

# Supplementary Methods

Anqi Zhu<sup>†</sup>, Nana Matoba<sup>†</sup>, Emmaleigh Wilson, Amanda L. Tapia,  
Yun Li, Joseph G. Ibrahim, Jason L. Stein, Michael I. Love

January 25, 2021

## 1 MRLocus statistical model

MRLocus proceeds in two separate hierarchical models, which are encoded in the Stan programming language and with posterior inference performed using the Stan and RStan software packages [Carpenter et al., 2017, Stan Development Team, 2020]. In section 1.1 we define the model for the colocalization step, and in section 1.2 we define the model for the slope fitting step.

### 1.1 Colocalization step

#### 1.1.1 Input data

The first step performs colocalization of eQTL (A) and GWAS (B) signals across a number of nearly-LD-independent signal clusters  $j \in 1, \dots, J$ , using summary statistics from both studies:  $\hat{\beta}_{i,j}^A$  and  $\text{se}(\hat{\beta}_{i,j}^A)$ ,  $\hat{\beta}_{i,j}^B$  and  $\text{se}(\hat{\beta}_{i,j}^B)$  for SNP  $i \in 1, \dots, n_j$  in cluster  $j$  for study A and B, and the respective LD matrices for each cluster  $j$ :  $\Sigma_j^A$  and  $\Sigma_j^B$ . Near LD independence is obtained by clumping with PLINK [Purcell et al., 2007] on the putative mediator A with  $r^2$  threshold of 0.1 and then trimming “signal clusters” (the term we use for PLINK clumps after they are created), in which the pairwise  $r^2$  of the index eSNP is greater than 0.05. Trimming is performed with the `trimClusters` function in MRLocus, which removes signal clusters with  $r^2 > 0.05$  to the first cluster, then the second cluster, and so on until no pairs remain with  $r^2 > 0.05$ . The clusters and the individual SNPs per cluster must be matched across study. The estimated coefficients  $\hat{\beta}_{i,j}^X$  for  $X \in \{A, B\}$  refer to either the estimated coefficients from a linear model of a continuous trait  $y$  on genotype dosages  $\{0, 1, 2\}$ , or the estimated log odds from a logistic regression of a binary trait  $y$  modeled on genotype dosages.

While it would be preferable to use allelic fold change (aFC) [Mohammadi et al., 2017] or ACME effect sizes [Palowitch et al., 2018] for the eQTL (A) study in MRLocus modeling, in practice we typically are provided with publicly available estimated coefficients representing inverse normal transformed (INT) or  $\log_2$  transformed expression values regressed on genotype dosages. For eQTL coefficients derived from INT expression data, the mediation effect estimated by MRLocus represents the effect on the trait from modifying gene expression by 1 SD, while for eQTL coefficients derived from  $\log_2$  transformed expression data, the mediation effect represents the effect on the trait from doubling gene expression.

The eQTL and GWAS studies are referred to as “A” and “B” in the formula and code below for generalization, for example, the eQTL study could be replaced with a pQTL (protein quantitative trait loci) study. “A” therefore refers to a study of a trait (or “exposure”) that is believed to be causally upstream of the trait examined in study “B” (or “outcome”).

### 1.1.2 Collapsing and allele flipping

MRLocus contains two convenience functions, `collapseHighCorSNPs` and `flipAllelesAndGather`, which are described briefly. The first function uses hierarchical clustering based on the LD matrix (the user must pick which to use if two are available), in order to collapse SNPs into groups using complete linkage, and thresholding the resulting dendrogram at 0.95 correlation. The SNP with the highest absolute  $z$ -score within a collapsed set is chosen as the representative SNP.

`flipAllelesAndGather` performs a number of allele flipping steps for assisting statistical modeling and visualization. The alleles are flipped such that the index SNP (as defined by its absolute value of  $z$ -score) for study A has a positive estimated coefficient. This is to simplify the interpretations of the plots – such that we are always describing the effects on downstream traits for expression increasing alleles. Additionally, we flip alleles such that SNPs with positive correlation of genotypes in either  $\Sigma_j^A$  or  $\Sigma_j^B$  (the user must pick which to use) are kept the same, while SNPs with negative correlation of genotypes have their alleles flipped. Allele flipping involves both keeping track of the reference and effect allele, as well as changing the sign of the estimated coefficient. This function also performs checks such that the A and B study agree in terms of the effect and reference allele. If two LD matrices are provided, one is prioritized for generating positive correlations of genotypes, while the other has its alleles flipped for consistency.

