in effect sizes.

not cis-eQTLs are annotated to the nearest gene. We then test gene sets defined in the Gene Ontology database [25, 26] for enrichment of genes associated with variants that have a high posterior probability of acting through U (see Methods). The most enriched gene sets include those related to cell division, to PIP2 binding, and to the integrin complex (Supplementary Table S2). The integrin pathway results are driven by several integrin subunit genes (Supplementary Table S3), one of which, ITGA1, is among the top loci associated with related traits including T2D, fasting insulin, β-cell function and glucose tolerance [27]. This analysis suggests specific genes and pathways that may drive shared genetic basis of T2D and birth weight.

2.5 Links between auto-immune disease and blood cell counts

There are a growing number of GWAS for intermediate or biomarker traits. Relationships between these traits and clinical outcomes are appealing targets for MR studies because biomarkers may be more likely than complex traits to causally affect disease risk, and may be accessible as drug targets. We use CAUSE to search for evidence of causal relationships between 13 measures of blood cell composition and eight auto-immune diseases with publicly available summary statistics (Supplementary Table S4).

At an FDR threshold of 0.05, only 2 of the 104 pairs of traits are found to be consistent with a causal effect using CAUSE (Supplementary Figures S6 and S7). These are positive effects of eosinophil count on asthma and allergy risk, both of which are supported by literature [28]. IVW regression also identifies these pairs at the same significance threshold. In addition IVW regression detects 20 other pairs (Supplementary Figure S6), including 12 causal effects of red blood cell traits, such as red cell distribution width (RDW) and mean corpuscular hemoglobin (MCH), on multiple diseases. Since red blood cells have no known direct role in autoimmunity, these results may be false positives due to sharing of variants that act through a common factor. Consistent with this, CAUSE estimates more than 10% of variants to act through a shared factor for 13 of the trait pairs found to be significant by IVW regression (Supplementary Figure S7).

3 Discussion

We have introduced CAUSE, a new approach to MR analysis that accounts for uncorrelated and correlated pleiotropy. In simulations CAUSE avoids false positives in the presence of pleiotropic effects that bias other methods. With highly powered GWAS, CAUSE tends to be more conservative than other methods, requiring more evidence for a causal effect. However, when GWAS are under-powered, CAUSE improves power by using all variants genome-wide rather than relying only on highly significant associations. Previous authors have identified uncorrelated horizontal pleiotropy as a pervasive phenomenon [6] that may adversely impact MR analyses. Our analyses demonstrate that correlated horizontal pleiotropy is also common and may explain a large fraction of positive results obtained using other MR methods.

Caution must be used when interpreting results of CAUSE, as well as any other MR method. CAUSE tests that the GWAS summary statistics for M and Y are consistent with a model in which every M effect variant has a correlated effect on Y. This pattern occurs when M has a causal effect on Y but can also occur in other circumstances. Notably, if most of the heritable variation of two traits is mediated by the same unobserved process, we expect to observe this pattern in both trait directions. This can be seen in our analysis of complex traits when CAUSE rejects the sharing model in both directions for pairs of traits like BMI and body fat percentage that are very closely related. With this in mind, it may be useful to test trait pairs in both directions and give more attention to effects that are seen in only one direction [13].

CAUSE has several limitations that provide interesting future research directions. First, it is not curently possible to account for known shared factors or confounders when they are measured. For example, when evaluating the relationship between HDL cholesterol and heart disease risk, it may be desirable to control for the effect of LDL cholesterol. Second, CAUSE models only a single shared factor, so it may not fully account for shared genetic components between two traits of interest. This problem is partially alleviated by the flexibility of the empirical prior distribution (see Methods). Finally, CAUSE simply prunes variants for LD rather than explicitly modeling variant correlation. This helps ensure that the problem is computationally tractable, however, using only one variant per LD block may lead to loss of information.

4 Methods

4.1 CAUSE model for GWAS Summary Statistics

We use effect estimates and standard errors measured in GWAS of traits M and Y (summary statistics) to look for evidence of a causal effect of M on Y. Let $(\hat{\beta}_{M,j}, \hat{s}_{M,j})$ and $(\hat{\beta}_{Y,j}, \hat{s}_{Y,j})$ be effect estimates and standard errors at variant G_j $(j = 1, \ldots, p)$ for traits M and Y respectively. Let $\beta_{M,j}$ and $\beta_{Y,j}$ be the true marginal associations of G_j with M and Y. We model effect estimates as normally distributed given the true effects, allowing for global correlation that can result from overlapping GWAS samples. We model

$$p\left(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j} \middle| \beta_{M,j}, \beta_{Y,j}, \rho, \hat{s}_{M,j}, \hat{s}_{Y,j}\right) = N_2\left(\begin{pmatrix}\hat{\beta}_{M,j}\\ \hat{\beta}_{Y,j}\end{pmatrix}; \begin{pmatrix}\beta_{M,j}\\ \beta_{Y,j}\end{pmatrix}, S_j(\rho)\right), \tag{3}$$

where where $S_j(\rho) = \begin{pmatrix}\hat{s}^2_{M,j} & \rho\hat{s}_{M,j}\hat{s}_{Y,j}\\ \rho\hat{s}_{M,j}\hat{s}_{Y,j} & \hat{s}^2_{Y,j}\end{pmatrix}$ and $N_2(x; \mu, \Sigma)$ is the bivariate normal density with mean μ and variance Σ evaluated at x. This model implicitly assumes that $\hat{s}_{M,j}$ and $\hat{s}_{Y,j}$ are measured without error. The correlation term, ρ, which accounts for the effects of sample overlap, is estimated empirically (Supplementary Note Section 5.1).

In the CAUSE model, a proportion, q of variants exhibit correlated pleiotropy, modeled as an effect on a shared factor, U (Figure 3b). The remaining proportion $1 - q$ of variants are independent of U (Figure 3a). All variants may have pleiotropic effects, θ_j on Y that are uncorrelated with their effects on M. Let Z_j be an indicator that variant j effects U. Then,

$$\begin{aligned}\beta_{Y,j}|Z_j = &\underbrace{\gamma\beta_{M,j}}_{\text{causal effect}} + \underbrace{Z_j\eta\beta_{M,j}}_{\substack{\text{correlated}\\ \text{pleiotropy}}} + \underbrace{\theta_j}_{\substack{\text{uncorrelated}\\ \text{pleiotropy}}}\\ Z_j \sim &\text{Bernoulli}(q),\end{aligned} \tag{4}$$

where η is the effect of U on Y and U is scaled so that the effect of U on M is 1. Note that if there are no uncorrelated or correlated pleiotropic effects, then Equation (4) reduces to $\beta_{Y,j} = \gamma\beta_{M,j}$, which is the relationship assumed by simple MR approaches such as IVW regression. We substitute the right side of Equation (4) into Equation (3) and integrate out Z_j to obtain

$$\begin{aligned}p\left(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j} \middle| \beta_{M,j}, \theta_j, \rho, \hat{s}_{M,j}, \hat{s}_{Y,j}, \gamma, \eta, q\right) =& qN_2\left(\begin{pmatrix}\hat{\beta}_{M,j}\\ \hat{\beta}_{Y,j}\end{pmatrix}; \begin{pmatrix}\beta_{M,j}\\ (\gamma+\eta)\beta_{M,j} + \theta_j\end{pmatrix}, S_j(\rho)\right) +\\ &(1-q)N_2\left(\begin{pmatrix}\hat{\beta}_{M,j}\\ \hat{\beta}_{Y,j}\end{pmatrix}; \begin{pmatrix}\beta_{M,j}\\ \gamma\beta_{M,j} + \theta_j\end{pmatrix}, S_j(\rho)\right).\end{aligned} \tag{5}$$

The parameters of interest in the CAUSE model are γ, η, and q. Rather than estimating individual variant effects $\beta_{M,j}$ and θ_j, we model their joint distribution empirically and integrate them out to obtain a marginal density for $\hat{\beta}_{M,j}$ and $\hat{\beta}_{Y,j}$. We model $\beta_{M,j}$ and θ_j as draws from a mixture of bivariate normal distributions. This strategy is based on Adaptive Shrinkage (ASH) approach for modeling univariate distributions [29] and provides a flexible unimodal distribution with mode at $(0, 0)$. We model

$$p\left(\beta_{M,j}, \theta_j | \pi_0, \ldots, \pi_K, \Sigma_0, \ldots, \Sigma_K\right) = \sum_{k=0}^{K} \pi_k N_2\left(\begin{pmatrix}\beta_{M,j}\\ \theta_j\end{pmatrix}; \begin{pmatrix}0\\ 0\end{pmatrix}, \Sigma_k\right), \tag{6}$$

where $\Sigma_0, \Sigma_1, \ldots, \Sigma_K$ are a set of pre-specified covariance matrices of the form $\Sigma_k = \begin{pmatrix}\sigma^2_{k,1} & 0\\ 0 & \sigma^2_{k,2}\end{pmatrix}$, and $\pi_0, \ldots, \pi_K$ are a set of mixing proportions that sum to 1. The set of parameters $\Omega = \{\pi_0, \ldots, \pi_K, \Sigma_0, \ldots, \Sigma_K\}$ are estimated from the data along with ρ in a single pre-processing step (Supplementary Note Section 5.1).

