

Supplementary Information

Douglas Yao, Luke O'Connor, Alkes Price, Alexander Gusev

Supplementary Note

Modes of expression causality

In Supplementary Figure 28, we depict 9 different causality scenarios between SNPs, gene expression levels, and complex trait. Scenarios A-D constitute the main causality scenarios described in the main manuscript text; scenarios E-I describe additional causality scenarios. We provide a description of each scenario and its contribution to estimates of h_{med}^2 .

A. Mediation

Here, the SNP affects the expression levels of the gene in cis, which then affect the complex trait. This is the desired scenario, since it is consistent with the hypothesis that SNPs exert their effects on complex traits via modulating gene expression levels. The presence of mediation will result in nonzero estimates of h_{med}^2 .

B. Pleiotropy

Here, the SNP independently affects the expression of the gene in cis and the complex trait. Under the assumption that the magnitude of pleiotropic effects is uncorrelated with the magnitude of eQTL effects (see “Model assumptions”), the presence of pleiotropy will not contribute to estimates of h_{med}^2 .

C. Linkage

Here, the SNP that affects gene expression in cis is in LD with another SNP that independently affects the trait. Under the assumption that the magnitude of linkage effects is uncorrelated with the magnitude of eQTL effects (see “Model assumptions”), the presence of linkage will not contribute to estimates of h_{med}^2 .

D. Reverse mediation

Here, the SNP directly affects the complex trait, which then affects the expression levels of the gene in cis. Although this scenario can in theory contribute to nonzero estimates of h_{med}^2 , the contribution will be negligible given that genetic effects on a complex (i.e. polygenic) trait are much smaller than genetic effects on gene expression (see “Reverse mediation”).

E. Mediation in unobserved cell type/context

Here, the SNP affects the expression levels of the gene in the *causal cell type/context* for the complex trait, which then affects the complex trait. In practice, we only have access to *assayed* expression levels. Estimates of h_{med}^2 using assayed expression levels will be nonzero if the assayed expression levels are correlated with expression levels in causal cell types/contexts (see “Assayed vs. total underlying h_{med}^2 ”).

F. Trans mediation

Here, the SNP affects the expression levels of the gene in *trans*, which then affects the complex trait. If trans-eQTL effect sizes are uncorrelated with cis-eQTL effect sizes, this scenario will not contribute to estimates of h_{med}^2 . Note that this scenario refers to *purely trans* effects that are not mediated in cis at any point. An alternative scenario is cis-by-trans mediation (see below), which is subsumed by scenario A.

G. Cis-by-trans mediation

Here, the SNP affects the expression levels of gene 1 in cis, gene 1 affects gene 2 in trans, and gene 2 affects the complex trait. Although the SNP acts as a trans-eQTL for gene 2, its effects are mediated in cis at some point, so this scenario is subsumed by scenario A (i.e. mediation).

H. Mediation by unobserved cis intermediary

Here, the SNP affects an unobserved intermediary in cis, which then has pleiotropic effects on a gene’s expression levels and the complex trait. The intermediary can be the expression levels/splicing/activity of another gene, and can also refer to any other molecular process. Note that this scenario refers specifically to *cis* intermediaries; the distinction between *cis* intermediaries and *trans* intermediaries is that SNP effect sizes on known molecular phenotypes are much larger in cis than in trans¹⁻³. In scenario H, it is not appropriate to assume that eQTL effect sizes and pleiotropic effect sizes are independent, as they are both affected by a common intermediary. Thus, this scenario can contribute to nonzero estimates of h_{med}^2 . Because this scenario does not strictly speaking involve mediation through the gene’s expression levels, its contribution to h_{med}^2 might be viewed as spurious.

However, there are several reasons why scenario H’s potential contribution to h_{med}^2 is likely not of major concern. First, if the intermediary represents the expression levels of another gene, then scenario H’s contribution to h_{med}^2 is justifiable given that mediation is actually occurring through the expression levels of the intermediary. Second, because we perform regression using all SNPs in the genome, scenario H must be pervasive across most loci in the genome in order for it to have a substantial impact on estimates of h_{med}^2 . Third, if the intermediary does not refer to the expression levels of a gene (e.g. it represents splicing or coding changes in a gene, or it represents some unknown molecular process), we argue that the contribution of the non-causal gene to h_{med}^2 is still of biological interest due to the fact that the gene’s expression levels are correlated with a truly causal intermediary.

Should the third scenario be pervasive across the genome and have a substantial contribution toward h_{med}^2 , we can amend our definition of h_{med}^2 in order to accommodate it. In “Definition of h_{med}^2 ,” we refer to our estimate of h_{med}^2 as the heritability mediated and/or “tagged” by assayed expression levels, and we state that “tagging” occurs due to assayed expression levels acting as a proxy to expression levels in causal cell types/contexts. We can extend this definition of h_{med}^2 so that tagging also refers to assayed expression levels acting as a proxy to unobserved intermediaries *other than* expression levels. We believe that this extended definition of h_{med}^2 is of equal biological interest to a definition of h_{med}^2 that excludes tagging of unobserved intermediaries other than expression levels.

I. Mediation by unobserved trans intermediary

Here, the SNP affects an unobserved intermediary in trans, with then has pleiotropic effects on gene expression levels in cis and the complex trait. The only difference between scenario I and scenario H is that here the intermediary is affected in trans rather than in cis. Because trans-effects on molecular phenotypes are much smaller than cis-effects, the contribution of scenario I to nonzero h_{med}^2 will be much smaller than the contribution of scenario H (see “Reverse mediation” for related intuition).

Reverse mediation

In our generative model, we do not model the effects of reverse mediation, which we define as the scenario in which a SNP influences the complex trait independently of the SNP’s effects on the expression of a gene, and the complex trait itself then influences the expression of the gene. Such a scenario will induce a genetic correlation between the gene’s expression and the complex trait and could potentially bias our estimates of h_{med}^2 . However, we posit that the bias (if present at all) is negligible for the following reasons. (1) Because we use an external expression panel to estimate eQTL effect sizes, the complex trait of interest must be represented in the expression panel samples in order for its effects on expression to be present in our analyses^{4,5}. Thus, we can rule out the possibility of reverse mediation influencing our results for any disease phenotypes. (2) Assuming that the complex trait of interest is represented in the expression panel samples, the total bias in estimates of h_{med}^2 caused by reverse mediation is guaranteed to be very small

under the assumption that SNP effect sizes on a trait are much smaller than eQTL effect sizes⁵, which is true for polygenic traits. To illustrate this point, it is useful to think of h_{med}^2 in terms of the covariance between eQTL effect sizes β and total SNP effect sizes ω (for simplicity, we assume that each gene has only one eQTL):

$$h_{med}^2 = \sum_i^G Cov(\beta_i, \omega_i)^2 / Var(\beta_i)$$

Assuming that $Cov(\beta_i, \omega_i)$ for SNP i is nonzero due to proper mediation, we have the following expression for $Cov(\beta_i, \omega_i)_{med}^2$:

$$\begin{aligned} \omega_i &= \beta_i \alpha_i + \gamma_i \\ Cov(\beta_i, \omega_i)_{med}^2 &= \alpha_i^2 Var(\beta_i)^2 \end{aligned}$$

A typical trait has $h_{med}^2 = 0.1$ and $Var(\beta_i) = 0.05$, so a realistic value for α_i^2 is $\frac{h_{med}^2}{GV ar(\beta_i)} = \frac{0.1}{20000 \cdot 0.05} = 0.0001$ (we assume here that all genes are causal; the proportion of causal genes will not affect the points conveyed by these calculations so long as the number of causal SNPs is at least as large as the number of causal genes; see below). Thus, we have $Cov(\beta_i, \omega_i)_{med}^2 \approx 0.0001 \cdot 0.05^2 = 2.5 \times 10^{-7}$. On the other hand, if we assume that $Cov(\beta_i, \omega_i)$ for SNP i is nonzero due to reverse mediation, we have the following expression for $Cov(\beta_i, \omega_i)_{revmed}^2$:

$$\begin{aligned} \beta_i &= \theta_i \sum_j^M \omega_j + \beta_{i(SNP)} \\ Cov(\beta_i, \omega_i)_{revmed}^2 &= \theta_i^2 Var(\omega_i)^2 \end{aligned}$$

Here, θ_i represents the effect size of the complex trait on the expression of gene i and $\beta_{i(SNP)}$ represents the direct effect size of SNP i on gene i without the effects of reverse mediation. The upper limit for θ_i^2 is β_i^2/h^2 , which occurs if $\beta_{i(SNP)} = 0$. Thus, we can rewrite the above as

$$Cov(\beta_i, \omega_i)_{revmed}^2 \leq \beta_i^2/h^2 Var(\omega_i)^2$$

A typical complex trait has $h^2 = 0.5$. If we assume that the number of causal SNPs is at least as large as the number of causal genes, we have $Var(\omega_i) \leq h^2/G = 2.5 \times 10^{-5}$. Thus, we have $Cov(\beta_i, \omega_i)_{revmed}^2 \leq 0.05/0.5 \cdot (2.5 \times 10^{-5})^2 = 6.25 \times 10^{-11}$. The reason why $Cov(\beta_i, \omega_i)_{revmed}^2$ is orders of magnitude smaller than $Cov(\beta_i, \omega_i)_{med}^2$ is that $Cov(\beta_i, \omega_i)_{revmed}^2$ involves the product of the *square* of the squared per-SNP effect on complex trait and the squared SNP effect on gene expression, while $Cov(\beta_i, \omega_i)_{med}^2$ involves the product of the squared per-gene effect on complex trait and the *square* of the squared SNP effect on gene expression. Because individual SNP effects on gene expression are much larger than individual SNP/gene effects on a polygenic trait, squaring the latter causes the overall magnitude of $Cov(\beta_i, \omega_i)^2$ to be much smaller than squaring the former.

Relationship between MESCS and stratified LD score regression

MESCS is similar in form to stratified LD score regression (S-LDSC), which aims to estimate total heritability partitioned across SNP categories from summary statistics^{6,7}. In particular, the τ coefficient estimated by S-LDSC is directly related to the π coefficient we obtain from MESCS in equation (20) (in Methods). In S-LDSC, the variance of total effect size of SNP k on the trait (ω_k) is modeled as follows:

$$Var(\omega_k) = \sum_c a_c(k) \tau_c$$

where a_c refers to a continuous-valued SNP annotation. Meanwhile, in MESCS, $Var(\omega_k)$ is modeled as follows:

$$Var(\omega_k) = \sum_d \pi_d \sum_{i \in D} \beta_{ik}^2 + \sum_{c: k \in C} \tau_c^{nonmed}$$

We label τ_c^{nonmed} as so in order to distinguish it from τ_c used in S-LDSC. Here, note that we can treat the value $\sum_{i \in D} \beta_{ik}^2$ as a continuous SNP annotation a_c , which means that the expression scores $\mathcal{L}_{k;d}$ used in equation (20) are equivalent to LD scores with continuous annotation $a_c(k) = \sum_{i \in D} \beta_{ik}^2$. Thus, π_d as defined above and in equation (20) is equivalent to τ_c as defined in S-LDSC for the SNP annotation that corresponds to $\sum_{i \in D} \beta_{ik}^2$.

The main implication for this equivalence between MESC and S-LDSC is that significantly nonzero π_d as estimated by MESC can be interpreted as significantly nonzero τ_c from S-LDSC. There is considerable interest in identifying SNP annotations with significantly nonzero τ_c conditional on the baselineLD model and other SNP annotations⁶⁻¹², since this means that the SNP annotation is informative for explaining trait heritability *beyond* the set of comprehensive but non-trait-specific SNP annotations contained in the baselineLD model (as well any additional SNP annotations included in overall model). When using MESC, we also include all SNP annotations in the baselineLD model in our analyses, albeit for a different purpose than in studies using S-LDSC; our reason for including the baselineLD model is to account for correlations between the magnitude of non-mediated effect sizes and eQTL effect sizes (see ‘‘Model assumptions’’). Nevertheless, we can still interpret significantly nonzero π_d for a given gene category D as implying that the SNP annotation corresponding to the eQTL effect sizes of all SNPs on genes in D is informative for explaining trait heritability beyond the baselineLD model.

Relationship between MESC and Mendelian randomization

In this section, we describe the motivation behind the regression procedure carried out in MESC and compare it to Mendelian randomization (MR). Our goal is to estimate h_{med}^2 , where $h_{med}^2 = \sum_i^G \sum_j^M \beta_{ij}^2 \alpha_i^2$. One way we could estimate h_{med}^2 would be to first estimate α_i^2 for each individual gene, then multiply α_i^2 by the cis-heritability of the gene and sum up this quantity across all genes to obtain h_{med}^2 . In principle, we could estimate α_i for each individual gene i using some type of MR approach, where the exposure of interest is the expression level of gene i , and the outcome is the trait. However, typical MR approaches are problematic for this aim. In the presence of non-mediated effects of genetic variants on the trait, MR is highly underpowered to estimate α_i with a small number of genetic instruments^{13,14}. This is a common scenario if we use gene expression as the exposure, since many genes have only a few detectable cis-eQTLs for their expression¹⁵. Alternatively, we could consider a MR approach with multiple genetic variants, which in principle can distinguish mediated from non-mediated effects so long as the InSIDE (instrument strength independent of direct effect) assumption holds¹⁴ (Note that the InSIDE assumption is essentially the same as the pleiotropy-eQTL independence assumption we describe in ‘‘Model assumptions’’). However, this approach is highly underpowered in the common scenario that genes have only a few detectable cis-eQTLs¹³, and this approach cannot be applied to genes with only one cis-eQTL. In summary, we cannot use typical MR approaches to estimate h_{med}^2 due to the sparse cis-genetic architecture of gene expression.

Unlike MR approaches, MESC is able to estimate h_{med}^2 in the presence of sparsity of eQTLs for individual genes by estimating gene-trait effects *across* many genes. To illustrate this, we contrast MESC and MR with multiple genetic variants (see Supplementary Figure 31 for an illustration). MR with multiple genetic variants essentially involves regressing SNP-trait effects on eQTL effects for a single gene. The slope from this regression will be the effect of the gene on the trait given that the InSIDE assumption is satisfied. Meanwhile, MESC essentially involves regressing *squared* SNP-trait effects on *squared* eQTL effects *summed across* a set of genes. The slope from this regression will be the *average squared* effect of all genes in the gene set on the trait given that both the pleiotropy-eQTL independence assumption (which is effectively the InSIDE assumption extended across eQTLs for all genes in the gene set) and an additional assumption are satisfied. This additional assumption is that eQTL effect sizes are independent of gene-trait effect sizes across genes (see ‘‘Model assumptions’’). We can then calculate h_{med}^2 with our average gene-trait effect estimate using an equivalent definition of h_{med}^2 that models β and α as random variables (See ‘‘Definition of h_{med}^2 ’’). Thus, MESC can still reliably estimate h_{med}^2 if individual genes have one or a small number of eQTLs, since it essentially aggregates information about gene-trait effects across many genes.

In summary, MESC can be conceptualized as an analogue to MR that models exposure effects as random and jointly estimates the average squared effect of multiple exposures on an outcome.

Use of eQTL summary statistics in MESC

We can use summary statistics from eQTL studies to estimate $h_{med}^2(\mathcal{D}_d)$ since expression score $\mathcal{L}_{k;d}$ is equivalent to the sum of marginal OLS estimates of eQTL effect sizes for SNP k on genes in \mathcal{D}_d ($\sum_{i \in \mathcal{D}_d} \hat{\beta}_{ik}^2(\text{sumstat})$) modulo an error term that depends on $|\mathcal{D}_d|$ and the sample size of the eQTL study. This error term will be captured by the intercept during regression. To illustrate this, we model the expression of gene i for N_{exp} expression panel samples as follows:

$$\mathbf{y}_{i(\text{exp})} = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_{i(\text{exp})}$$

where $\mathbf{y}_{i(\text{exp})}$ is an N_{exp} -vector of gene expression measurements (standardized to mean 0 and variance 1), \mathbf{X} is an $N_{exp} \times M$ genotype for M SNPs (standardized to mean 0 and variance 1), $\boldsymbol{\beta}_i$ is an M -vector of eQTL effect sizes, and $\boldsymbol{\epsilon}_{i(\text{exp})}$ is an N_{exp} -vector of environmental effects. Under this model, we have

$$\begin{aligned} E \left[\sum_{i \in \mathcal{D}_d} \hat{\beta}_{ik}^2(\text{sumstat}) \right] &= \sum_{i \in \mathcal{D}_d} \left(\sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + \frac{E[\epsilon_{i(\text{exp})}^2]}{N_{exp}} \right) \\ &= \sum_{i \in \mathcal{D}_d} \sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + \sum_{i \in \mathcal{D}_d} \frac{1 - E[h_{cis}^2]}{N_{exp}} \\ &= \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2 + \frac{|\mathcal{D}_d| E[h_{cis}^2]}{N_{exp}} + \frac{|\mathcal{D}_d| (1 - E[h_{cis}^2])}{N_{exp}} \\ &= \mathcal{L}_{k;d} + \frac{|\mathcal{D}_d|}{N_{exp}} \end{aligned}$$

Thus, we can use the following alternate form of equation (20) (in Methods) to perform regression:

$$E[\chi_k^2] = N \sum_c \tau_c \ell_{k;c} + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \hat{\beta}_{ik}^2(\text{sumstat}) + 1 + \frac{N h_{med;causal}^2}{N_{exp} E[h_{cis}^2]}$$

MESC with assayed expression scores

In this section, we show that when we carry out the regression procedure described in ‘‘MESC with summary statistics’’ using expression scores in assayed tissues T rather than in causal cell types/contexts, we obtain an estimate of $h_{med;assayed}^2(T)$.

Let β represent cis-eQTL effect sizes in causal cell types/contexts for the trait, and β' represent cis-eQTL effect sizes in assayed tissues T . For simplicity, assume no sampling noise and non-mediated effects in GWAS χ^2 statistics. Upon regressing GWAS χ^2 statistics on expression scores in assayed tissues (see equation (17)), we have

$$\begin{aligned} \alpha'^2 &\approx \frac{1}{G} \sum_i^G \frac{\text{Cov}(\chi^2, \sum_j^M r_j^2 \beta_{ij}'^2)}{\text{Var}(\sum_j^M r_j^2 \beta_{ij}'^2)} \\ &\approx \frac{1}{G} \sum_i^G \frac{\text{Cov}(\sum_j^M r_j^2 \alpha_i^2 \beta_{ij}^2, \sum_j^M r_j^2 \beta_{ij}'^2)}{\text{Var}(\sum_j^M r_j^2 \beta_{ij}'^2)} \\ &\approx E[\alpha^2] \frac{1}{G} \sum_i^G \frac{\ell^2 \text{Cov}(\beta_i^2, \beta_i'^2)}{\ell^2 \text{Var}(\beta_i'^2)} \\ &\approx E[\alpha^2] \frac{1}{G} \sum_i^G \frac{\text{Cov}(\beta_i^2, \beta_i'^2)}{\text{Var}(\beta_i'^2)} \end{aligned}$$

Here, $\ell^2 = \text{Var}(\sum_j^M r_j^2)$. The third and fourth line follow given that r^2 is independent of α , β , and β' . See ‘‘MESC with assayed eQTL effect sizes’’ for the remainder of the derivation.