### 1.1.3 Scaling

In practice, before supplying  $\hat{\beta}_{i,j}^A$  and  $\hat{\beta}_{i,j}^B$  and the associated standard errors to the colocalization hierarchical model, the values are scaled such that the index SNP (as defined by its absolute value of  $z$ -score) for study A has estimated coefficient of  $\pm 1$  for both study A and B. If two or more SNPs have the same  $z$ -score, the first is chosen. This simplifies the Stan code and improves model fit, as the two studies are then at comparable scale. The scaling is reversed after the Stan model is fit. For the user-input estimated coefficients and standard errors for study  $X \in \{A, B\}$  and cluster  $j$ ,  $\hat{\beta}_{i,j}^{X,\text{input}}$  and  $\text{se}(\hat{\beta}_{i,j}^{X,\text{input}})$ , in the first step we create scaled estimated coefficients:

$$z_j^* = \max_i \left( |\hat{\beta}_{i,j}^{A,\text{input}}| / \text{se}(\hat{\beta}_{i,j}^{A,\text{input}}) \right) \quad (1)$$

$$i_j^* = \min (i \in 1, \dots, n_j) \text{ s.t. } |\hat{\beta}_{i,j}^{A,\text{input}}| / \text{se}(\hat{\beta}_{i,j}^{A,\text{input}}) = z_j^* \quad (2)$$

$$s_j^X = 1 / |\hat{\beta}_{i^*,j}^{X,\text{input}}| \quad (3)$$

$$\hat{\beta}_{i,j}^X = s_j^X \hat{\beta}_{i,j}^{X,\text{input}}, \quad i \in 1, \dots, n_j \quad (4)$$

$$\text{se}(\hat{\beta}_{i,j}^X) = s_j^X \text{se}(\hat{\beta}_{i,j}^{X,\text{input}}), \quad i \in 1, \dots, n_j \quad (5)$$

Note that equations (1-2) refer specifically to study A, while equations (3-5) refer to steps that are repeated for  $X = A$  and  $X = B$ . Again, in words,  $i_j^*$  is the index of the first occurrence of the maximal value of absolute value of  $z$ -score in study A and cluster  $j$ .

### 1.1.4 Colocalization

Colocalization refers to the task of determining if the same signal in eQTL and GWAS summary statistics arise from the same causal variant(s), considering the correlation of genotypes in a locus (the LD matrix). Here we perform colocalization using a generative model for the estimated coefficients, where the true coefficients will be modeled and their posterior distribution used for inference. In the following equations,  $\hat{\beta}_{i,j}^X$  and  $\text{se}(\hat{\beta}_{i,j}^X)$  refer to these scaled estimates as described

in the previous section, but in the following section 1.2 on slope fitting, the values are transformed back to the original scale by multiplying by  $1/s_j^X$  for  $X \in \{A, B\}$  respectively.

We use a statistical model for the summary statistics motivated by the eCAVIAR model [Hormozdiari et al., 2016]. In eCAVIAR, the summary statistic  $z$ -scores in a vector  $S$  are modeled as a multivariate normal distribution with mean vector  $\Sigma\Lambda$  and covariance matrix  $\Sigma$  (the LD matrix), where  $\Lambda$  is a vector giving the true standardized effect sizes. In a locus with a single causal SNP producing the observed enrichment of signal, and if the causal SNP is in the set modeled by eCAVIAR,  $\Lambda$  would consist of a vector with all 0 values except for one SNP with non-zero value. eCAVIAR then models  $\Lambda$  with a multivariate normal prior distribution centered on 0 with covariance matrix based on a vector of 0's and 1's giving the true causal status of the SNPs in the locus and a preset scale parameter determined from previous studies.

Here MRLocus diverges from eCAVIAR in two ways. First, we will use a generative model for the elements in the vector  $\widehat{\beta}_{\cdot,j}^X$  conditional on the true effect sizes  $\beta_{\cdot,j}^X$  within nearly-LD-independent signal cluster  $j$  (that is, MRLocus models effect sizes rather than  $z$ -scores). Second, we will use a univariate distribution for each  $\widehat{\beta}_{i,j}^X$  instead of a multivariate distribution for the entire vector. This second difference is motivated by practical concerns of model fitting and specification; the univariate modeling provided more efficient model fitting with Stan (higher effective sample sizes and R-hat values near 1), and allowed for more flexible choice of priors, as described below.

Instead of a multivariate normal prior on the estimated coefficients, MRLocus uses a horseshoe prior [Carvalho et al., 2009, 2010], a type of hierarchical shrinkage prior that provides sparsity in the posterior estimates [Piironen and Vehtari, 2017]. In the following, we use the notation of Carvalho et al. [2009], where  $\lambda_i$  provides the *local* shrinkage parameters and  $\tau$  provides the *global* shrinkage parameter.