Integrating the distribution of $\beta_{M,j}$ and θ_j out of Equation (5), we obtain

$$p(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j}|\gamma, \eta, q, \hat{s}_{M,j}, \hat{s}_{Y,j}, \Omega, \rho) = q \sum_{k=0}^{K} \pi_k N\left(\begin{pmatrix}\hat{\beta}_{M,j}\\ \hat{\beta}_{Y,j}\end{pmatrix}; \begin{pmatrix}0\\0\end{pmatrix}, A(\gamma+\eta)\Sigma_k A(\gamma+\eta)^\top + S_j(\rho)\right) + (1-q)\sum_{k=0}^{K} \pi_k N\left(\begin{pmatrix}\hat{\beta}_{M,j}\\ \hat{\beta}_{Y,j}\end{pmatrix}; \begin{pmatrix}0\\0\end{pmatrix}, A(\gamma)\Sigma_k A(\gamma)^\top + S_j(\rho)\right), \quad (7)$$

where $A(x) = \begin{pmatrix}1 & 0\\ x & 1\end{pmatrix}$. Treating variants as independent, we obtain a joint density for the entire set of summary statistics as a product over variants. We place independent prior distributions on γ, η, and q, (Section 4.3), and estimate posterior distributions via adaptive grid approximation (Supplementary Note Section 5.3). CAUSE is implemented in an open source R package (https://github.com/jean997/cause). A flow-chart illustrating a CAUSE analysis including parameter estimation is shown in Supplementary Figure S1.

4.2 Accounting for Linkage Disequilibrium

We treat variants as independent when we compute the joint density of summary statistics. In reality, variants are correlated due to linkage disequilibrium (LD). We define the *LD-transformed effects* $\beta^*_M = S_M R S_M^{-1}\beta_M$ and $\beta^*_Y = S_Y R S_Y^{-1}\beta_Y$, where R is the variant correlation matrix matrix and S_M and S_Y are diagonal matrices with diagonal elements $(s_{M,j})$ and $(s_{Y,j})$ [16]. We assume that R is the same in both GWAS. The pair of estimates $(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j})$ can be approximated as normally distributed with mean $(\beta^*_{M,j}, \beta^*_{Y,j})$ and variance $S(\rho)$. If the relationship in Equation (4) holds for true effects, $\beta_{M,j}$ and $\beta_{Y,j}$, and either M effects are sparse relative to LD structure (most M effect variants are independent) or q is small, then the relationship between the LD-transformed effects is

$$\beta^*_{Y,j} = \gamma\beta^*_{M,j} + Z_j\eta\beta^*_{M,j} + \theta^*_j, \quad (8)$$

where $\theta^* = S_Y R S_Y^{-1}\theta$ (see Supplementary Note Section 5.4). In this case, the mean relationship for summary statistics at a single locus is the same with or without LD and Equation (5) and (7) remain valid for variants in LD if $\beta_{M,j}$ and θ_j are replaced with $\beta^*_{M,j}$ and θ^*_j.

Correlations among variants can affect the joint density of all summary statistics, which we compute as a product of the densities at each variant. To account for this, we use a subset of variants with low mutual LD ($r^2 < 0.1$ by default), prioritizing variants with low trait M p-values ito improve power.

If M effect variants are dense relative to LD structure and q is large, LD can induce a positive correlation between Y and M effect estimates for all variants even when there is no causal effect, leading to false positives (see Supplementary Note Section 5.4). In this case, CAUSE may have an inflated false positive rate.

4.3 Prior Distributions for γ, η, and q

We place normal prior distributions with mean zero on γ and η. We find in simulations that results are robust to different choices of prior variance for γ and η, as long the same value is used for both parameters (Supplementary Note Section 5.2). Since the magnitude of a possible causal effect is difficult to know a priori due to differences in trait scaling and covariate adjustment across GWAS, we use the data to suggest a prior variance (Supplementary Note Section 5.2).

By default, and in all analyses presented here, we use a Beta$(1, 10)$ prior distribution for q. This distribution gives a prior probability of 0.001 that $q > 0.5$ and 0.056 that $q > 0.25$. In our implementation of CAUSE, the parameters of the Beta distribution can be adjusted by the user to reflect different prior beliefs about the size of q. For example, when one uses a stronger prior with prior mean closer to 0, q is restricted to be nearly equal to 0. In the limit, this reduces to standard MR, while still accounting for polygenicity of traits, sample overlap, and uncorrelated pleiotropy.

4.4 Model Comparison Using ELPD

To determine whether GWAS summary statistics are consistent with a causal effect of M on Y we compare a model in which the causal effect is fixed at zero (the *sharing model*) to a model that allows a non-zero causal effect (the *causal model*). To compare the fit of these models, we estimate the difference in the *expected log pointwise posterior density* (ΔELPD) [15]. The ELPD measures how well the posterior distributions estimated under a given model are expected to predict a hypothetical new set of summary statistics obtained from GWAS of M and Y in different samples.

Let Θ be the set of parameters (γ, η, q). Let $p_C(\Theta|\text{Data})$ and $p_S(\Theta|\text{Data})$ be the posterior density of Θ given the observed summary statistics under the causal model and sharing model respectively. Let y'_j denote a new observation of $(\hat{\beta}_{M,j}, \hat{\beta}_{Y,j}, \hat{s}_{M,j}, \hat{s}_{Y,j})$. The ELPD for model $m \in \{C, S\}$ (for causal and sharing) is

$$\text{ELPD}_m = \sum_{j=1}^{p} \left(\int \left(\int p(y'_j|\Theta, \Omega, \rho) p_m(\Theta|\text{Data}) d\Theta \right) p_{\text{true}}(y'_j) dy'_j \right). \tag{9}$$

where $p(y'_j|\Theta, \Omega, \rho)$ is given in Equation (7) and p_{true} is the probability density under the true data generating mechanism. If $\Delta\text{ELPD} = \text{ELPD}_C - \text{ELPD}_S$ is positive, then the posteriors from the causal model predict the data better so the causal model is a better fit. If $\Delta\text{ELPD} \leq 0$, then the sharing model fits at least as well, indicating that the data are not consistent with a causal effect.

We estimate ΔELPD and a standard error of the estimator using the Pareto-smoothed importance sampling method described by Vehtari, Gelman, and Gabry [15] and implemented in the `R` package `loo`. We then compute a z-score, $z_{elpd} = \frac{\widehat{\Delta\text{elpd}}}{\hat{se}\left(\widehat{\Delta\text{elpd}}\right)}$, that is larger when posteriors estimated under the causal model fit the data better than the posteriors estimated under the sharing model. We compute a one-sided p-value by comparing the z-score to a standard normal distribution. The p-value estimates the probability of obtaining a z-score larger than the one observed if the true value of ΔELPD were less than or equal to zero.

4.5 Generating Simulated Summary Statistics

To create data with a realistic correlation structure, we estimate LD for 19,490 HapMap variants on chromosome 19 in the CEU 1,000 Genomes population using LDShrink ([30]; `https://github.com/stephenslab/ldshrink`) and replicate this pattern 30 times to create a genome sized data set of $p = 584,700$ variants. We set the heritability of each trait, h^2_M and h^2_Y, the sample size for each GWAS, N_M and N_Y, and the expected number of effect variants m_M and m_θ. Additionally, we set γ, η, and q and generate effect estimates from the CAUSE model in Equation (4). We note that m_θ is the expected number of variants with $\theta_j \neq 0$, rather than the expected number of variants with non-zero effect on Y.

We simulate an effect estimate and standard error for each variant using the following procedure. First, standardized effects $\tilde{\beta}_{M,j} = \beta_{M,j}\sqrt{2f_j(1-f_j)}$ and $\tilde{\theta}_j = \theta_j\sqrt{2f_j(1-f_j)}$are drawn from a mixture distribution:

$$\begin{aligned} \tilde{\beta}_{M,j} \sim & (1-\pi_M)\delta_0 + \pi_M N(0, \sigma^2_M) \\ \tilde{\theta}_j \sim & (1-\pi_\theta)\delta_0 + \pi_\theta N(0, \sigma^2_\theta), \end{aligned} \tag{10}$$

where f_j is the allele frequency of variant j in the 1,000 Genomes CEU population. Variances σ^2_M and σ^2_θ are chosen to give the desired expected heritability as $\sigma^2_M = \frac{h^2_M}{m_M}$ and $\sigma^2_\theta = \frac{h^2_Y - (\gamma^2 + q\eta^2)h^2_M}{m_\theta}$. Mixing parameters are $\pi_M = m_M/p$ and $\pi_\theta = m_\theta/p$. Note that $\gamma^2 h^2_M/h^2_Y$ is the proportion of trait Y heritability mediated by the causal effect and $q\eta^2 h^2_M/h^2_Y$ is the proportion of trait Y heritability mediated by U.

Second, standardized effects are converted to non-standardized effects and standard errors are computed as as $s_{\cdot j} = \sqrt{\frac{1}{(2*N_{\cdot}*f_j*(1-f_j)}}$, where $\cdot$ may be M or Y. Third, Z_j are drawn from a Bernoulli(q) distribution and true effects $\beta_{\gamma,j}$ are computed using Equation (4). Finally, effect estimates are simulated from true effects as

$$\hat{\boldsymbol{\beta}}_{\cdot} \sim N_p(\boldsymbol{S}_{\cdot}RS_{\cdot}\boldsymbol{\beta}_{\cdot}, S_{\cdot}RS_{\cdot}^{-1}) \tag{11}$$

where $S_\cdot$ is the diagonal matrix of standard errors and R is the variant correlation matrix [16]. Simulations can be replicated using the online tutorial (`https://jean997.github.io/cause/simulations.html`).

4.6 Existing MR Methods

We compare the performance of CAUSE in simulated data with five other MR methods. These are implemented as follows.

- IVW regression: Implemented by us in R.
- Egger regression: Implemented by us in R.
- MR-PRESSO: Performed using `MRPRESSO` R package ([6], `https://github.com/rondolab/MR-PRESSO`) with outlier and distortion tests.
- GSMR: Performed using `gsmr` R package ([10], `http://cnsgenomics.com/software/gsmr/`) with the Heidi outlier test, default threshold 0.01, and minimum number of instruments lowered to 1.
- Weighted median: Performed using R code available in [31].