Stratified MESc derivation

Starting from equation (14):

$$E[\chi_k^2 | \mathbf{R}, \mathbf{B}] = N \sum_j^M E[\gamma_j^2 | \mathbf{R}, \mathbf{B}] \hat{r}_{jk}^2 + N \sum_i^G E[\alpha_i^2 | \mathbf{R}, \mathbf{B}] \sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + NE[(\epsilon')^2] \quad (7.1)$$

$$= N \sum_j^M \left(\sum_{c:j \in \mathcal{C}_c} \tau_c | \mathbf{R}, \mathbf{B} \right) \hat{r}_{jk}^2 + N \sum_i^G \left(\sum_{d:i \in \mathcal{D}_d} \pi_d | \mathbf{R}, \mathbf{B} \right) \sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + NE[(\epsilon')^2] \quad (7.2)$$

$$E[\chi_k^2] = N \sum_c \tau_c \sum_{j \in \mathcal{C}_c} \hat{r}_{jk}^2 + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + NE[(\epsilon')^2] \quad (7.3)$$

In order for (7.3) to be true, we must make the following assumptions:

- Within each gene category \mathcal{D}_d , π_d is uncorrelated with the magnitude of eQTL effect sizes
- Within each SNP category \mathcal{C}_c , τ_c is uncorrelated with the magnitude of eQTL effect sizes
- π_d is uncorrelated with the LD scores of eQTLs that affect genes in \mathcal{D}_d
- τ_c is uncorrelated with the LD scores of SNPs in \mathcal{C}_c

Since $E[\hat{r}_{jk}^2] \approx r_{jk}^2 + \frac{1}{N}$, we have

$$\begin{aligned} E[\chi_k^2] &= N \sum_c \tau_c \sum_{j \in \mathcal{C}_c} \left(r_{jk}^2 + \frac{1}{N} \right) + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M \left(r_{jk}^2 \beta_{ij}^2 + \frac{\beta_{ij}^2}{N} \right) + NE[(\epsilon')^2] \\ &= N \sum_c \tau_c \sum_{j \in \mathcal{C}_c} r_{jk}^2 + \sum_c \sum_{j \in \mathcal{C}_c} \tau_c + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2 + \\ &\quad \sum_d \sum_{i \in \mathcal{D}_d} \sum_j^M (\pi_d |\mathcal{D}_d| E[h_{cis}^2(\mathcal{D}_d)]) + NE[(\epsilon')^2] \\ &= N \sum_c \tau_c \sum_{j \in \mathcal{C}_c} r_{jk}^2 + h_{nonmed;causal}^2 + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2 + h_{med;causal}^2 + 1 - h_{nonmed;causal}^2 - h_{med;causal}^2 \\ &= N \sum_c \tau_c \sum_{j \in \mathcal{C}_c} r_{jk}^2 + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2 + 1 \end{aligned}$$

Letting $\ell_{k;c} = \sum_{j \in \mathcal{C}_c} r_{jk}^2$ and $\mathcal{L}_{k;d} = \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2$, we arrive at our main equation for stratified MESc:

$$E[\chi_k^2] = N \sum_c \tau_c \ell_{k;c} + N \sum_d \pi_d \mathcal{L}_{k;d} + 1$$

Impact of systemic differences in eQTL effect size magnitude between assayed vs. causal tissues

According to equation (13), we have

$$\alpha^2 = r_g^2(T) E[\alpha^2] \frac{1}{G} \sum_i^G \sqrt{\frac{\text{Var}(\beta_i^2)}{\text{Var}(\beta_i'^2)}}$$

There are reasonable scenarios in which $Var(\beta_i^2)$ can be systemically larger or smaller than $Var(\beta_i'^2)$ across all genes (e.g. if causal genes for the trait are primarily influenced by cell type-specific eQTLs that are weaker/absent in assayed tissues). This will cause $\alpha'^2 \neq r_g^2(T)E[\alpha^2]$. However, because we multiply α'^2 by $GE[h_{cis}^2]$ to obtain $h_{med;assayed}^2(T)$, we will still have $h_{med;assayed}^2(T) = r_g^2(T)h_{med;causal}^2$.

To illustrate this, consider a scenario where β'^2 is both correlated with β^2 and scaled by a factor c relative to β^2 . (13) thus becomes

$$\begin{aligned}\alpha'^2 &= r_g^2(T)E[\alpha^2] \frac{1}{G} \sum_i^G \sqrt{\frac{Var(\beta_i^2)}{Var(c\beta_i^2)}} \\ &= \frac{1}{c} r_g^2(T)E[\alpha^2]\end{aligned}$$

Note that scaling β'^2 by c will not change the average squared correlation $r_g^2(T)$ between β'^2 and β . We then have

$$\begin{aligned}h_{med;assayed}^2(T) &= GE[\beta'^2] \frac{1}{c} r_g^2(T)E[\alpha^2] \\ &= GE[c\beta^2] \frac{1}{c} r_g^2(T)E[\alpha^2] \\ &= r_g^2(T)h_{med;causal}^2\end{aligned}$$

Comparing different methods of estimating expression scores

In this set of simulations, we evaluated the prediction accuracy and bias of different methods of estimating expression scores from simulated expression data with varying numbers of samples (Methods). Note that this set of simulations does not involve complex trait phenotypes; see Figure 2a and Supplementary Figure 1 for simulation results involving complex trait phenotypes.

In total, we compared five different methods of estimating expression scores \mathcal{L}_k for SNP k . Here, G represents genes within 1 Mb of SNP k , while M represents SNPs within 1 Mb of SNP k :

1. *eQTL summary statistics*. $\hat{\mathcal{L}}_k = \sum_i^G \hat{\beta}_{ik}^2(\text{sumstat})$, where $\hat{\beta}_{ik}^2(\text{sumstat})$ represents the squared eQTL summary statistic of SNP k for gene i .
2. *LASSO*. $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r_{jk}^2 \hat{\beta}_{ij}^2(\text{LASSO})$, where r_{jk}^2 represents the squared correlation between SNP j and SNP k , and $\hat{\beta}_{ij}^2(\text{LASSO})$ represents squared *causal* eQTL effect sizes of SNP j on gene i estimated by LASSO¹⁶.
3. *LASSO with REML correction*. $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r_{jk}^2 c_i \hat{\beta}_{ij}^2(\text{LASSO})$. This method is identical to *LASSO* except that we scale $\hat{\beta}_{ij}^2(\text{LASSO})$ by a factor c_i . We define $c_i = \hat{h}_{cis;i}^2 / \sum_j^M \hat{\beta}_{ij}^2(\text{LASSO})$, where $\hat{h}_{cis;i}^2$ is the expression cis-heritability of gene i predicted by REML. This approach is the same as the one described in ‘‘Estimation of expression scores.’’ For computational ease, we did not actually use REML to predict expression cis-heritability for each gene in each simulation, but rather we took the true expression cis-heritability of the gene and added a realistic amount of noise in order to simulate REML prediction error.
4. *BLUP*. $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r_{jk}^2 \hat{\beta}_{ij}^2(\text{BLUP})$. Here, $\hat{\beta}_{ij}^2(\text{BLUP})$ represents squared *causal* eQTL effect sizes of SNP j on gene i estimated by best linear unbiased predictor (BLUP)¹⁷.
5. *BLUP with REML correction*. $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r_{jk}^2 c_i \hat{\beta}_{ij}^2(\text{BLUP})$. Same as *LASSO with REML correction*, but using BLUP rather than LASSO.

We report mean prediction accuracy (in terms of R^2 between predicted and true expression scores) and bias (in terms of the slope from regressing predicted expression scores on true expression scores) with mean standard errors over 100 independent simulations. Across all expression panel sample sizes, we obtained

the best prediction R^2 when using LASSO with REML correction (Supplementary Figure 2). The superior prediction R^2 of LASSO compared to other methods can be attributed to the fact that LASSO enforces a sparsity prior on effect sizes, which matches the sparse nature of cis-eQTL effect sizes. However, LASSO also produces biased estimates of effect sizes, which is why we need to scale the effects to match the expression cis-heritability estimated by REML (which is unbiased) to obtain unbiased estimates of expression scores. LASSO with REML correction gave us approximately unbiased estimates of expression scores at large enough sample sizes (>500). For OLS summary statistics, we observed unbiased estimates of expression scores at all sample sizes but inferior prediction R^2 to LASSO with REML correction. The remainder of the estimation methods were biased at all sample sizes and had comparable prediction R^2 to OLS summary statistics. We observed concordant results when varying the number of eQTLs per gene (Supplementary Figure 2a-c), eQTL window size (Supplementary Figure 2d), and REML prediction error (Supplementary Figure 2f). Prediction R^2 and bias across all methods varied when changing the mean cis-heritability of expression, though the relative performance of the five methods compared to each other remained consistent across different mean cis-heritability values (Supplementary Figure 2e).

Notably, we observed poor prediction R^2 of all methods for expression data sets of size 100-200, which is comparable to the size of most individual tissue expression panel data sets. This result suggests that we cannot reliably predict expression scores using available individual tissue expression panel data sets. In real data, this result is corroborated by the fact that we obtain low estimates of h_{med}^2/h_g^2 across all traits when using individual tissue expression panel data sets to estimate expression scores (Figure 5b).

Rare vs. common variant h_{med}^2

In all our analyses, we restrict the regression SNPs used by MESC to only Hapmap3 SNPs¹⁸. Because Hapmap3 SNPs essentially only tag common variants, by restricting to Hapmap3 SNPs we estimate the proportion of *common* disease heritability mediated by the cis-genetic component of gene expression ($h_{med(common)}^2/h_{common}^2$) (see following section for simulation results). We define $h_{med(common)}^2$ as $\sum_{j \in C} \sum_i \beta_{ij}^2 \alpha_i^2$ (given standardized genotypes and phenotypes), where C represents the set of all SNPs with minor allele frequency (MAF) > 0.05 , β_{ij} represents the cis-eQTL effect size of SNP j on gene i , and α_i represents the effect size of gene i on disease. This quantity differs from the *total* disease heritability mediated by gene expression ($h_{med(common)}^2 + h_{med(rare)}^2$), where $h_{med(rare)}^2$ is defined in the same manner as $h_{med(common)}^2$ but C is replaced with the set of all SNPs with MAF < 0.05 . We do not aim to estimate $h_{med(rare)}^2$ because doing so requires eQTL effect size estimates for rare variants, which cannot be reliably obtained from current expression panel data sets. Even if we had the data to estimate $h_{med(rare)}^2$ (i.e. many thousands of whole-genome sequencing expression samples), there are several reasons why we would expect the proportion of total disease heritability mediated by gene expression ($h_{med(common)}^2 + h_{med(rare)}^2$)/($h_{common}^2 + h_{rare}^2$) to be either similar or smaller than the quantity $h_{med(common)}^2/h_{common}^2$ that we estimate:

1. Most SNP heritability is explained by common variants^{19,20}. Thus, we can expect the quantity ($h_{med(common)}^2 + h_{med(rare)}^2$)/($h_{common}^2 + h_{rare}^2$) to depend mostly on $h_{med(common)}^2$ and h_{common}^2 .
2. Rare variant heritability has a much larger enrichment in coding regions than common variant heritability²¹, suggesting that the effects of rare variants on disease tend to be mediated by protein-coding changes rather than changes in gene expression. Protein-coding changes are not reflected in h_{med}^2 , so we would expect that $h_{med(rare)}^2/h_{rare}^2 < h_{med(common)}^2/h_{common}^2$.

Role of singletons in h_{med}^2

A recent study²² has shown that a substantial proportion of total expression cis-heritability (around 20%) in an expression panel of 360 individuals is explained by singletons with MAF < 0.0001 (i.e. singletons that are not observed in large genome reference panels). However, as mentioned above, rare variant effects do not contribute to $h_{med(common)}^2$ since they are not tagged by Hapmap3 SNPs. Furthermore, even if we had the data to estimate eQTL effect sizes for singletons, there is evidence that singletons contribute very little or nothing to disease heritability, as a recent study²³ has shown that virtually all narrow-sense heritability for height and BMI can be explained by SNPs with MAF > 0.0001 (which excludes the class of ultra-rare

SNPs defined as singletons in Hernandez et al.). In light of this result, we would not expect singleton effects on expression to substantially mediate any disease heritability.

Simulations under frequency-dependent genetic architectures

To evaluate the bias of MESC in estimating $h_{med(common)}^2$ in the presence of frequency-dependent genetic architectures (including rare and low-frequency variants), we conducted simulations in which both eQTL and GWAS per-allele effect size magnitude were inversely proportional to minor allele frequency, consistent with purifying selection acting on gene expression^{22,24} and complex trait^{19,20}. We conducted our simulations using real genotypes imputed to include rare and low-frequency variants from UK Biobank²⁵ ($N_{GWAS} = 100,000$; $M = 1,539,668$ SNPs from chromosome 1). We simulated cis-eQTLs for $G = 1000$ genes with variance of per-allele effect sizes proportional to $[p_i(1 - p_i)]^\alpha$, where p is the minor allele frequency of SNP i and α is a parameter ranging from -0.33 (corresponding to 5% of heritability explained by rare variants with MAF < 0.01 in our data set) to -1.33 (corresponding to 50% heritability explained by rare variants). We simulated frequency dependent non-mediated SNP effect sizes in a similar fashion as eQTL effect sizes. Finally, we simulated gene effect sizes on complex trait corresponding to $h_{med(common)}^2/h_{common}^2 = 0.2$ and $h^2 = 0.5$. From these effect sizes, we simulated GWAS summary statistics, as well as eQTL summary statistics using a separate set of genotypes ($N_{eQTL} = 10,000$).

We applied MESC to these summary statistics while performing the regression using only Hapmap3 SNPs (consistent with what we do in practice). In order to capture dependence between LD scores and GWAS effect sizes (which constitutes a model violation and leads to biased h_g^2 estimates if uncorrected^{26,27}), we stratified regression SNPs by 10 MAF bins, which has been shown to adequately account for this dependence^{7,27}. In all simulations, we obtained unbiased/slightly conservative estimates of $h_{med(common)}^2/h_{common}^2$ across diverse values of α , including scenarios in which α for non-mediated effect sizes was different than α for eQTL effect sizes (Supplementary Figure 5). Thus, MESC is robust to frequency-dependent genetic architectures for both gene expression and disease.

Role of tissue specificity in explaining low heritability genes

Because we define the h_{cis}^2 of a gene by averaging individual-tissue h_{cis}^2 estimates across all tissues, a gene with low meta-tissue h_{cis}^2 can reflect two different scenarios: the gene has low individual-tissue h_{cis}^2 across many tissues, or the gene has high individual-tissue h_{cis}^2 in only one or a small number of tissues (i.e. the gene has tissue-specific eQTLs). To investigate the potential role of tissue-specific eQTLs in explaining low meta-tissue h_{cis}^2 , we obtained three quantities for each gene: (1) the number of tissues in which the gene had a significantly nonzero h_{cis}^2 ($p < 0.05$), (2) the max h_{cis}^2 of the gene across all tissues, and (3) the average h_{cis}^2 of the gene across the tissues for which h_{cis}^2 was significantly nonzero. If the number of tissues in which the gene has truly nonzero h_{cis}^2 (an indicator of the tissue specificity of the gene) is the primary factor in determining the magnitude of the meta-tissue h_{cis}^2 , we would expect that (1) be proportional to the magnitude of the meta-tissue h_{cis}^2 , while (2) and (3) not be proportional to the magnitude of the meta-tissue h_{cis}^2 . We observed that (1) was indeed proportional to the magnitude of the meta-tissue h_{cis}^2 (Supplementary Figure 20a); however, (2) and (3) were also proportional to the magnitude of the meta-tissue h_{cis}^2 (Supplementary Fig 20b,c), suggesting that statistical power due to the magnitude of h_{cis}^2 , rather than tissue specificity, was primarily responsible for the fact that (1) was proportional to the meta-tissue h_{cis}^2 . In summary, these results suggest that low h_{cis}^2 genes are not primarily genes with highly tissue-specific eQTLs, though we cannot rule out the possibility of tissue-specific eQTLs having some contribution to low h_{cis}^2 genes.

Impact of adding window around gene set to model

When estimating the h_{med}^2 enrichment of a given gene set, we stratify SNPs by the baselineLD model v2.0 in order to account for correlations between eQTL effects and non-mediated effects, and we stratify genes within the gene set into three bins to account for correlations between eQTL effects and gene effects. In addition to the 72 baselineLD model SNP annotations, we might consider adding an additional annotation that corresponds to all SNPs within a certain genomic distance (e.g. 100 Kb) of genes in the gene category. By including this annotation, we impose a stricter standard for identifying h_{med}^2 enrichment of the gene category.

To illustrate this, consider a scenario in which GWAS signal tends to physically localize around genes in a given gene set, but that none of the GWAS signal is actually mediated by the expression levels of those genes. Because cis-eQTLs also (by definition) physically localize around those genes, by chance we will observe that eQTLs for those genes will have larger GWAS effect sizes compared to the genomic background, in which case we will likely spuriously identify the gene set as having significant h_{med}^2 enrichment. By including a 100 Kb window around each gene, we require that eQTL effect size magnitude is correlated with GWAS effect size magnitude *within* the 100 Kb windows to detect significant h_{med}^2 enrichment, which will not occur if the GWAS signal is not mediated by the expression levels of the genes. In summary, including a SNP annotation corresponding to a window around each gene can eliminate false positive h_{med}^2 enrichment estimates that arise due to localization of GWAS signal around genes that is not mediated by gene expression.

In practice, we chose to include a 100 Kb window around genes, given precedence in the literature^{9,12,28,29}. These studies report large heritability enrichment of a window of this size around many functional gene sets. When including this annotation for each gene set, we observed that the h_{med}^2 enrichment estimate for the gene set was very similar for most gene sets (Supplementary Figure 23), demonstrating that the h_{med}^2 enrichment of these gene sets was not due to coincidental overlap between non-mediated effects and eQTL effects near these genes. Given this result, we did not include the window around genes in any of our subsequent analyses, nor in any of the results we report in the manuscript.

Comparing MESC to other gene set enrichment methods

MESC can be used to prioritize disease-relevant gene sets using the h_{med}^2 enrichment of a gene set, defined as (proportion of h_{med}^2 in gene set) / (proportion of genes in gene set). Larger h_{med}^2 enrichment of the gene set suggests that the expression of genes in the gene set have larger causal effects on disease. Many other methods exist that also aim to prioritize causal gene sets using GWAS data³⁰⁻³⁶. MESC primarily differs from these other methods in that (1) it utilizes eQTL data, and (2) it specifically estimates causal effects of gene expression on disease, under a generative model for disease that connects SNP effects on gene expression to gene effects on disease. On the other hand, other popular methods for gene set enrichment analysis (e.g. MAGMA, DEPICT) are not based on eQTL data and do not model gene effects on disease. Instead, these methods prioritize gene sets under the assumption that causal genes should have more GWAS signal in close genomic proximity to them, which may not be true in some cases^{37,38}. Thus, the two qualities above can make MESC desirable as a discovery tool, especially since eQTLs have been useful in elucidating the mechanistic basis of disease in many other settings^{4,5,13,39-49}.

However, there are also scenarios in which MESC will miss gene sets that play a causal role in disease. In particular, MESC focuses only on genes whose expression levels mediate the effects of GWAS hits, to the extent that can be detected in existing eQTL studies such as GTEx. SNP effects on disease might be mediated by mechanisms other than gene expression levels (e.g. protein-coding changes), or they may be mediated by gene expression levels in specific cell types or contexts that are not captured by existing eQTL studies. Moreover, a key drawback of MESC is that it produces large standard errors for small gene sets and thus can only be applied to large gene sets with more than 200 genes, whereas other methods can analyze gene sets of any size. Thus, we propose MESC as a complementary approach rather than replacement for other pathway enrichment methods.

To compare MESC to other pathway enrichment methods, we applied MAGMA³³ and DEPICT³² to the same GWAS summary statistics for 26 traits with nominally significant h_{med}^2 . We analyzed a total of 501 gene sets, which represent the intersection of gene sets we analyzed using MESC in our study and gene sets built into the DEPICT software. We ran MAGMA with default parameters. DEPICT requires that we specify a p-value threshold for defining significant GWAS loci; however, the recommended thresholds of 1×10^{-5} and 5×10^{-8} caused DEPICT to exceed its maximum number of loci for many traits. Thus, for each trait, we set the p-value threshold to the maximum of the following values that did not cause DEPICT to exceed its maximum number of loci: 1×10^{-5} , 5×10^{-8} , 5×10^{-15} , 5×10^{-20} , 5×10^{-25} , 5×10^{-30} , 5×10^{-35} , 5×10^{-40} , 5×10^{-45} , 5×10^{-50} . We then compared gene sets identified as significantly enriched by MAGMA and DEPICT to gene sets with significant h_{med}^2 enrichment (see Supplementary Table 9 for all estimates). Out of 13,206 total trait-gene set pairs, MESC identified 106 with significant h_{med}^2 enrichment (FDR < 0.05 while correcting for $13,206 \times 3 = 39,078$ hypotheses), compared to 85 for MAGMA and 957 for DEPICT (Supplementary Figure 27a). We observed correlated enrichment p-values across the three

methods (MESC vs. MAGMA $R^2 = 0.14$, MESC vs. DEPICT $R^2 = 0.20$, MAGMA vs. DEPICT $R^2 = 0.10$) (Supplementary Figure 27b). Of the 106 significant trait-gene sets pairs identified by MESC, 32 were not detected as significant by either MAGMA or DEPICT (Supplementary Figure 27c), including biologically plausible trait-gene sets pairs such as “phospholipid metabolic process” for high-density lipoprotein level and “synapse part” for schizophrenia. These results demonstrate that MESC produces broadly concordant gene set enrichment estimates as the other methods, while also capturing unique signal that is present in only eQTL data.

Impact of environmental noise in gene expression measurements on expression-mediated heritability estimates

In this section, we show that the level of environmental noise in gene expression measurements (which differs across assays and affects both standardized eQTL effect sizes and the magnitude of expression cis-heritability, h_{cis}^2) does not impact our estimates of expression-mediated heritability h_{med}^2 . In other words, h_{med}^2 depends on only the genetic component of gene expression levels. One consequence of this fact is that the magnitude of h_{med}^2 does not *a priori* depend on the magnitude of h_{cis}^2 . For example, mean h_{cis}^2 can be very low in a given gene expression data set due to e.g. large stochastic fluctuations in gene expression levels or other sources of technical noise specific to the data set, while estimated h_{med}^2 from this gene expression data set can in principle be very high (i.e. close to total SNP heritability h_g^2).

