Supplementary Information for "Quantifying genetic effects on disease mediated by assayed gene expression levels"

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

Contents

1 Supplementary Note

1.1 Derivations

1.1.1 Unstratified MESC with non-causal eQTL effect sizes

In this section, we show that when we carry out the regression procedure described in "Unstratified MESC" (Methods) using eQTL effect sizes assayed in non-causal tissues T, we obtain an estimate of the quantity $h^2_{med;assayed}(T)$ as defined in "Definition of expression-mediated heritability" (Methods).

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. We start with regression equation (2) from Methods:

$$E[\omega_k^2] = E[\alpha^2]\sum_i^G \beta_{ik}^2 + E[\gamma^2]$$

The ordinary least squares estimate of the coefficient from regressing ω^2 on $\sum_i^G \beta'^2_i$ is

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

The third line follows given the $Cov(\gamma^2, \beta'^2) = 0$ and $Cov(\alpha^2, \beta'^2) = 0$. Let $r_g^2(T) = \frac{1}{G}\sum_i^G \frac{Cov(\beta_i^2, \beta'^2_i)}{\sqrt{Var(\beta_i^2)Var(\beta'^2_i)}}$ represent the average squared genetic correlation between expression in T vs. in causal cell types. Given this definition, we have

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

For simplicity, we make the assumption that $Var(\beta_i^2) \approx Var(\beta'^2_i)$ across genes. Note that violations to this assumption can realistically occur in practice but will not bias our estimate of $h^2_{med;assayed}(T)$ (see below). Given this assumption, we have

$$\alpha'^2 \approx r_g^2(T)E[\alpha^2]$$

We can then multiply α'^2 by $GE[h'^2_{cis}]$ to obtain an unbiased estimate of $h^2_{med;assayed}(T)$, where $E[h'^2_{cis}]$ is the average expression cis-heritability of genes in T.

Estimates of $h^2_{med;assayed}(T)$ when $Var(\beta_i^2) \neq Var(\beta_i'^2)$. There are reasonable scenarios in which $Var(\beta_i^2)$ can be systemically larger or smaller than $Var(\beta'^2_i)$ 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 $GE[h'^2_{cis}]$ to obtain $h^2_{med;assayed}$(T), we will still have $h^2_{med;assayed}(T) = r_g^2(T)h^2_{med;causal}$.

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. (1) thus becomes

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

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

$$h^2_{med;assayed}(T) = GE[\beta'^2]\frac{1}{c}r_g^2(T)E[\alpha^2]$$
$$= GE[c\beta^2]\frac{1}{c}r_g^2(T)E[\alpha^2]$$
$$= r_g^2(T)h^2_{med;causal}$$

1.1.2 Unstratified MESC with summary statistics

We have previously derived an estimator for h^2_{med} in the idealized scenario that SNP effect sizes are given (Methods). In practice, we use GWAS summary statistics, which are affected by sampling noise and by LD. It has previously been shown that LD and sampling noise can be accounted for by regressing GWAS χ^2 statistics on LD scores, which measure the total of LD for each SNP[1,2]. Under our generative model (see equation (1) in Methods), the marginal OLS estimate of the total effect size of a SNP k on the trait is given by

$$\hat{\omega}_k = \frac{1}{N}\boldsymbol{X_k}^T\boldsymbol{y}$$
$$= \frac{1}{N}(\boldsymbol{X_k}^T(\boldsymbol{X\gamma} + \boldsymbol{XB\alpha} + \boldsymbol{\epsilon}))$$
$$= \frac{1}{N}\boldsymbol{X_k}^T\boldsymbol{X\gamma} + \frac{1}{N}\boldsymbol{X_k}^T\boldsymbol{XB\alpha} + \frac{1}{N}\boldsymbol{X_k}^T\boldsymbol{\epsilon}$$
$$= \sum_j^M \gamma_j\hat{r}_{jk} + \sum_i^G \alpha_i \sum_j^M \hat{r}_{jk}\beta_{ij} + \boldsymbol{\epsilon'}$$

Let $\hat{\boldsymbol{R}} = \frac{1}{N}\boldsymbol{X}^T\boldsymbol{X}$ denote the in-sample LD matrix, and let $\boldsymbol{\epsilon'} = \frac{1}{N}\boldsymbol{X}^T\boldsymbol{\epsilon}$ denote the noise term in the summary statistics. The χ^2 statistic for SNP k (defined as $N\hat{\omega}_k^2$) is:

$$E[\chi_k^2 \mid \hat{\boldsymbol{R}}, \boldsymbol{B}] = N\sum_j^M E[\gamma_j^2 \mid \hat{\boldsymbol{R}}, \boldsymbol{B}]\hat{r}_{jk}^2 + N\sum_i^G E[\alpha_i^2 \mid \hat{\boldsymbol{R}}, \boldsymbol{B}]\sum_j^M \hat{r}_{jk}^2\beta_{ij}^2 + NE[(\boldsymbol{\epsilon'})^2] \quad (2)$$
$$= NE[\gamma^2]\sum_j^M \hat{r}_{jk}^2 + NE[\alpha^2]\sum_i^G\sum_j^M \hat{r}_{jk}^2\beta_{ij}^2 + NE[(\boldsymbol{\epsilon'})^2] \quad (3)$$
$$= \frac{Nh^2_{nonmed;causal}}{M}\sum_j^M \hat{r}_{jk}^2 + \frac{Nh^2_{med;causal}}{GE[h^2_{cis}]}\sum_i^G\sum_j^M \hat{r}_{jk}^2\beta_{ij}^2 + 1 - h^2_{med;causal} - h^2_{nonmed;causal} \quad (4)$$

In order for the equation regarding the unconditional expectation of χ_k^2 to hold true (3), we must make two independence assumptions involving LD-dependent genetic architecture, in addition to the independence assumptions described in "Model assumptions" (Methods). LD-dependent architecture, if not accounted for, is known to produce bias in heritability estimates[3,4]. These assumptions are:

- Across all genes (indexed by i), the magnitude of α_i is uncorrelated with the LD scores of eQTLs for gene i

- Across all SNPs (indexed by k), the magnitude of γ_k is uncorrelated with the LD score of SNP k

(4) follows (3) from our definitions of $h^2_{med;causal}$ and $h^2_{nonmed;causal}$. Since $E[\hat{r}^2_{jk}] \approx r^2_{jk} + \frac{1}{N}$, we have

$$E\left[\sum_j^M \hat{r}^2_{jk}\right] \approx \sum_j^M r^2_{jk} + \frac{M}{N}$$

and

$$\begin{aligned} E\left[\sum_i^G \sum_j^M \hat{r}^2_{jk}\beta^2_{ij}\right] &\approx \sum_i^G \sum_j^M \left(r^2_{jk}\beta^2_{ij} + \frac{\beta^2_{ij}}{N}\right) \\ &\approx \sum_i^G \sum_j^M r^2_{jk}\beta^2_{ij} + \frac{GE[h^2_{cis}]}{N} \end{aligned}$$

Thus,

$$\begin{aligned} E[\chi^2_k] &\approx \frac{Nh^2_{nonmed;causal}}{M}\left(\sum_j^M r^2_{jk} + \frac{M}{N}\right) + \frac{Nh^2_{med;causal}}{GE[h^2_{cis}]}\left(\sum_i^G \sum_j^M r^2_{jk}\beta^2_{ij} + \frac{GE[h^2_{cis}]}{N}\right) + 1 - h^2_{nonmed;causal} - h^2_{med;causal} \\ &\approx \frac{Nh^2_{nonmed;causal}}{M}\sum_j^M r^2_{jk} + \frac{Nh^2_{med;causal}}{GE[h^2_{cis}]}\sum_i^G \sum_j^M r^2_{jk}\beta^2_{ij} + 1 \end{aligned}$$

Defining LD scores $\ell_k = \sum_j^M r^2_{jk}$ and expression scores $\mathcal{L}_k = \sum_i^G \sum_j^M r^2_{jk}\beta^2_{ij}$, we arrive at our main equation for summary MESC regression:

$$E[\chi^2_k] \approx \frac{Nh^2_{nonmed;causal}}{M}\ell_k + \frac{Nh^2_{med;causal}}{GE[h^2_{cis}]}\mathcal{L}_k + 1 \tag{5}$$

Analogous to the derivation in "Unstratified MESC with non-causal eQTL effect sizes" (see above), we show that if we perform this regression using expression scores in assayed tissues T rather than expression scores in causal cell types/contexts, we will estimate $h^2_{med;assayed}(T)$ rather than $h^2_{med;causal}$ (see below).

1.1.3 Unstratified MESC with non-causal expression scores

In this section, we show that when we carry out the regression procedure described above using expression scores in assayed tissues T rather than in causal cell types/contexts, we obtain an estimate of $h^2_{med;assayed}(T)$.

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 (5)), we have

$$\begin{aligned} \alpha'^2 &\approx \frac{1}{G}\sum_i^G \frac{Cov(\chi^2, \sum_j^M r^2_j \beta'^2_{ij})}{Var(\sum_j^M r^2_j \beta'^2_{ij})} \\ &\approx \frac{1}{G}\sum_i^G \frac{Cov(\sum_j^M r^2_j \alpha^2_i \beta^2_{ij}, \sum_j^M r^2_j \beta'^2_{ij})}{Var(\sum_j^M r^2_j \beta'^2_{ij})} \\ &\approx E[\alpha^2]\frac{1}{G}\sum_i^G \frac{\ell^2 Cov(\beta^2_i, \beta'^2_i)}{\ell^2 Var(\beta'^2_i)} \\ &\approx E[\alpha^2]\frac{1}{G}\sum_i^G \frac{Cov(\beta^2_i, \beta'^2_i)}{Var(\beta'^2_i)} \end{aligned}$$

Here, $\ell^2 = Var(\sum_j^M r^2_j)$. The third and fourth line follow given that r^2 is independent of α, β, and β'. See "Unstratified MESC with non-causal eQTL effect sizes" (above) for the remainder of the derivation.

1.1.4 Stratified MESC

Starting from equation (2):

$$E[\chi_k^2 \mid \hat{\boldsymbol{R}}, \boldsymbol{B}] = N \sum_j^M E[\gamma_j^2 \mid \hat{\boldsymbol{R}}, \boldsymbol{B}] \hat{r}_{jk}^2 + N \sum_i^G E[\alpha_i^2 \mid \hat{\boldsymbol{R}}, \boldsymbol{B}] \sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + NE[(\boldsymbol{\epsilon}')^2] \tag{6}$$

$$= N \sum_j^M \left(\sum_{c: j \in \mathcal{C}_c} \tau_c \mid \hat{\boldsymbol{R}}, \boldsymbol{B} \right) \hat{r}_{jk}^2 + N \sum_i^G \left(\sum_{d: i \in \mathcal{D}_d} \pi_d \mid \hat{\boldsymbol{R}}, \boldsymbol{B} \right) \sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + NE[(\boldsymbol{\epsilon}')^2] \tag{7}$$

$$E[\chi_k^2] = N \sum_c \tau_c \sum_{j \in \mathcal{C}_c} \hat{r}_{jk}^2 + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M \hat{r}_{jk}^2 \beta_{ij}^2 + NE[(\boldsymbol{\epsilon}')^2] \tag{8}$$

In order for (8) 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

$$E[\chi_k^2] = N \sum_c \tau_c \sum_{j \in C_c} \left(r_{jk}^2 + \frac{1}{N} \right) + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M \left(r_{jk}^2 \beta_{ij}^2 + \frac{\beta_{ij}^2}{N} \right) + NE[(\boldsymbol{\epsilon}')^2]$$

$$= N \sum_c \tau_c \sum_{j \in C_c} r_{jk}^2 + \sum_c \sum_{j \in C_c} \tau_c + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2 +$$

$$\sum_d \sum_{i \in \mathcal{D}_d} \sum_j^M \left(\pi_d |\mathcal{D}_d| E[h_{cis}^2(\mathcal{D}_d)] \right) + NE[(\boldsymbol{\epsilon}')^2]$$

$$= N \sum_c \tau_c \sum_{j \in C_c} r_{jk}^2 + h_{nonmed;causal}^2 + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2 + h_{med;causal}^2 + 1 - h_{nonmed;causal}^2 - h_{med;causal}^2$$

$$= N \sum_c \tau_c \sum_{j \in C_c} r_{jk}^2 + N \sum_d \pi_d \sum_{i \in \mathcal{D}_d} \sum_j^M r_{jk}^2 \beta_{ij}^2 + 1$$

Letting $\ell_{k;c} = \sum_{j \in C_c} r_{jk}^2$ 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:

$$E[\chi_k^2] = N \sum_c \tau_c \ell_{k;c} + N \sum_d \pi_d \mathcal{L}_{k;d} + 1 \tag{9}$$

1.1.5 Estimating expression scores from eQTL summary statistics

We can use summary statistics from eQTL studies to estimate expression scores $\mathcal{L}_{k;d}$, which are equivalent to the sum of marginal OLS estimates of eQTL effect sizes for SNP k on genes in $\mathcal{D}_d$ ($\sum_{i \in \mathcal{D}_d} \hat{\beta}_{ik(sumstat)}^2$) 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:

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

where $\boldsymbol{y}_{i(exp)}$ is an N_{exp}-vector of gene expression measurements (standardized to mean 0 and variance 1), $\boldsymbol{X}$ 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 $\boldsymbol{\epsilon}_{i(exp)}$ is an N_{exp}-vector of environmental effects. Under this model, we have

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

Thus, we can use the following alternate form of equation (9) to perform regression:

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

1.1.6 Impact of environmental noise in gene expression measurements on h^2_{med} 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^2_{cis}) does not impact our estimates of expression-mediated heritability h^2_{med}. In other words, h^2_{med} depends on only the genetic component of gene expression levels. One consequence of this fact is that the magnitude of h^2_{med} does not *a priori* depend on the magnitude of h^2_{cis}. For example, mean h^2_{cis} 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^2_{med} from this gene expression data set can in principle be very high (i.e. close to total SNP heritability h^2_g).