4.7 Gene set enrichment analysis with shared variants of trait pairs

When CAUSE results indicate that summary statistics for a pair of traits are not consistent with a causal effect, but there is a substantial amount of correlated pleiotropy (i.e. the posterior distribution of q is large), CAUSE posterior estimates can be used to identify which pathways contribute to this pleiotropy. We use a four step procedure to identify pathways enriched for shared variants. First, we assign variants to genes. We include only variants with p-value for association with $M < 0.001$, pruned for LD with a threshold of $r^2 < 0.1$. These are the variants used to compute the CAUSE posterior distributions. Variants that are significantly associated with gene expression (q-value < 0.05) in any of 48 tissues in the GTEx catalog (cis-eQTLs) are assigned to the genes they are associated with. Variants that are not cis-eQTLs are assigned to the nearest gene. Some variants may be assigned to multiple genes. Second, genes are assigned to gene sets defined by the Gene Ontology database ([25, 26]; `http://geneontology.org`). Variants are assigned to gene sets transitively. Third, for each gene, i we compute the median posterior probability of acting through U for all variants assigned to that gene, denoted P_i. Finally, for each gene set j, we assess evidence for an enrichment of genes with variants that are likely to be shared using logistic regression. We model

$$\text{logit}(\text{Gene } i \text{ in set } j) = \alpha_{0,j} + \alpha_{e,j} P_i + \epsilon_{ij}. \tag{12}$$

For each pathway we estimate $\alpha_{e,j}$ and $\alpha_{0,j}$ using logistic regression and test the hypothesis that $\alpha_{e,j}$ is equal to zero. We use a Benjamini-Hochberg correction to account for multiple testing.

4.8 URLs

- Website: `https://jean997.github.io/cause/`
- CAUSE Software (R package): `https://github.com/jean997/cause`
- Simulations Software (R package): `https://github.com/jean997/causeSims`

4.9 Acknowledgements

Expression quantitative trait loci in Section 2.4 were obtained from the Genotype-Tissue Expression (GTEx) project accessed on Thursday, April 25 2019 02:10:21 PM CDT. This work was supported by National Institutes of Health (NIH) grants MH110531 (to XH) and HG002585 (to MS), and a Research Grant from the March of Dimes (XH).

5 Supplementary Note (Appendix)

5.1 Empirical Parameter Estimation

CAUSE analysis involves two steps (see Supplementary Figure S1). We first empirically estimate ρ, the global correlation of summary statistics that can result from GWAS sample overlap and the parameters that define the joint prior distribution of $\beta_{M,j}$ and θ_j. We do this in two sub-steps: 1) Select a panel of candidate covariance matrices $\Sigma_0, \ldots, \Sigma_K$ and 2) Fix $\gamma = \eta = 0$ and calculate the maximum a posteriori (MAP) values of ρ and $\pi_0, \ldots, \pi_K$.

The set of candidate covariance matrices should be large enough that a flexible joint distribution can be fit for $\beta_{M,j}$ and θ_j, but not so large that evaluating the likelihood in (7) becomes burdensome. To choose this set, we first apply the Adaptive Shrinkage (ASH) method proposed by [29] to estimate the distributions of $\boldsymbol{\beta}_M$ and $\boldsymbol{\beta}_Y$ separately. Briefly, given a set of summary statistics for a single study, ASH estimates a sparse unimodal distribution for the marginal effects. This distribution is flexible and parameterized as a mixture of univariate normal distributions centered at 0. ASH uses the model

$$\beta_{\cdot,j} \sim \sum_{l=0}^{L} \varpi_l N(0, \varsigma_l^2),$$

where $\varsigma_0, \ldots, \varsigma_L$ are a fixed grid of variances with $\varsigma_0 = 0$. ASH estimates the mixing proportions $\varpi_0, \ldots, \varpi_L$ using a prior on $\boldsymbol{\varpi}$ that encourages more weight to be given to ϖ_0, the proportion of effects equal to 0. Despite starting with a large number of candidate variances, ASH solutions tend to place most of the weight on only a few values. The resulting solution is sparse (most of the estimated effects are 0) and parsimonious (there are few components in the model with non-zero mixing proportion).

Let $\varsigma_{M,0}, \ldots, \varsigma_{M,l_M}$ and $\varsigma_{Y,0}, \ldots, \varsigma_{Y,l_Y}$ be the set of variances with non-zero weight in the ASH estimates for traits M and Y respectively. Because ASH encourages sparsity, in all cases $\varsigma_{M,0} = \varsigma_{L,0} = 0$. We construct the panel of candidate 2×2 covariance matrices by taking all pairs of these variances as diagonal elements and setting the off diagonal elements to be 0. Thus if $l_M = 4$ and $l_Y = 3$, our method produces a set of $(4+1)(3+1) = 20$ candidate covariance matrices.

In the second step, we fix $\gamma = \eta = 0$ and calculate the MAP values of ρ and $\pi_0, \ldots, \pi_K$. We use a Dirichlet$(10, 1, \ldots, 1)$ prior on $\pi_0, \ldots, \pi_K$ with π_0 corresponding to the covariance matrix of all zeros. This prior is the same prior used by default in ASH and encourages a sparse solution, however the weights may be adjusted by the user in the CAUSE software. We use a prior on $z = \tanh^{-1}(\rho) = \frac{log(1+\rho)}{2(1+\rho)}$ of $z \sim N(0, 0.25)$, which is a weak prior encouraging ρ to be close to zero. To calculate the MAP estimate, we use coordinate descent, alternating between optimization of ρ with $\boldsymbol{\pi}$ fixed and optimization of $\boldsymbol{\pi}$ with ρ fixed. As observed by [29] and others, maximization in $\boldsymbol{\pi}$ is a convex optimization problem that can be completed quickly. In practice, we find that convergence is usually reached within five iterations.

5.2 Prior Distributions for γ and η

In most cases, little prior information is available about the size of the causal or confounding effect. Differences in variable scaling and covariate adjustment across GWAS may make it difficult to predict the magnitude of these effects. Fortunately, we find that CAUSE results are robust to a wide range in prior distributions for these parameters. We require that the same prior be used for γ and η. If this is not the case false positives can arise when the true shared factor effect is better represented by the prior on γ than the prior on η. We use normal prior distributions with mean 0 and variance $\sigma^2_{\gamma\eta}$ for γ and η.

To assess the robustness of CAUSE to changes in $\sigma^2_{\gamma\eta}$, we analyze data simulated from three scenarios using a range of values for $\sigma^2_{\gamma\eta}$. The three scenarios are 1) a setting with a causal effect ($\gamma = \sqrt{0.05}$, $\eta = 0$, $q = 0$), 2) a setting with no causal effect but some correlated pleiotropy ($\gamma = 0$, $\eta = \sqrt{0.05}$, $q = 0.3$), and 3) a setting with neither a casual effect or correlated pleiotropy ($\gamma = 0$, $\eta = 0$, $q = 0$). We analyze simulated data using three values of $\sigma^2_{\gamma\eta}$. These are chosen so that $\sqrt{0.05}$, is at the 80th, 65th, and 51st quantile of the $N(0, \sigma^2_{\gamma\eta})$ distribution, giving small, medium, and large values of $\sigma_{\gamma\eta}$ respectively. All simulations are conducted with sample sizes $N_M = N_Y = 40,000$.

We compare p-values and posterior median estimates for γ, η, and q under the causal and sharing models across different values of $\sigma^2_{\gamma\eta}$. The Pearson correlation in p-values was greater than 0.95 for all settings and

between all pairs of values of $\sigma^2_{\gamma\eta}$. The correlation in posterior medians was higher than 0.8 for all parameters and all settings except for estimates of q in setting 1 under the causal model. These had a somewhat lower correlation (minimum correlation 0.51). However, in all cases the posterior median of q in the causal model was very low – less than 0.05. These results demonstrate that very similar inference can be obtained using a wide range of prior distributions for γ and η.

By default, $\sigma^2_{\gamma\eta}$ is chosen using the data. We use a set of variants with trait M p-value $< 10^{-3}$ and compute $\hat{\gamma}_{max} = \max|\frac{\hat{\beta}_2}{\hat{\beta}_1}|$. This is the largest magnitude of causal estimate that could be achieved using only one variant. We then choose $\sigma^2_{\gamma\eta}$ so that the prior probability that γ or η has magnitude larger than $\hat{\gamma}_{max}$ is 0.05.

5.3 Approximating posteriors of γ, η, and q

We use an adaptive variation of a simple grid posterior approximation[32] to approximate the joint posterior distribution of γ, η, and q. To compute this approximation, we begin with initial bounds on γ and η of $(-1, 1)$. These will be adaptively expanded as needed. The bounds on q are fixed at $(0, 1)$.

The approximation proceeds as follows:

1. The domain of (γ, η, q) is divided into a coarse set of cubes. The approximate posterior probability of each cube is computed by approximating the likelihood within the cube as constant and equal to the likelihood at the midpoint of the cube.

2. After the first rough approximation, the bounds of γ and η are expanded so that less that 0.001 of the posterior mass falls in the cubes closest to the boundary. These bounds are then fixed.

3. The grid is then iteratively refined until no cube contains more than 1% of the posterior density. At each iteration, all cubes containing more than this are subdivided into nine smaller cubes and the posterior is re-estimated.