The following hierarchical model is fit separately across nearly-LD-independent signal clusters  $j \in 1, \dots, J$ , and so in the following equations, the subscript for  $j$  is omitted for clarity. In all equations below but the last,  $i \in 1, \dots, n_j$ . Here, for consistency with Stan code, the normal distribution is written as  $N(\mu, \sigma)$  where the second element  $\sigma$  provides the standard deviation instead of the variance.

$$\widehat{\beta}_i^A \sim N([\Sigma^A \beta^A]_i, \text{se}(\widehat{\beta}_i^A)) \quad (6)$$

$$\widehat{\beta}_i^B \sim N([\Sigma^B \beta^B]_i, \text{se}(\widehat{\beta}_i^B)) \quad (7)$$

$$\beta_i^A \sim N(0, \lambda_i \tau) \quad (8)$$

$$\beta_i^B \sim N(0, \lambda_i \tau) \quad (9)$$

$$\lambda_i \sim \text{Cauchy}(0, 1) \quad (10)$$

$$\tau \sim \text{Cauchy}(0, 1) \quad (11)$$

Of note,  $\beta_i^A$  and  $\beta_i^B$  share a prior involving  $\lambda_i$ , such that evidence from study A and B supporting a SNP  $i^\dagger$  that is causal both for eQTL and GWAS signal ( $\beta_{i^\dagger}^A \neq 0, \beta_{i^\dagger}^B \neq 0$ ), will be reflected in larger posterior draws for  $\lambda_{i^\dagger}$  compared to other  $\lambda_{i'}$  for  $i' \neq i^\dagger$ .

The posterior mean for  $\beta_{i,j}^A$  and  $\beta_{i,j}^B$  are the parameters of interest from this step of model fitting, and passed along to the next step after scaling back to the original scale by  $1/s_j^A$  and  $1/s_j^B$ , respectively, as described earlier. We also considered making use of the posterior standard deviation for these two parameters, but found MRLocus gave better performance in terms of accuracy and stability in the Stan fitting procedure if only the posterior mean was kept from the colocalization step. The use of the horseshoe prior in this colocalization step is distinct from other uses of the

horseshoe prior in Bayesian Mendelian Randomization methods, Berzuini et al. [2018] and Uche-Ikonne et al. [2019], where it is used as a prior for the pleiotropic effects (effects not mediated by the exposure) or on the mediation slope, respectively.

### 1.1.5 Colocalization with eCAVIAR (optional alternative)

Alternatively, MRLocus can accept colocalization results from eCAVIAR as input to the slope fitting step, as explored in the main text. eCAVIAR is run with default options on each nearly-LD-independent signal cluster separately, supplying the LD matrix, and z-scores for study A and B, and  $-c 1$ , i.e. setting the maximum number of causal SNPs to 1. The following section describes how the output from eCAVIAR colocalization can be used alternatively to the MRLocus colocalization output.

## 1.2 Slope fitting step

The second step of MRLocus is to estimate the gene-to-trait effect  $\alpha$  (the slope in a Mendelian Randomization analysis) using the posterior mean values from the colocalization step. For each nearly-LD-independent signal cluster  $j$ , MRLocus extracts one SNP, based on the largest posterior mean value for  $\beta_{i,j}^A$  (prioritizing the putative mediator A for SNP selection). MRLocus also has the non-default option to perform model-based clustering using an EM algorithm to determine how many SNPs per nearly-LD-independent signal cluster to pass to the slope fitting step [Scrucca et al., 2016], but this option was not extensively evaluated here. For eCAVIAR colocalization, the SNP with the largest colocalization posterior probability (CLPP) is selected from each signal cluster, among those SNPs which are valid instruments, i.e. with an absolute value of z-score as large as the value used for PLINK clumping [Purcell et al., 2007]. eCAVIAR alleles are flipped such that the alternate allele is the one corresponding to an increase in the measured phenotype in A (e.g. gene expression). Another round of  $r^2$  trimming is performed, as in section 1.1.1, so that all chosen SNPs for slope fitting have pairwise  $r^2 < 0.05$ .

By default,  $J$  SNPs are passed from the colocalization step to the slope fitting step, which should represent the candidate causal SNPs from each nearly-LD-independent signal cluster. The posterior mean for  $\beta_{i,j}^A$  and  $\beta_{i,j}^B$  for these SNPs is then provided to the slope fitting step as variables  $\hat{\beta}_j^A$  and  $\hat{\beta}_j^B$ . These two variables are modeled as normally distributed variables centered on true values  $\beta_j^A$  and  $\beta_j^B$  with standard deviation according to the original standard errors  $se(\hat{\beta}_{i,j}^A)$  and  $se(\hat{\beta}_{i,j}^B)$ . We found this incorporation of the original standard errors into the slope fitting procedure helped to accurately estimate the uncertainty on the slope (the gene-to-trait effect).