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_i 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 $\frac{\text{std(expr)}}{\text{std(geno)}}$), while α_i is expressed in terms of *additive increase in standardized phenotype per unit increase of standardized expression levels* (which we abbreviate as $\frac{\text{std(pheno)}}{\text{std(expr)}}$). When we multiply α^2_i by $\sum_i \beta^2_{ij}$, we obtain a quantity corresponding to the heritability mediated by gene expression levels for SNP j in units of $\left(\frac{\text{std(expr)}}{\text{std(geno)}}\right)^2\left(\frac{\text{std(pheno)}}{\text{std(expr)}}\right)^2 = \left(\frac{\text{std(pheno)}}{\text{std(geno)}}\right)^2$.

Because std(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^2_j on $\sum_i \beta^2_{ij}$, we obtain an estimate of $E[\alpha^2]$ in units of $\left(\frac{\text{std(pheno)}}{\text{std(expr)}}\right)^2$. To obtain an estimate of per-SNP expression-mediated heritability, we then multiply $E[\alpha^2]$ by $E\left[\sum_i \beta^2_{ij}\right]$ (or equivalently $E[h^2_{cis}]$), which is in units of $\left(\frac{\text{std(expr)}}{\text{std(geno)}}\right)^2$. Suppose we were to scale $\sum_i \beta^2_{ij}$ by an arbitrary factor c, in which case our estimate of $E[\alpha^2]$

would be in units of $\left(\frac{\text{std(pheno)}}{c\cdot\text{std(expr)}}\right)^2$ and $E\left[c\sum_i \beta_{ij}^2\right]$ would be in units of $\left(\frac{c\cdot\text{std(expr)}}{\text{std(geno)}}\right)^2$. When multiplying $E[\alpha^2]$ by $E\left[c\sum_i \beta_{ij}^2\right]$, the product would be expressed in units of $\left(\frac{c\cdot\text{std(expr)}}{\text{std(geno)}}\right)^2\left(\frac{\text{std(pheno)}}{c\cdot\text{std(expr)}}\right)^2 = \left(\frac{\text{std(pheno)}}{\text{std(geno)}}\right)^2$ and would thus be unchanged compared to before. This is essentially the same argument as made in the section "Estimates of $h^2_{med;assayed}(T)$ when $Var(\beta_i^2) \neq Var(\beta_i'^2)$" (see above), 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^2_{cis} by a constant factor, which we have shown above does not affect estimates of h^2_{med}. 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:

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

where $\boldsymbol{y}_{i(exp)}$ is a vector of *non-standardized* gene expression levels for gene i, $\boldsymbol{X}$ 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 $\sum \beta^2_{i(std)} = h^2_{cis}$), 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^2_{cis} and $\beta^2_{i(std)}$ 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^2_{cis} and $\beta^2_{i(std)}$ 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^2_{cis} and $\beta^2_{i(std)}$.

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^2_{med}.

1.1.7 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^2_{med}. In this section, we show that h^2_{med} estimates produced by MESC are bounded by the average genetic prediction r^2 of gene expression multiplied by true h^2_{med}, so $r^2 < 1$ results in downward bias in estimated h^2_{med} (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-eQTLs[5]. 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[6–8]. 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 (eQTLGen[7], N = 31,684) is only 0.026, which is far too low to yield meaningful estimates of h^2_{med}. 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^2_{med}. 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 $\sum_i^G \hat{\beta}'^2_i$, the estimate of the coefficient $\hat{\alpha}'^2$ is:

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

The first line follows from the same derivation as "Unstratified MESC with non-causal eQTL effect sizes." 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 $Var(\beta_i^2) \approx Var(\beta'^2_i)$. To illustrate this, see that the numerators of both sides of the inequality are equivalent: $Cov(\hat{\beta}_i'^2, \beta_i'^2) = Cov(\beta'^2_i + \epsilon_i^2, \beta_i'^2) = Cov(\beta'^2_i, \beta_i'^2) = Var(\beta'^2_i) \approx Var(\beta_i^2)$ 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 $Var(\hat{\beta}_i'^2) \geq Var(\beta'^2_i)$. 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 $h^2_{med;assayed}(T)$ that has the following property:

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

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)[9,10]:

$$r^2_{pred;trans}(T) = \frac{1}{G}\sum_i^G \frac{h^2_{i;trans}}{h^2_{i;trans} + \frac{M}{N}(1 - r^2_{pred;trans}(T))}$$
$$\approx \frac{N}{M}\frac{1}{G}\sum_i^G h^2_{i;trans}$$

where $h^2_{i;trans}$ is the expression trans-heritability of gene i, M is the effective number of independent SNPs (approximately 60,000[11]), and N is expression panel sample size. The largest expression panel available to date is from eQTLGen[7], with N = 31,684 in blood. The average h^2_{trans} of expression is around 0.05[5,8,12]. Thus, we can expect $r^2_{pred;trans}(T)$ 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^2_{med}. In order to obtain $r^2_{pred;trans}(T)$ 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.

1.1.8 Relationship between MESC and stratified LD score regression

MESC is similar in form to stratified LD score regression (S-LDSC), which aims to estimate total heritability partitioned across SNP categories from summary statistics[2,13]. In particular, the τ coefficient estimated by S-LDSC is directly related to the π coefficient we obtain from MESC in equation (9). 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 MESC, $Var(\omega_k)$ 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 (9) 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 (9) 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 annotations[2,13–18], 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 "Violations of effect size independence assumptions" below). 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.

1.1.9 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[19,20]. 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 expression[6]. 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 holds[20] (Note that the InSIDE assumption is essentially the same as the pleiotropy-eQTL independence assumption we describe in "Model assumptions" in Methods). However, this approach is highly underpowered in the common scenario that genes have only a few detectable cis-eQTLs[19], 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 12 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" in Methods). We can then calculate h^2_{med} with our average gene-trait effect estimate using an equivalent definition of h^2_{med} that models β and α as random variables (See "Definition of expression-mediated heritability" in Methods). Thus, MESC can still reliably estimate h^2_{med} 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.

1.2 Modes of expression causality

In Supplementary Figure 8, 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^2_{med}.

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^2_{med}.

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" in Methods), the presence of pleiotropy will not contribute to estimates of h^2_{med}.

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" in Methods), the presence of linkage will not contribute to estimates of h^2_{med}.

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^2_{med}, 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 below for justification).

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^2_{med} using assayed expression levels will be nonzero if the assayed expression levels are correlated with expression levels in causal cell types/contexts (see "Definition of expression-mediated heritability" in Methods).

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^2_{med}. 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 trans[8,21,22]. 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^2_{med}. Because this scenario does not strictly speaking involve mediation through the gene's expression levels, its contribution to h^2_{med} might be viewed as spurious.

However, there are several reasons why scenario H's potential contribution to h^2_{med} is likely not of major concern. First, if the intermediary represents the expression levels of another gene, then scenario H's contribution to h^2_{med} 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^2_{med}. 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^2_{med} is still of biological interest due to the fact that the gene's expression levels are correlated with a truly causal intermediary.

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^2_{med} will be much smaller than the contribution of scenario H (see "Impact of reverse mediation on h^2_{med} estimates" below for related intuition).

1.2.1 Impact of reverse mediation on h^2_{med} estimates

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^2_{med}. 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[5,23]. 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^2_{med} 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 sizes[23], which is true for polygenic traits. To illustrate this point, it is useful to think of h^2_{med} 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):

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

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

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

A typical trait has $h^2_{med} = 0.1$ and $Var(\beta_i) = 0.05$, so a realistic value for α_i^2 is $\frac{h^2_{med}}{GVar(\beta_i)} = \frac{0.1}{20000 \cdot 0.05} = 0.0001$ (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 $Cov(\beta_i, \omega_i)^2_{med} \approx 0.0001 \cdot 0.05^2 = 2.5 \times 10^{-7}$. 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 $Cov(\beta_i, \omega_i)^2_{revmed}$:

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

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

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

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 $Cov(\beta_i, \omega_i)^2_{revmed} \leq 0.05/0.5 \cdot (2.5 \times 10^{-5})^2 = 6.25 \times 10^{-11}$. The reason why $Cov(\beta_i, \omega_i)^2_{revmed}$ is orders of magnitude smaller than $Cov(\beta_i, \omega_i)^2_{med}$ is that $Cov(\beta_i, \omega_i)^2_{revmed}$ involves the product of the *square* of the squared per-SNP effect on complex trait and the squared SNP effect on gene expression, while $Cov(\beta_i, \omega_i)^2_{med}$ 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.

1.3 Violations of effect size independence assumptions

In this section, we describe realistic scenarios in which the two main effect size independence assumptions (see "Model assumptions" in Methods) might be violated, and we describe how conditioning on SNP- and gene-level annotations can ameliorate any resulting bias.

1.3.1 Gene-eQTL effect size independence

Gene-eQTL effect size independence is violated in the scenario that less heritable genes have larger causal effect sizes on the trait, which is supported by evidence that evolutionarily constrained genes tend to have fewer eQTLs[6]. A negative correlation between the magnitude of gene effect sizes and eQTL effect sizes across the genome will result in downwardly biased estimates of $E[\alpha^2]$ and upwardly biased estimates of $E[\gamma^2]$, as illustrated in Supplementary Figure 13a. The downward bias in h^2_{med} arises due to the fact that during regression, SNPs with larger eQTL effect sizes are implicitly weighted more than SNPs with smaller eQTL effect sizes, so the average slope will be biased toward the value of α^2 for high heritability genes.

In order to account for violations to gene-eQTL effect size independence, we can stratify genes by the magnitude of their cis-heritability so that within each gene category, gene-eQTL effect size independence approximately holds. We can then obtain unbiased estimates of $E[\alpha_D^2]$ for each gene category D as illustrated in Supplementary Figure 13b. In practice, we stratify genes by 5 bins according to their cis-heritability, which we show adequately captures genome-wide dependence of gene effect sizes on eQTL effect sizes in simulations (Figure 2d).

1.3.2 Pleiotropy-eQTL effect size independence

Pleiotropy-eQTL effect size independence is violated in the presence of regulatory hotspots with high biological activity in the genome[2,24–27], resulting in an increased number and/or magnitude of both eQTLs and pleiotropic/linkage effects in these hotspots. A positive correlation between the magnitude of non-mediated effect sizes and eQTL effect sizes across the genome will result in upwardly biased estimates of $E[\alpha^2]$ and downwardly biased estimates of $E[\gamma^2]$, as illustrated in Supplementary Figure 14a.

In order to account for violations to pleiotropy-eQTL effect size independence, we can stratify SNPs by the magnitude of their eQTL effect sizes so that within each SNP category, pleiotropy-eQTL effect size independence approximately holds. We can then obtain unbiased estimates of overall $E[\alpha^2]$ and $E[\gamma_C^2]$ for each SNP category C as illustrated in Supplementary Figure 14b. Note that in practice, LD between SNPs makes it is difficult or impossible to identify the exact SNPs that act as eQTLs. This in turn makes it impractical to stratify SNPs according to eQTL effect size. Therefore, we instead stratify SNPs by a set of comprehensive functional SNP annotations, the baselineLD model[2,13], which should capture most known regulatory hotspots in the genome and act as a reasonable proxy to eQTL effect sizes (see Figure 2e for simulation results)

1.4 Rare vs. common variant h^2_{med}

In all our analyses, we restrict the regression SNPs used by MESC to only Hapmap3 SNPs[28]. 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 ($h^2_{med(common)}/h^2_{common}$). We define $h^2_{med(common)}$ as $\sum_{j\in C}\sum_i \beta^2_{ij}\alpha^2_i$ (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 ($h^2_{med(common)} + h^2_{med(rare)}$), where $h^2_{med(rare)}$ is defined in the same manner as $h^2_{med(common)}$ but C is replaced with the set of all SNPs with MAF < 0.05. We do not aim to estimate $h^2_{med(rare)}$ 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 $h^2_{med(rare)}$ (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 $(h^2_{med(common)} + h^2_{med(rare)})/(h^2_{common} + h^2_{rare})$ to be either similar or smaller than the quantity $h^2_{med(common)}/h^2_{common}$ that we estimate:

1. Most SNP heritability is explained by common variants[29,30]. Thus, we can expect the quantity $(h^2_{med(common)} + h^2_{med(rare)})/(h^2_{common} + h^2_{rare})$ to depend mostly on $h^2_{med(common)}$ and h^2_{common}.
2. Rare variant heritability has a much larger enrichment in coding regions than common variant heritability[31], 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^2_{med}, so we would expect that $h^2_{med(rare)}/h^2_{rare} < h^2_{med(common)}/h^2_{common}$.

Role of singletons in h^2_{med}. A recent study[32] 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 $h^2_{med(common)}$ 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 study[33] 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.

1.5 Additional simulation results

1.5.1 Additional details on simulations under frequency-dependent genetic architectures

We sought to assess the bias of MESC in estimating $h^2_{med(common)}$ (see "Rare vs. common variant h^2_{med}" above for definition) in the presence of frequency-dependent genetic architectures (including rare and low-frequency variants). To this end, 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[32,34] and complex trait[29,30]. We conducted our simulations using real genotypes

imputed to include rare and low-frequency variants from UK Biobank[35] (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 $h^2_{med(common)}/h^2_{common} = 0.2$ 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^2_g estimates if uncorrected[3,4]), we stratified regression SNPs by 10 MAF bins, which has been shown to adequately account for this dependence[4,13]. In all simulations, we obtained unbiased/slightly conservative estimates of $h^2_{med(common)}/h^2_{common}$ 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 3). Thus, MESC is robust to frequency-dependent genetic architectures for both gene expression and disease.