5.4 Effects of LD

In Sections 2.1 and 4.2 and we briefly describe the effects of LD. Here, we discuss these in greater details. CAUSE relies on two assumptions. The first is that the joint likelihood of all pairs of summary statistics can be factorized into the product of the likelihood for each variant. Variants in LD are not independent, however, by pruning variants so that the set is nearly independent, we can approximate this condition. The second is that

$$\mathrm{Cov}(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}|Z_j) = (\gamma + Z_j\eta)\mathrm{Var}(\beta_{M,j}) = (\gamma + Z_j\eta)\left(\mathrm{Var}(\hat{\beta}_{M,j}) - s^2_{M,j}\right) \tag{13}$$

Without LD, this is a consequence of Equation (4) and the assumption that $\hat{\beta}_{M,j}$ and $\hat{\beta}_{Y,j}$ are unbiased estimators for $\beta_{M,j}$ and $\beta_{Y,j}$. In the presence of LD, the effect estimate $\hat{\beta}_{\cdot,j}$ ($\cdot$ may be M or Y) is not an estimator of $\beta_{\cdot,j}$, but instead estimates $\beta_{\cdot,j}$ plus a contribution from each of the neighbors of variant j. Using results from [16],

$$E[\hat{\beta}_{\cdot,j}] = \sum_k \frac{r_{jk}s_{\cdot,j}}{s_{\cdot,k}}\beta_{\cdot,k} \equiv \beta^*_{\cdot,j}, \tag{14}$$

where r_{jk} is the correlation between variant j and variant k and we assume that the LD structure in the samples used for the two GWAS are the same. We refer to $\beta^*_{\cdot,j}$ as the *LD-transformed effects*. Note that if allele frequencies are the same in the two GWAS populations then $s_{M,j} = cs_{Y,j}$ where c is a constant depending on sample size. Thus

$$\frac{r_{jk}s_{M,j}}{s_{M,k}} = \frac{r_{jk}s_{Y,j}}{s_{Y,k}} \equiv h_{j,k}. \tag{15}$$

We now derive $\mathrm{Cov}(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}|\mathbf{Z})$ in the presence of LD, where $\mathbf{Z}$ is a vector with jth element equal to Z_j. We assume that direct effects are independent so $\mathrm{Cov}(\beta_{M,j}, \beta_{M,k}) = 0$ if $j \neq k$ and $\mathrm{Cov}(\beta_{M,j}, \theta_k) = 0$ for

all j and k. We also assume that $\hat{\beta}_{M,j}$ and $\hat{\beta}_{Y,j}$ are independent conditional on $\beta^*_{M,j}$ and $\beta^*_{Y,j}$. With these assumptions,

$$\begin{aligned}\mathrm{Cov}(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}|\mathbf{Z}) =& \mathrm{Cov}(\sum_k h_{j,k}\beta_{Y,k}, \sum_k h_{j,k}\beta_{M,k}|\mathbf{Z}) \\ =& \sum_k h^2_{j,k}\mathrm{Cov}(\beta_{Y,k}, \beta_{M,k}|Z_k) \\ =& \sum_k h^2_{j,k}(\gamma + Z_k\eta)\mathrm{Var}(\beta_{M,k}). \end{aligned} \tag{16}$$

Note that

$$\mathrm{Var}(\hat{\beta}_{M,j}|\mathbf{Z}) = \mathrm{Var}(\beta^*_{M,j}) + s^2_{M,j} = \sum_k h^2_{j,k}\mathrm{Var}(\beta_{M,k}) + s^2_{M,j}. \tag{17}$$

Suppose that the variant correlation structure can be decomposed into independent LD blocks and that there is at most one M effect variant per block. If this variant has index k' then, for any variant in the block

$$\begin{aligned}\mathrm{Cov}(\hat{\beta}_{Y,j}, \hat{\beta}_{M,j}|\mathbf{Z}) =& h^2_{j,k'}(\gamma + Z_{k'}\eta)\mathrm{Var}(\beta_{M,k'}) \\ =& (\gamma + Z_{k'}\eta)\left(\mathrm{Var}(\hat{\beta}_{M,j}) - s^2_{M,j}\right). \end{aligned} \tag{18}$$

This means that, if $Z_{k'} = 1$, then variant k' induces correlation between effect estimates for other variants in the block, even when $\gamma = 0$. However, if we use only one variant per LD block to estimate parameters then there is no distortion in the proportion of correlated variants. If $\gamma = 0$, then the proportion of variants with correlated effect estimates will be equal q, the proportion of true effect variants acting through U. In this case, CAUSE will not have an increased false positive rate but may have lower power if the variants selected to use in estimation are far from the true causal variants. To maximize power, we prune for LD prioritizing variants with low trait M p-values.

More generally, a block may contain multiple causal variants for M, with some acting on U (shared variants) and others not. Equation 16 suggests that a shared variant in LD with variant j will induce non-zero correlation between $\hat{\beta}_{Y,j}$ and $\hat{\beta}_{M,j}$ even if $\gamma = 0$. The presence of non-shared variants in a block reduces the correlation, but will not eliminate it. When M is highly polygenic and q is large, the proportion of blocks containing at least one shared variant may be much larger than the true value of q. In this case, LD can create an impression that a larger propotion of variants have correlated effects, which can lead to inflated estimates of q and higher false positive rates using CAUSE. CAUSE assumes that these settings are rare. However, our simulations are conducted under moderately dense conditions using 1000 effect variants and only 1,170 LD blocks, demonstrating that CAUSE still performs well even under somewhat unfavorable conditions.

5.5 Connections with LCV

O'Connor and Price [14] propose an approach to identifying pairs of traits with causal relationships that uses a latent causal variable (LCV) model. This model is similar to the CAUSE model (Figure 3) with $\gamma = 0$. Rather than explicitly modeling both correlated pleiotropy and a causal effect, the LCV model includes only a shared factor (U in the CAUSE model) and estimates the "genetic causality proportion" (GCP). The GCP reflects the relative proportions of heritability of each trait that are explained by a shared factor. A causal effect (with no additional correlated pleiotropy from other sources) is equivalent to a model in which all variants act through a shared factor. In this case, the GCP is equal to 1 or -1 depending on the directionality of the effect.

LCV estimates the GCP and computes a test statistic testing whether GCP$= 0$. However, models with non-zero GCP are not necessarily causal. In fact, many non-causal models have large magnitude GCPs, so the LCV test cannot be interpreted as a test of causality. O'Connor and Price [14] use an estimated GCP larger than 0.6 to suggest a possible causal relationship, but this cutoff is somewhat arbitrary.

Because the LCV model is similar to the CAUSE model with $\gamma = 0$, we can derive an expression for GCP in terms of CAUSE parameters under this condition. The LCV model uses two parameters q^{lcv}_M and q^{lcv}_Y, the square root of the proportions of trait M and Y heritability explained by the shared factor. Here

we use the same trait M and Y notation used in our discussion of CAUSE for simplicity. In terms of CAUSE parameters these are

$$q_M^{lcv} = \sqrt{q} \tag{19}$$

$$q_Y^{lcv} = \frac{\sqrt{q}\eta h_M}{h_Y} \tag{20}$$

The GCP is then defined as

$$GCP = \frac{\log|q_Y^{lcv}| - \log|q_M^{lcv}|}{\log|q_M^{lcv}| + \log|q_Y^{lcv}|} = \frac{\log(\eta h_M/h_Y)}{\log(q\eta h_M/h_Y)}. \tag{21}$$

From this formula, we see that GCP is non-zero if η and q are both non-zero. For example, if the heritability of the two traits is equal, $q = 0.3$ and $\eta = \sqrt{0.05}$ then GCP$= 0.55$. From this we see that testing whether GCP$= 0$ is very different from testing whether $\gamma = 0$ as CAUSE and other MR methods do.

6 Supplementary Figures

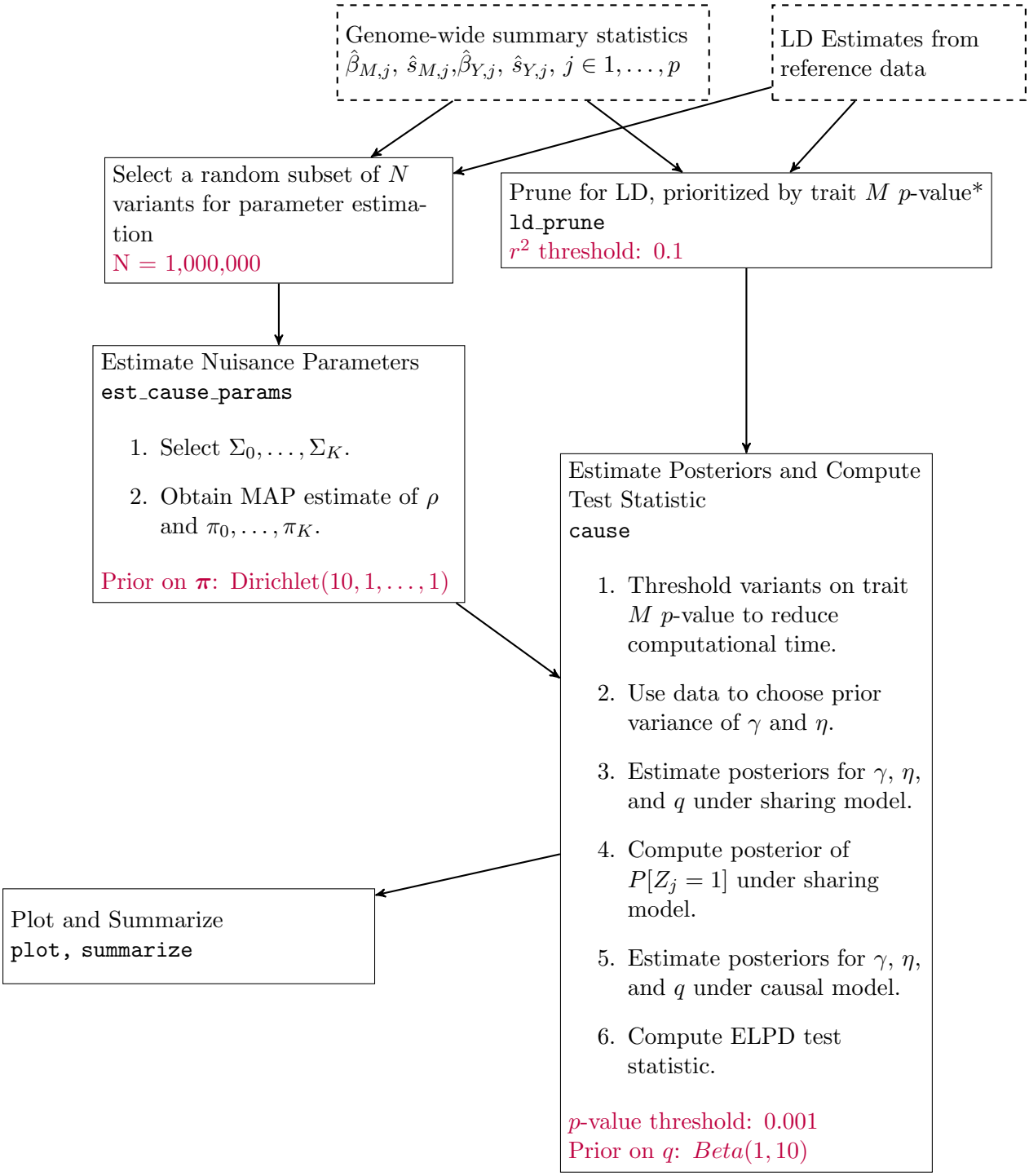


Figure S1: Workflow of a CAUSE analysis. Dashed boxes represent input data. Each solid box is an analysis step completed by the given function in the `cause` R package. LD pruning (*) can be parallelized over chromosomes. Purple text indicates user provided parameters and their default values. All analyses presented are run with default parameters.