To understand this intuitively, one can think of the units in which all SNP effect sizes operate. Recall our model for the effect size of SNP j on complex trait:

$$\omega_j = \sum_i \beta_{ij} \alpha_i + \gamma_j$$

where ω_j represents the total effect size of SNP j on the complex trait, β_{ij} represents the effect size of SNP j on the expression levels of gene i , α_i represents the effect size of gene i on the complex trait, and γ_j represents the non-mediated effect size of SNP j on the complex trait. When complex trait and gene expression levels are standardized to zero mean and unit variance, β_{ij} is expressed in terms of *additive increase in standardized expression levels per unit increase in standardized genotype* (which we abbreviate as $\frac{\text{std}(\text{expr})}{\text{std}(\text{geno})}$), while α_i is expressed in terms of *additive increase in standardized phenotype per unit increase of standardized expression levels* (which we abbreviate as $\frac{\text{std}(\text{pheno})}{\text{std}(\text{expr})}$). When we multiply α_i^2 by $\sum_i \beta_{ij}^2$, we obtain a quantity corresponding to the heritability mediated by gene expression levels for SNP j in units of $\left(\frac{\text{std}(\text{expr})}{\text{std}(\text{geno})}\right)^2 \left(\frac{\text{std}(\text{pheno})}{\text{std}(\text{expr})}\right)^2 = \left(\frac{\text{std}(\text{pheno})}{\text{std}(\text{geno})}\right)^2$.

Because $\text{std}(\text{expr})$ cancels out in this above product, the units in which gene expression levels are represented does not actually affect our final estimate of the heritability mediated by gene expression levels for SNP j (provided that both α_i and β_{ij} use the same units of expression). To elaborate, when we regress ω_j^2 on $\sum_i \beta_{ij}^2$, we obtain an estimate of $E[\alpha^2]$ in units of $\left(\frac{\text{std}(\text{pheno})}{\text{std}(\text{expr})}\right)^2$. To obtain an estimate of per-SNP expression-mediated heritability, we then multiply $E[\alpha^2]$ by $E[\sum_i \beta_{ij}^2]$ (or equivalently $E[h_{cis}^2]$), which is in units of $\left(\frac{\text{std}(\text{expr})}{\text{std}(\text{geno})}\right)^2$. Suppose we were to scale $\sum_i \beta_{ij}^2$ by an arbitrary factor c , in which case our estimate of $E[\alpha^2]$ would be in units of $\left(\frac{\text{std}(\text{pheno})}{c \cdot \text{std}(\text{expr})}\right)^2$ and $E[c \sum_i \beta_{ij}^2]$ would be in units of $\left(\frac{c \cdot \text{std}(\text{expr})}{\text{std}(\text{geno})}\right)^2$. When multiplying $E[\alpha^2]$ by $E[c \sum_i \beta_{ij}^2]$, the product would be expressed in units of $\left(\frac{c \cdot \text{std}(\text{expr})}{\text{std}(\text{geno})}\right)^2 \left(\frac{\text{std}(\text{pheno})}{c \cdot \text{std}(\text{expr})}\right)^2 = \left(\frac{\text{std}(\text{pheno})}{\text{std}(\text{geno})}\right)^2$ and would thus be unchanged compared to before. This is essentially the same argument as made in the section “Impact of systemic differences in eQTL effect size magnitude between assayed vs. causal tissues” (Supplementary Note), in which we show that true differences in eQTL effect size magnitude agnostic of environmental noise in expression assays also do not affect our estimates of expression-mediated heritability.

Adding environmental noise to gene expression levels has the effect of scaling both squared standardized eQTL effect sizes and h_{cis}^2 by a constant factor, which we have shown above does not affect estimates of h_{med}^2 . To illustrate this, consider the following generative model for the expression levels of gene i , in which genotypes are standardized to zero mean and unit variance but gene expression levels are *not* standardized:

$$\mathbf{y}_{i(exp)} = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_{i(exp)}$$

where $\mathbf{y}_{i(exp)}$ is a vector of *non-standardized* gene expression levels for gene i , \mathbf{X} is a matrix of *standardized* genotypes, $\boldsymbol{\beta}_i$ is a vector of *non-standardized* cis-eQTL effect sizes for gene i , and $\boldsymbol{\epsilon}_{i(exp)}$ is a vector of environmental effects. In order to standardize cis-eQTL effect sizes (so that $\sum \beta_{i(std)}^2 = h_{cis}^2$), we divide all non-standardized cis-eQTL effect sizes β_i by $\sqrt{\sum \beta_i^2 + Var(\epsilon_{i(exp)})}$ to obtain $\beta_{i(std)}$. Now, note that adjusting the variance of the noise term $\epsilon_{i(exp)}$ is akin to scaling both h_{cis}^2 and $\beta_{i(std)}^2$ by the same constant factor. For example, let $Var(\epsilon_{i(exp)})$ be the original environmental variance, and let $Var(\epsilon'_{i(exp)})$ be the new environmental variance. Both original h_{cis}^2 and $\beta_{i(std)}^2$ are multiplied by the factor $\frac{\sum \beta_i^2 + Var(\epsilon_{i(exp)})}{\sum \beta_i^2 + Var(\epsilon'_{i(exp)})}$ to obtain the new h_{cis}^2 and $\beta_{i(std)}^2$.

In summary, we show that the level of environmental noise in gene expression panels (due to e.g. stochastic fluctuations in gene expression levels or other sources of assay noise) does not impact our estimates of h_{med}^2 .

Prospects for estimating disease heritability mediated by trans-eQTLs

In all our analyses, we aim to estimate disease heritability mediated by gene expression *in cis*, rather than the full genetic component of gene expression that includes trans effects. In theory, we can also estimate disease heritability mediated by gene expression *in trans* using MESC, where we would simply replace cis-eQTL effect sizes with trans-eQTL effect sizes in all our analyses. However, trans-eQTLs are much more difficult to estimate than cis-eQTLs due to their much smaller effect sizes, impacting resulting estimates of h_{med}^2 . In this section, we show that h_{med}^2 estimates produced by MESC are bounded by the average genetic prediction r^2 of gene expression multiplied by true h_{med}^2 , so $r^2 < 1$ results in downward bias in estimated h_{med}^2 (note that here r^2 refers to the prediction accuracy of the only the genetic component of gene expression, which does not include environmental effects). For gene expression *in cis*, this downward bias is minimal at current sample sizes (see simulation result in Figure 2a), as we can obtain a prediction r^2 close to 1 for cis-eQTLs⁴ (also see simulation in Supplementary Figure 2). However, for gene expression *in trans*, this downward bias becomes problematic at current sample sizes, since trans-eQTLs are highly polygenic and thus more difficult to estimate^{3,15,50}. In order to obtain a reliable estimate of disease heritability mediated by trans-eQTLs, we would ideally want a genetic prediction r^2 of 0.8 or greater (comparable to the prediction r^2 of cis-eQTLs from currently available gene expression data sets), which we show requires expression panels on the order of 1,000,000 or more samples (see below). Note that these sample sizes are comparable to those needed for very accurate polygenic disease risk prediction. We also show that the expected prediction r^2 of trans-eQTLs from the largest gene expression data set (eQTLGen⁵⁰, $N = 31,684$) is only 0.026, which is far too low to yield meaningful estimates of h_{med}^2 . Thus, estimating disease heritability mediated by trans-eQTLs using MESC is not feasible with currently available gene expression data sets.

Relationship between genetic prediction r^2 of gene expression and magnitude of estimated h_{med}^2

Let β represent eQTL effect sizes (either cis or trans) in causal cell types/contexts for the trait, let β' represent eQTL effect sizes in assayed tissues T , and let $\hat{\beta}' = \beta' + \epsilon$ represent *estimated* eQTL effect sizes in assayed tissues T , where ϵ is a noise term. We assume that ϵ is independent of β' . Upon regressing squared GWAS effects sizes ω^2 on squared estimated eQTL effect sizes $\sum_i^G \hat{\beta}'_i^2$, the estimate of the coefficient $\hat{\alpha}'^2$ is:

$$\begin{aligned}
\hat{\alpha}'^2 &\approx E[\alpha^2] \frac{1}{G} \sum_i^G \frac{Cov(\beta_i^2, \hat{\beta}_i'^2)}{Var(\hat{\beta}_i'^2)} \\
&\approx E[\alpha^2] \frac{1}{G} \sum_i^G \frac{Cov(\beta_i^2, \beta_i'^2 + \epsilon_i^2)}{Var(\hat{\beta}_i'^2)} \\
&\approx E[\alpha^2] \frac{1}{G} \sum_i^G \frac{Cov(\beta_i^2, \beta_i'^2)}{Var(\hat{\beta}_i'^2)}
\end{aligned}$$

The first line follows from the same derivation as ‘‘MESC with eQTL effect sizes in non-causal tissues.’’ As before, we define $r_g^2(T) = \frac{1}{G} \sum_i^G \frac{Cov(\beta_i^2, \beta_i'^2)}{\sqrt{Var(\beta_i^2)Var(\beta_i'^2)}}$ as the average squared genetic correlation between expression in assayed tissues T vs. in causal cell types/contexts. Given this definition, we have:

$$\hat{\alpha}'^2 \approx r_g^2(T) E[\alpha^2] \frac{1}{G} \sum_i^G \frac{\sqrt{Var(\beta_i^2)Var(\beta_i'^2)}}{Var(\hat{\beta}_i'^2)}$$

We can establish an upper bound for $\hat{\alpha}'^2$ in terms of the genetic prediction accuracy of expression. Let $r_{pred}^2(T) = \frac{1}{G} \sum_i^G \frac{Cov(\hat{\beta}_i'^2, \beta_i'^2)}{\sqrt{Var(\hat{\beta}_i'^2)Var(\beta_i'^2)}}$ represent the average squared genetic prediction accuracy of expression across genes in tissues T . Note that $\frac{\sqrt{Var(\beta_i^2)Var(\beta_i'^2)}}{Var(\hat{\beta}_i'^2)} \leq \frac{Cov(\hat{\beta}_i'^2, \beta_i'^2)}{\sqrt{Var(\hat{\beta}_i'^2)Var(\beta_i'^2)}}$ under the assumption that $Var(\beta_i^2) \approx Var(\beta_i'^2)$. To illustrate this, see that the numerators of both sides of the inequality are equivalent: $Cov(\hat{\beta}_i'^2, \beta_i'^2) = Cov(\beta_i'^2 + \epsilon_i^2, \beta_i'^2) = Cov(\beta_i'^2, \beta_i'^2) = Var(\beta_i'^2) \approx Var(\beta_i^2)$ on the left side, and $\sqrt{Var(\beta_i^2)Var(\beta_i'^2)} \approx \sqrt{Var(\beta_i^2)Var(\beta_i'^2)} \approx Var(\beta_i^2)$ on the right side. However, in the denominators, $Var(\hat{\beta}_i'^2) \geq \sqrt{Var(\hat{\beta}_i'^2)Var(\beta_i'^2)}$, since $Var(\hat{\beta}_i'^2) \geq Var(\beta_i'^2)$. Thus, we have:

$$\hat{\alpha}'^2 \leq r_{pred}^2(T) r_g^2(T) E[\alpha^2]$$

We can multiply $\hat{\alpha}'^2$ by $GE[h_{cis}^2]$ to obtain an estimate of $h_{med;assayed}^2(T)$ that has the following property:

$$\hat{h}_{med;assayed}^2(T) \leq r_{pred}^2(T) h_{med;assayed}^2(T)$$

Expected genetic prediction r^2 of gene expression in trans

Unlike cis-eQTLs, trans-eQTLs are known to be polygenic. Thus, we can invoke the following equation that relates sample size to polygenic prediction accuracy for gene expression in trans using the best linear unbiased predictor (BLUP)^{51,52}:

$$\begin{aligned}
r_{pred;trans}^2(T) &= \frac{1}{G} \sum_i^G \frac{h_{i;trans}^2}{h_{i;trans}^2 + \frac{M}{N}(1 - r_{pred;trans}^2(T))} \\
&\approx \frac{N}{M} \frac{1}{G} \sum_i^G h_{i;trans}^2
\end{aligned}$$

where $h_{i;trans}^2$ is the expression trans-heritability of gene i , M is the effective number of independent SNPs (approximately 60,000⁵³), and N is expression panel sample size. The largest expression panel available to date is from eQTLGen⁵⁰, with $N = 31,684$ in blood. The average h_{trans}^2 of expression is around 0.05^{3,4,54}. Thus, we can expect $r_{pred;trans}^2(T)$ trained on eQTLGen data to be around $\frac{31,684 \cdot 0.05}{60,000} = 0.026$, which is far too low to yield meaningful estimates of h_{med}^2 . In order to obtain $r_{pred;trans}^2(T)$ of 0.8 (comparable to the prediction r^2 of gene expression in cis from current expression panels), we would need $\frac{0.8 \cdot 60,000}{0.05} = 960,000$ samples.

Choice of 10 traits for display in Figure 5

Starting with the full set of 26 traits with h_{med}^2/h_g^2 greater than 0 with nominal significance ($p < 0.05$), we pruned genetically correlated traits as follows. First, we selected the pair of traits with the greatest genetic correlation (estimated using cross-trait LD score regression⁵⁵). Between the pair of selected traits, we then retained the trait with the larger h_g^2 z-score (estimated using stratified LD score regression⁶). We repeated this procedure until 10 traits were left.

Simulation: Supplementary Figure 2

We selected a 20 Mb region on chromosome 1 (base pair coordinates 60,000,000 to 80,000,000), which contained 8,604 SNPs. 100 genes were simulated within this region, with the average cis-heritability across all genes set at $\overline{h_{cis}^2} = 0.05$ or 0.01. For each gene, we simulated 1, 5, or 10 cis-eQTLs with locations randomly selected within a random 1 Mb or 10 Kb window within the overall 20 Mb region. For simulations with one eQTL per gene, the effect size for the eQTL was drawn from $\mathcal{N}(0, \overline{h_{cis}^2})$. For simulations with more than one eQTL per gene, one eQTL was randomly selected to explain 80% of the total heritability of the gene and had an effect size drawn from $\mathcal{N}(0, 0.8\overline{h_{cis}^2})$. The remaining eQTLs had effect sizes drawn from $\mathcal{N}(0, 0.2\overline{h_{cis}^2}/(N_{eQTL} - 1))$, where N_{eQTL} is the total number of eQTLs per gene. Expression values were simulated for each gene using an additive generative model with previously simulated effect sizes and environmental noise drawn from $\mathcal{N}(0, 1 - \overline{h_{cis}^2})$. LASSO and BLUP prediction of eQTL effect sizes was performed using all SNPs within 1Mb of each simulated gene. To simulate REML prediction error in expression heritability estimates, we added noise drawn from $\mathcal{N}(0, 0.01^2)$ to the true heritability values, which is consistent with the average standard error of GCTA estimates of expression cis-heritability across all GTEx samples (Supplementary Figure 32).

Simulation: Supplementary Figure 3

All effect sizes and complex trait phenotypes were simulated in the same manner as Figure 2a. We simulated expression phenotypes for 1, 5, or 10 tissues with 200 samples per tissue using the same eQTL effect sizes used to generate complex trait phenotypes. We estimated expression scores in each individual tissue. We meta-analyzed expression scores across tissues by averaging the causal squared LASSO-predicted eQTL effect sizes across all tissues for each gene (after scaling the effect sizes to the estimated expression cis-heritability). We then used these averaged causal eQTL effect sizes to compute expression scores by multiplying them by the element-wise squared LD matrix. This meta-analysis procedure is the same as the one described in “Meta-analysis of expression scores” in Methods.

Simulation: Supplementary Figure 5

For all simulations, we set $\overline{h_{cis}^2} = 0.05$, $h_{med}^2 = 0.1$, and $h^2 = 0.5$. For each of 1000 genes, we randomly selected 5 SNPs within a random 1 Mb window to act as cis-eQTLs with effect sizes drawn from $\mathcal{N}(0, [p_i(1-p_i)]^{\alpha_{eQTL}})$, where p_i is the MAF of the given SNP i . To avoid extreme effect sizes for singletons or doubletons, effect sizes for SNPs with MAF < 0.01 were drawn from $\mathcal{N}(0, [0.01(1-0.01)]^{\alpha_{eQTL}})$. α_{eQTL} was set to either -0.33, -0.60, -0.99, or -1.33. Finally, we scaled the effect sizes of the cis-eQTLs so that the sum of their squared effects equalled $\overline{h_{cis}^2}$. Expression phenotypes were simulated for each gene with environmental noise drawn from $\mathcal{N}(0, 1 - \overline{h_{cis}^2})$.

Similarly, we simulated non-mediated effect sizes for each SNP from $\mathcal{N}(0, [p_i(1-p_i)]^{\alpha_{GWAS}})$, or from $\mathcal{N}(0, [0.01(1-0.01)]^{\alpha_{GWAS}})$ for SNPs with MAF < 0.01 . α_{GWAS} was set to any of the same four values α_{eQTL} could take on (without necessarily being equal to α_{eQTL}). We then scaled these effect sizes so that the sum of the squared effects equalled $h^2 - h_{med}^2$. We simulated gene-trait effect sizes for each gene from $\mathcal{N}(0, (h_{med}^2/(G\overline{h_{cis}^2}))$. Complex trait phenotypes were simulated with environmental noise drawn from $\mathcal{N}(0, 1 - h_g^2)$.

Expression scores were estimated by computing eQTL summary statistics from the simulated expression panel. In-sample LD scores were computed for all 1,539,668 SNPs from the 100,000 GWAS samples. Regression was performed using only Hapmap3 SNPs.

Simulation: Supplementary Figure 6

We set the overall $h_{med}^2/h_g^2 = 0.4$. We simulated a gene category containing 200 random genes from the 1000 total genes. For the null scenario, we drew gene-trait effect sizes for all genes, including genes in the gene category, from $\mathcal{N}(0, h_{med}^2/(G\overline{h_{cis}^2}))$. For the enriched scenario, we simulated gene-trait effect sizes so that h_{med}^2 within the gene category was 2x the h_{med}^2 of genes outside of the category. eQTL effect sizes, non-mediated SNP effect sizes, expression phenotypes, and complex trait phenotypes were simulated in the same manner as in Figure 2a.

Simulation: Supplementary Figure 8

We set $\overline{h_{cis}^2} = 0.05$. We simulated 1 eQTL per gene with effect size drawn from $\mathcal{N}(0, \overline{h_{cis}^2})$. To simulate 10x enrichment of eQTLs in coding, TSS, and conserved regions, we selected eQTL locations so that 10x more eQTLs per SNP were located in the three SNP categories than the remainder of the genome. We simulated non-mediated effect sizes so that the heritability enrichment of the three SNP categories was 10x. Gene-trait effect sizes, expression phenotypes, and complex trait phenotypes were simulated in the same manner as in Figure 2a.

References

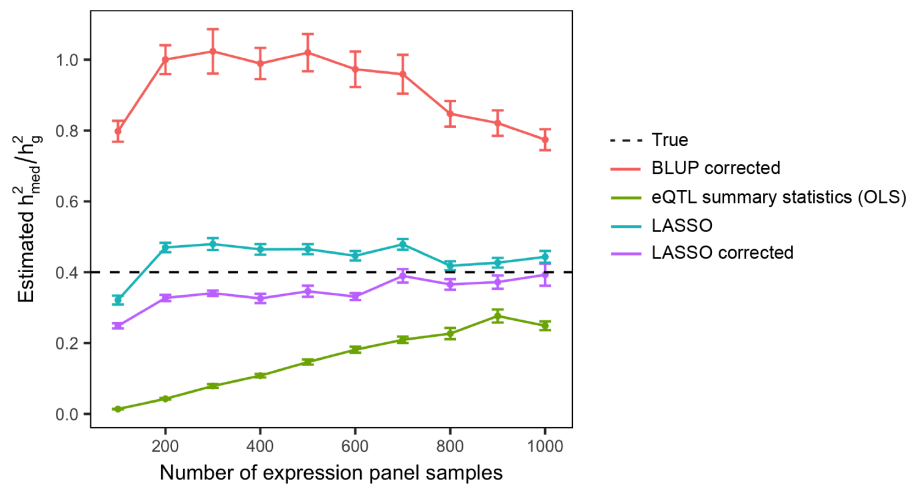
- ¹ Price, A. L. *et al.* Single-Tissue and Cross-Tissue Heritability of Gene Expression Via Identity-by-Descent in Related or Unrelated Individuals. *PLoS Genet.* **7**, e1001317, DOI: [10.1371/journal.pgen.1001317](https://doi.org/10.1371/journal.pgen.1001317) (2011).
- ² Liu, X. *et al.* Functional Architectures of Local and Distal Regulation of Gene Expression in Multiple Human Tissues. *The Am. J. Hum. Genet.* **100**, 605–616, DOI: [10.1016/j.ajhg.2017.03.002](https://doi.org/10.1016/j.ajhg.2017.03.002) (2017).
- ³ Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022–1034.e6, DOI: [10.1016/j.cell.2019.04.014](https://doi.org/10.1016/j.cell.2019.04.014) (2019).
- ⁴ Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252, DOI: [10.1038/ng.3506](https://doi.org/10.1038/ng.3506) (2016).
- ⁵ Gusev, A. *et al.* A transcriptome-wide association study of high-grade serous epithelial ovarian cancer identifies new susceptibility genes and splice variants. *Nat. Genet.* **51**, 815, DOI: [10.1038/s41588-019-0395-x](https://doi.org/10.1038/s41588-019-0395-x) (2019).
- ⁶ Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235, DOI: [10.1038/ng.3404](https://doi.org/10.1038/ng.3404) (2015).
- ⁷ Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427, DOI: [10.1038/ng.3954](https://doi.org/10.1038/ng.3954) (2017).
- ⁸ Hormozdiari, F. *et al.* Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat. Genet.* **50**, 1041, DOI: [10.1038/s41588-018-0148-2](https://doi.org/10.1038/s41588-018-0148-2) (2018).
- ⁹ Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621, DOI: [10.1038/s41588-018-0081-4](https://doi.org/10.1038/s41588-018-0081-4) (2018).
- ¹⁰ Amariuta, T. *et al.* IMPACT: Genomic Annotation of Cell-State-Specific Regulatory Elements Inferred from the Epigenome of Bound Transcription Factors. *The Am. J. Hum. Genet.* **104**, 879–895, DOI: [10.1016/j.ajhg.2019.03.012](https://doi.org/10.1016/j.ajhg.2019.03.012) (2019).
- ¹¹ Hujoel, M. L. A., Gazal, S., Hormozdiari, F., Geijn, B. v. d. & Price, A. L. Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species. *The Am. J. Hum. Genet.* **104**, 611–624, DOI: [10.1016/j.ajhg.2019.02.008](https://doi.org/10.1016/j.ajhg.2019.02.008) (2019).
- ¹² Kim, S. S. *et al.* Genes with High Network Connectivity Are Enriched for Disease Heritability. *The Am. J. Hum. Genet.* **104**, 896–913, DOI: [10.1016/j.ajhg.2019.03.020](https://doi.org/10.1016/j.ajhg.2019.03.020) (2019).