1.5.2 Additional details on simulations under violations of effect size independence assumptions

We sought to assess the bias of MESC in estimating h^2_{med} under violations of gene-eQTL independence and pleiotropy-eQTL independence. Correlations between eQTL effect size magnitude and gene effect size magnitude or non-mediated effect size magnitude will result in biases in h^2_{med} estimates in the same direction of the correlation. In other words, if eQTL effect size magnitude is *positively* correlated with gene effect size magnitude and/or non-mediated effect size magnitude, h^2_{med} estimates will be *upwardly* biased; whereas if eQTL effect size magnitude is *negatively* correlated with gene effect size magnitude and/or non-mediated effect size magnitude, h^2_{med} estimates will be *downwardly* biased. Given our biological knowledge, we would expect in practice that eQTL effect size magnitude is *negatively* correlated with gene effect size magnitude due to negative selection removing large-effect eQTLs for large-effect genes, while we would expect that eQTL effect size magnitude is *positively* correlated with non-mediated effect size magnitude due to the presence of regulatory hotspots in the genome. Thus, we impose these specific correlations in our simulations.

Violation of pleiotropy-eQTL independence. To simulate violations of pleiotropy-eQTL independence, we initially attempted to emulate realistic levels of overlap between expression heritability enrichment and disease heritability enrichment. We selected three functional SNP categories from the baselineLD model[2,13]—transcription start sites, coding regions, and conserved regions—that have been shown to be highly enriched for both expression cis-heritability[22] and complex trait heritability[2]. We then simulated 10x expression cis-heritability enrichment and 10x total heritability enrichment for SNPs in any of these three categories, which is close to empirical estimates[2,22]. However, in this particular simulation, we observed no discernible upward bias in our estimates of h^2_{med}/h^2_g (Supplementary Figure 15), demonstrating that stronger violations to pleiotropy-eQTL effect size independence were necessary to induce bias in estimates of h^2_{med}/h^2_g.

We only observed upward bias in h^2_{med}/h^2_g when simulating more extreme patterns of colocalization between eQTL effects and non-mediated effects, i.e. in which 100% of eQTLs and disease heritability were entirely concentrated in one of transcription start sites, coding regions, or conserved regions respectively. We estimated h^2_{med}/h^2_g for each simulation in two ways: (1) with SNPs stratified by the full baselineLD model, and (2) with SNPs stratified by a misspecified baselineLD model missing the causal category and window around the causal category, emulating a realistic scenario in which the SNP model does not fully capture all sources of non-mediated effects. The results from these simulations are reported in Figure 2e and Supplementary Figure 6. We note these particular simulations are not meant to be realistic, but rather serve to illustrate the flexibility of the baselineLD model in correcting for biases under even extreme violations of pleiotropy-eQTL independence.

1.5.3 Additional details on simulations comparing MESC to stratified LD score regression

We sought to compare the performance of MESC to that of stratified LD score regression (S-LDSC). Although the two methods do not estimate the same quantity, h^2_{med} (as estimated by MESC) shares some similarities with the h^2_g enrichment of the SNP category corresponding to the set of eQTLs (as estimated by S-LDSC) in that $h^2_{med} > 0$ implies h^2_g enrichment > 1. The converse is true only when pleiotropy-eQTL independence is satisfied, but we expect that pleiotropy-eQTL independence is almost certainly violated in practice.

First, we show that S-LDSC indeed has a well-calibrated false positive rate for detecting heritability of the eQTL category when $h^2_{med} = 0$, with increasing power to detect heritability enrichment of the eQTL category as h^2_{med} increased (Supplementary Figure 16).

However, in the presence of violations to pleiotropy-eQTL effect size independence, S-LDSC will detect significant heritability enrichment of the eQTL category even in the absence of mediation. Like the simulation performed in Supplementary Figure 15, we simulated 10x enrichment of eQTL effect sizes in three SNP categories (coding regions, transcription start sites, and conserved regions). With h^2_{med} fixed at 0, we then varied the heritability enrichment of the three SNP categories from 2.5x to 10x. We observed that S-LDSC detected significant heritability enrichment of the eQTL category in the absence of mediation when total heritability enrichment was 5x or greater, whereas MESC had a well-calibrated false positive rate at all levels of enrichment (Figure 2f). Note that this result does not imply that S-LDSC or other heritability partitioning methods are flawed, but rather that they cannot specifically distinguish mediated effects from non-mediated effects when they are applied to annotations generated from eQTL data. We did not compare MESC to the colocalization methods of Chun et al.[36] or Ongen et al.[37], since these methods only operate on genome-wide significant GWAS loci and also do not attempt to distinguish pleiotropy from mediation.

1.5.4 Simulations 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. Note that this set of simulations does not involve 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.* $\hat{\mathcal{L}}_k = \sum_i^G \hat{\beta}^2_{ik(sumstat)}$, where $\hat{\beta}^2_{ik(sumstat)}$ represents the squared eQTL summary statistic of SNP k for gene i.
2. *LASSO.* $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r^2_{jk} \hat{\beta}^2_{ij(LASSO)}$, where r^2_{jk} represents the squared correlation between SNP j and SNP k, and $\hat{\beta}^2_{ij(LASSO)}$ represents squared *causal* eQTL effect sizes of SNP j on gene i estimated by LASSO[38].
3. *LASSO with REML correction.* $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r^2_{jk} c_i \hat{\beta}^2_{ij(LASSO)}$. This method is identical to *LASSO* except that we scale $\hat{\beta}^2_{ij(LASSO)}$ by a factor c_i. We define $c_i = \hat{h}^2_{cis;i} / \sum_j^M \hat{\beta}^2_{ij(LASSO)}$, where $\hat{h}^2_{cis;i}$ 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" in Methods. 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.* $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r^2_{jk} \hat{\beta}^2_{ij(BLUP)}$. Here, $\hat{\beta}^2_{ij(BLUP)}$ represents squared *causal* eQTL effect sizes of SNP j on gene i estimated by best linear unbiased predictor (BLUP)[39].
5. *BLUP with REML correction.* $\hat{\mathcal{L}}_k = \sum_i^G \sum_j^M r^2_{jk} c_i \hat{\beta}^2_{ij(BLUP)}$. 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 9). 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 9a-c), eQTL window size (Supplementary Figure 9d), and REML prediction error (Supplementary Figure 9f). 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 9e).

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.

1.5.5 Simulations assessing calibration of standard errors for h^2_{med} enrichment estimates

We conducted an additional simulation aiming to test the calibration of jackknife standard errors computed by MESC for the h^2_{med} enrichment of a gene category D, defined as (proportion of h^2_{med} in D) / (proportion of genes in D). h^2_{med} enrichment is the main quantity that we aim to estimate for analyses involving gene sets. We simulated a gene category containing 200 genes. We then performed two sets of simulations in which this gene category was either enriched or not enriched for h^2_{med}. We observed well-calibrated h^2_{med} enrichment standard errors for the gene category under both the null and causal enrichment scenarios (Supplementary Figure 17).

1.5.6 Simulations assessing REML prediction error

In this section, we justify that our simulation procedure which takes the true expression cis-heritability (h^2_{cis}) of each gene and adds noise drawn from $N(0, 0.01^2)$ is a reasonable proxy for estimates of expression cis-heritability obtained from REML. We did not directly use REML to estimate h^2_{cis} in our simulations for computational ease.

First, we sought to confirm that REML (as implemented in GCTA[40]) produces unbiased estimates of h^2_{cis} in sparse genetic architectures. GCTA models all SNP effects as coming from a one-component Gaussian distribution, which differs from the sparse cis-genetic architecture of gene expression. However, we conducted simulations involving sparse genetic architectures consisting of a single large effect and many small ones (mimicking the cis-genetic architecture of gene expression) and found that GCTA still produced unbiased estimates of h^2_{cis} (Supplementary Figure 18).

Next, we observed that REML prediction error appears to be heteroskadastic as a function of h^2_{cis} (Supplementary Figure 11), so we conducted simulations in which we modeled the standard error of h^2_{cis} according to the best quadratic fit line between h^2_{cis} and $SE(h^2_{cis})$ (Supplementary Figure 19). We found that we obtained virtually identical estimates of h^2_{med}/h^2_g when modeling the standard error of h^2_{cis} in this fashion compared to simply drawing noise from $N(0, 0.01^2)$ regardless of the magnitude of h^2_{cis}.

These two results demonstrate that our simulation procedure which adds normally distributed noise to true h^2_{cis} is reasonable.

1.6 Additional real data analyses

1.6.1 h^2_{med}/h^2_g estimates with other choices of SNP/gene categories

First, we sought to investigate whether modifying the expression cis-heritability bins and baselineLD model influenced our h^2_{med}/h^2_g estimates. When we stratified genes in 10 bins rather than 5, we obtained very similar h^2_{med}/h^2_g estimates (Supplementary Figure 20). Moreover, we observed that our h^2_{med}/h^2_g estimates were robust when making small changes to the baselineLD model (mean h^2_{med}/h^2_g when individually removing

each SNP category = 0.11) and quite robust to even large changes in the baselineLD model (mean h^2_{med}/h^2_g when removing 50% of all SNP categories = 0.13) (Supplementary Figure 21). These results demonstrate that our choice of stratifying genes by 5 expression cis-heritability bins and stratifying SNPs by the baselineLD model can robustly correct for biases in h^2_{med}/h^2_g estimates.

Next, we sought to evaluate the effect of not stratifying genes and/or SNPs on h^2_{med}/h^2_g estimates. When we did not stratify genes by expression cis-heritability bins, we obtained much lower estimates of h^2_{med}/h^2_g (mean estimate 0.011 with S.E. 0.003) (Supplementary Figure 22). This result is consistent with our simulation result in Figure 2d, which shows that not stratifying genes in the presence of a genome-wide negative correlation between the magnitude of eQTL effect sizes and gene-trait effect sizes leads to downwardly biased estimates of h^2_{med}/h^2_g. Moreover, when we did not stratify SNPs by the baselineLD model, we obtained much higher estimates of h^2_{med}/h^2_g (mean estimate 0.45 with S.E 0.03). This result is consistent with our simulation result in Figure 2e, which shows that not partitioning SNPs in the presence of genome-wide positive correlation between the magnitude of eQTL effect sizes and non-mediated effect sizes leads to upwardly biased estimates of h^2_{med}/h^2_g. Together, these results demonstrate that both gene-eQTL independence and pleiotropy-eQTL independence are strongly violated in practice, justifying the necessity of stratifying genes by expression cis-heritability bins and stratifying SNPs by the baselineLD model to obtain unbiased estimates of h^2_{med}/h^2_g.

1.6.2 Role of tissue specificity in explaining low heritability genes

Because we define the h^2_{cis} of a gene by averaging individual-tissue h^2_{cis} estimates across all tissues, a gene with low meta-tissue h^2_{cis} can reflect two different scenarios: the gene has low individual-tissue h^2_{cis} across many tissues, or the gene has high individual-tissue h^2_{cis} 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^2_{cis}, we obtained three quantities for each gene: (1) the number of tissues in which the gene had a significantly nonzero h^2_{cis} ($p < 0.05$), (2) the max h^2_{cis} of the gene across all tissues, and (3) the average h^2_{cis} of the gene across the tissues for which h^2_{cis} was significantly nonzero. If the number of tissues in which the gene has truly nonzero h^2_{cis} (an indicator of the tissue specificity of the gene) is the primary factor in determining the magnitude of the meta-tissue h^2_{cis}, we would expect that (1) be proportional to the magnitude of the meta-tissue h^2_{cis}, while (2) and (3) not be proportional to the magnitude of the meta-tissue h^2_{cis}. We observed that (1) was indeed proportional to the magnitude of the meta-tissue h^2_{cis} (Supplementary Figure 23a); however, (2) and (3) were also proportional to the magnitude of the meta-tissue h^2_{cis} (Supplementary Fig 23b,c), suggesting that statistical power due to the magnitude of h^2_{cis}, rather than tissue specificity, was primarily responsible for the fact that (1) was proportional to the meta-tissue h^2_{cis}. In summary, these results suggest that low h^2_{cis} 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^2_{cis} genes.

We also investigated the relationship between h^2_{cis} and expression levels, but did not observe a strong relationship between h^2_{cis} and average expression level across individuals, nor the number of tissues in which the gene is expressed (Supplementary Figure 24).

1.6.3 Impact of adding window around gene set when estimating h^2_{med} enrichment

When estimating the h^2_{med} 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^2_{med} 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^2_{med} 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^2_{med} 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^2_{med} 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[15,18,41,42]. 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^2_{med} enrichment estimate for the gene set was very similar for most gene sets (Supplementary Figure 25), demonstrating that the h^2_{med} 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.

1.6.4 Comparing MESC to other gene set enrichment methods

MESC can be used to prioritize disease-relevant gene sets using the h^2_{med} enrichment of a gene set, defined as (proportion of h^2_{med} in gene set) / (proportion of genes in gene set). Larger h^2_{med} 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 data[43–49]. 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[50,51]. 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[5,19,23,52–62].

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 MAGMA[46] and DEPICT[45] to the same GWAS summary statistics for 26 traits with nominally significant h^2_{med}. 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^2_{med} enrichment (see Supplementary Table 9 for all estimates). Out of 13,206 total trait-gene set pairs, MESC identified 106 with significant h^2_{med} enrichment (FDR < 0.05 while correcting for $13{,}206 \times 3 = 39{,}078$ hypotheses), compared to 85 for MAGMA and 957 for DEPICT (Extended Data 8a). 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 $R^2 = 0.10$) (Extended Data 8b). Of the 106 significant trait-gene sets pairs identified by MESC, 32 were not detected as significant by either MAGMA or DEPICT (Extended Data 8c), 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.

1.7 Simulation parameters

Figure 2a

For this simulation, we varied h^2_{med}/h^2_g from 0 to 1. For each gene, we selected 5 SNPs to act as cis-eQTLs with locations randomly selected within a random 1 Mb window. The average cis-heritability across all genes was set at $\overline{h^2_{cis}} = 0.05$. One eQTL for each gene was randomly selected to explain 80% of the total cis-heritability of the gene and had an effect size drawn from $\mathcal{N}(0,\, 0.8\overline{h^2_{cis}})$. The remaining eQTLs had effect sizes drawn from $\mathcal{N}(0,\, 0.2\overline{h^2_{cis}}/4)$. Expression phenotypes were simulated for each gene with environmental noise drawn from $\mathcal{N}(0,\, 1 - \overline{h^2_{cis}})$. Non-mediated effect sizes were simulated for each SNP from $\mathcal{N}(0,\, (h^2_g - h^2_{med})/M)$. Gene-trait effect sizes simulated for each gene from $\mathcal{N}(0,\, h^2_{med}/(G\overline{h^2_{cis}}))$. Complex trait phenotypes were simulated with environmental noise drawn from $\mathcal{N}(0, 1 - h^2_g)$.