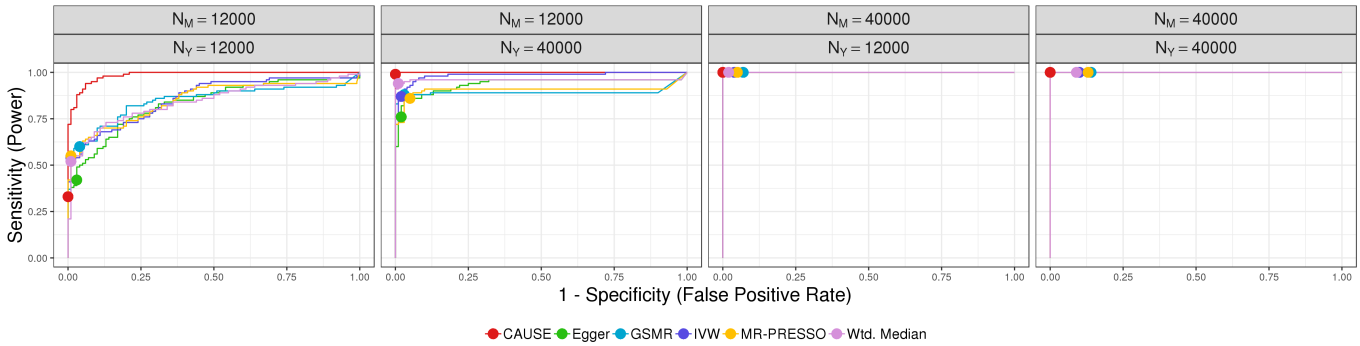


Figure S2: We compare the power in simulated data when $\gamma = \sqrt{0.05}$ to the false positive rate when there is no causal effect and no correlated pleiotropy. There are 100 simulations each in the causal and non-causal scenarios. Curves are created by varying the significance threshold from very stringent (lower left) to very permissive (upper right). Points indicate the power and false positive rate achieved at a threshold of $p \leq 0.05$.

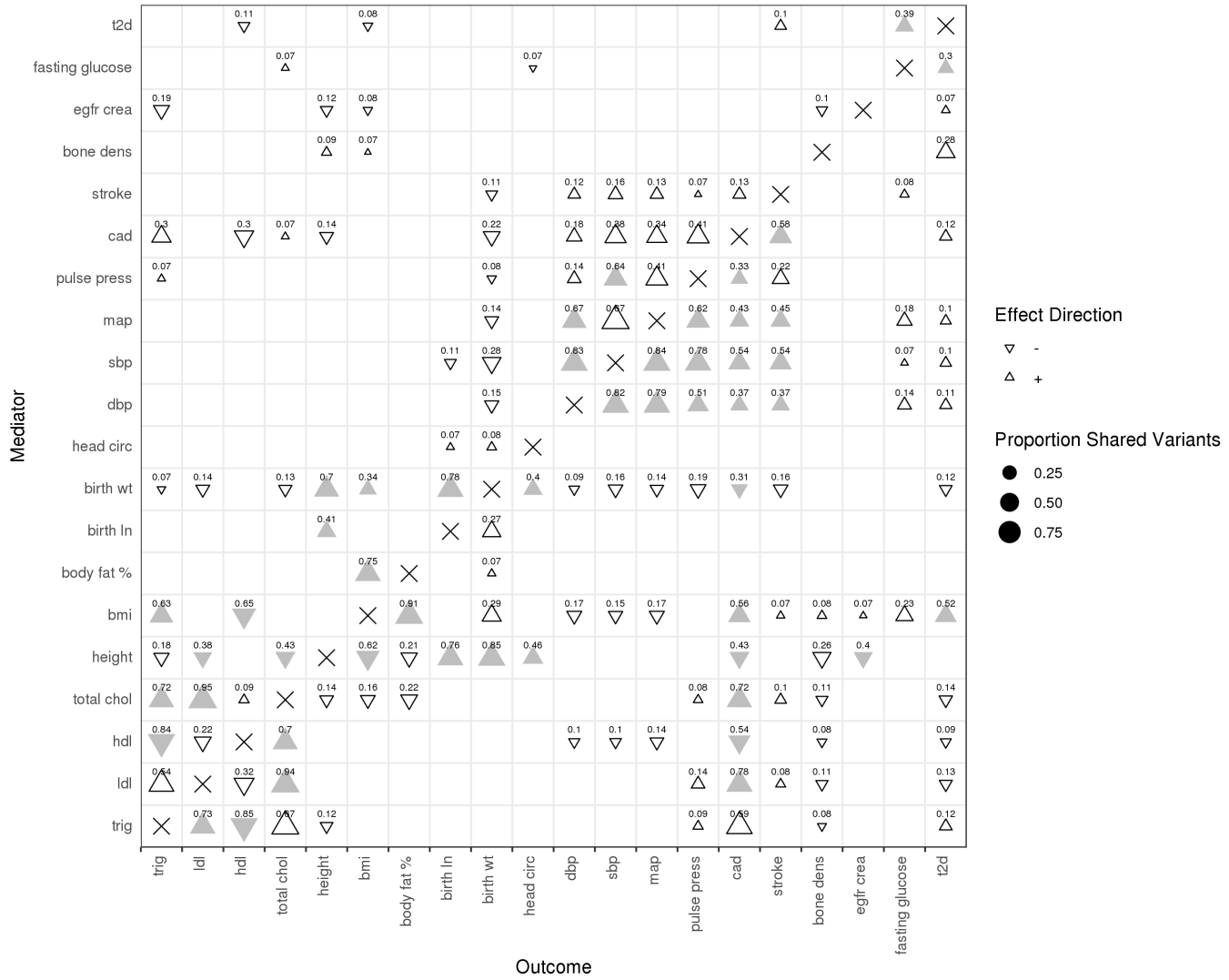


Figure S3: Posterior estimates of q from the sharing model for pairs of GWAS triats (Section 2.3). The parameter q is the proportion of trait 1 variants that effect trait 2 through a shared factor. Shape indicates effect direction of U on Y (the sign of posterior median of η). Shaded symbols indicate that the ELPD test statistic favors the causal model at an FDR threshold of 0.05. Point size is proportionate to the posterior median of q under the sharing model. This value is also printed above each point. Points are shown only for pairs with posterior median of q greater than the prior median of q.