The primary two parameters of interest are the slope  $\alpha$ , i.e. the predominant gene-to-trait effect demonstrated by the nearly-LD-independent signal clusters in the locus, and  $\sigma$ , the dispersion of gene-to-trait effects from individual signal clusters around the predominant slope. The hierarchical model for the slope fitting step is given by:

$$\widehat{\beta}_j^A \sim N(\beta_j^A, se_j^A) \quad (12)$$

$$\widehat{\beta}_j^B \sim N(\beta_j^B, se_j^B) \quad (13)$$

$$\beta_j^A \sim N(0, SD_{\beta}) \quad (14)$$

$$\beta_j^B \sim N(\alpha\beta_j^A, \sigma) \quad (15)$$

$$\alpha \sim N(\mu_{\alpha}, SD_{\alpha}) \quad (16)$$

$$\sigma \sim HN(0, SD_{\sigma}) \quad (17)$$

where the first four equations are defined for  $j \in 1, \dots, J$ , and where  $HN$  specifies the Half-Normal distribution. In words,  $\beta_j^B$  is assumed to follow a normal distribution centered on the predicted value from a line with slope  $\alpha$  and no intercept, and with dispersion  $\sigma$  around the fitted line. We are both interested in the posterior mean of  $\alpha$  and  $\sigma$  as well as our uncertainty regarding the point estimates. In particular we focus on a quantile-based credible interval for  $\alpha$ , which is used for determining our confidence in a gene being a causal mediator for a trait. Large values of  $\sigma$  relative to  $\alpha$  times typical mediator perturbation sizes ( $\beta_j^A$ ) reflect significant heterogeneity of the gene-to-trait effect.

### 1.2.1 Choice of hyperparameters

The hyperparameter for the prior for  $\beta_j^A$  is  $SD_{\beta}$ , which is set to 2 times the largest absolute value of  $\widehat{\beta}_j^A$ . The hyperparameters for the prior for  $\alpha$  are  $\mu_{\alpha}$  and  $SD_{\alpha}$ , and are set based on a simple un-weighted linear model of  $\widehat{\beta}_j^B$  on  $\widehat{\beta}_j^A$  without an intercept.  $\mu_{\alpha}$  is set to the estimated slope coefficient, and  $SD_{\alpha}$  is set to 2 times the absolute value of this estimated slope coefficient.

The hyperparameter for the prior for  $\sigma$  is  $SD_{\sigma}$ , which is set to  $\max(2sd(\widehat{\beta}_j^B), \max_j(|\widehat{\beta}_j^B|))$ , where  $sd(.)$  denotes the sample standard deviation. That is, the larger value between (i) 2 times the standard deviation of the coefficients for B and (ii) the largest absolute value of coefficient for B. When the coefficients for B are widely dispersed, then taking two times the SD results in a wide prior for  $\sigma$ , and when the coefficients are not widely dispersed, then taking the absolute value of the largest coefficient ensures the prior width for  $\sigma$  does not become too small.

Prior predictive checks [Gabry et al., 2019], in which data is generated from the estimated prior using the data generating mechanism, are provided in the MRLocus package and demonstrated in the vignette.

### 1.2.2 Loci with only one signal cluster ( $J = 1$ )

It is not recommended to use MRLocus with only one signal cluster.

### 1.2.3 Variation in $\alpha$ in simulations

We note that the value for true slope parameter  $\alpha$  (x-axis in the simulation assessment plots) varies across iterations in the simulation for fixed h2med and h2g. The reason for this variation is that the  $\alpha$  provided by the `twas_sim` simulation framework corresponds to the gene-to-trait mediated effect, when eQTL and GWAS effect sizes are standardized using the SD of gene expression and trait in the eQTL and GWAS samples (as would occur in real data analysis). The h2g and h2med in the `twas_sim` simulator describe the population-level expression heritability and mediated trait

heritability. As the eQTL and GWAS sample sizes grow (e.g. already seen in the eQTL  $N = 1,000$  simulation), the variation is reduced and we have  $\alpha$  values converging to  $\sqrt{\frac{h2med}{h2g}}$ .

## 2 MRLocus Stan code

### 2.1 Colocalization step

inst/stan/beta\_coloc.stan

```
1 data {
2     int n;
3     vector[n] beta_hat_a;
4     vector[n] beta_hat_b;
5     vector[n] se_a;
6     vector[n] se_b;
7     matrix[n,n] Sigma_a;
8     matrix[n,n] Sigma_b;
9 }
10 parameters {
11     vector[n] beta_a;
12     vector[n] beta_b;
13     vector<lower=0>[n] lambda;
14     real<lower=0> tau;
15 }
16 model {
17     tau ~ cauchy(0, 1);
18     lambda ~ cauchy(0, 1);
19     beta_hat_a ~ normal(Sigma_a * beta_a, se_a);
20     beta_hat_b ~ normal(Sigma_b * beta_b, se_b);
21     for (i in 1:n) {
22         beta_a[i] ~ normal(0, lambda[i] * tau);
23         beta_b[i] ~ normal(0, lambda[i] * tau);
24     }
25 }
```

## 2.2 Slope fitting step

inst/stan/slope.stan

```
1 data {
2     int n;
3     vector[n] beta_hat_a;
4     vector[n] beta_hat_b;
5     vector[n] sd_a;
6     vector[n] sd_b;
7     real sd_beta;
8     real mu_alpha;
9     real sd_alpha;
10    real sd_sigma;
11 }
12 parameters {
13     real alpha;
14     real<lower=0> sigma;
15     vector[n] beta_a;
16     vector[n] beta_b;
17 }
18 model {
19     beta_hat_a ~ normal(beta_a, sd_a);
20     beta_hat_b ~ normal(beta_b, sd_b);
21     beta_a ~ normal(0, sd_beta);
22     beta_b ~ normal(alpha * beta_a, sigma);
23     alpha ~ normal(mu_alpha, sd_alpha);
24     sigma ~ normal(0, sd_sigma);
25 }
```