Figure 2b

We set $\overline{h^2_{cis}} = 0.05$. For the simulation involving 1 eQTL per gene, we drew its effect size from $\mathcal{N}(0,\, \overline{h^2_{cis}})$. For the simulation involving 10% of genes being causal, we simulated gene-trait effect sizes from a point-normal distribution, with nonzero effects drawn from $\mathcal{N}(0,\, h^2_{med}/(0.1G\overline{h^2_{cis}}))$. For the simulation involving 10% of SNPs being causal, we simulation non-mediated SNP effect sizes from a point-normal distribution, with nonzero effects drawn from $\mathcal{N}(0,\, (h^2_g - h^2_{med})/0.1M)$. The remainder of simulation parameters were the same as in Figure 2a.

Figure 2c

We set $\overline{h^2_{cis}} = 0.05$. We simulated one eQTL per gene, with its effect size in the assayed tissue and causal tissue drawn from a multivariate normal distribution with mean $\mathbf{0}$ and 2×2 covariance matrix Σ with diagonal values $\Sigma_{ii} = \overline{h^2_{cis}}$ and off-diagonal values $\Sigma_{ij} = \overline{h^2_{cis}}\sqrt{\frac{1}{2}(r^2_g(T)(3-p) - (1-p))}$, where p represents the proportion of SNPs with non-zero eQTL effect sizes on any gene (i.e. $\frac{G}{M}$). Simulating eQTL effect sizes in this fashion results in an average squared genetic correlation of $r^2_g(T)$ across all genes. We simulated non-mediated effect sizes, expression phenotypes for the assayed tissues, and complex trait phenotypes in the same manner as in Figure 2a. eQTL effect sizes used to generate complex trait phenotypes were taken in the causal tissue. Expression scores were estimated from only the assayed expression phenotypes.

Figure 2d

We simulated 1 eQTL for each gene with effect size drawn from a normal distribution with mean 0 and variance $100 \cdot 2^{k/200}/(\sum_k 2^{k/200})$, where k randomly indexes the genes from 1 to 1000. Simulating eQTL effect sizes in this fashion results in a realistic continuous distribution of eQTL effect sizes, where the quintiles for expression cis-heritability across genes are 0.016, 0.032, 0.064, 0.13, and 0.26. Next, we simulated gene-trait effect sizes from a normal distribution with mean 0 and variance $h^2_{med}/(10000 \cdot 2^{k/200}/(\sum_k 2^{k/200}))$. This causes the magnitude of gene-trait effects to be strongly inversely correlated with the magnitude of eQTL effect sizes across genes, but the per-gene h^2_{med} remains constant. We simulated non-mediated effect sizes, expression phenotypes, and complex trait phenotypes in the same manner as in Figure 2a.

Figure 2e

We set $\overline{h^2_{cis}} = 0.05$. We simulated 1 eQTL per gene with effect size drawn from $\mathcal{N}(0,\, \overline{h^2_{cis}})$. All eQTLs were selected to fall in coding regions. Next, for all SNPs in coding regions, we simulated non-mediated SNP effect sizes from $\mathcal{N}(0,\, (h^2_g - h^2_{med})/M_{coding})$, where M_{coding} is the number of SNPs that fall into coding regions. We simulated gene-trait effect sizes, expression phenotypes, and complex trait phenotypes in the same manner as in Fig 2a.

Figure 2f

We set $\overline{h^2_{cis}} = 0.05$. We simulated 1 eQTL per gene with effect size drawn from $\mathcal{N}(0, \overline{h^2_{cis}})$. 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 2.5x, 5x, or 10x. We simulated expression phenotypes and complex trait phenotypes in the same manner as Figure 2a. For stratified LD-score regression, we defined the eQTL category as the set of all true eQTLs. For MESC, we stratified SNPs by the baselineLD model with the three SNP categories and windows around the SNP categories removed from the model.

Supplementary Figure 3

For all simulations, we set $\overline{h^2_{cis}} = 0.05$, $h^2_{med} = 0.1$, 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 $\overline{h^2_{cis}}$. Expression phenotypes were simulated for each gene with environmental noise drawn from $\mathcal{N}(0, 1 - \overline{h^2_{cis}})$.

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 $h^2 - h^2_{med}$. We simulated gene-trait effect sizes for each gene from $\mathcal{N}(0, (h^2_{med}/(G\overline{h^2_{cis}}))$. Complex trait phenotypes were simulated with environmental noise drawn from $\mathcal{N}(0, 1 - h^2_g)$.

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.

Supplementary Figure 9

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 $\overline{h^2_{cis}} = 0.05$ 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 $\mathcal{N}(0, \overline{h^2_{cis}})$. 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 $\mathcal{N}(0, 0.8\overline{h^2_{cis}})$. The remaining eQTLs had effect sizes drawn from $\mathcal{N}(0, 0.2\overline{h^2_{cis}}/(N_{eQTL} - 1))$, 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 $\mathcal{N}(0, 1 - \overline{h^2_{cis}})$. 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 11).

Supplementary Figure 10

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.

Supplementary Figure 15

We set $\overline{h^2_{cis}} = 0.05$. We simulated 1 eQTL per gene with effect size drawn from $\mathcal{N}(0, \overline{h^2_{cis}})$. 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.

Supplementary Figure 17

We set the overall $h^2_{med}/h^2_g = 0.4$. 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 $\mathcal{N}(0, h^2_{med}/(G\overline{h^2_{cis}}))$. For the enriched scenario, we simulated gene-trait effect sizes so that h^2_{med} within the gene category was 2x the h^2_{med} 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.

1.8 Choice of 10 traits for display in Figure 3

Starting with the full set of 26 traits with h^2_{med}/h^2_g 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 regression[63]). Between the pair of selected traits, we then retained the trait with the larger h^2_g z-score (estimated using stratified LD score regression[2]). We repeated this procedure until 10 traits were left.

References

[1] 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 (2015).

[2] 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 (2015).

[3] Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992, DOI: 10.1038/ng.3865 (2017).

[4] Gazal, S., Marquez-Luna, C., Finucane, H. K. & Price, A. L. Reconciling S-LDSC and LDAK functional enrichment estimates. *Nat. Genet.* **51**, 1202–1204, DOI: 10.1038/s41588-019-0464-1 (2019).

[5] Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252, DOI: 10.1038/ng.3506 (2016).

[6] GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213, DOI: 10.1038/nature24277 (2017).

[7] Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 447367, DOI: 10.1101/447367 (2018).

[8] 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 (2019).

[9] 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 (2008).

[10] Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515, DOI: 10.1038/nrg3457 (2013).

[11] 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 (2014).

[12] 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 (2017).

[13] 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 (2017).

[14] 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 (2018).

[15] 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 (2018).

[16] 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 (2019).

[17] 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 (2019).

[18] 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 (2019).

[19] Barfield, R. *et al.* Transcriptome-wide association studies accounting for colocalization using Egger regression. *Genet. Epidemiol.* **42**, 418–433, DOI: 10.1002/gepi.22131 (2018).

[20] 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 (2018).

[21] 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 (2011).

[22] 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 (2017).

[23] 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 (2019).

[24] Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212, DOI: 10.1038/nrg3891 (2015).

[25] Albert, F. W., Bloom, J. S., Siegel, J., Day, L. & Kruglyak, L. Genetics of trans-regulatory variation in gene expression. *eLife* **7**, e35471, DOI: 10.7554/eLife.35471 (2018).

[26] Yao, C. *et al.* Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *The Am. J. Hum. Genet.* **100**, 571–580, DOI: 10.1016/j.ajhg.2017.02.003 (2017).

[27] Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186, DOI: 10.1016/j.cell.2017.05.038 (2017).

[28] The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58, DOI: 10.1038/nature09298 (2010).

[29] 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 (2018).

[30] 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 (2019).

[31] 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 (2018).

[32] 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 (2019).

[33] Wainschtein, P. *et al.* Recovery of trait heritability from whole genome sequence data. *bioRxiv* 588020, DOI: 10.1101/588020 (2019).

[34] 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 (2019).

[35] Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203, DOI: 10.1038/s41586-018-0579-z (2018).

[36] Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600, DOI: 10.1038/ng.3795 (2017).

[37] Ongen, H. *et al.* Estimating the causal tissues for complex traits and diseases. *Nat. Genet.* **49**, 1676–1683, DOI: 10.1038/ng.3981 (2017).

[38] Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. Royal Stat. Soc. Ser. B (Methodological)* **58**, 267–288 (1996).

[39] Robinson, G. K. That BLUP is a Good Thing: The Estimation of Random Effects. *Stat. Sci.* **6**, 15–32, DOI: 10.1214/ss/1177011926 (1991).

[40] Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **88**, 76–82, DOI: 10.1016/j.ajhg.2010.11.011 (2011).

[41] 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 (2016).

[42] Fine, R. S., Pers, T. H., Amariuta, T., Raychaudhuri, S. & Hirschhorn, J. N. Benchmarker: 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 (2019).

[43] 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 (2010).

[44] 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 (2012).

[45] Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890, DOI: 10.1038/ncomms6890 (2015).

[46] 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 (2015).

[47] 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 (2016).

[48] 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 (2018).

[49] 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 (2018).

[50] Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375, DOI: 10.1038/nature13138 (2014).

[51] Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *New Engl. J. Medicine* **373**, 895–907, DOI: 10.1056/NEJMoa1502214 (2015).

[52] 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 (2015).

[53] 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 (2016).

[54] 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 (2018).

[55] Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408, DOI: 10.7554/eLife.34408 (2018).

[56] 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 (2018).

[57] 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 (2018).

[58] 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 (2018).

[59] Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127, DOI: 10.1126/science.aat8127 (2018).

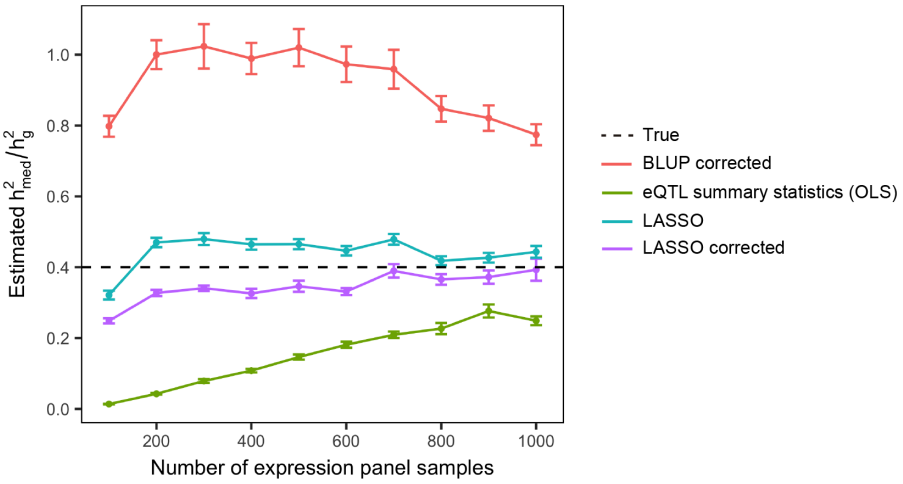
[60] Huckins, L. M. *et al.* Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. *Nat. Genet.* **51**, 659, DOI: 10.1038/s41588-019-0364-4 (2019).

[61] 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 (2019).

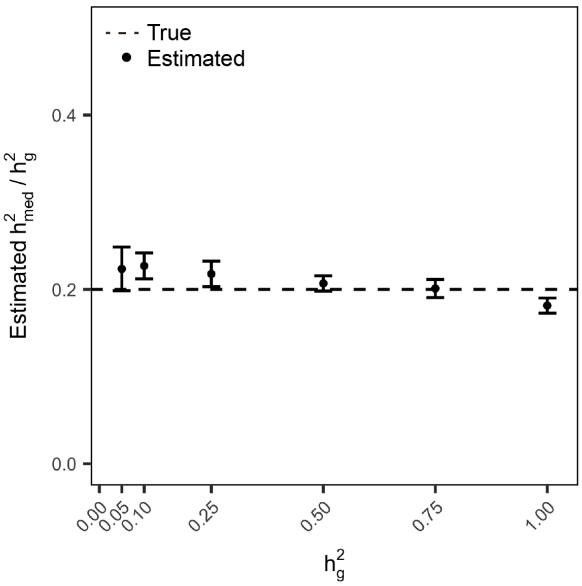
[62] Porcu, E. *et al.* Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat. Commun.* **10**, 1–12, DOI: 10.1038/s41467-019-10936-0 (2019).

[63] 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 (2015).