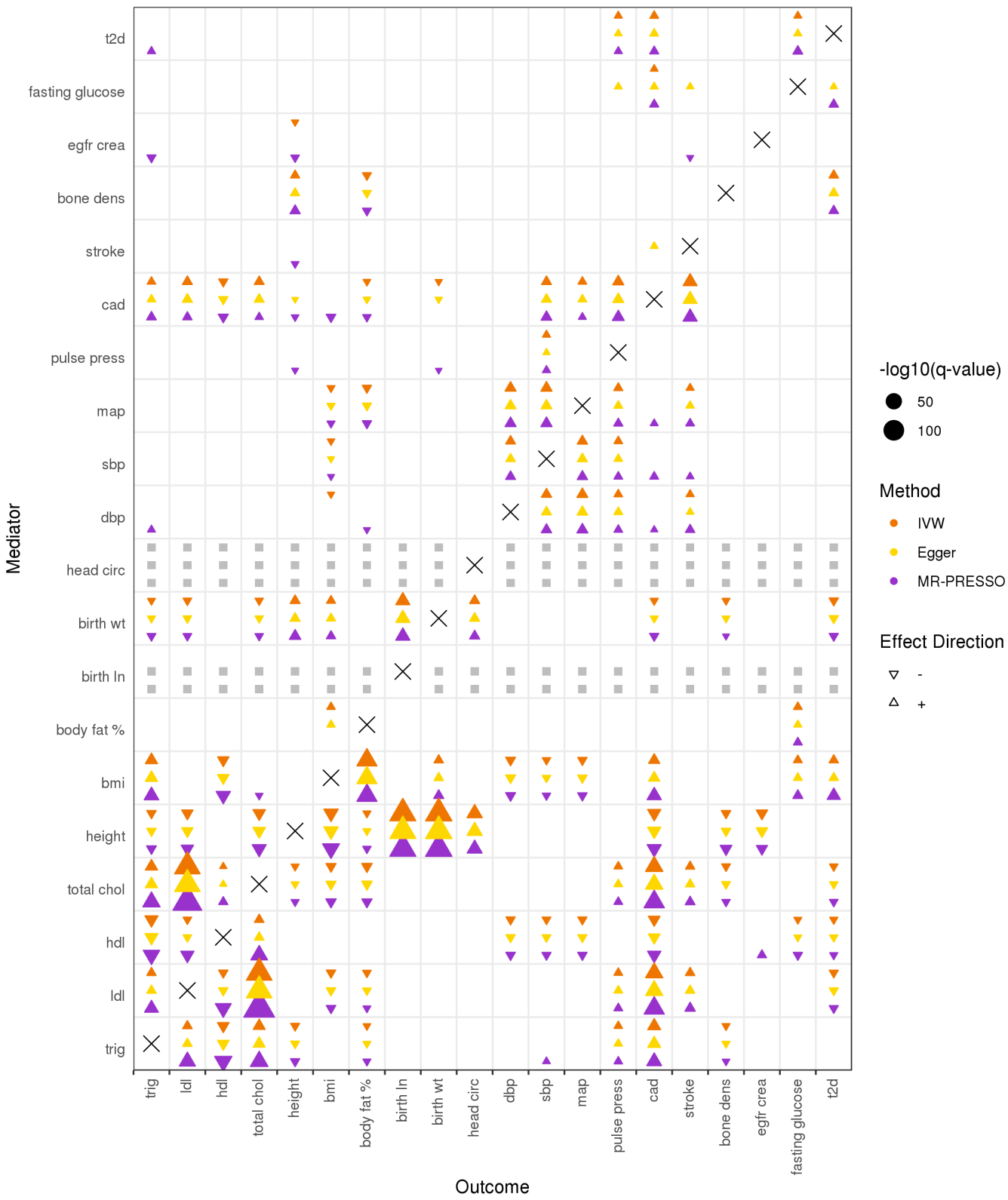


Figure S4: Each cell summarizes the IVW (upper, orange), Egger regression (middle, yellow) and MR-PRESSO (lower, purple) results for a pair of traits. Points are shown only for pairs significant at an FDR threshold of 0.05. Grey squares indicates that the method returned no results. This occurs when there are fewer than 2 geonome-wide significant loci for IVW regression or 3 for Egger regression and MR-PRESSO.

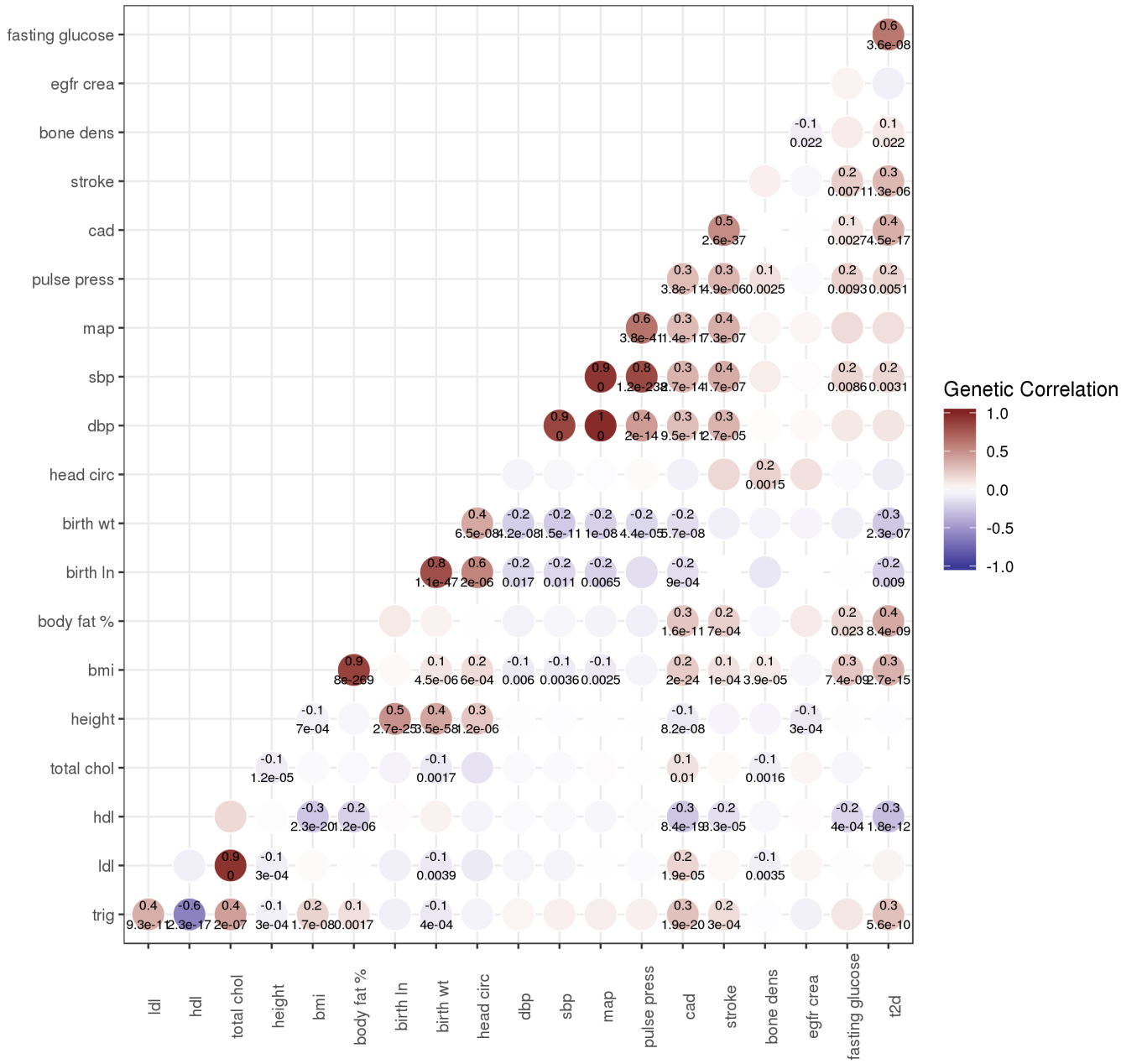


Figure S5: Estimates of genetic correlation from LD score regression for pairs of GWAS traits. Text over each point gives the estimated genetic correlation (top) and q-value (bottom) for the null hypothesis of no genetic correlation. Text is shown only for pairs with FDR < 0.05.

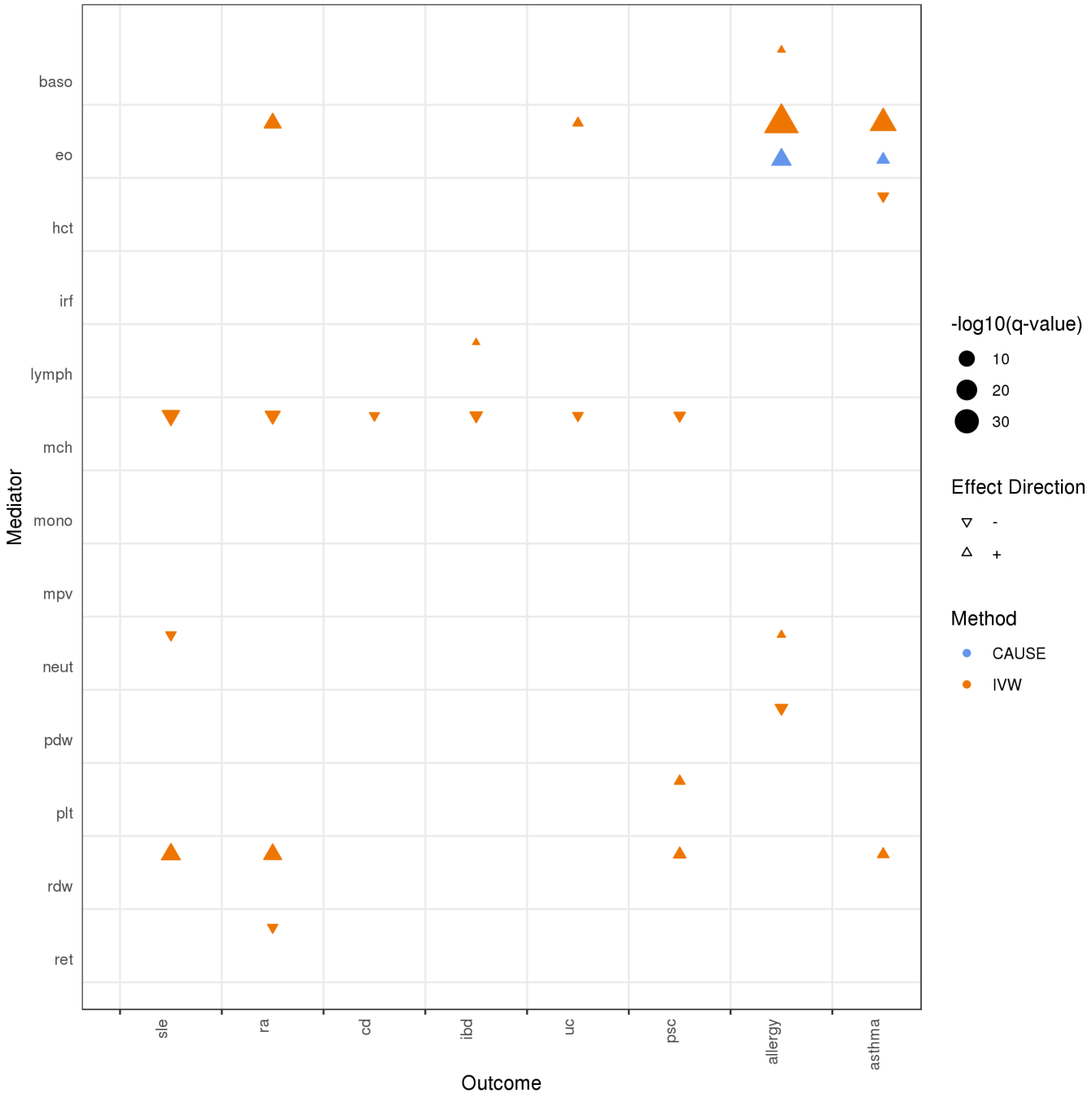


Figure S6: CAUSE and IVW results for blood cell and autoimmune traits (Section 2.5). Each cell summarizes the IVW (upper, orange) and CAUSE (lower, blue) results for a pair of traits. Points are shown only for pairs significant at an FDR threshold of 0.05.