- ¹³ Barfield, R. *et al.* Transcriptome-wide association studies accounting for colocalization using Egger regression. *Genet. Epidemiol.* **42**, 418–433, DOI: [10.1002/gepi.22131](https://doi.org/10.1002/gepi.22131) (2018).
- ¹⁴ Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27**, R195–R208, DOI: [10.1093/hmg/ddy163](https://doi.org/10.1093/hmg/ddy163) (2018).
- ¹⁵ GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213, DOI: [10.1038/nature24277](https://doi.org/10.1038/nature24277) (2017).
- ¹⁶ Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. Royal Stat. Soc. Ser. B (Methodological)* **58**, 267–288 (1996).
- ¹⁷ Robinson, G. K. That BLUP is a Good Thing: The Estimation of Random Effects. *Stat. Sci.* **6**, 15–32, DOI: [10.1214/ss/1177011926](https://doi.org/10.1214/ss/1177011926) (1991).
- ¹⁸ The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58, DOI: [10.1038/nature09298](https://doi.org/10.1038/nature09298) (2010).
- ¹⁹ Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50**, 746–753, DOI: [10.1038/s41588-018-0101-4](https://doi.org/10.1038/s41588-018-0101-4) (2018).
- ²⁰ Schoech, A. P. *et al.* Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* **10**, 1–10, DOI: [10.1038/s41467-019-08424-6](https://doi.org/10.1038/s41467-019-08424-6) (2019).
- ²¹ Gazal, S. *et al.* Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet.* **50**, 1600–1607, DOI: [10.1038/s41588-018-0231-8](https://doi.org/10.1038/s41588-018-0231-8) (2018).
- ²² Hernandez, R. D. *et al.* Ultrarare variants drive substantial cis heritability of human gene expression. *Nat. Genet.* **51**, 1349–1355, DOI: [10.1038/s41588-019-0487-7](https://doi.org/10.1038/s41588-019-0487-7) (2019).
- ²³ Wainschtein, P. *et al.* Recovery of trait heritability from whole genome sequence data. *bioRxiv* 588020, DOI: [10.1101/588020](https://doi.org/10.1101/588020) (2019).
- ²⁴ Glassberg, E. C., Gao, Z., Harpak, A., Lan, X. & Pritchard, J. K. Evidence for Weak Selective Constraint on Human Gene Expression. *Genetics* **211**, 757–772, DOI: [10.1534/genetics.118.301833](https://doi.org/10.1534/genetics.118.301833) (2019).
- ²⁵ Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203, DOI: [10.1038/s41586-018-0579-z](https://doi.org/10.1038/s41586-018-0579-z) (2018).
- ²⁶ Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992, DOI: [10.1038/ng.3865](https://doi.org/10.1038/ng.3865) (2017).
- ²⁷ Gazal, S., Marquez-Luna, C., Finucane, H. K. & Price, A. L. Reconciling S-LDSC and LDK functional enrichment estimates. *Nat. Genet.* **51**, 1202–1204, DOI: [10.1038/s41588-019-0464-1](https://doi.org/10.1038/s41588-019-0464-1) (2019).
- ²⁸ Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLOS Comput. Biol.* **12**, e1004714, DOI: [10.1371/journal.pcbi.1004714](https://doi.org/10.1371/journal.pcbi.1004714) (2016).
- ²⁹ Fine, R. S., Pers, T. H., Amariuta, T., Raychaudhuri, S. & Hirschhorn, J. N. Benchmark: An Unbiased, Association-Data-Driven Strategy to Evaluate Gene Prioritization Algorithms. *The Am. J. Hum. Genet.* **104**, 1025–1039, DOI: [10.1016/j.ajhg.2019.03.027](https://doi.org/10.1016/j.ajhg.2019.03.027) (2019).
- ³⁰ Segrè, A. V. *et al.* Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. *PLOS Genet.* **6**, e1001058, DOI: [10.1371/journal.pgen.1001058](https://doi.org/10.1371/journal.pgen.1001058) (2010).
- ³¹ Lee, P. H., O’Dushlaine, C., Thomas, B. & Purcell, S. M. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28**, 1797–1799, DOI: [10.1093/bioinformatics/bts191](https://doi.org/10.1093/bioinformatics/bts191) (2012).

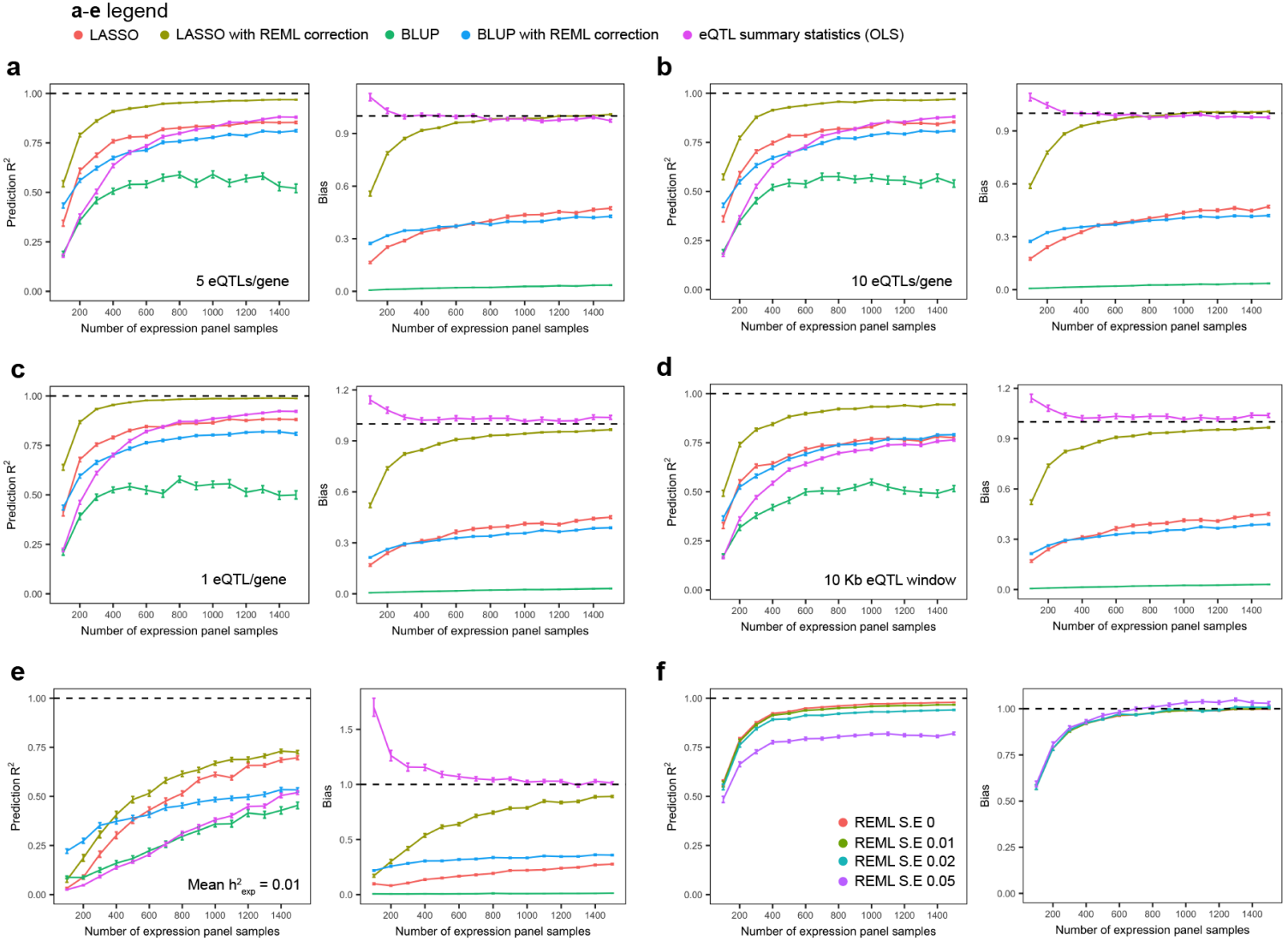
- ³² Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890, DOI: [10.1038/ncomms6890](https://doi.org/10.1038/ncomms6890) (2015).
- ³³ Leeuw, C. A. d., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, e1004219, DOI: [10.1371/journal.pcbi.1004219](https://doi.org/10.1371/journal.pcbi.1004219) (2015).
- ³⁴ de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364, DOI: [10.1038/nrg.2016.29](https://doi.org/10.1038/nrg.2016.29) (2016).
- ³⁵ Yoon, S. *et al.* Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. *Nucleic Acids Res.* **46**, e60–e60, DOI: [10.1093/nar/gky175](https://doi.org/10.1093/nar/gky175) (2018).
- ³⁶ Zhu, X. & Stephens, M. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat. Commun.* **9**, 4361, DOI: [10.1038/s41467-018-06805-x](https://doi.org/10.1038/s41467-018-06805-x) (2018).
- ³⁷ Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375, DOI: [10.1038/nature13138](https://doi.org/10.1038/nature13138) (2014).
- ³⁸ Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *New Engl. J. Medicine* **373**, 895–907, DOI: [10.1056/NEJMoa1502214](https://doi.org/10.1056/NEJMoa1502214) (2015).
- ³⁹ Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098, DOI: [10.1038/ng.3367](https://doi.org/10.1038/ng.3367) (2015).
- ⁴⁰ Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487, DOI: [10.1038/ng.3538](https://doi.org/10.1038/ng.3538) (2016).
- ⁴¹ Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825, DOI: [10.1038/s41467-018-03621-1](https://doi.org/10.1038/s41467-018-03621-1) (2018).
- ⁴² Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408, DOI: [10.7554/eLife.34408](https://doi.org/10.7554/eLife.34408) (2018).
- ⁴³ Mancuso, N. *et al.* Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat. Commun.* **9**, 4079, DOI: [10.1038/s41467-018-06302-1](https://doi.org/10.1038/s41467-018-06302-1) (2018).
- ⁴⁴ Wu, L. *et al.* A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nat. Genet.* **50**, 968, DOI: [10.1038/s41588-018-0132-x](https://doi.org/10.1038/s41588-018-0132-x) (2018).
- ⁴⁵ Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668, DOI: [10.1038/s41588-018-0090-3](https://doi.org/10.1038/s41588-018-0090-3) (2018).
- ⁴⁶ Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127, DOI: [10.1126/science.aat8127](https://doi.org/10.1126/science.aat8127) (2018).
- ⁴⁷ Huckins, L. M. *et al.* Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. *Nat. Genet.* **51**, 659, DOI: [10.1038/s41588-019-0364-4](https://doi.org/10.1038/s41588-019-0364-4) (2019).
- ⁴⁸ Gamazon, E. R., Zwinderman, A. H., Cox, N. J., Denys, D. & Derks, E. M. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits. *Nat. Genet.* **51**, 933, DOI: [10.1038/s41588-019-0409-8](https://doi.org/10.1038/s41588-019-0409-8) (2019).
- ⁴⁹ Porcu, E. *et al.* Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat. Commun.* **10**, 1–12, DOI: [10.1038/s41467-019-10936-0](https://doi.org/10.1038/s41467-019-10936-0) (2019).
- ⁵⁰ Vösa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 447367, DOI: [10.1101/447367](https://doi.org/10.1101/447367) (2018).

- ⁵¹ Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. *PLOS ONE* **3**, e3395, DOI: [10.1371/journal.pone.0003395](https://doi.org/10.1371/journal.pone.0003395) (2008).
- ⁵² Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515, DOI: [10.1038/nrg3457](https://doi.org/10.1038/nrg3457) (2013).
- ⁵³ Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106, DOI: [10.1038/ng.2876](https://doi.org/10.1038/ng.2876) (2014).
- ⁵⁴ Lloyd-Jones, L. R. *et al.* The Genetic Architecture of Gene Expression in Peripheral Blood. *The Am. J. Hum. Genet.* **100**, 228–237, DOI: [10.1016/j.ajhg.2016.12.008](https://doi.org/10.1016/j.ajhg.2016.12.008) (2017).
- ⁵⁵ Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241, DOI: [10.1038/ng.3406](https://doi.org/10.1038/ng.3406) (2015).
- ⁵⁶ Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295, DOI: [10.1038/ng.3211](https://doi.org/10.1038/ng.3211) (2015).

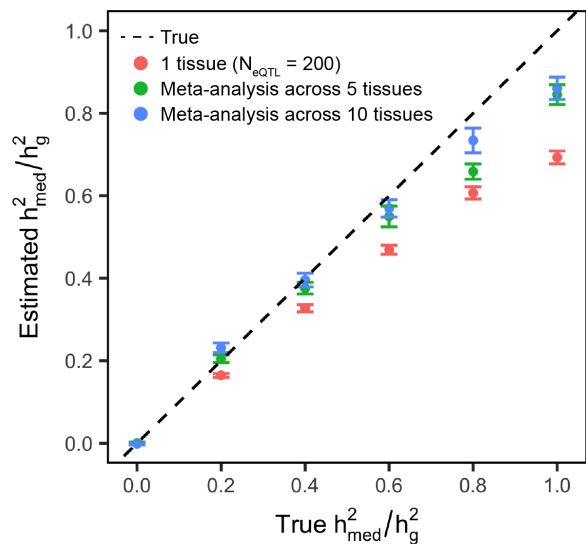
Supplementary Figures



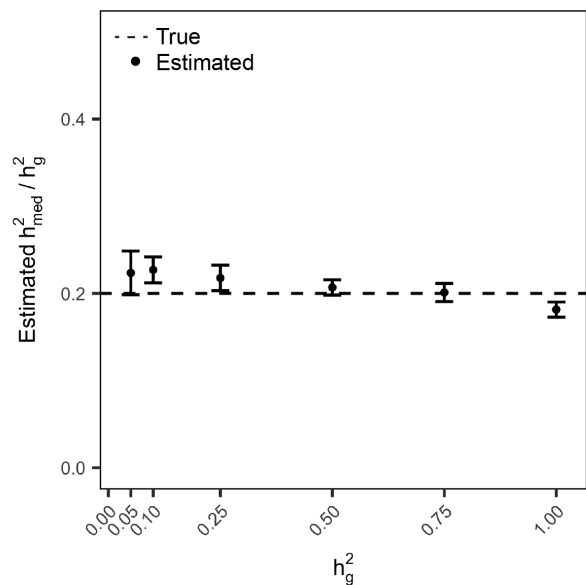
Supplementary Figure 1. Impact of different methods of estimating expression scores on estimates of h^2_{med}/h_g^2 . For this simulation, $h^2_{med}/h_g^2 = 0.4$. All other simulation parameters were the same as in Figure 2a. We exclude results for expression scores estimated using BLUP, since h^2_{med}/h_g^2 estimates obtained from these expression scores were severely upwardly biased (i.e. greater than 1 at all sample sizes).



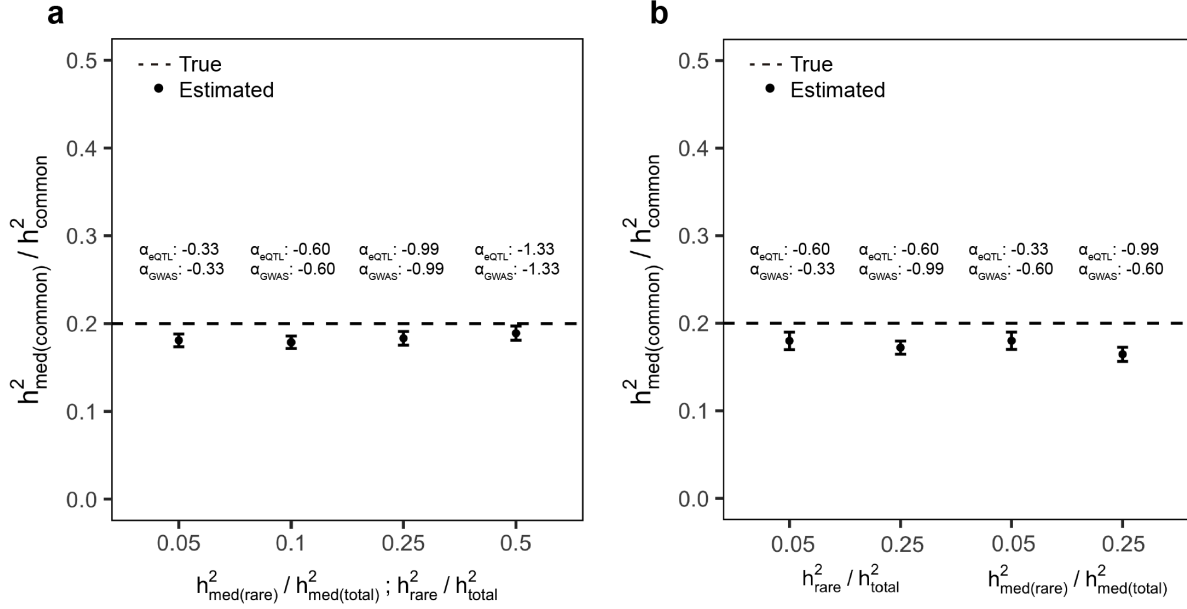
Supplementary Figure 2. Accuracy and bias of different methods of estimating expression scores in various simulated cis-genetic architectures. Left: R^2 between predicted and true expression scores at different expression panel sample sizes. Right: Bias of predicted expression scores (slope from regressing predicted expression scores on true expression scores). Default settings: 5 cis-eQTLs per gene, cis-eQTLs randomly selected within 1 Mb window, mean expression cis-heritability = 0.05. (a-c) 5, 10, and 1 simulated cis-eQTL per gene respectively. (d) eQTLs randomly selected within 10 Kb window. (e) Mean expression cis-heritability = 0.01. (f) LASSO with REML correction results for various levels of REML noise. Error bars for all plots represent mean standard errors across 100 simulations. See Supplementary Note for remaining details on this simulation.



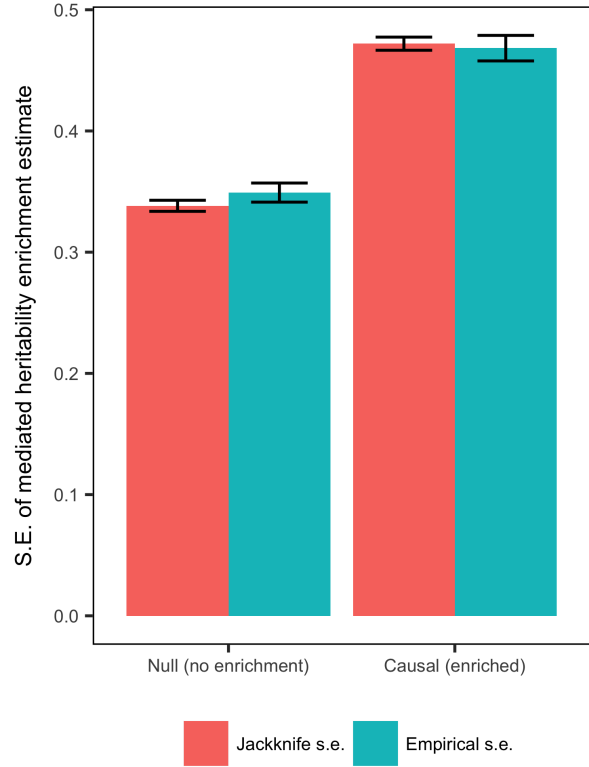
Supplementary Figure 3. h_{med}^2/h_g^2 estimates from expression scores meta-analyzed across tissues. We simulated expression phenotypes in multiple tissues, then estimated expression scores in individual tissues and meta-analyzed expression scores across tissues. Error bars represent mean standard errors across 100 simulations. See Supplementary Note for remaining details on this simulation.