## References

- Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32, 2017. doi:10.18637/jss.v076.i01.
- Stan Development Team. RStan: the R interface to Stan, 2020. URL <http://mc-stan.org/>. R package version 2.21.1.
- Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I.W. de Bakker, Mark J. Daly, and Pak C. Sham. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81(3), 2007. doi:10.1086/519795.
- Pejman Mohammadi, Stephane E. Castel, Andrew A. Brown, and Tuuli Lappalainen. Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. *Genome Research*, 27(11):1872–1884, 2017. doi:10.1101/gr.216747.116.
- John Palowitch, Andrey Shabalin, Yi-Hui Zhou, Andrew B. Nobel, and Fred A. Wright. Estimation of cis-eQTL effect sizes using a log of linear model. *Biometrics*, 74(2):616–625, 2018. doi:10.1111/biom.12810.
- Farhad Hormozdiari, Martijn van de Bunt, Ayellet V. Segrè, Xiao Li, Jong Wha J. Joo, Michael Bilow, Jae Hoon Sul, Sriram Sankararaman, Bogdan Pasaniuc, and Eleazar Eskin. Colocalization of gwas and eqtl signals detects target genes. *The American Journal of Human Genetics*, 99(6):1245–1260, 2016. doi:10.1016/j.ajhg.2016.10.003.
- Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. Handling sparsity via the horseshoe. *Proceedings of Machine Learning Research*, 5:73–80, 2009. URL <http://proceedings.mlr.press/v5/carvalho09a.html>.
- Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010. doi:10.1093/biomet/asq017.
- Juho Piironen and Aki Vehtari. On the Hyperprior Choice for the Global Shrinkage Parameter in the Horseshoe Prior. *Proceedings of Machine Learning Research*, 54:905–913, 2017. URL <http://proceedings.mlr.press/v54/piironen17a.html>.
- Carlo Berzuini, Hui Guo, Stephen Burgess, and Luisa Bernardinelli. A Bayesian approach to Mendelian randomization with multiple pleiotropic variants. *Biostatistics*, 21(1):86–101, 2018. doi:10.1093/biostatistics/kxy027.
- Okezie O Uche-Ikonne, Frank Dondelinger, and Tom Palmer. Bayesian estimation of IVW and MR-Egger models for two-sample Mendelian randomization studies. *medRxiv*, 2019. doi:10.1101/19005868.
- Luca Scrucca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1):289–317, 2016. doi:10.32614/RJ-2016-021.

Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402, 2019. doi:10.1111/rssa.12378.

## Supplementary Tables and Figures

Anqi Zhu<sup>†</sup>, Nana Matoba<sup>†</sup>, Emmaleigh Wilson, Amanda L. Tapia,  
Yun Li, Joseph G. Ibrahim, Jason L. Stein, Michael I. Love