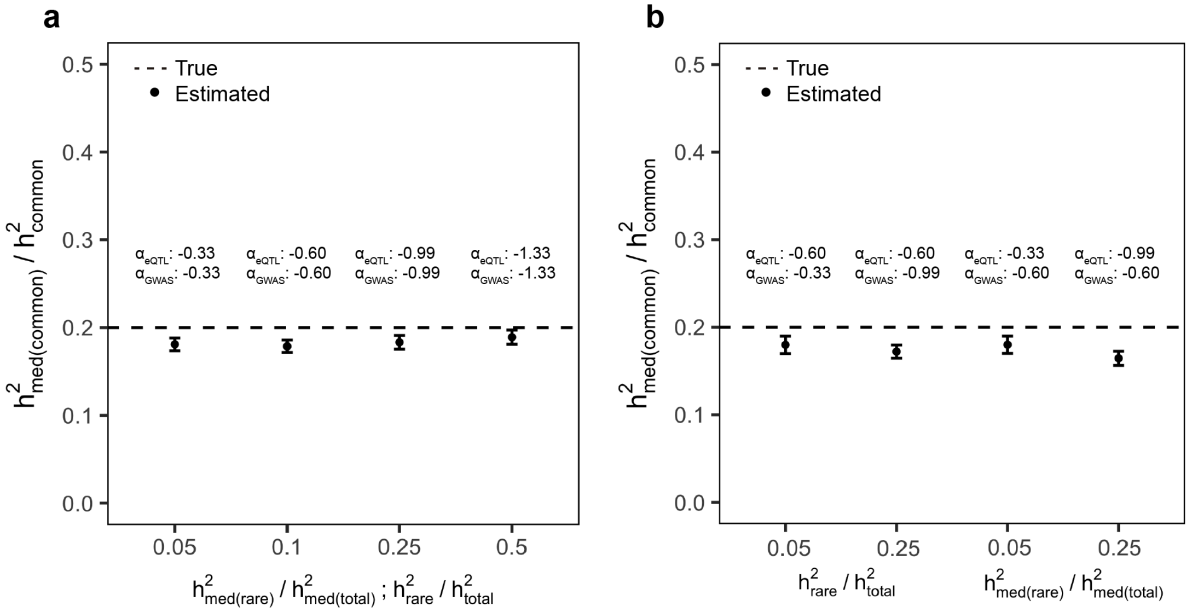
2 Supplementary Figures



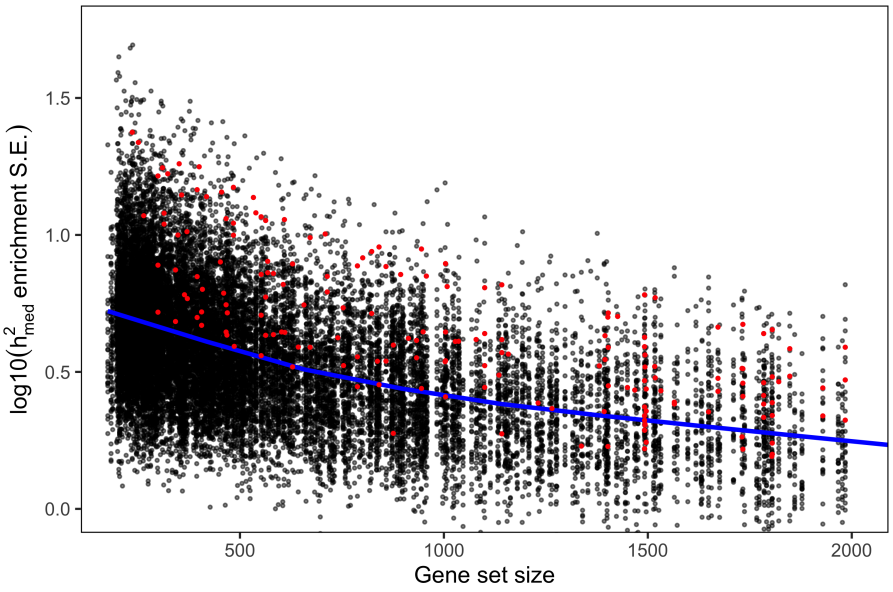
Supplementary Figure 1. Simulations assessing impact of different methods of estimating expression scores on estimates of h^2_{med}/h^2_g. See "Simulations comparing different methods of estimating expression scores" (Supplementary Note) for description of methods. For this simulation, $h^2_{med}/h^2_g = 0.4$. 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^2_g estimates obtained from these expression scores were severely upwardly biased (i.e. greater than 1 at all sample sizes). Error bars represent mean standard errors across 100 simulations.



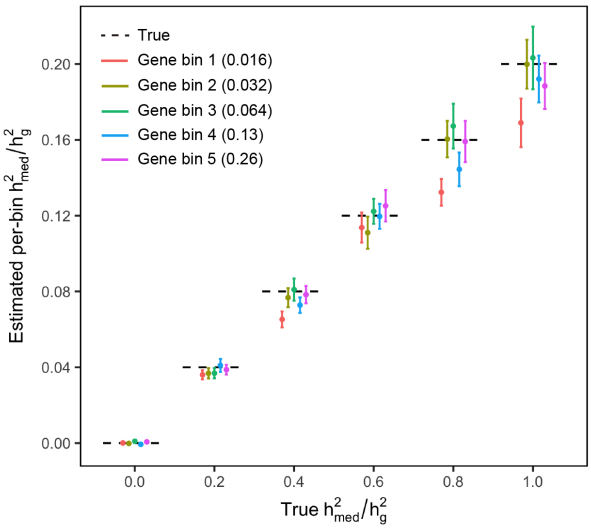
Supplementary Figure 2. Simulations assessing h^2_{med}/h^2_g estimates when varying h^2_g. Simulation was performed in the same manner as in Figure 2a (with expression panel size fixed at 1000). h^2_{med}/h^2_g was fixed at 0.2 for all simulations. Error bars represent mean standard errors across 100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



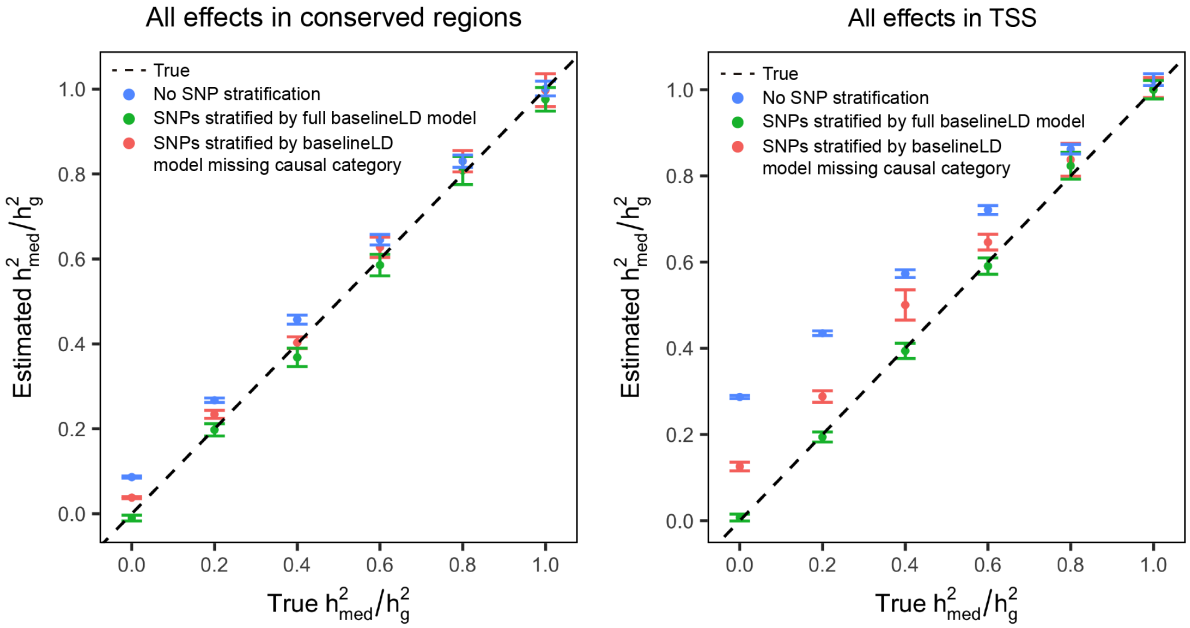
Supplementary Figure 3. Simulations assessing h^2_{med}/h^2_g estimates under frequency-dependent genetic architectures. Error bars represent mean standard errors across 100 simulations. With $h^2_{med(common)}/h^2_{common}$ fixed at 0.2 and h^2 fixed at 0.5, we simulated eQTL effect sizes with variance 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**) $h^2_{med(common)}/h^2_{common}$ estimates with $\alpha_{eQTL} = \alpha_{GWAS}$, in which case the proportion of rare h^2_{med} and rare h^2 is the same (x-axis). (**b**) $h^2_{med(common)}/h^2_{common}$ estimates with $\alpha_{eQTL} \neq \alpha_{GWAS}$. Left: proportion of rare h^2_{med} 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^2_{med} is varied from 0.05 to 0.25. See "Additional details on simulations under frequency-dependent architectures" in Supplementary Note for remaining details on this simulation.



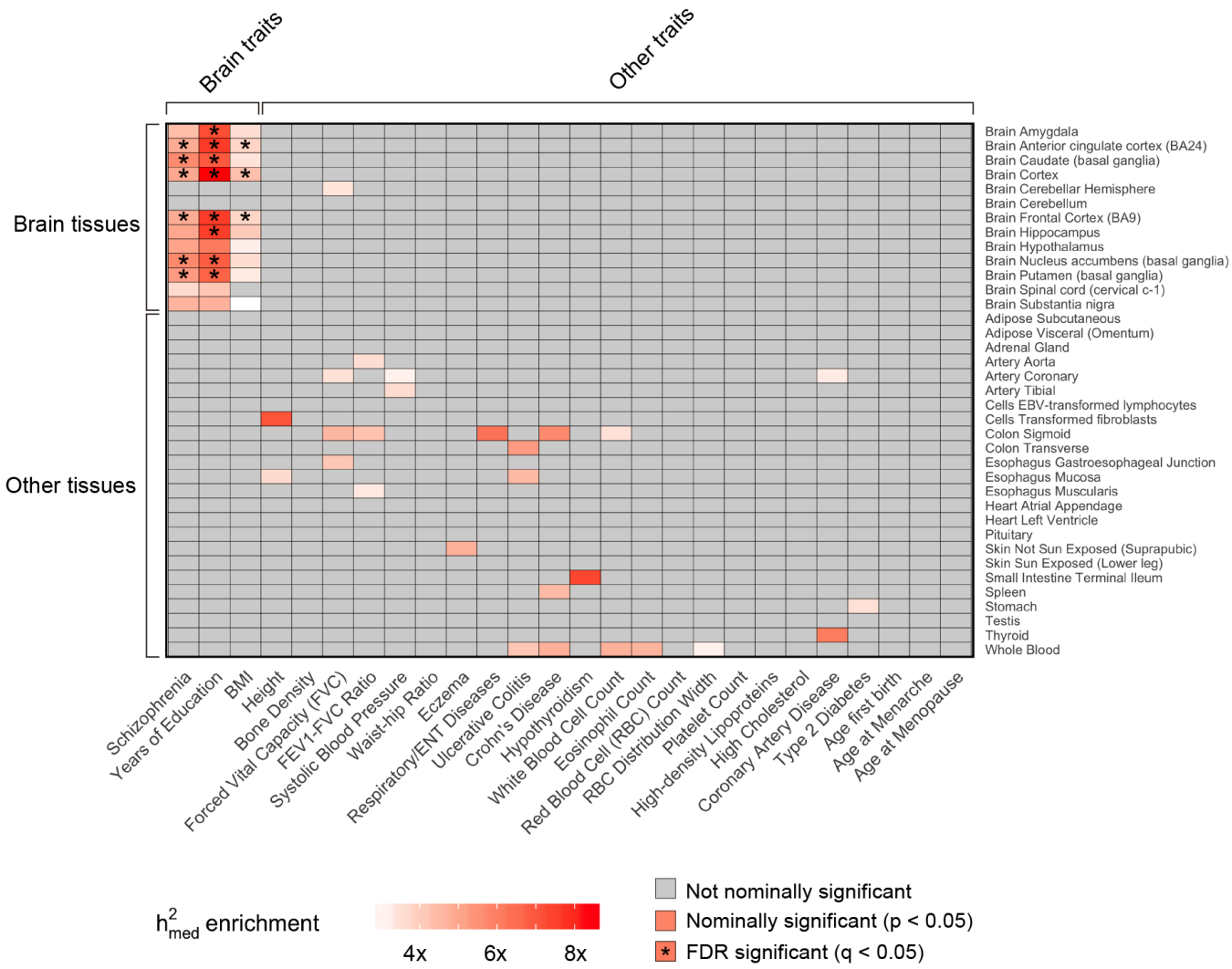
Supplementary Figure 4. Relationship between gene set size and log10 h^2_{med} enrichment standard error. Each point represent a gene set-complex trait pair. Points highlighted in red indicate gene set-complex trait pairs with FDR < 0.05 (after accounting for 21,502 hypotheses tested). Blue line indicates the LOESS best fit line.