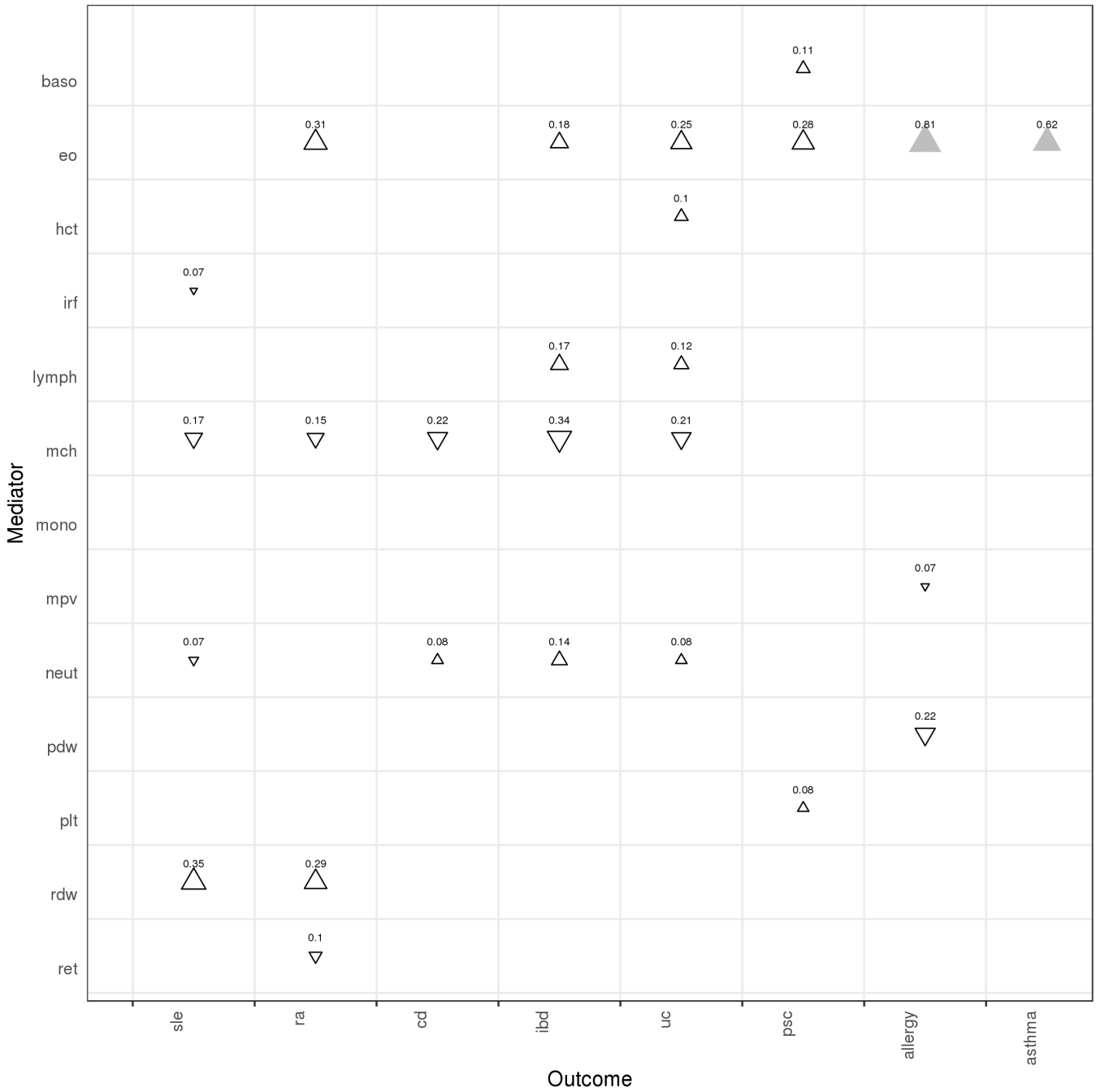


Figure S7: Posterior estimates of q from the sharing model for blood cell and autoimmune triats (Section 2.5). The parameter q is the proportion of trait 1 variants that effect trait 2 through the confounder or shared factor. Shape indicates effect direction of U on Y (the sign of posterior median of η). Shaded symbols indicate that the ELPD test statistic favors the causal model at an FDR threshold of 0.05. Point size is proportionate to the posterior median of q under the sharing model. This value is also printed above each point. Points are shown only for pairs with posterior median of q greater than the prior median of q.

Abbreviation	Trait	Sample Size	Cases	Controls	PMID
trig	Triglycerides	188577			24097068
ldl	LDL Cholesterol	188577			24097068
hdl	HDL Cholesterol	188577			24097068
total chol	Total Cholesterol	188577			24097068
height	Height	253288			25282103
bmi	Body Mass Index	339224			25673413
body fat %	Body Fat Percentage	100716			26833246
birth ln	Birth Length	28459			25281659
birth wt	Birth Weight	153781			27680694
head circ	Head Circumference at Birth	10678			22504419
dbp	Diastolic Blood Pressure	203056			21909115
sbp	Systolic Blood Pressure	203056			21909115
map	Mean Arterial Pressure	203056			21909115
pulse press	Pulse Pressure	203056			21909115
cad	Coronary Artery Disease	547261	122733	424528	29212778
stroke	Stroke	446696	40585	406111	29531354
bone dens	Bone Density	66628			29304378
egfr crea	eGFR Creatinine	133814			26831199
fasting glucose	Fasting Glucose	46186			20081858
t2d	Type 2 Diabetes	69033	12171	56862	22885922

Table S1: Genome wide association studies analyzed in Section 2.3

7 Supplementary Tables

GO term	GO Description	Enrichment Estimate	p-value	B-H q-value
GO:0007049	cell cycle	11.50	$2.13 \cdot 10^{-4}$	$8.36 \cdot 10^{-4}$
GO:0051301	cell division	13.40	$1.50 \cdot 10^{-4}$	$8.55 \cdot 10^{-4}$
GO:0005546	phosphatidylinositol-4,5-bisphosphate binding	26.37	$9.70 \cdot 10^{-4}$	$4.16 \cdot 10^{-3}$
GO:0000775	chromosome, centromeric region	17.72	$2.92 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$
GO:0001895	retina homeostasis	31.53	$3.88 \cdot 10^{-3}$	$1.35 \cdot 10^{-2}$
GO:0048471	perinuclear region of cytoplasm	9.78	$3.66 \cdot 10^{-3}$	$1.44 \cdot 10^{-2}$
GO:0008305	integrin complex	28.19	$4.45 \cdot 10^{-3}$	$1.54 \cdot 10^{-2}$
GO:0070182	DNA polymerase binding	36.20	$4.20 \cdot 10^{-3}$	$1.64 \cdot 10^{-2}$
GO:0000777	condensed chromosome kinetochore	18.13	$5.02 \cdot 10^{-3}$	$1.72 \cdot 10^{-2}$
GO:0001917	photoreceptor inner segment	49.58	$5.99 \cdot 10^{-3}$	$2.02 \cdot 10^{-2}$
GO:0018026	peptidyl-lysine monomethylation	49.31	$6.02 \cdot 10^{-3}$	$2.08 \cdot 10^{-2}$
GO:0016605	PML body	19.39	$6.09 \cdot 10^{-3}$	$2.08 \cdot 10^{-2}$
GO:0032587	ruffle membrane	17.73	$6.35 \cdot 10^{-3}$	$2.09 \cdot 10^{-2}$
GO:0009408	response to heat	27.18	$6.42 \cdot 10^{-3}$	$2.10 \cdot 10^{-2}$
GO:0016458	gene silencing	47.79	$6.28 \cdot 10^{-3}$	$2.12 \cdot 10^{-2}$
GO:0051726	regulation of cell cycle	14.36	$6.19 \cdot 10^{-3}$	$2.22 \cdot 10^{-2}$
GO:0036297	interstrand cross-link repair	27.48	$5.76 \cdot 10^{-3}$	$2.38 \cdot 10^{-2}$
GO:0043325	phosphatidylinositol-3,4-bisphosphate binding	63.69	$8.60 \cdot 10^{-3}$	$2.48 \cdot 10^{-2}$
GO:0000132	establishment of mitotic spindle orientation	32.99	$8.56 \cdot 10^{-3}$	$2.71 \cdot 10^{-2}$
GO:0070192	chromosome organization involved in meiotic cell cycle	42.94	$8.42 \cdot 10^{-3}$	$2.81 \cdot 10^{-2}$
GO:0051382	kinetochore assembly	26.15	$9.23 \cdot 10^{-3}$	$3.03 \cdot 10^{-2}$
GO:0031345	negative regulation of cell projection organization	41.38	$9.75 \cdot 10^{-3}$	$3.06 \cdot 10^{-2}$
GO:0034260	negative regulation of GTPase activity	29.08	$8.19 \cdot 10^{-3}$	$3.15 \cdot 10^{-2}$
GO:0000723	telomere maintenance	31.94	$1.09 \cdot 10^{-2}$	$3.26 \cdot 10^{-2}$
GO:0060231	mesenchymal to epithelial transition	32.13	$1.05 \cdot 10^{-2}$	$3.31 \cdot 10^{-2}$
GO:0014065	phosphatidylinositol 3-kinase signaling	21.18	$1.18 \cdot 10^{-2}$	$3.73 \cdot 10^{-2}$
GO:0045022	early endosome to late endosome transport	24.97	$1.38 \cdot 10^{-2}$	$3.77 \cdot 10^{-2}$
GO:0017148	negative regulation of translation	25.23	$1.27 \cdot 10^{-2}$	$3.80 \cdot 10^{-2}$
GO:0005694	chromosome	9.67	$1.45 \cdot 10^{-2}$	$4.08 \cdot 10^{-2}$
GO:0060421	positive regulation of heart growth	38.45	$1.37 \cdot 10^{-2}$	$4.12 \cdot 10^{-2}$
GO:0040008	regulation of growth	22.32	$1.17 \cdot 10^{-2}$	$4.12 \cdot 10^{-2}$
GO:0005768	endosome	8.76	$1.52 \cdot 10^{-2}$	$4.26 \cdot 10^{-2}$
GO:0032039	integrator complex	37.71	$1.52 \cdot 10^{-2}$	$4.43 \cdot 10^{-2}$
GO:0006415	translational termination	36.85	$1.71 \cdot 10^{-2}$	$4.69 \cdot 10^{-2}$
GO:0042059	negative regulation of epidermal growth factor receptor signaling pathway	27.28	$1.41 \cdot 10^{-2}$	$4.79 \cdot 10^{-2}$
GO:0030695	GTPase regulator activity	36.96	$1.68 \cdot 10^{-2}$	$4.80 \cdot 10^{-2}$
GO:0030291	protein serine/threonine kinase inhibitor activity	36.76	$1.73 \cdot 10^{-2}$	$4.87 \cdot 10^{-2}$

Table S2: Gene set enrichment results discussed in Section 2.4. Gene sets are ranked by evidence of enrichment for genes associated with variants with high posterior probability of acting through a shared factor affecting both T2D risk and birth weight. Gene sets with Benjamini-Hochberg adjusted q-value less than 0.05 and positive enrichment estimates are shown.