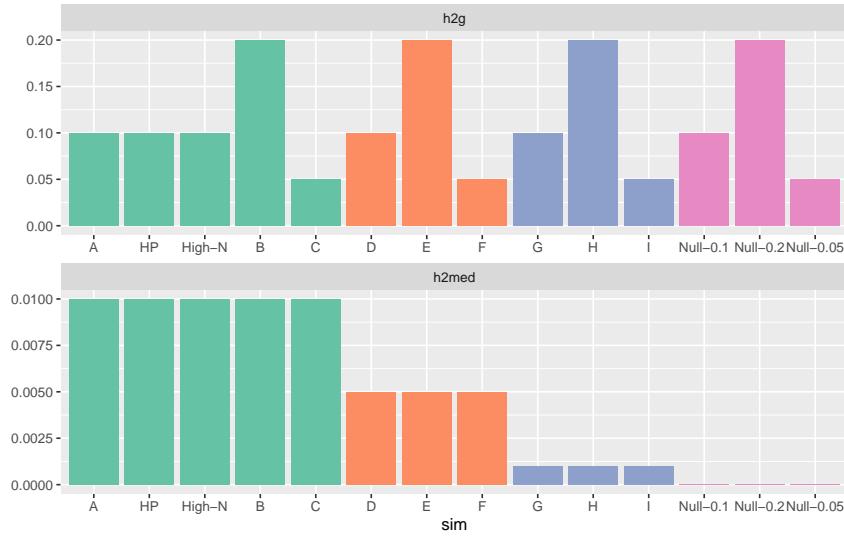
January 25, 2021

### Supplementary Tables

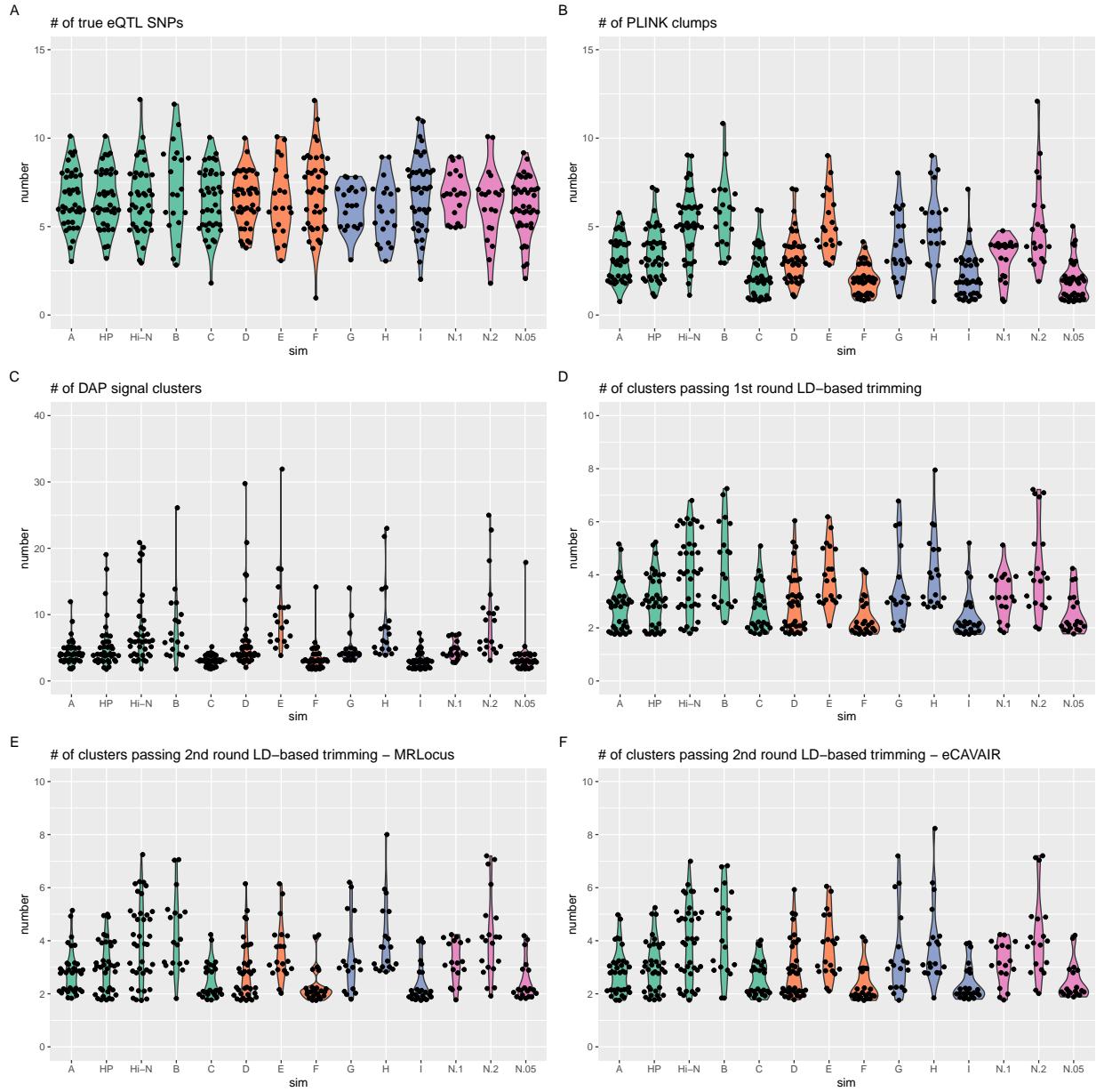
eQTL Tissue	Resource
Artery Tibial (GTEx v8)	<a href="https://console.cloud.google.com/storage/browser/gtex-resources">https://console.cloud.google.com/storage/browser/gtex-resources</a>
Blood (eQTLGen 2019-12-23 release)	<a href="https://www.eqtlgen.org/cis-eqtls.html">https://www.eqtlgen.org/cis-eqtls.html</a>
Liver	<a href="https://www.nature.com/articles/s41598-018-24219-z">https://www.nature.com/articles/s41598-018-24219-z</a>
GWAS Phenotype	Resource
CAD (CARDIoGRAMplusC4D)	<a href="http://www.cardiogramplusc4d.org/media/cardioimageplusc4d-consortium/data-downloads/cad.additive.Oct2015.pub.zip">http://www.cardiogramplusc4d.org/media/cardioimageplusc4d-consortium/data-downloads/cad.additive.Oct2015.pub.zip</a>
HDL (UKBB)	<a href="https://www.dropbox.com/s/65jisgxwbbdrkaw/30760_irnt.gwas.imputed_v3.both_sexes.tsv.bgz">https://www.dropbox.com/s/65jisgxwbbdrkaw/30760_irnt.gwas.imputed_v3.both_sexes.tsv.bgz</a>
LDL (UKBB)	<a href="https://www.dropbox.com/s/4rnjzcwjgs5pg1/30780_irnt.gwas.imputed_v3.both_sexes.tsv.bgz">https://www.dropbox.com/s/4rnjzcwjgs5pg1/30780_irnt.gwas.imputed_v3.both_sexes.tsv.bgz</a>

Supplementary Table 1: Links to URLs for eQTL and GWAS summary data employed in MRLocus real data evaluation.