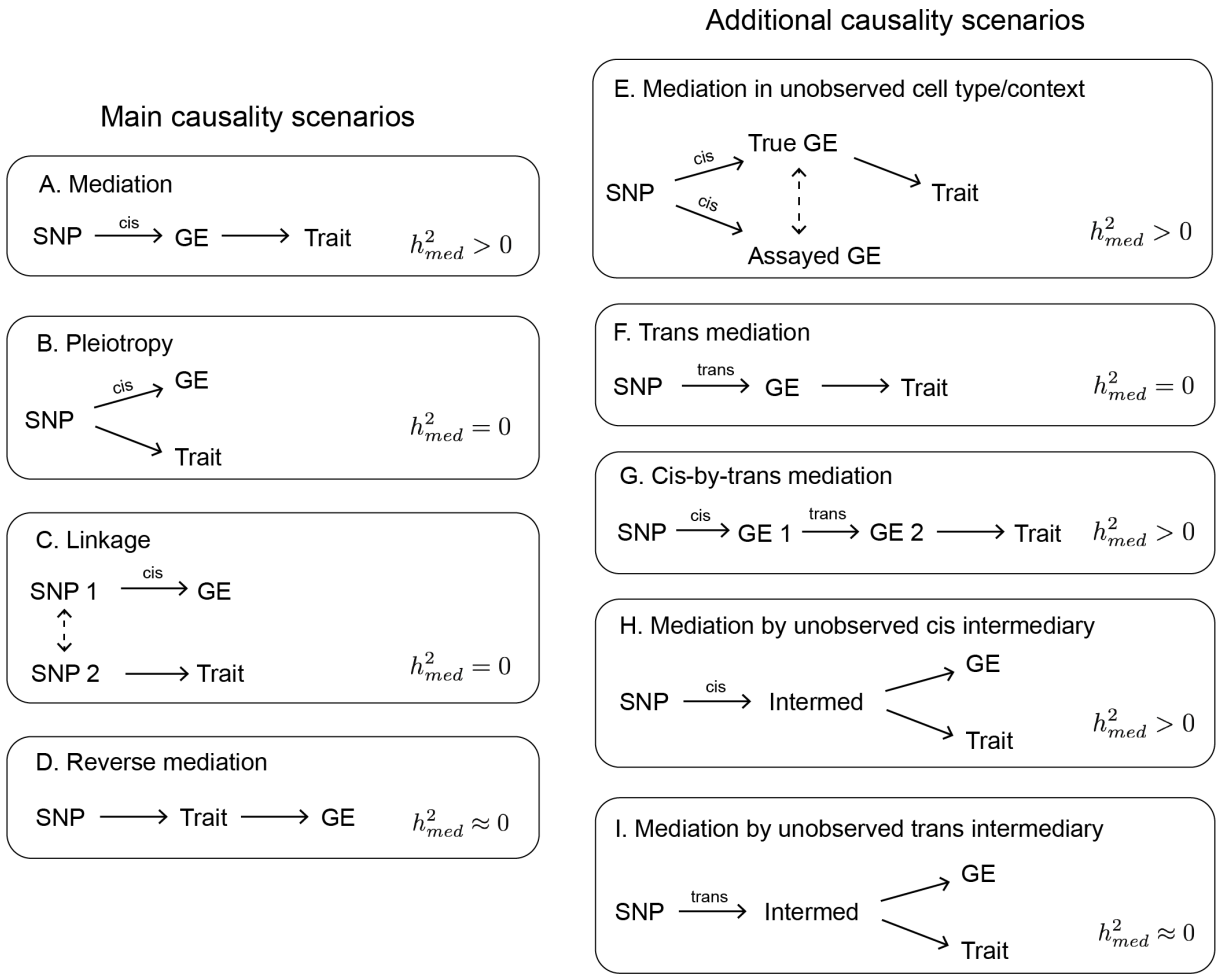
Supplementary Figure 5. Per-bin h^2_{med}/h^2_g estimates for simulation in Fig 2d. In the legend, the number in parentheses indicates the average expression cis-heritability of genes in a given gene bin. Error bars represent mean standard errors across 300 simulations.



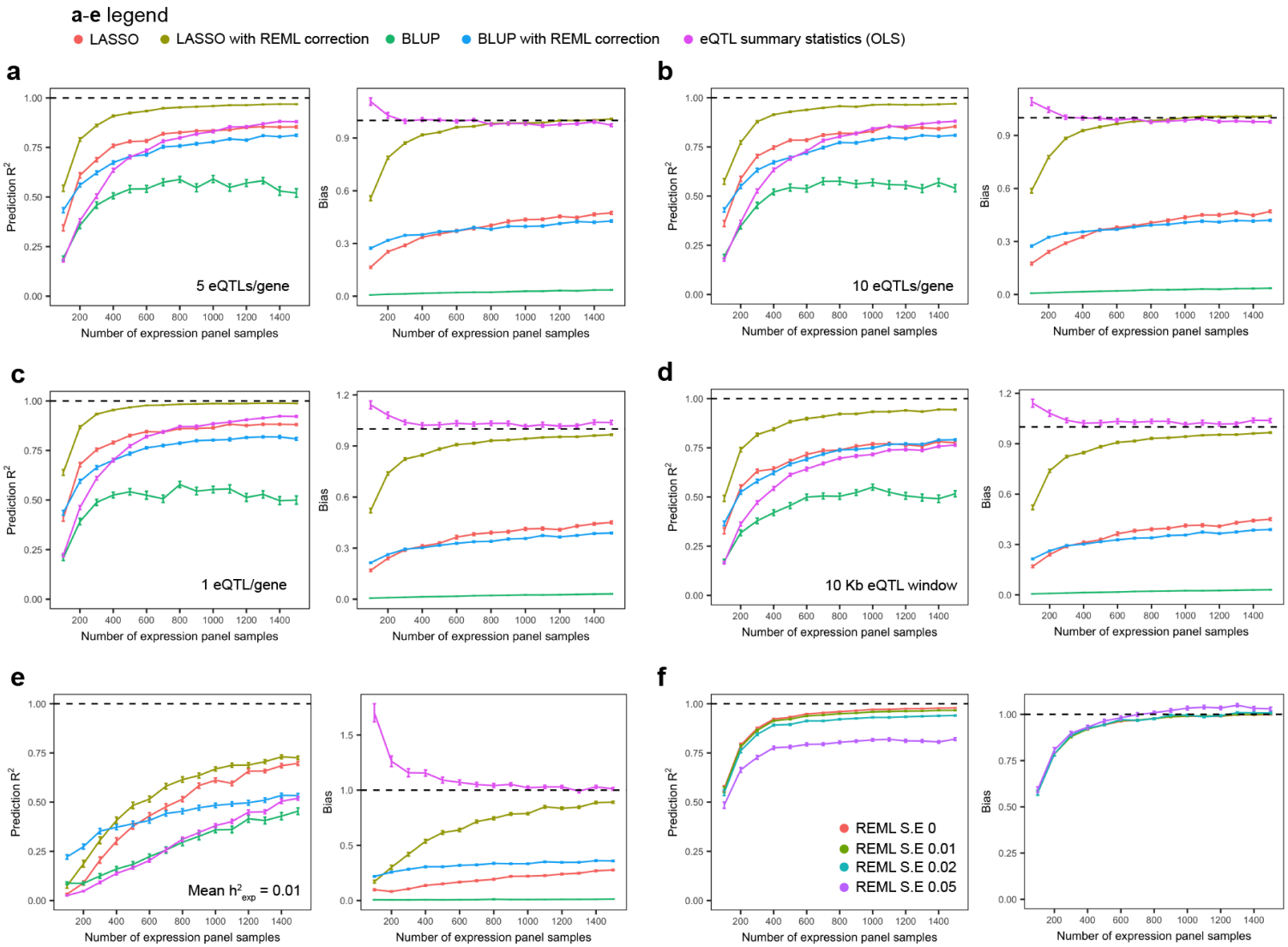
Supplementary Figure 6. Same simulation parameters as Figure 2e, but all non-mediated effects and eQTL effects localized in conserved regions and transcription start sites (TSS) respectively. Error bars represent mean standard errors across 300 simulations.



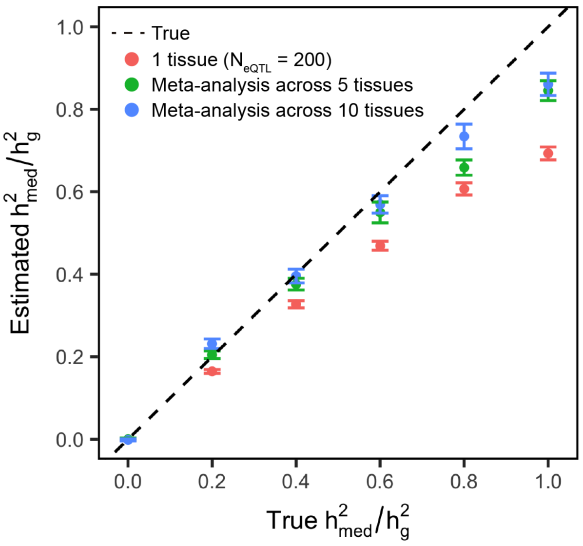
Supplementary Figure 7. Same as Figure 5c, but including h^2_{med} enrichment estimates for all 26 traits and with individual GTEx tissues labelled.



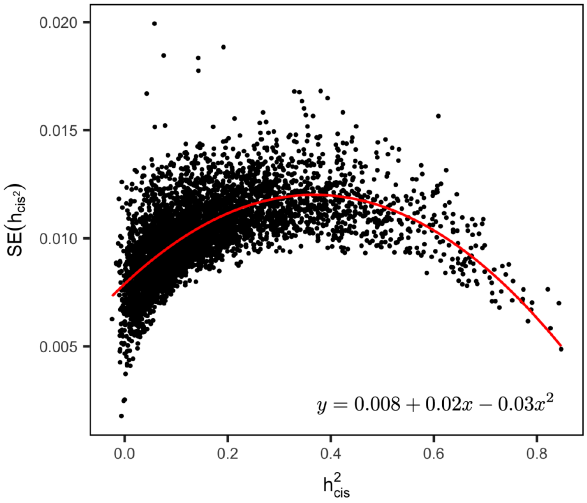
Supplementary Figure 8. Modes of expression causality. See "Modes of expression causality" (Supplementary Note) for a description of each scenario and its contribution to estimates of h^2_{med}.



Supplementary Figure 9. Simulations assessing accuracy and bias of different methods of estimating expression scores in various cis-genetic architectures. See "Comparing different methods of estimating expression scores" (Supplementary Note) for details on these simulations. 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.

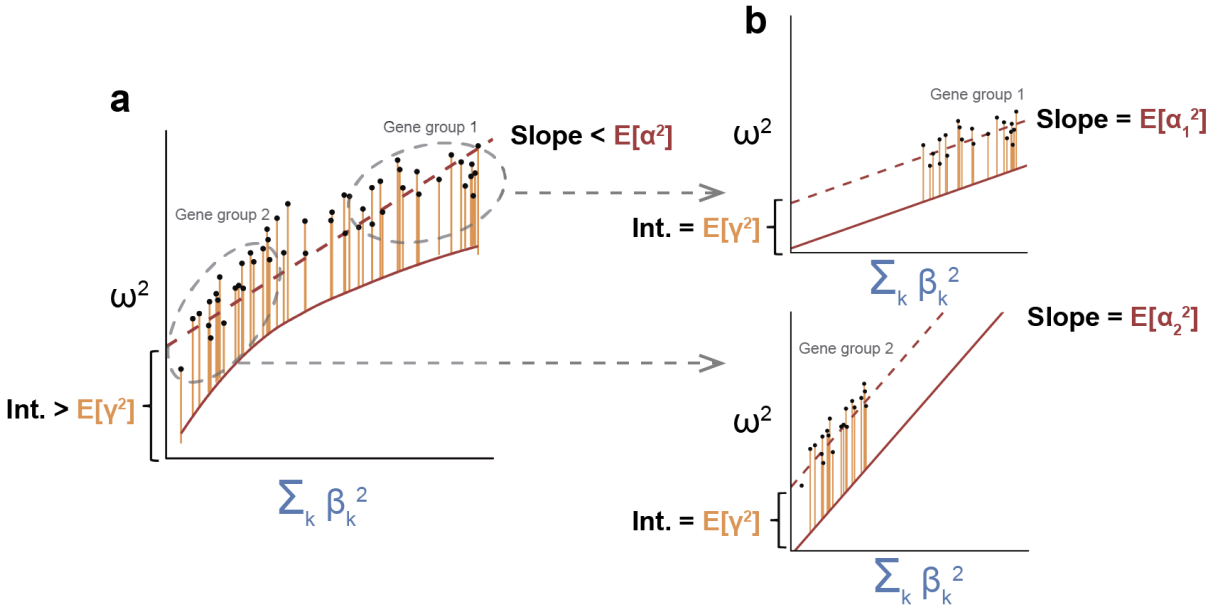


Supplementary Figure 10. Simulations assessing h^2_{med}/h^2_g estimates from expression scores meta-analyzed across tissues. We simulated expression phenotypes in multiple tissues, then estimated expression scores in individual tissues and meta-analyzed expression scores across tissues. Error bars represent mean standard errors across 100 simulations. See "Simulations parameters" in Supplementary Note for specific details on this simulation.

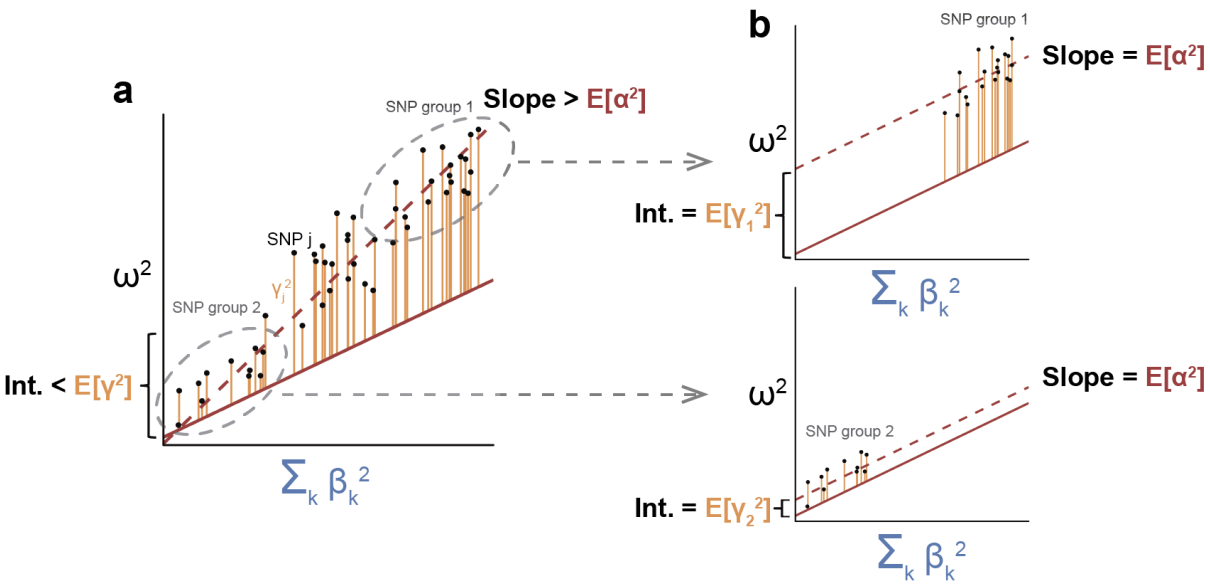


Supplementary Figure 11. Standard errors of expression cis-heritability estimates across GTEx tissues. Expression cis-heritability estimates are obtained for each gene in each of 48 GTEx tissues using GCTA. Standard error represents the mean standard error of expression cis-heritability estimates for each gene across all 48 tissues. Red line denotes the best quadratic fit.