Gene	SNP	Chromosome	Position	P_i
ITGA6	rs10209443	2	173201204	0.13
ITGA1	rs1812128	5	52000024	0.18
ITGB6	rs3772071	2	161135544	0.19
ITGB5	rs4234221	3	124438329	0.10
ITGA3	rs4793636	17	48139500	0.14

Table S3: Genes in the integrin pathway discussed in Section 2.4. Each gene is associated with only one variant. P_i gives the posterior probability that each variant affects a factor shared by T2D and birth weight. For this pair of traits, the posterior median of q under the sharing model is 0.12.

Abbreviation	Trait	Sample Size	Cases	Controls	PMID
sle	Systemic Lupus Erythematosus	23210	7219	15991	26502338
ra	Rheumatoid Arthritis	103638	29880	73758	24390342
cd	Crohns Disease	69268	22575	46693	26192919
ibd	Irritable Bowel Disease	96486	42950	53536	26192919
uc	Ulcerative Colitis	72647	20417	52230	26192919
psc	Primary Sclerosing Cholangitis	24751	4796	19955	27992413
allergy	Allergic Disease	360838	180129	180709	29083406
asthma	Asthma	142486	23948	118538	29273806
baso	Basophil Count	173480			27863252
eo	Eosinophil Count	173480			27863252
hct	Hematocrit	173480			27863252
irf	Immature Fraction of Reticulocytes	173480			27863252
lymph	Lymphocyte Count	173480			27863252
mch	Mean corpuscular hemoglobin	173480			27863252
mono	Monocyte Count	173480			27863252
mpv	Mean Platelet Volume	173480			27863252
neut	Neutrophil Count	173480			27863252
pdw	Platelet Distribution Width	173480			27863252
plt	Platelet Count	173480			27863252
rdw	Red Cell Distribution Width	173480			27863252
ret	Reticulocyte Count	173480			27863252

Table S4: Genome wide association studies analyzed in Section 2.5

References

[1] George Davey Smith and Gibran Hemani. "Mendelian randomization: genetic anchors for causal inference in epidemiological studies". In: *Human Molecular Genetics* 23.1 (2014), pp. 89–98.

[2] Anna G.C. Boef, Olaf M. Dekkers, and Saskia Le Cessie. "Mendelian randomization studies: A review of the approaches used and the quality of reporting". In: *International Journal of Epidemiology* 44.2 (2015), pp. 496–511.

[3] Jie Zheng, Denis Baird, Maria-Carolina Borges, Jack Bowden, Gibran Hemani, et al. "Recent Developments in Mendelian Randomization Studies". In: *Current Epidemiology Reports* 4.4 (2017), pp. 330–345.

[4] Stephen Burgess, Frank Dudbridge, and Simon G. Thompson. "Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods". In: *Statistics in Medicine* 35.11 (2016), pp. 1880–1906.

[5] Gibran Hemani, Jack Bowden, and George Davey Smith. "Evaluating the potential role of pleiotropy in Mendelian randomization studies". In: *Human Molecular Genetics* 27.R2 (2018), R195–R208.

[6] Marie Verbanck, Chia-yen Chen, Benjamin Neale, and Ron Do. "Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases". In: *Nature Genetics* 50.May (2018).

[7] Jack Bowden, George Davey Smith, and Stephen Burgess. "Mendelian Randomization Methodology Mendelian randomization with invalid instruments : effect estimation and bias detection through Egger regression". In: *International Journal of Epidemiology* June (2015), pp. 512–525.

[8] Richard Barfield, Helian Feng, Alexander Gusev, Lang Wu, and Wei Zheng. "Transcriptome-wide association studies accounting for co- localization using Egger regression". In: *Genetic Epidemiology* 42 (2018), pp. 418–433.

[9] Jack Bowden, M. Fabiola Del Greco, Cosetta Minelli, George Davey Smith, Nuala A. Sheehan, et al. "Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: The role of the I2statistic". In: *International Journal of Epidemiology* 45.6 (2016), pp. 1961–1974.

[10] Zhihong Zhu, Zhili Zheng, Futao Zhang, Yang Wu, Maciej Trzaskowski, et al. "Causal associations between risk factors and common diseases inferred from GWAS summary data". In: *Nature Communications* 9.1 (2018), p. 224.

[11] Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies". In: *Nature Genetics* advance on.3 (2015), pp. 291–295.

[12] Verneri Anttila, Brendan Bulik-Sullivan, Hilary K. Finucane, Raymond K. Walters, Jose Bras, et al. "Analysis of shared heritability in common disorders of the brain". In: *Science* 360.6395 (2018).

[13] Joseph K Pickrell. "Fulfilling the promise of Mendelian randomization". In: *bioRxiv* (2015). doi: `10.1101/018150`.

[14] Luke J O'Connor and Alkes L. Price. "Distinguishing genetic correlation from causation across 52 diseases and complex traits". In: *Nature Genetics* 50 (2018), pp. 1728–1734.

[15] Aki Vehtari, Andrew Gelman, and Jonah Gabry. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". In: *Statistics and Computing* September (2016), pp. 1–20.

[16] Xiang Zhu and Matthew Stephens. "Bayesian large-scale multiple regression with summary statistics from genome-wide association studies". In: *Annals of Applied Statistics* 11.3 (2017), pp. 1561–1592.

[17] The 1000 Genomes Project Consortium. "A global reference for human genetic variation". In: *Nature* 526.7571 (2015), pp. 68–74.

[18] Y Benjamini and Y Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal Statistical Society B* 57.1 (1995), pp. 289–300.

[19] John D. Storey. "The Positive False Discover Rate: A Bayesian Interpretation and the q-Value". In: *The Annals of Statistics* 31.6 (2003), pp. 2013–2035.

[20] Bharathi Upadhya, Michael Rocco, Cora E. Lewis, Suzanne Oparil, Laura C. Lovato, et al. "Effect of Intensive Blood Pressure Treatment on Heart Failure Events in the Systolic Blood Pressure Reduction Intervention Trial". In: *Circulation: Heart Failure* 10.4 (2017), pp. 1–10.

[21] Anna Helgadottir, Solveig Gretarsdottir, Gudmar Thorleifsson, Eirikur Hjartarson, Asgeir Sigurdsson, et al. "Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease". In: *Nature Genetics* 48.6 (2016), pp. 634–639.

[22] Benjamin F. Voight, Gina M. Peloso, Marju Orho-Melander, Ruth Frikke-Schmidt, Maja Barbalic, et al. "Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study". In: *The Lancet* 380.9841 (2012), pp. 572–580.

[23] Michael V. Holmes, Folkert W. Asselbergs, Tom M. Palmer, Fotios Drenos, Matthew B. Lanktree, et al. "Mendelian randomization of blood lipids for coronary heart disease". In: *European Heart Journal* 36.9 (2015), pp. 539–550.

[24] Brendan Bulik-Sullivan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, et al. "An Atlas of Genetic Correlations across Human Diseases and Traits". In: *Nature Genetics* 47.11 (2015), pp. 1236–1241.

[25] Gene Ontology Consortium. "Gene Ontology : tool for the unification of biology". In: *Nature Genetics* 25.may (2000), pp. 25–29. arXiv: `10614036`.

[26] S. Carbon, E. Douglass, N. Dunn, B. Good, N. L. Harris, et al. "The Gene Ontology Resource: 20 years and still GOing strong". In: *Nucleic Acids Research* 47.D1 (2019), pp. D330–D338.

[27] Liana K. Billings, Yi Hsiang Hsu, Rachel J. Ackerman, Josée Dupuis, Benjamin F. Voight, et al. "Impact of common variation in bone-related genes on type 2 diabetes and related traits". In: *Diabetes* 61.8 (2012), pp. 2176–2186.

[28] Patricia C. Fulkerson and Marc E. Rothenberg. "Targeting Eosinophils in Allergy, Inflammation and Beyond Patricia". In: *Nature Reviews Drug Discovery* 12.2 (2013).

[29] Matthew Stephens. "False discovery rates: a new deal". In: *Biostatistics* 18.2 (2017), pp. 275–295.

[30] Xiaoquan Wen and Matthew Stephens. "Using linear predictors to impute allele frequencies from summary or pooled genotype data". In: *Annals of Applied Statistics* 4.3 (2010), pp. 1158–1182.

[31] Jack Bowden, George Davey Smith, Philip C. Haycock, and Stephen Burgess. "Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator". In: *Genetic Epidemiology* 40.4 (2016), pp. 304–314.

[32] Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* Boca Raton, FL: CRC Press, Taylor and Francis Group, 2016.