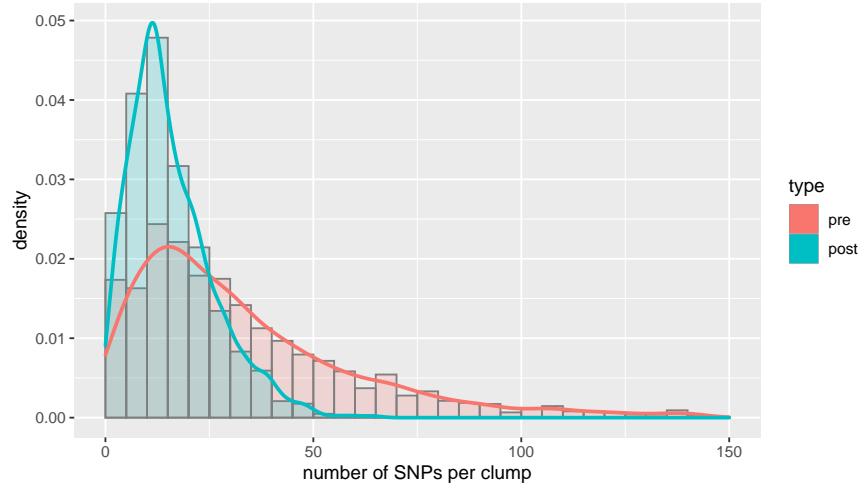
## Supplementary Figures



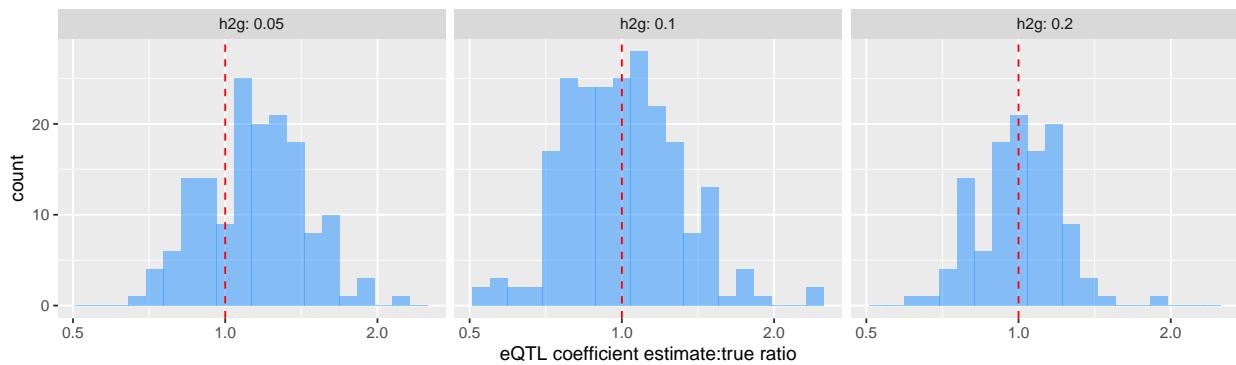
Supplementary Figure 1: Diagram of the 14 types of `twas_sim` simulations performed, varying expression heritability  $h^2g$  (top row) and expression mediated heritability  $h^2med$  (bottom row). Results for simulation A are presented in Figure 2 in the main text, while results for the remaining 13 simulation settings follow in Supplementary Figures.



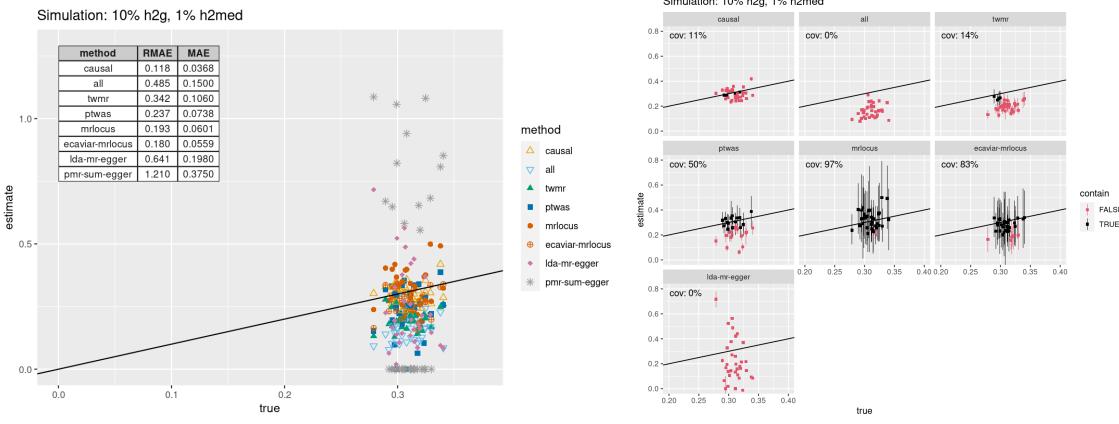
Supplementary Figure 2: Number of (A) true causal eQTL SNPs, (B) PLINK clumps, (C) DAP signal clusters per simulation, and number of clusters passing (D) 1st round LD trimming, (E) 2nd round LD trimming for MRLocus colocalization, and (F) 2nd round LD trimming for eCAVIAR colocalization.