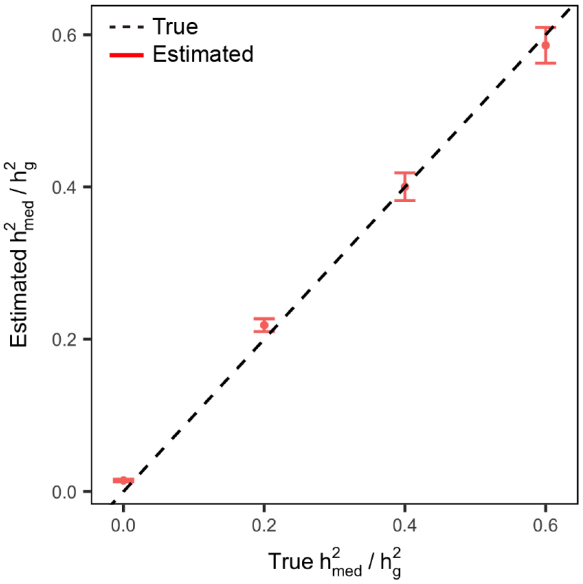
Supplementary Figure 12. Comparison between Mendelian randomization with multiple genetic variants and MESC.



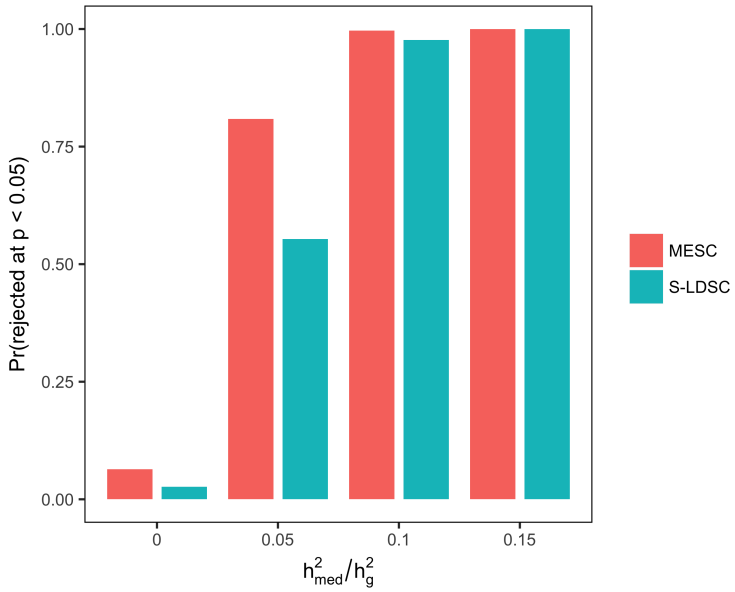
Supplementary Figure 13. Illustration of impact of violations to gene-eQTL effect size independence on estimates of $E[\alpha^2]$. In the figure, we depict a scenario where the magnitude of α is negatively correlated with the magnitude of β. (**a**) If we perform the regression using all genes, the slope 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.



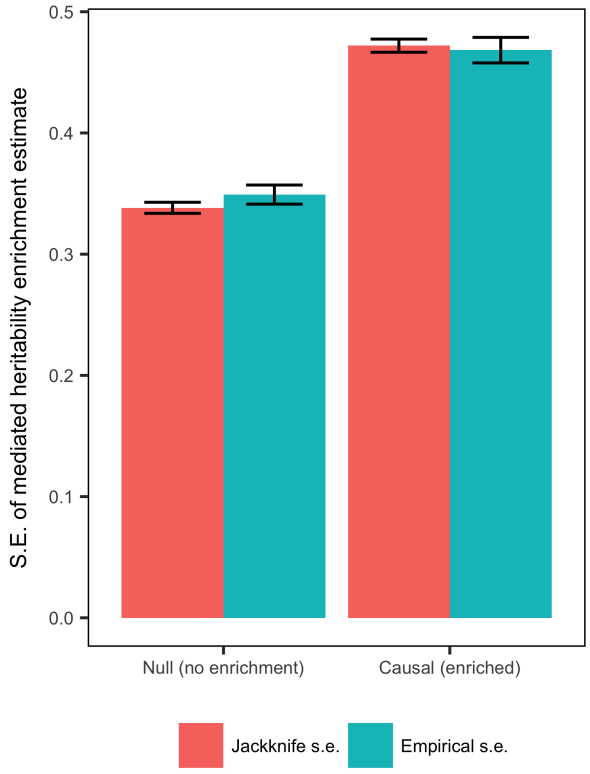
Supplementary Figure 14. Illustration of impact of violations to pleiotropy-eQTL effect size independence on estimates of $E[\alpha^2]$. In the figure, we depict a scenario where the magnitude of γ is positively correlated with the magnitude of β. (**a**) If we perform the regression using all SNPs, 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 an approximately unbiased estimate of $E[\alpha^2]$.



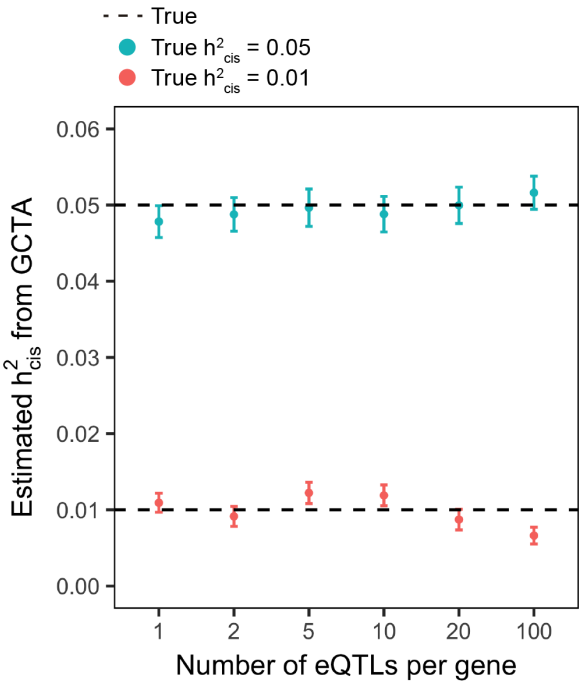
Supplementary Figure 15. Simulations assessing h^2_{med}/h^2_g estimates involving realistic violation to pleiotropy-eQTL independence. h^2_{med}/h^2_g was varied from 0 to 0.6. Because the union of the three SNP categories with eQTL effect size enrichment (coding regions, conserved regions, and transcription start sites) comprises around 6% of the genome, the maximum value that h^2_{med}/h^2_g can be is 0.6 if we have the condition that h^2_g enrichment of the three SNP categories is 10x. Error bars represent mean standard errors across 100 simulations. See Supplementary Note for more details on this simulation.



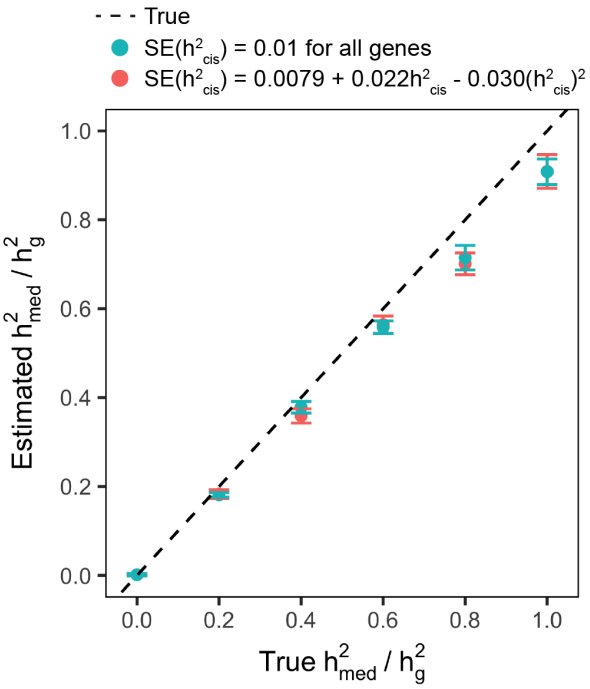
Supplementary Figure 16. Simulations assessing null calibration and power of MESC and stratified LD-score regression given pleiotropy-eQTL effect size independence. For various levels of h^2_{med}/h^2_g, this figure reports the proportion of simulations which the null hypothesis that $h^2_{med}/h^2_g = 0$ is rejected by MESC, and the proportion of simulations in which the null hypothesis of no h^2_g 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.



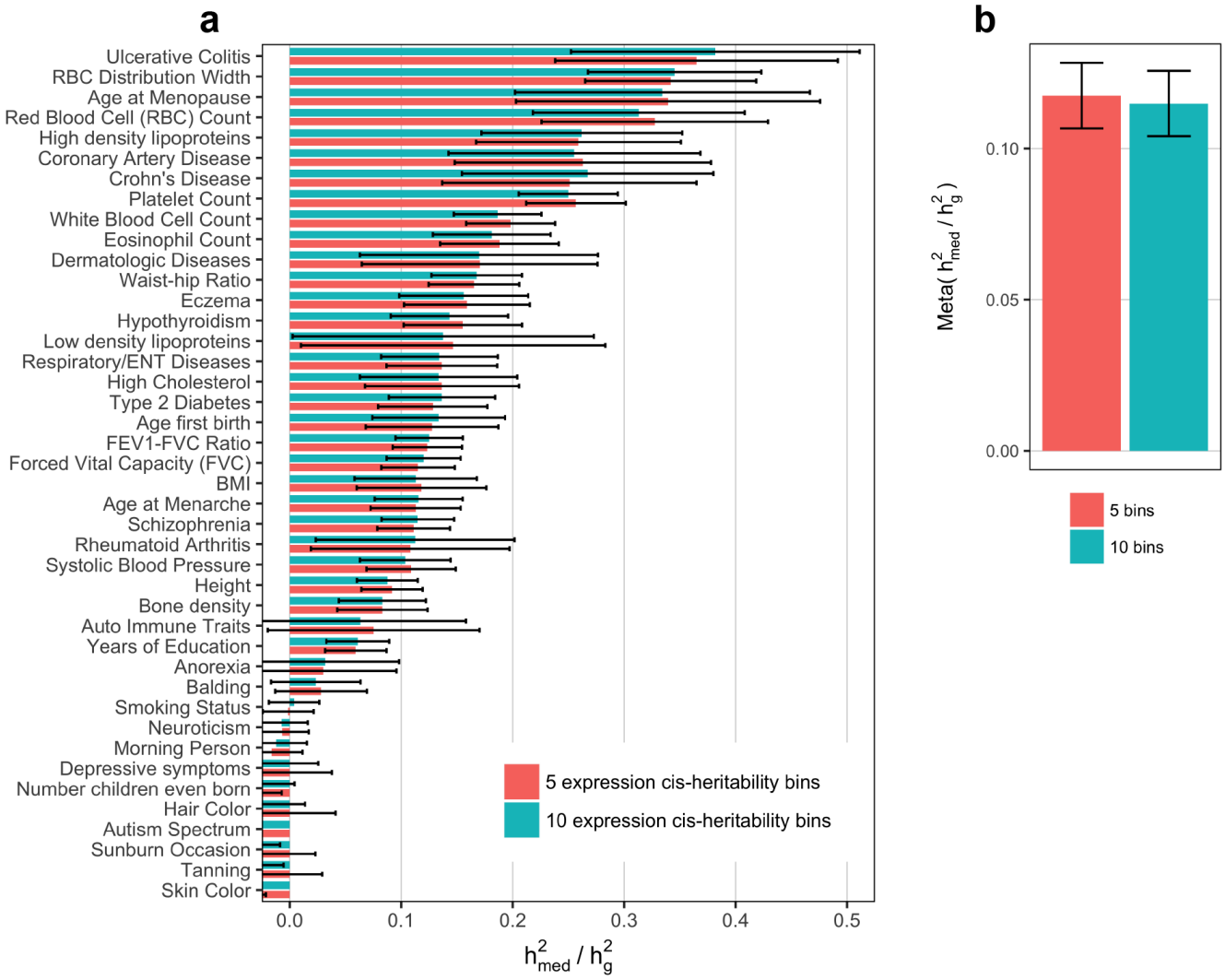
Supplementary Figure 17. Simulations assessing calibration of jackknife standard errors for h^2_{med} enrichment. We simulated a gene category without h^2_{med} enrichment (null) and a gene category with 2x h^2_{med} enrichment (causal). Jackknife standard errors for h^2_{med} enrichment are compared to empirical standard errors of h^2_{med} enrichment estimates across 1000 simulations. See Supplementary Note for more details on this simulation.



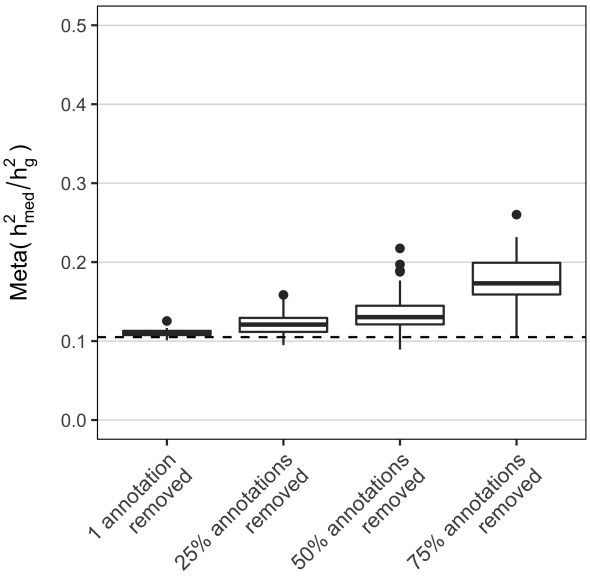
Supplementary Figure 18. Simulations assessing GCTA estimates of heritability under sparse cis-eQTL genetic architectures. We simulated cis-eQTLs for a gene by selecting x random SNPs within a random 1 Mb window on chromosome 1 (restricting to Hapmap3 SNPs). One cis-eQTL was randomly selected to be the 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.