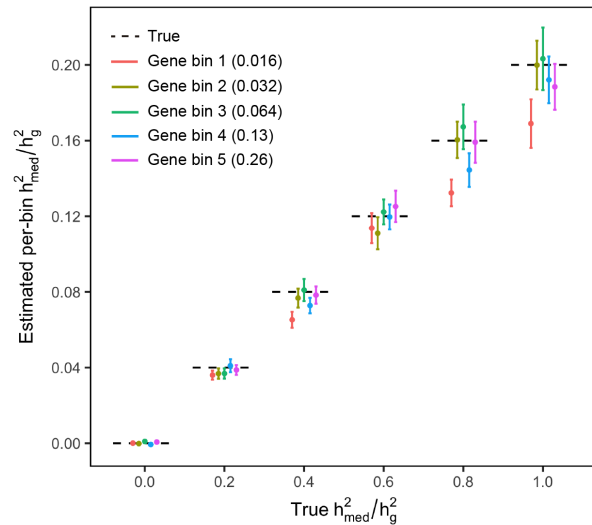
Supplementary Figure 4. h_{med}^2/h_g^2 estimates when varying h_g^2 . Simulation was performed in the same manner as in Figure 2a (with expression panel size fixed at 1000). h_{med}^2/h_g^2 was fixed at 0.2 for all simulations. Error bars represent mean standard errors across 100 simulations.



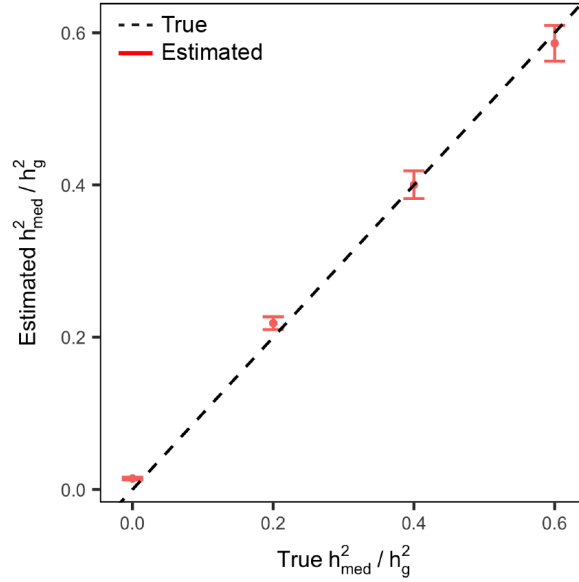
Supplementary Figure 5. h_{med}^2/h_g^2 estimates under simulated frequency-dependent genetic architectures. Error bars represent mean standard errors across 100 simulations. With $h_{med(common)}^2/h_{common}^2$ fixed at 0.2 and h^2 fixed at 0.5, we simulated eQTL effect sizes with variance of per-allele effect sizes proportional to $[p_i(1-p_i)]^{\alpha_{eQTL}}$ and non-mediated effect sizes with variance proportional to $[p_i(1-p_i)]^{\alpha_{GWAS}}$. We selected values of α_{eQTL} and α_{GWAS} to have the following property: $\alpha = -0.33$: 5% of total heritability explained by rare variants with MAF < 0.01 ; $\alpha = -0.60$: 10% heritability explained by rare variants; $\alpha = -0.99$: 25% heritability explained by rare variants; $\alpha = -1.33$: 50% heritability explained by rare variants. **(a)** $h_{med(common)}^2/h_{common}^2$ estimates with $\alpha_{eQTL} = \alpha_{GWAS}$, in which case the proportion of rare h_{med}^2 and rare h^2 is the same (x-axis). **(b)** $h_{med(common)}^2/h_{common}^2$ estimates with $\alpha_{eQTL} \neq \alpha_{GWAS}$. Left: proportion of rare h_{med}^2 is fixed at 0.1, while proportion of rare h^2 is varied from 0.05 to 0.25. Right: proportion of rare h^2 is fixed at 0.1, while proportion of rare h_{med}^2 is varied from 0.05 to 0.25. See Supplementary Note for remaining details on this simulation.



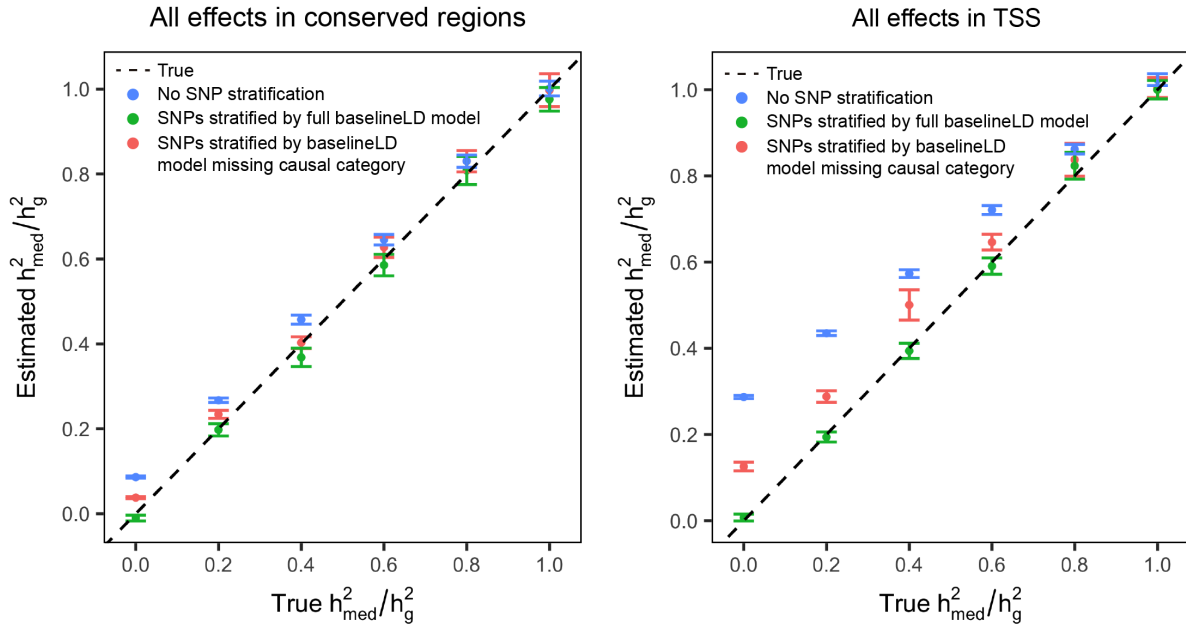
Supplementary Figure 6. Calibration of jackknife standard errors for h_{med}^2 enrichment. We simulated a gene category without h_{med}^2 enrichment (null) and a gene category with 2x h_{med}^2 enrichment (causal). Jackknife standard errors for h_{med}^2 enrichment are compared to empirical standard errors of h_{med}^2 enrichment estimates across 1000 simulations. See Supplementary Note for remaining details on this simulation.



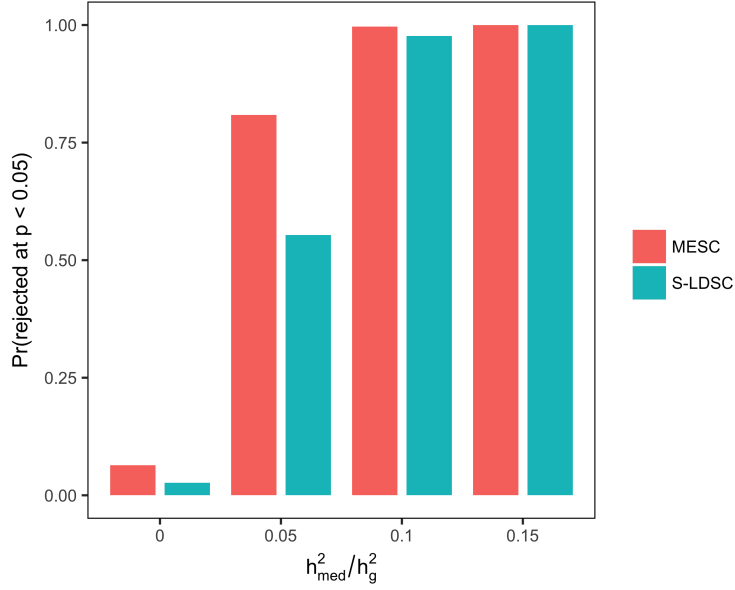
Supplementary Figure 7. Per-bin h_{med}^2/h_g^2 estimates for simulation in Fig 3a. In the legend, the number in parentheses indicates the average expression cis-heritability of genes in a given gene bin.



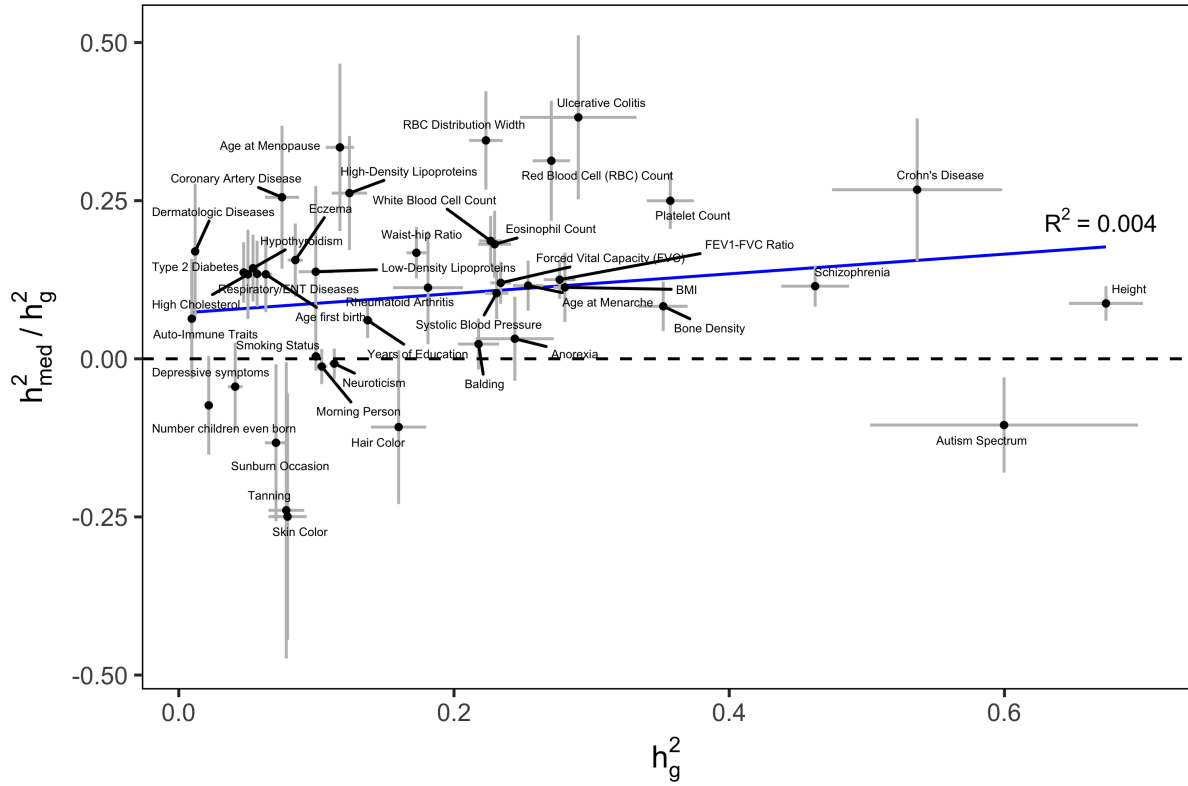
Supplementary Figure 8. h_{med}^2/h_g^2 estimates for simulation involving realistic violation to pleiotropy-eQTL independence. h_{med}^2/h_g^2 was varied from 0 to 0.6. Because the union of the three SNP categories with eQTL effect size enrichment (coding regions, conserved regions, and transcription start sites) comprises around 6% of the genome, the maximum value that h_{med}^2/h_g^2 can be is 0.6 if we have the condition that h_g^2 enrichment of the three SNP categories is 10x. 100 simulations were performed. See Supplementary Note for remaining details on this simulation.



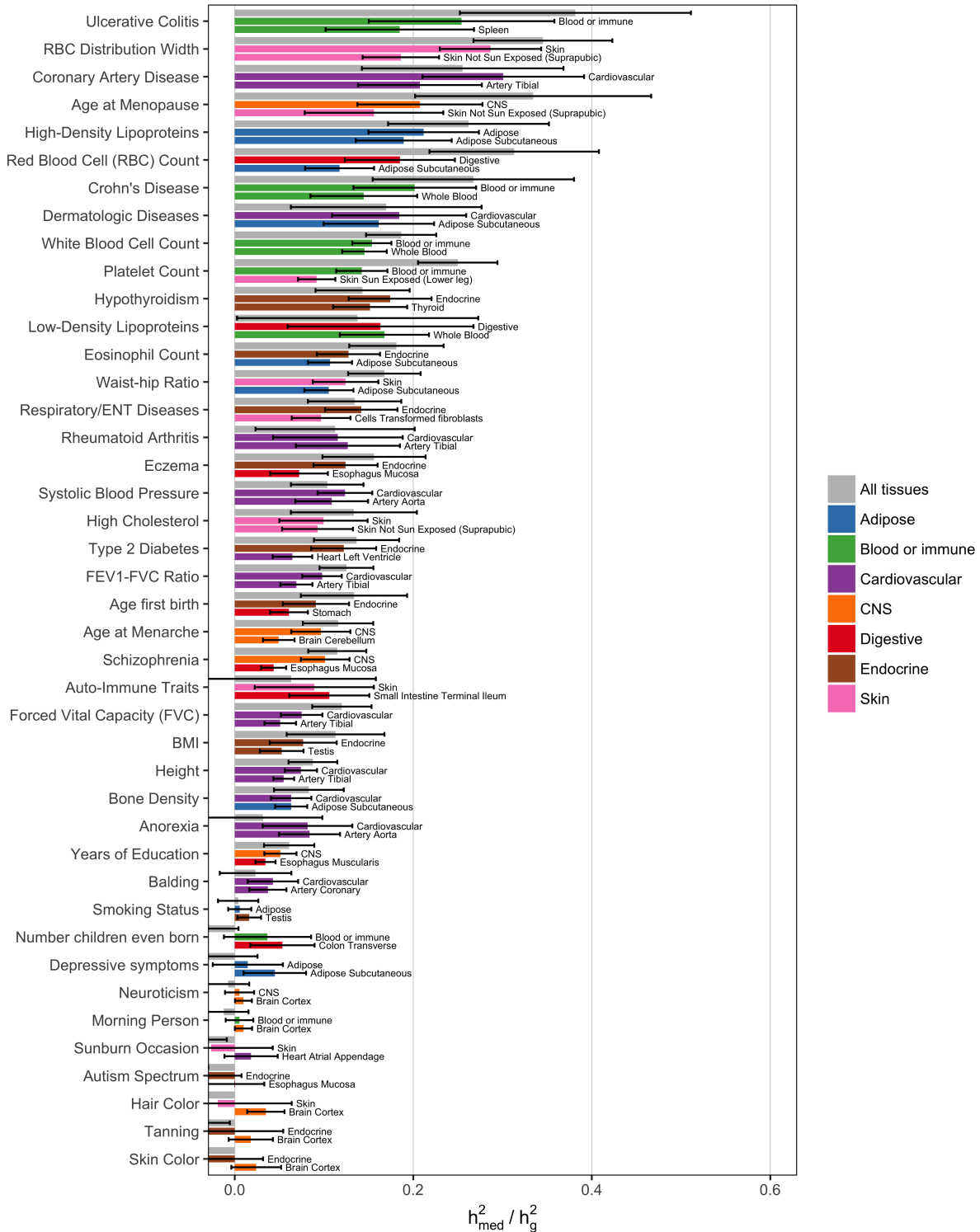
Supplementary Figure 9. Same simulation parameters as Figure 3b, but all non-mediated effects and eQTL effects localized in conserved regions and transcription start sites (TSS) respectively.



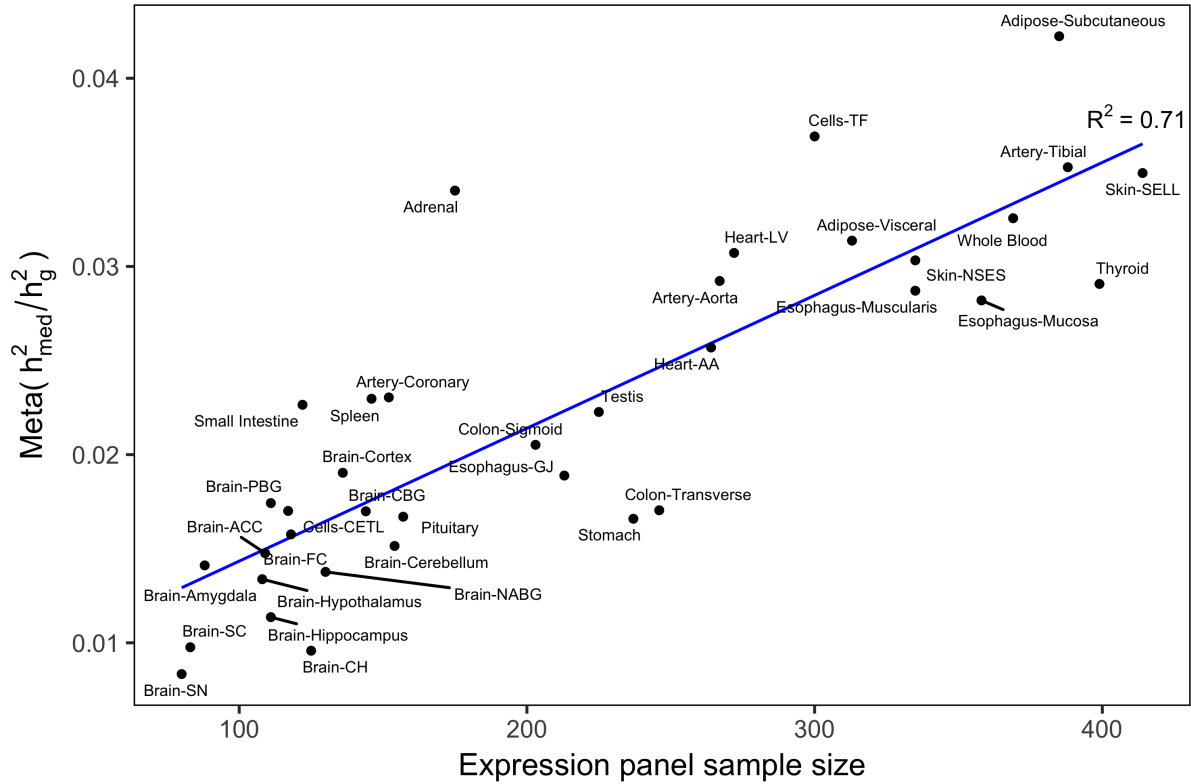
Supplementary Figure 10. Null calibration and power of MESC and stratified LD-score regression given pleiotropy-eQTL effect size independence. For various levels of h_{med}^2/h_g^2 , this figure reports the proportion of simulations which the null hypothesis that $h_{med}^2/h_g^2 = 0$ is rejected by MESC, and the proportion of simulations in which the null hypothesis of no h_g^2 enrichment for the set of all eQTLs is rejected by stratified LD-score regression (S-LDSC). All effect sizes, expression phenotypes, and complex trait phenotypes were simulated in the same manner as Figure 2a. For stratified LD-score regression, we defined the eQTL category as the set of all true eQTLs. 300 simulations were performed.



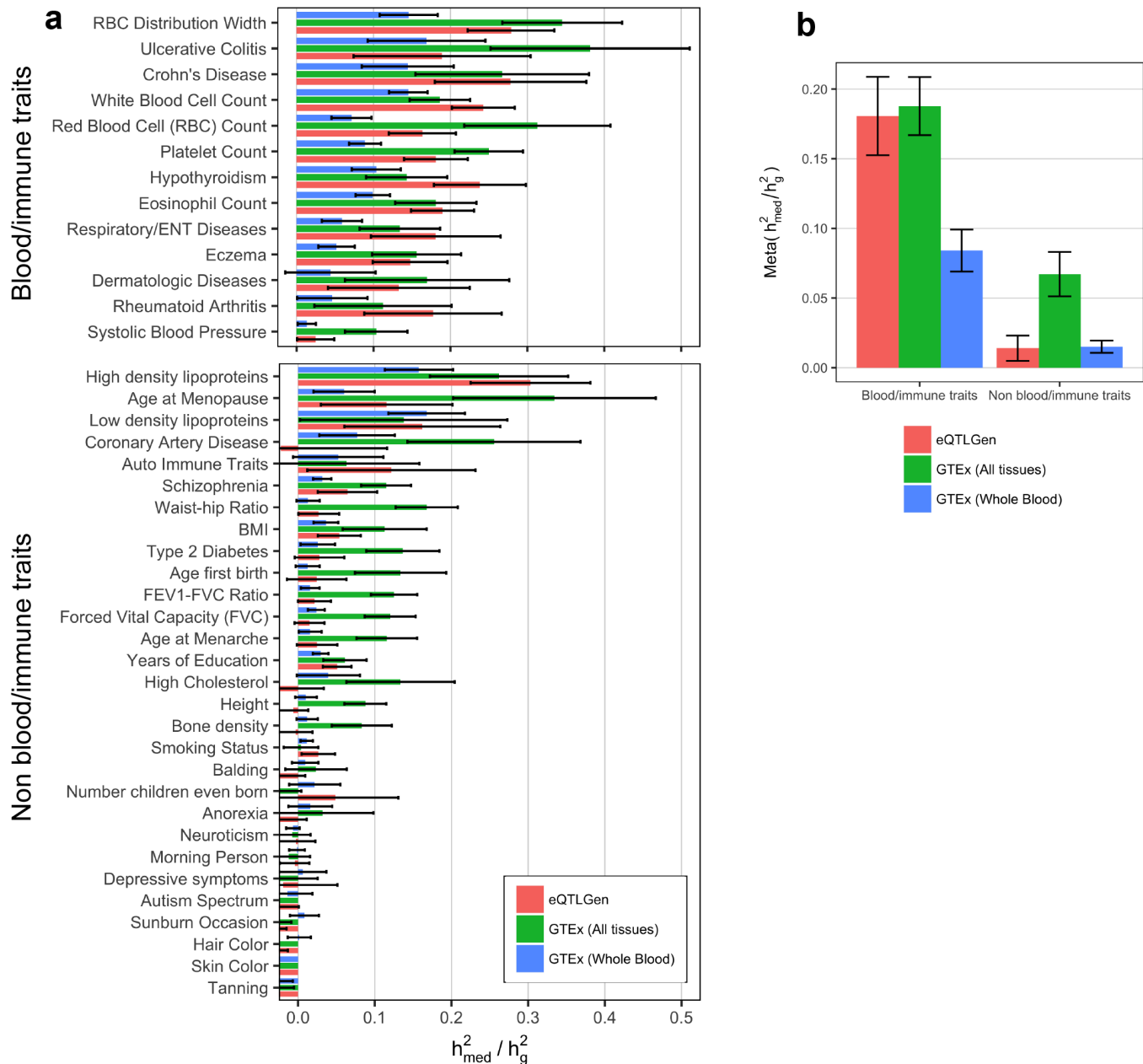
Supplementary Figure 11. Relationship between h^2_{med}/h^2_g and h^2_g . h^2_{med}/h^2_g estimates were obtained using all-tissue meta-analyzed expression scores. h^2_g estimates were obtained using stratified LD-score regression. Error bars represent jackknife standard errors.



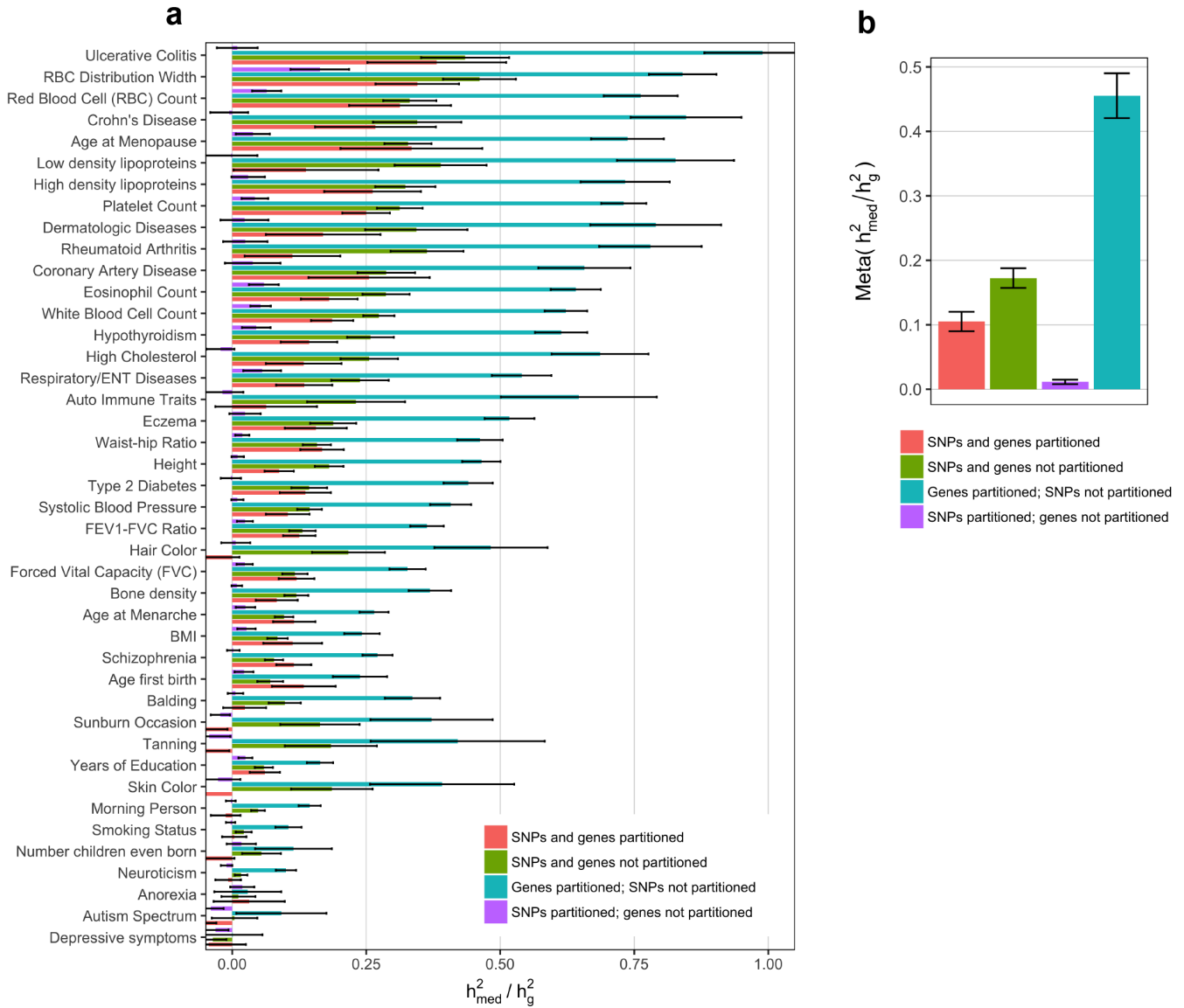
Supplementary Figure 12. Same as Figure 5a, but containing h^2_{med}/h^2_g estimates for all 42 traits from all three types of expression scores: “All tissues” (expression scores meta-analyzed across all 48 GTEx tissues), “Best tissue group” (expression scores meta-analyzed within 7 tissue groups), and “Best tissue” (expression scores computed within individual tissues). Here, “best” refers to the tissue/tissue group resulting in the highest estimates of h^2_{med}/h^2_g compared to all other tissues/tissue groups.



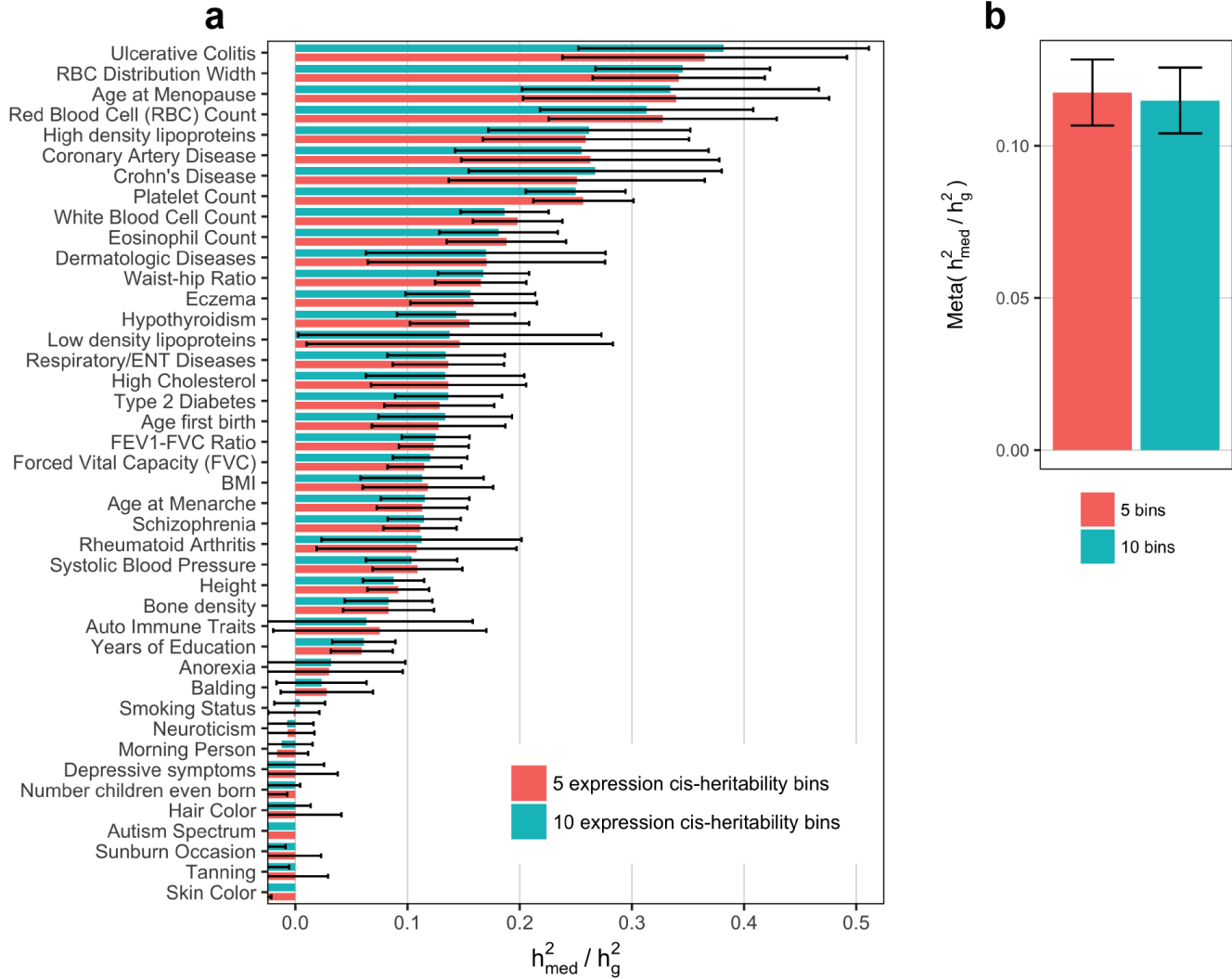
Supplementary Figure 13. Relationship between individual tissue sample size and magnitude of h^2_{med}/h^2_g . h^2_{med}/h^2_g estimates from expression scores estimated in each of 48 individual GTEx tissues were meta-analyzed across 42 complex traits, then plotted against the number of samples in each tissue. We use the following abbreviations: adipose visceral, adipose visceral omentum; brain ACC, brain anterior cingulate cortex BA24; brain CBG, brain caudate basal ganglia; brain CH, brain cerebellar hemisphere; brain FC, brain frontal cortex BA9; brain NABG, brain nucleus accumbens basal ganglia; brain PBG brain putamen basal ganglia; cells CETL, cells EBV-transformed lymphocytes; cells TF, cells transformed fibroblasts; esophagus-GJ, esophagus gastroesophageal junction; heart AA, heart atrial appendage; heart LV, heart left ventricle; skin NSES, skin not sun exposed suprapubic; skin SELL, skin sun exposed lower leg; small intestine, small intestine terminal ileum.



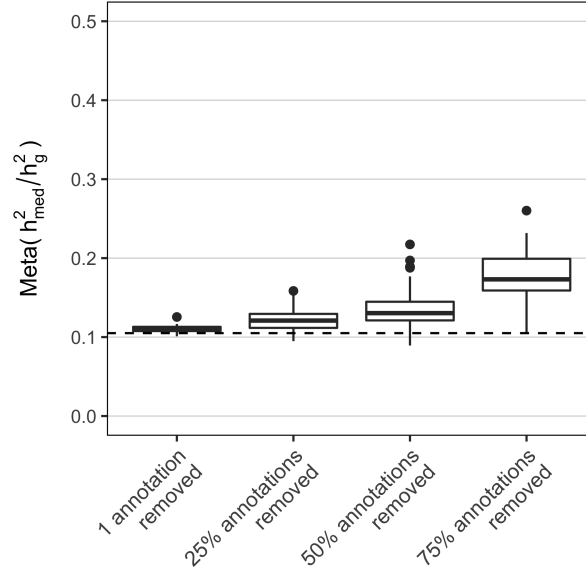
Supplementary Figure 14. h_{med}^2/h_g^2 estimates for 42 diseases and complex traits using data from eQTLGen. We estimated expression scores for all SNPs using cis-eQTL summary statistics from eQTLGen ($N = 31,684$ blood samples), then estimated h_{med}^2/h_g^2 using GWAS summary statistics for the same 42 traits analyzed in the main text. In line with our previous analyses using GTEx data, we stratified expression scores by 5 expression cis-heritability bins. Expression cis-heritability estimates for eQTLGen data were obtained using LD-score regression⁵⁶. For sake of comparison, we also display h_{med}^2/h_g^2 estimates obtained from expression scores from GTEx all-tissue meta-analysis and GTEx whole blood only. (a) h_{med}^2/h_g^2 estimates for 42 individual traits, organized into blood/immune and non-blood/immune traits. (b) Results from a meta-analyzed across traits. Note that low estimates of h_{med}^2/h_g^2 for GTEx whole blood expression scores are caused by the small sample size of the GTEx whole blood data set ($N = 369$).



Supplementary Figure 15. h_{med}^2/h_g^2 estimates without stratifying genes/SNPs. All estimates were obtained using all-tissue meta-analyzed expression scores. (a) We estimated h_{med}^2/h_g^2 for 42 traits without stratifying genes by 5 expression cis-heritability and/or without stratifying SNPs by the baselineLD model. (b) Estimates from a meta-analyzed across all 42 traits.

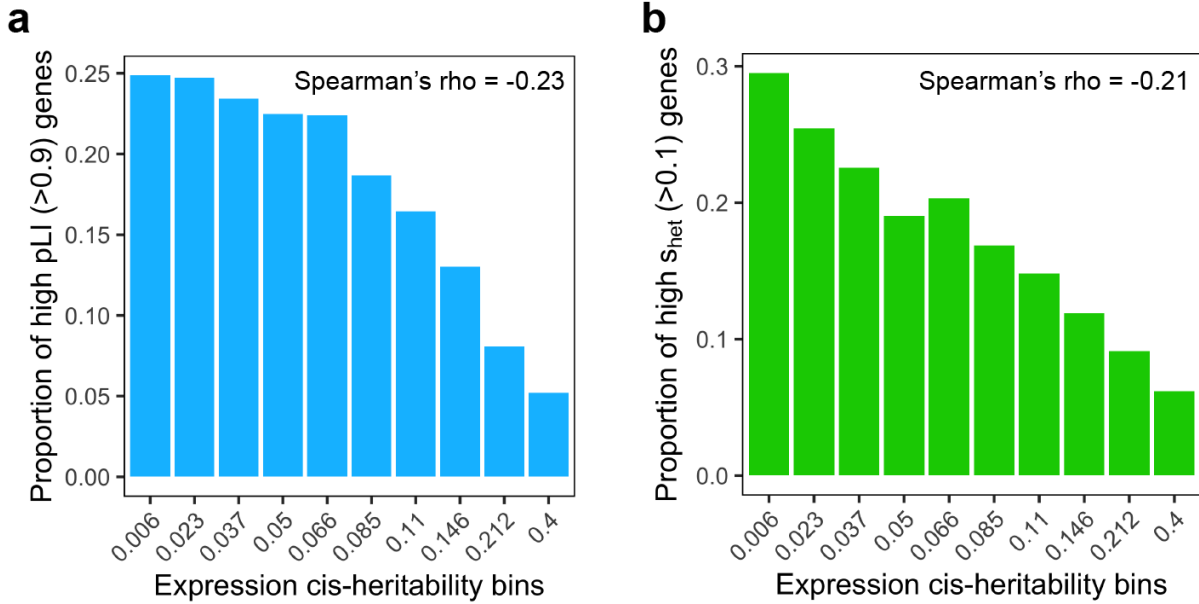


Supplementary Figure 16. h_{med}^2/h_g^2 estimates when varying number of expression cis-heritability bins. All estimates were obtained using all-tissue meta-analyzed expression scores. (a) We estimated h_{med}^2/h_g^2 for 42 traits while stratifying genes by either 5 or 10 expression cis-heritability bins. (b) Estimates from a meta-analyzed across all 42 traits.

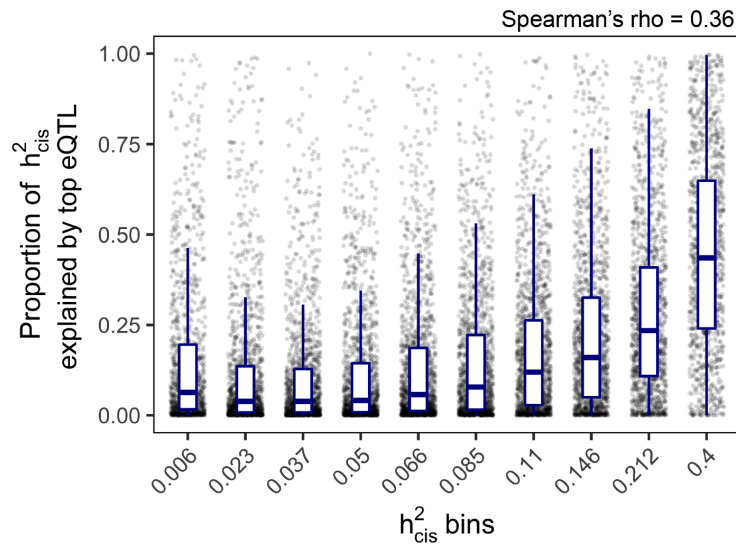


Supplementary Figure 17. h_{med}^2/h_g^2 estimates when removing subsets of the baselineLD model.