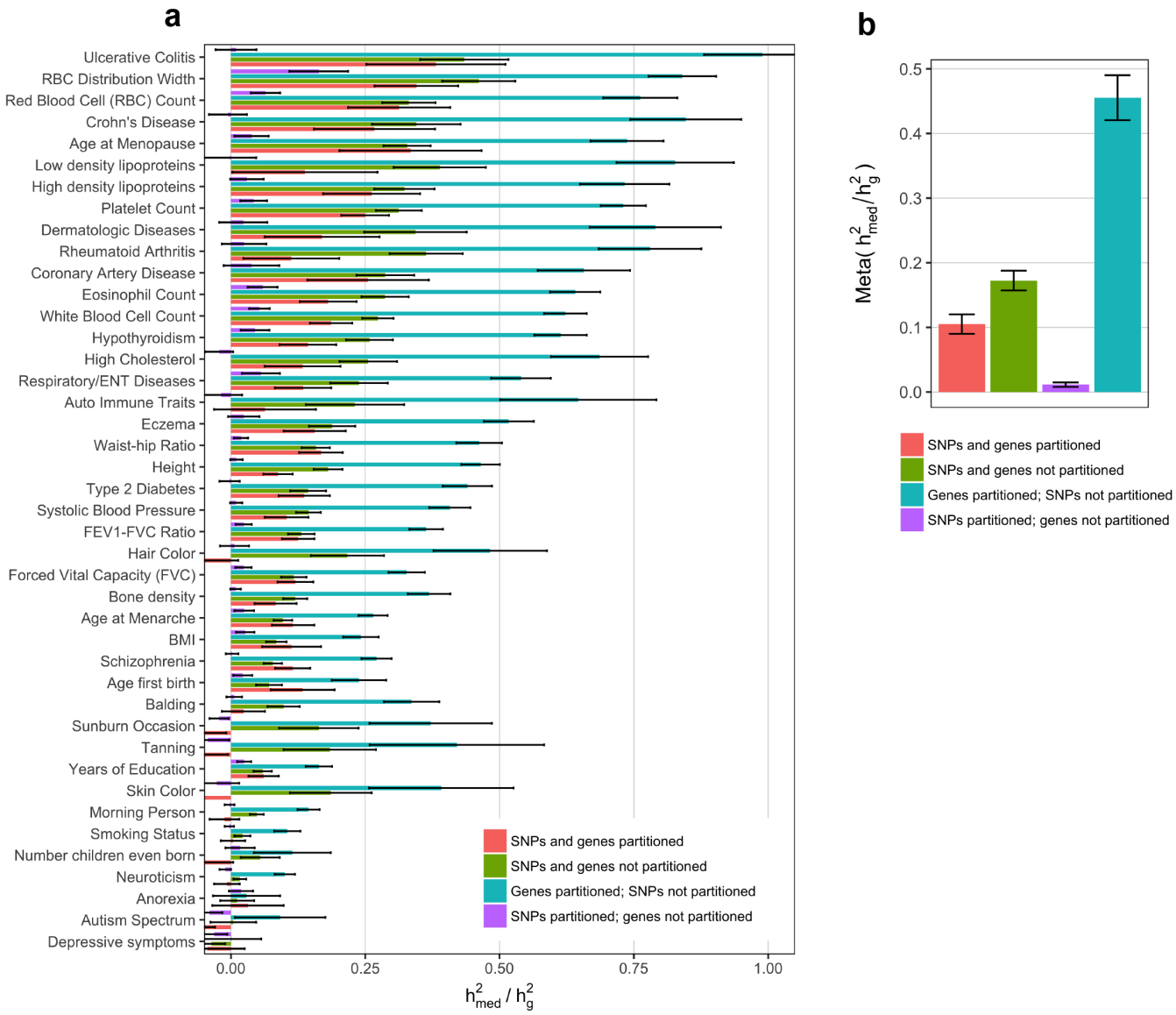
Supplementary Figure 19. Simulations assessing h^2_{med}/h^2_g estimates with varying $SE(h^2_{cis})$ as a function of h^2_{cis}. Simulation was performed in the same manner as in Figure 2a (with expression panel size fixed at 1000). Standard error for h^2_{cis} was simulated from either $\mathcal{N}(0, 0.01^2)$ (consistent the mean of empirical $SE(h^2_{cis})$ estimates across all GTEx samples) or from $\mathcal{N}(0, (0.0079 + 0.22h^2_{cis} - 0.03(h^2_{cis})^2)^2)$ (consistent with the best quadratic fit line relating empirical h^2_{cis} to $SE(h^2_{cis})$ estimates across all GTEx samples, see Supplementary Figure 11). Error bars represent mean standard errors across 100 simulations.



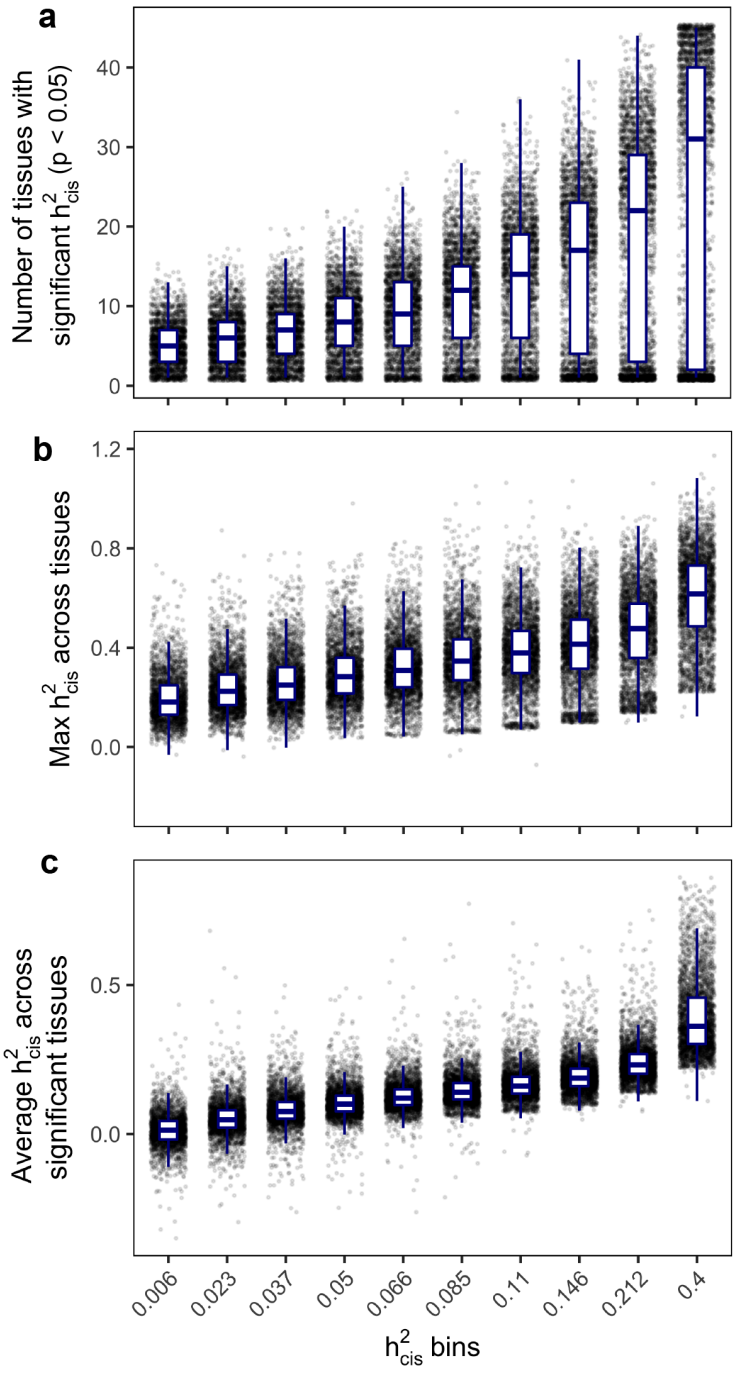
Supplementary Figure 20. h^2_{med}/h^2_g **estimates when varying number of expression cis-heritability bins**. All estimates were obtained using all-tissue meta-analyzed expression scores. (**a**) We estimated h^2_{med}/h^2_g for 42 traits while stratifying genes by either 5 or 10 expression cis-heritability bins. Error bars represent jackknife standard errors. (**b**) Estimates from **a** meta-analyzed across all 42 traits. Error bars represent standard errors from random-effect meta analysis.



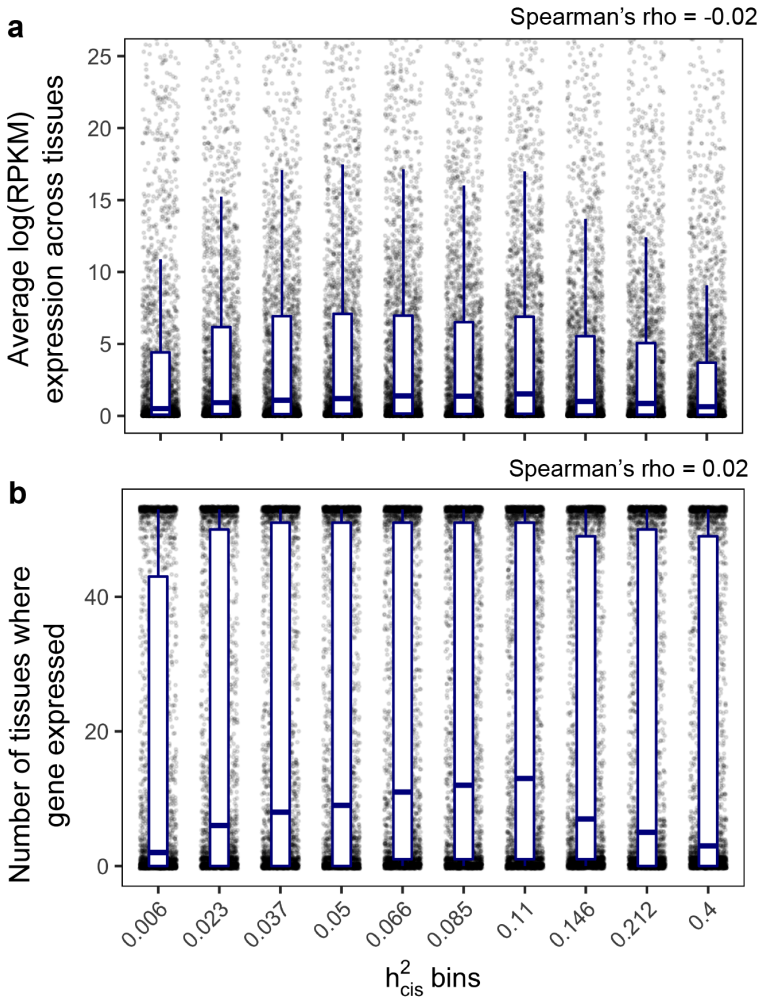
Supplementary Figure 21. h^2_{med}/h^2_g estimates when removing subsets of the baselineLD model. In total, the baselineLD model v2.0 we used in our main analyses contains 72 SNP annotations. We grouped together related baselineLD annotations (typically consisting of a main annotation and the same annotation 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^2_{med}/h^2_g estimate meta-analyzed over 42 traits and estimated from GTEx all-tissue expression scores. For the boxplot labelled "1 annotation removed," we show h^2_{med}/h^2_g estimates when removing each of the 29 individual categories. For the boxplots labelled "25%/50%/75% annotations removed," we show h^2_{med}/h^2_g estimates from 100 random subsets of the categories corresponding to the percentage of annotations removed. Dotted line indicates the h^2_{med}/h^2_g estimate when using the full baselineLD model.



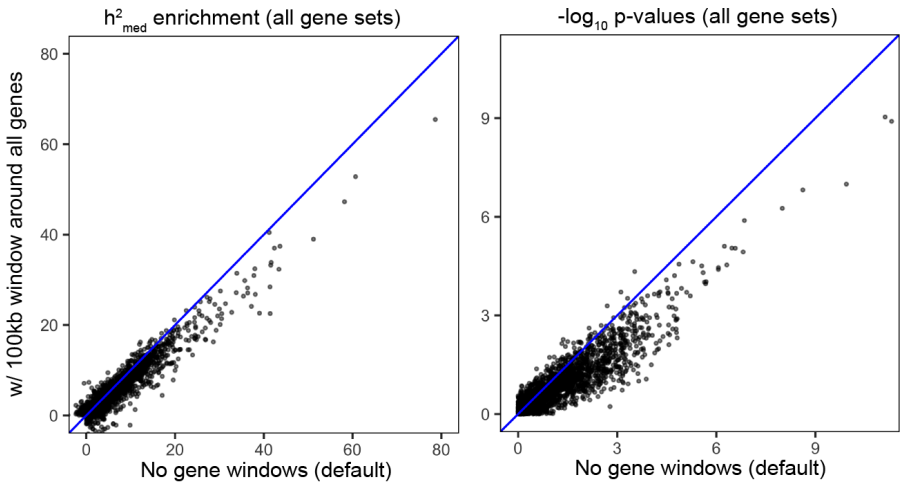
Supplementary Figure 22. h^2_{med}/h^2_g **estimates without stratifying genes/SNPs.** All estimates were obtained using all-tissue meta-analyzed expression scores. (**a**) We estimated h^2_{med}/h^2_g for 42 traits without stratifying genes by 5 expression cis-heritability and/or without stratifying SNPs by the baselineLD model. Error bars represent jackknife standard errors. (**b**) Estimates from **a** meta-analyzed across all 42 traits. Error bars represent standard error from random-effect meta-analysis.



Supplementary Figure 23. 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} deciles 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 24. 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 25. h^2_{med} **enrichment estimates with 100 Kb window around genes.** (**Left**) h^2_{med} 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^2_{med} enrichment p-values.