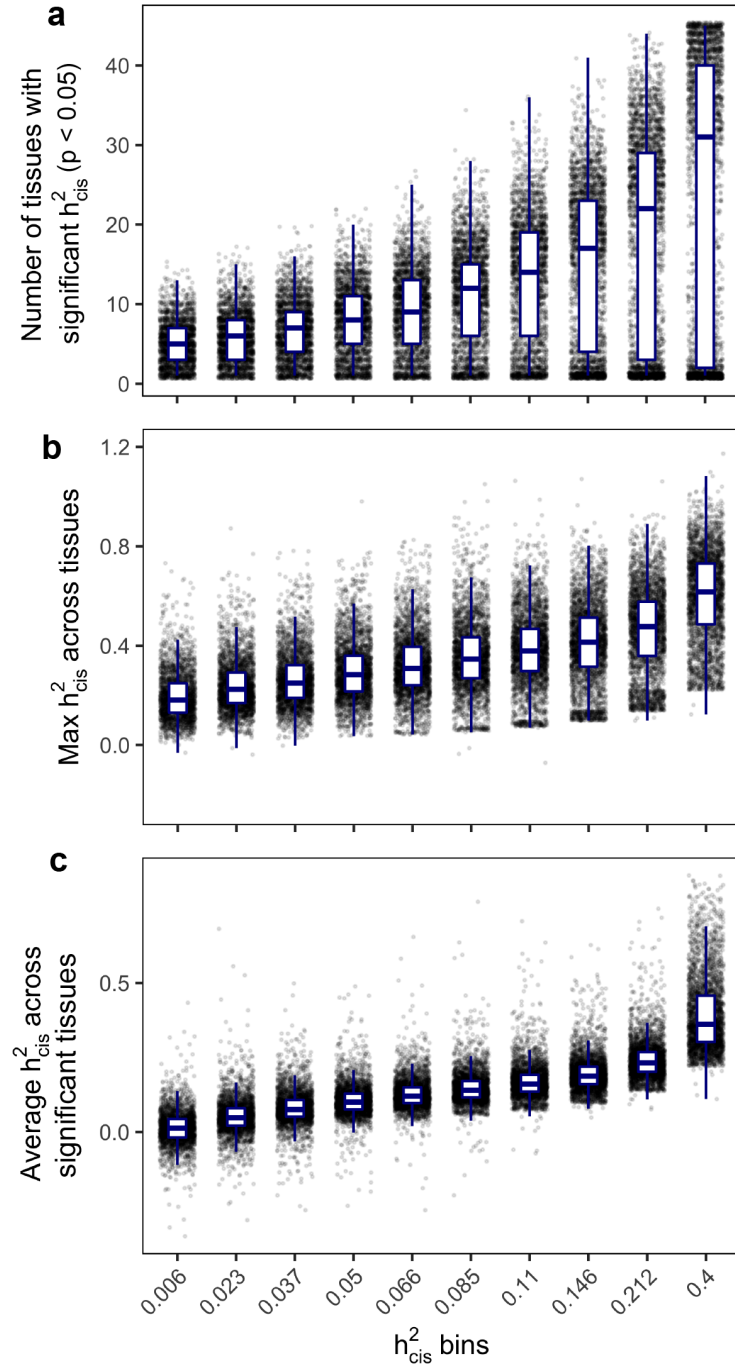
In total, the baselineLD model v2.0 we used in our main analyses contains 72 SNP annotations. We grouped together related baselineLD annotations (typically consisting of a main annotation and the same annotation with 100-500 bp flanking windows), producing the following 29 categories: Coding, Conserved, CTCF, DGF, DHS, Enhancer, Fetal DHS, H3K27ac, H3K4me1, H3K4me3, H3K9ac, Intron, Promoter, Repressed, Super Enhancer, TFBS, Transcribed, TSS, 3' UTR, 5' UTR, Weak Enhancer, GERP, Allele Age, LLD, Recombination Rate, Nucleotide Diversity, Background Selection Statistic, CpG Content, and ASMC. When removing annotations, we remove all related annotations that fall into one of the 29 categories. Each data point represents an h_{med}^2/h_g^2 estimate meta-analyzed over 42 traits and estimated from GTEx all-tissue expression scores. For the boxplot labelled “1 annotation removed,” we show h_{med}^2/h_g^2 estimates when removing each of the 29 individual categories. For the boxplots labelled “25%/50%/75% annotations removed,” we show h_{med}^2/h_g^2 estimates from 100 random subsets of the categories corresponding to the percentage of annotations removed. Dotted line indicates the h_{med}^2/h_g^2 estimate when using the full baselineLD model.



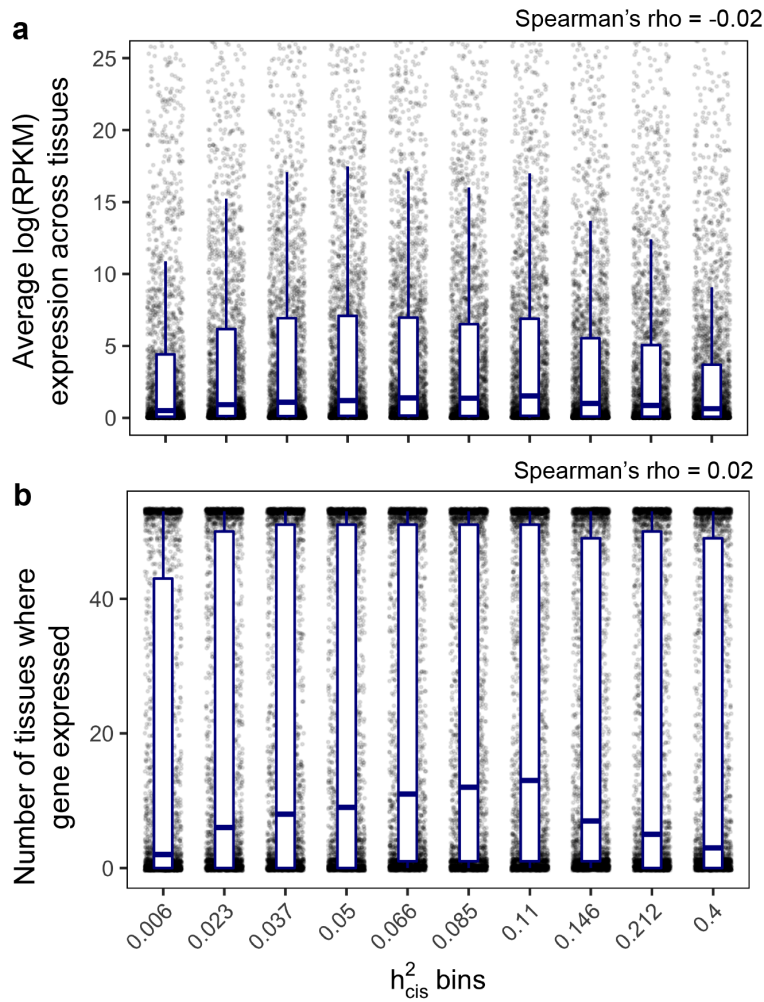
Supplementary Figure 18. Relationship between expression cis-heritability and metrics of gene essentiality. For each gene, pLI (probability of loss-of-function intolerance) was obtained from Lek et al. 2016 Nature and s_{het} (selection against protein-truncating variants) was obtained from Cassa et al. 2017 Nature Genetics.



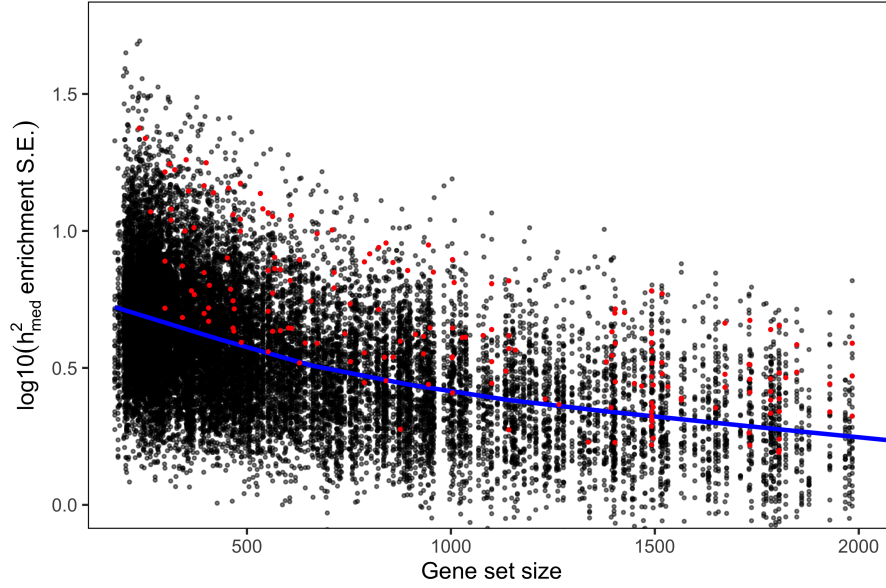
Supplementary Figure 19. Relationship between expression cis-heritability and proportion of expression cis-heritability explained by top cis-eQTL. Top cis-eQTL effect size is meta-analyzed using fixed effects meta-analysis across all 48 GTEx tissues.



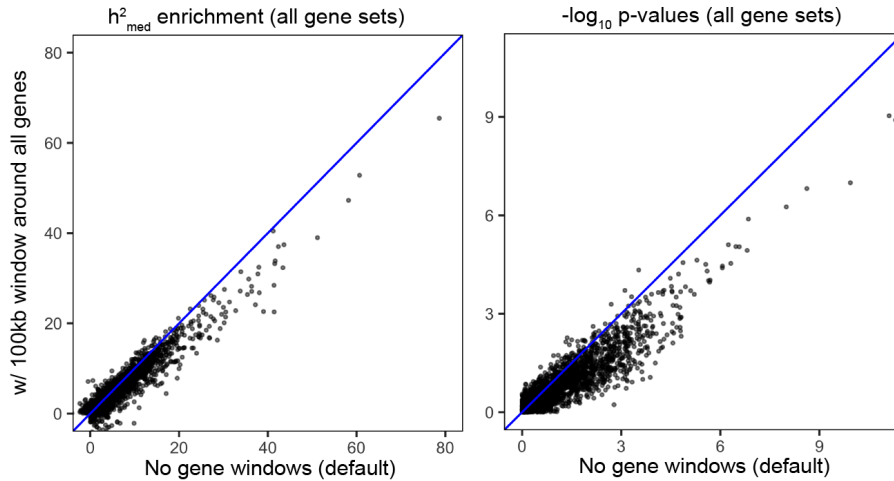
Supplementary Figure 20. Relationship between expression cis-heritability and metrics of tissue specificity. Meta-tissue h^2_{cis} (x-axis) is computed for each gene by averaging h^2_{cis} across all tissues. x-axis labels indicate the average meta-tissue h^2_{cis} of genes within each decile. h^2_{cis} (y-axis) refers to estimates within individual tissues. (a) Relationship between meta-tissue h^2_{cis} deciles and the number of tissues with significantly nonzero h^2_{cis} ($p < 0.05$) for each gene. (b) Relationship between meta-tissue h^2_{cis} ($p < 0.05$) and the max h^2_{cis} across tissues for each gene. (c) Relationship between meta-tissue h^2_{cis} deciles and the average h^2_{cis} across tissues with significantly nonzero h^2_{cis} ($p < 0.05$) for each gene.



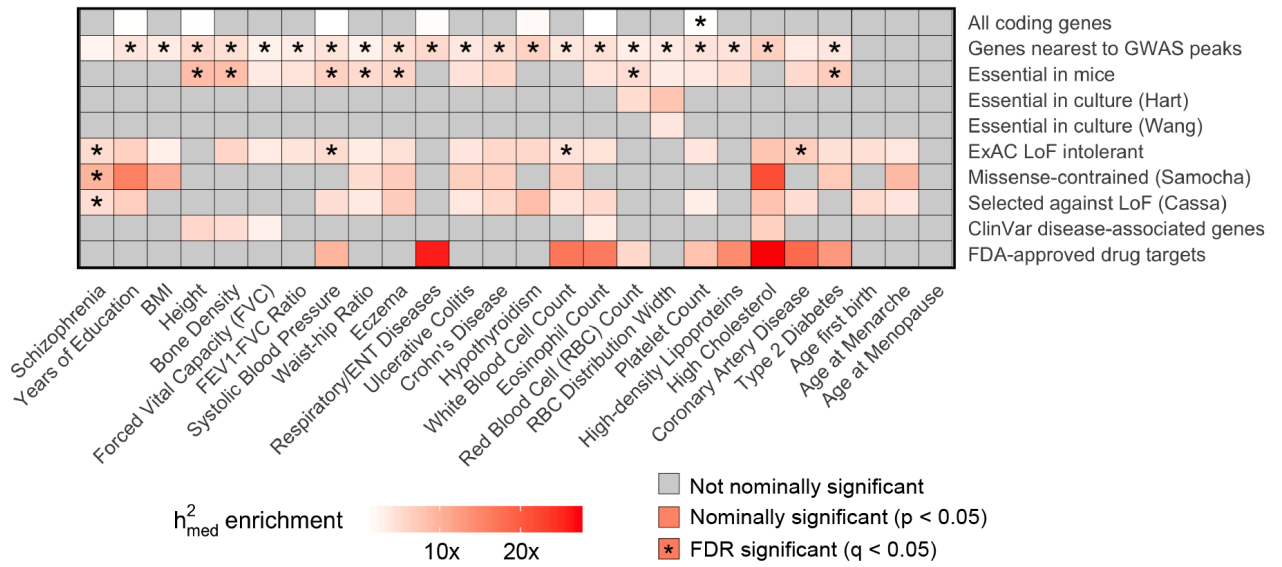
Supplementary Figure 21. Relationship between expression cis-heritability and expression levels. (a) Expression levels represent the median log(RPKM) expression across individuals, which are then averaged across tissues. (b) A gene is expressed in a tissue if RPKM > 0.3 in that tissue.



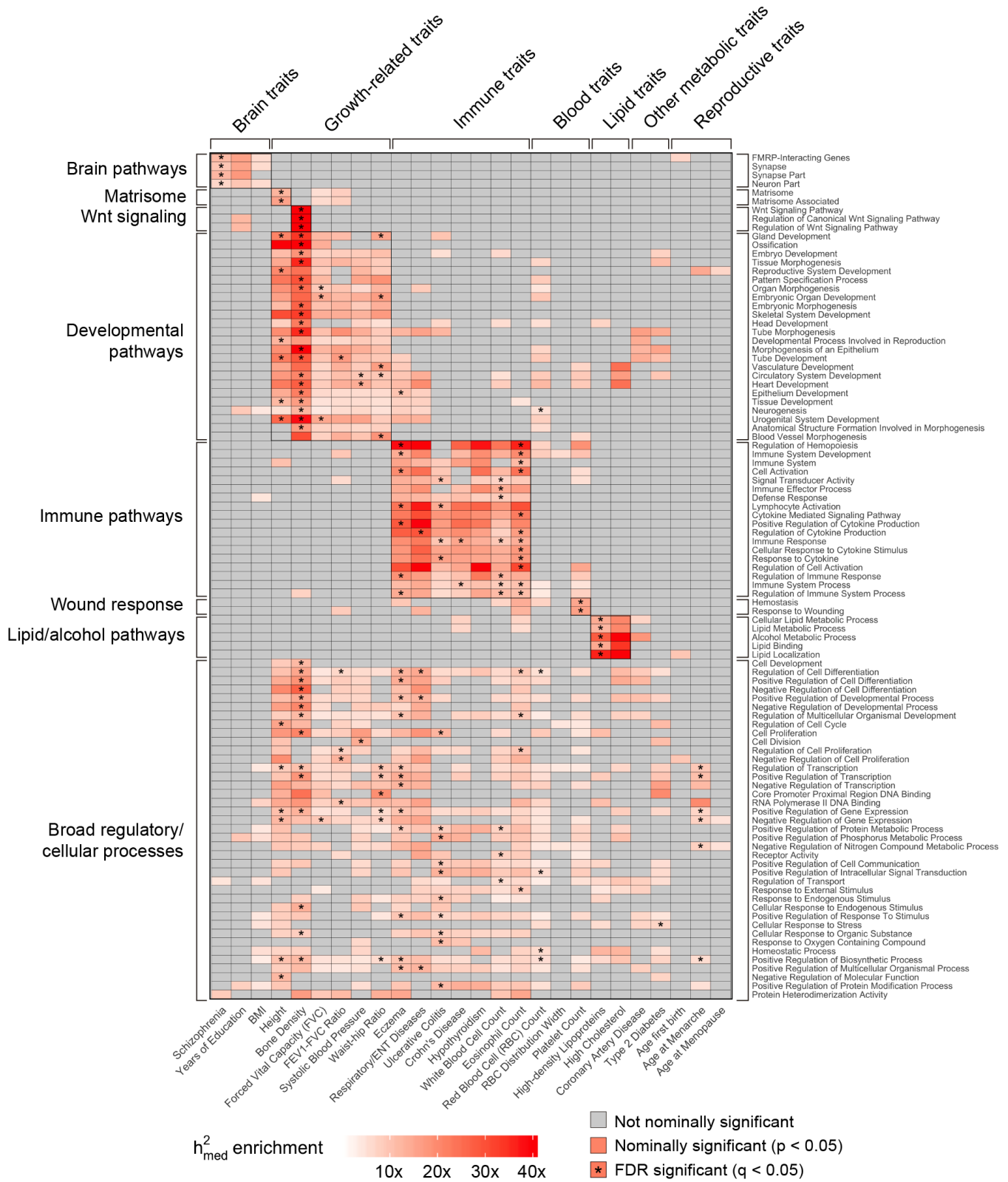
Supplementary Figure 22. Relationship between gene set size and $\log_{10} h_{med}^2$ enrichment standard error. Each point represent a gene set-complex trait pair. Points highlighted in red indicate gene set-complex trait pairs with $FDR < 0.05$ (after accounting for 21,502 hypotheses tested). Blue line indicates the LOESS best fit line.



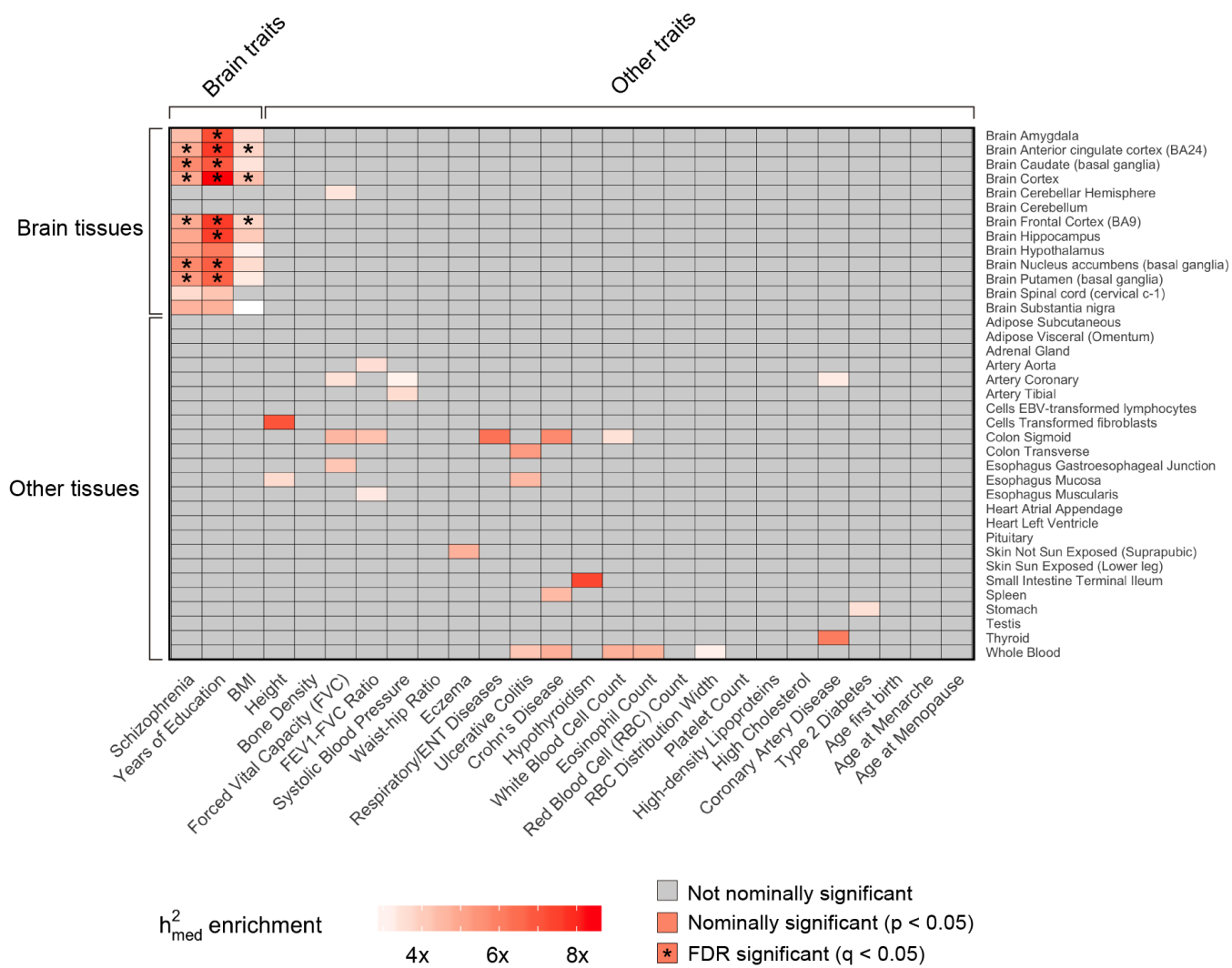
Supplementary Figure 23. h_{med}^2 enrichment estimates with 100 Kb window around genes. (Left) h_{med}^2 enrichment estimates for all 21,502 trait-gene sets pairs analyzed in the main text when including a SNP annotation corresponding to 100 Kb windows around each gene in each gene set. (Right) Same as left, but showing h_{med}^2 enrichment p-values.



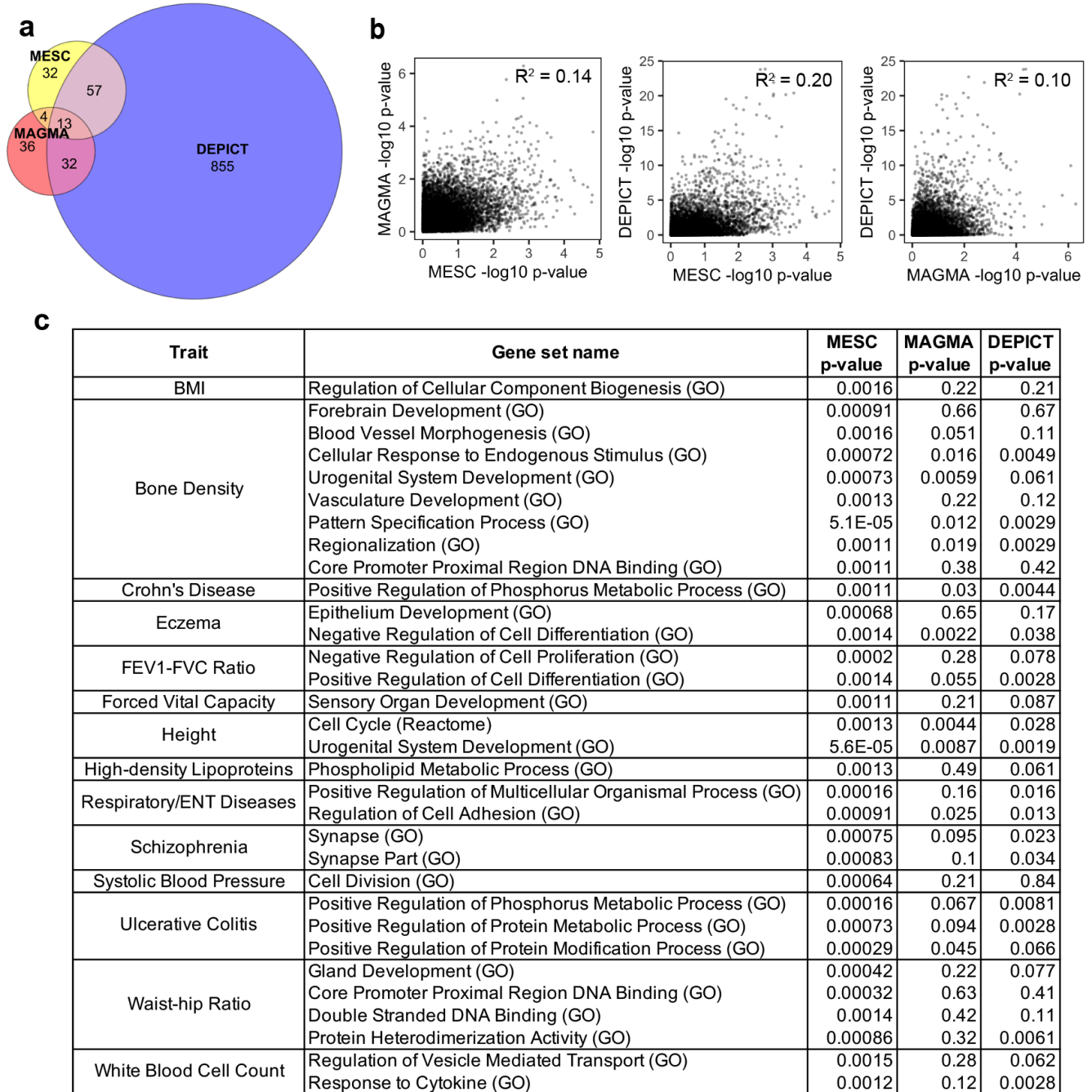
Supplementary Figure 24. h^2_{med} enrichment estimates for all 10 broadly essential gene sets across all 26 complex traits.



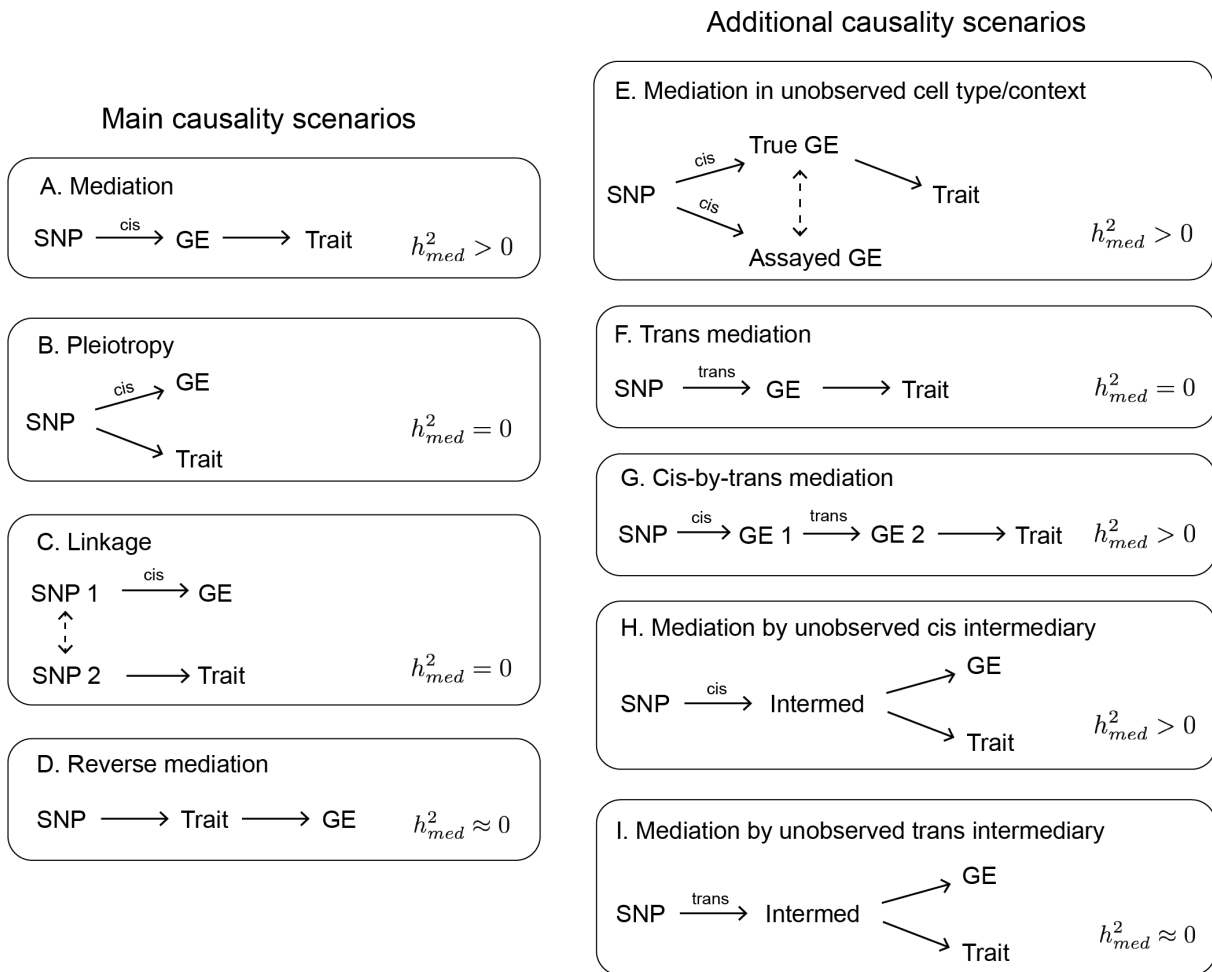
Supplementary Figure 25. h^2_{med} enrichment estimates for 97 pathway-specific gene sets across all 26 complex traits. 97 pathway-specific gene sets represent all pathway-specific gene sets (out of 780 total) with FDR-significant h^2_{med} enrichment in at least one of the 26 complex traits. For sake of display, we grouped together related traits and gene sets.



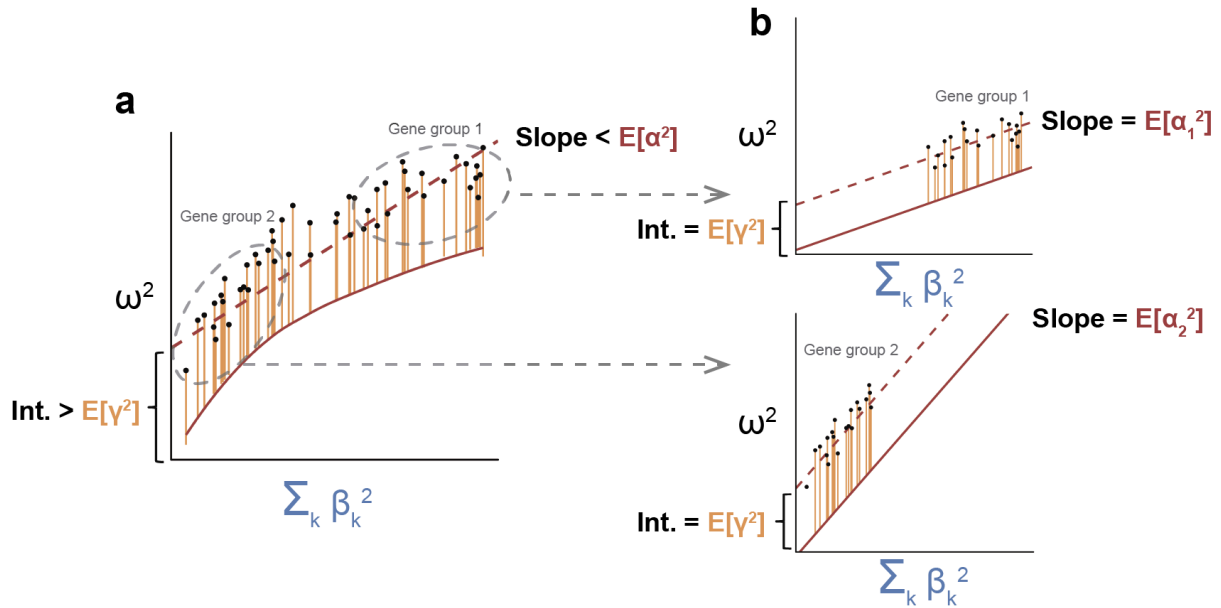
Supplementary Figure 26. Same as Figure 7b, but including h_{med}^2 enrichment estimates for all 26 traits and with individual GTEx tissues labelled.



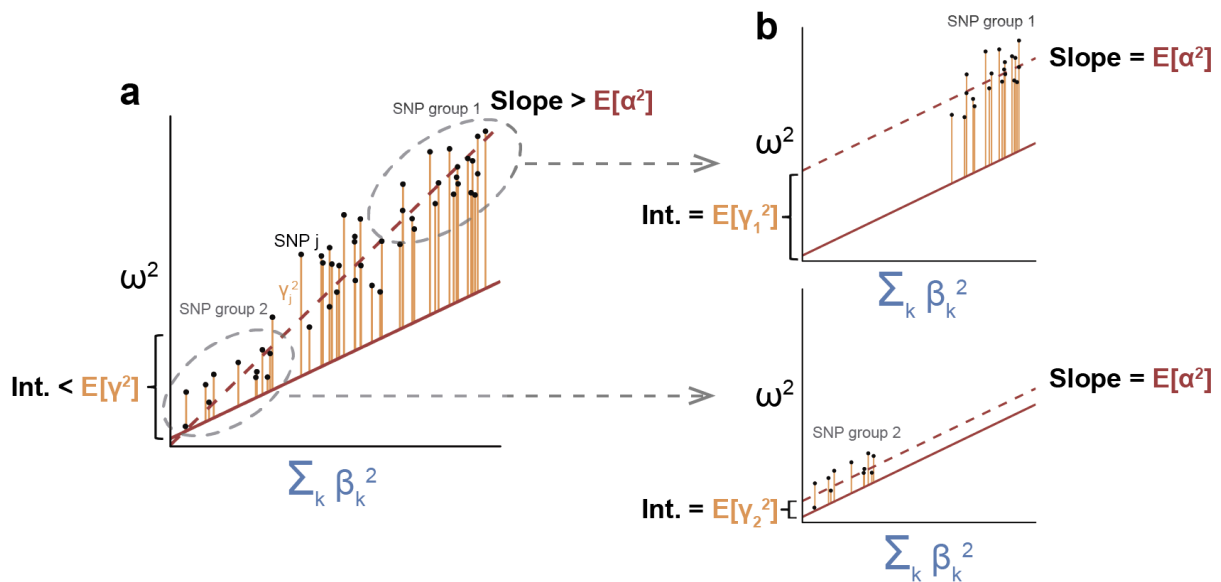
Supplementary Figure 27. Comparison between gene set enrichment estimates from MESC, MAGMA, and DEPICT. (a) Venn diagram showing the overlap between significantly enriched trait-gene set pairs (FDR < 0.05) identified by the three methods. (b) Scatterplots of $-\log_{10}$ enrichment p-values from MESC vs. MAGMA (left), MESC vs. DEPICT (middle), and MAGMA vs. DEPICT (right). Each point represents a trait-gene set pair. (c) List of all 32 gene sets-complex traits pairs detected as significant by MESC (FDR q -value < 0.05) that are not detected as significant by MAGMA or DEPICT. See Supplementary Table 9 for enrichment estimates for all gene set-complex traits pairs.



Supplementary Figure 28. Modes of expression causality. See “Modes of expression causality” for a description of each scenario and its contribution to estimates of h_{med}^2 .



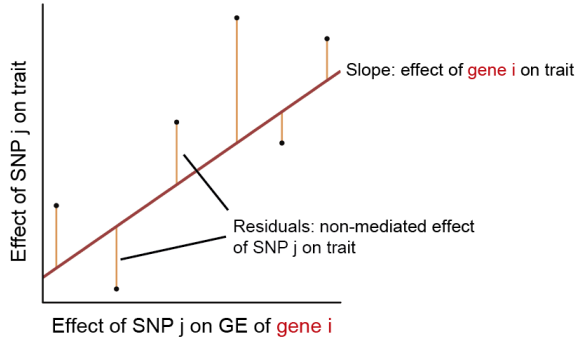
Supplementary Figure 29. Illustration of impact of violations to gene-eQTL effect size independence on estimates of $E[\alpha^2]$. See “Model assumptions” for context for this figure. In the figure, we depict a scenario where the magnitude of α is negatively correlated with the magnitude of β . (a) If we perform the regression using all genes, the slope from the regression will be downwardly biased relative to the true $E[\alpha^2]$. (b) If we stratify the regression across genes by the magnitude of their expression cis-heritability, we can obtain approximately unbiased estimates of $E[\alpha_D^2]$ for each gene category D .



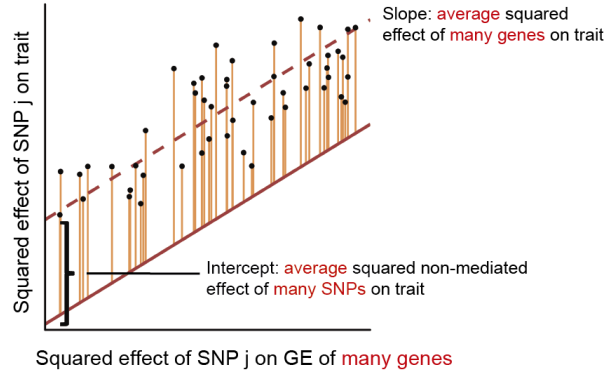
Supplementary Figure 30. Illustration of impact of violations to pleiotropy-eQTL effect size independence on estimates of $E[\alpha^2]$. See “Model assumptions” for context for this figure. In the figure, we depict a scenario where the magnitude of γ is positively correlated with the magnitude of β . (a) If we perform the regression using all SNPs, the slope from the regression will be upwardly biased relative to the true $E[\alpha^2]$. (b) If we stratify the regression across SNPs by the magnitude of their eQTL effect sizes, we can obtain approximately unbiased estimates of $E[\alpha_D^2]$ for each SNP category D .

can obtain an approximately unbiased estimate of $E[\alpha^2]$.

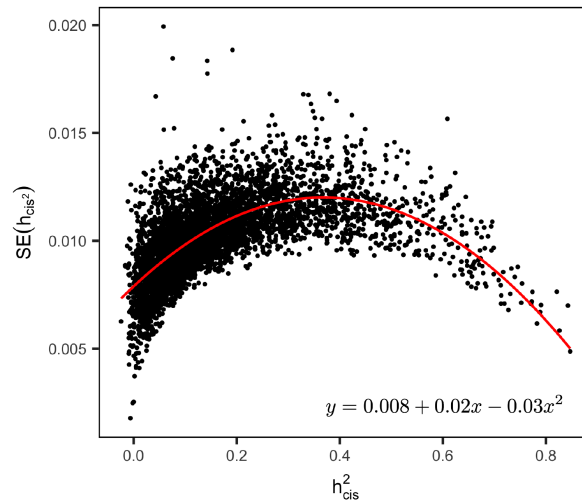
a Mendelian randomization with multiple genetic instruments



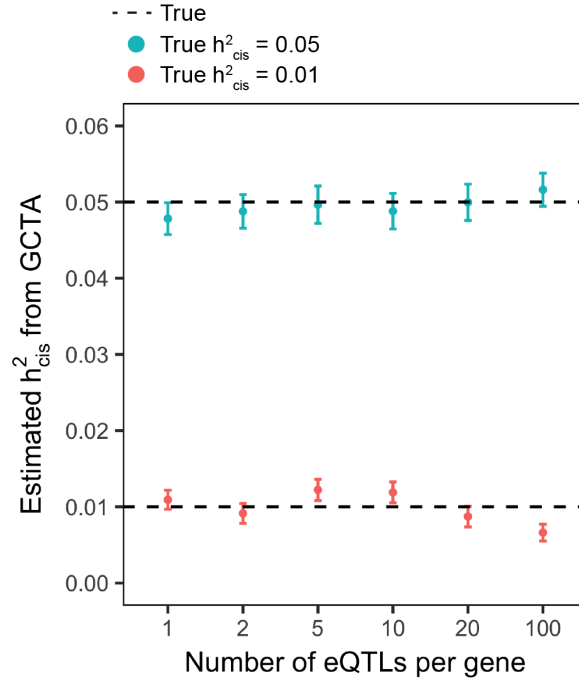
b Mediated expression score regression



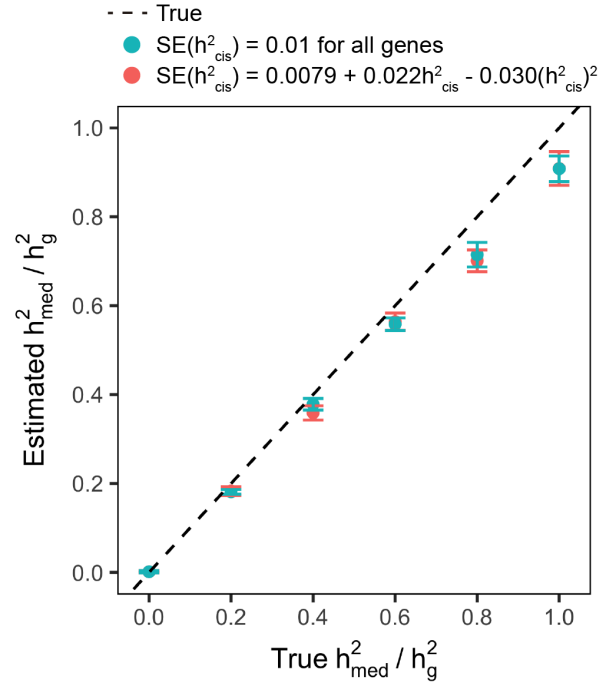
Supplementary Figure 31. Comparison between Mendelian randomization with multiple genetic variants and MESC.



Supplementary Figure 32. Standard errors of expression cis-heritability estimates across GTEx tissues. Expression cis-heritability estimates are obtained for each gene in each of 48 GTEx tissues using GCTA. Standard error represents the mean standard error of expression cis-heritability estimates for each gene across all 48 tissues. Red line denotes the best quadratic fit.



Supplementary Figure 33. GCTA estimates of heritability under simulated cis-eQTL genetic architectures. We simulated cis-eQTLs for a gene by selecting x random SNPs within a random 1 Mb window on chromosome 1 (restricting to Hapmap3 SNPs). One cis-eQTL was randomly selected to be the lead eQTL with effect size drawn from $\mathcal{N}(0, 0.8h^2_{cis})$. The remaining $x - 1$ cis-eQTLs had effect sizes drawn from $\mathcal{N}(0, 0.2h^2_{cis}/(x - 1))$. Expression phenotypes for the gene were simulated for 1000 individuals (genotypes randomly selected from UK Biobank) with environmental noise drawn from $\mathcal{N}(0, 1 - h^2_{cis})$. GCTA was used to predict h^2_{cis} from the expression phenotypes and genotypes within the 1 Mb window. Error bars represent mean standard errors across 100 simulations.



Supplementary Figure 34. h_{med}^2/h_g^2 estimates with varying $SE(h_{cis}^2)$ as a function of h_{cis}^2 . Simulation was performed in the same manner as in Figure 2a (with expression panel size fixed at 1000). Standard error for h_{cis}^2 was simulated from either $\mathcal{N}(0, 0.01^2)$ (consistent the mean of empirical $SE(h_{cis}^2)$ estimates across all GTEx samples) or from $\mathcal{N}(0, (0.0079 + 0.22h_{cis}^2 - 0.03(h_{cis}^2)^2)^2)$ (consistent with the best quadratic fit line relating empirical h_{cis}^2 to $SE(h_{cis}^2)$ estimates across all GTEx samples, see Supplementary Figure 32). Error bars represent mean standard errors across 100 simulations.