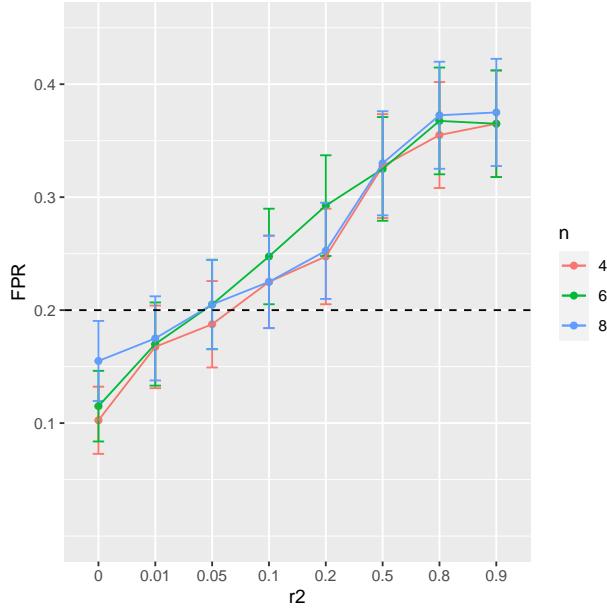
Supplementary Figure 3: SNPs per clump in the 240 simulations, before and after collapsing with MRLocus `collapseHighCorSNPs` function. The mean number of SNPs per clump was 19.9 and 8.8, and the median number of SNPs per clump was 15 and 8 (before and after, respectively).



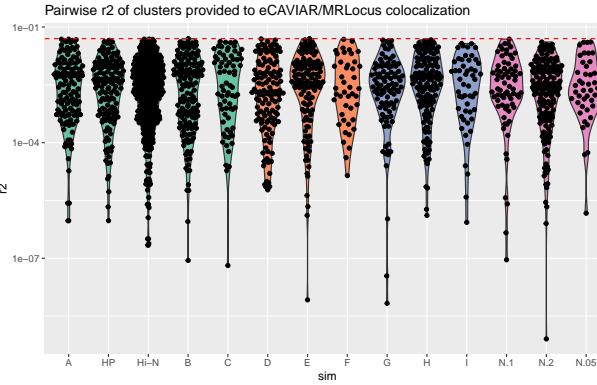
Supplementary Figure 4: Ratio of estimated eQTL coefficient over true value across simulations, grouped by  $h^2g$  parameter. Histograms represent the ratio for all causal eSNPs with un-adjusted eQTL p-value less than the clumping threshold (0.001). The median and interquartile range (IQR) of the distributions (on the original ratio scale) are (1.15, 0.37), (1.00, 0.35), and (1.01, 0.25), for the 5%, 10%, and 20%  $h^2g$  settings, respectively.



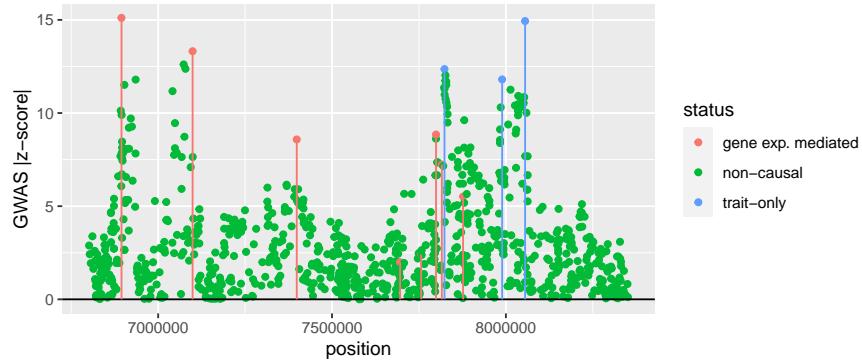
Supplementary Figure 5: Accuracy and interval coverage of gene-to-trait effect estimation in simulation A, including LDA-MR-Egger and PMR-Summary-Egger in comparisons. As PMR-Summary-Egger does not provide a standard error for the causal effect, interval coverage was only computed for LDA-MR-Egger.



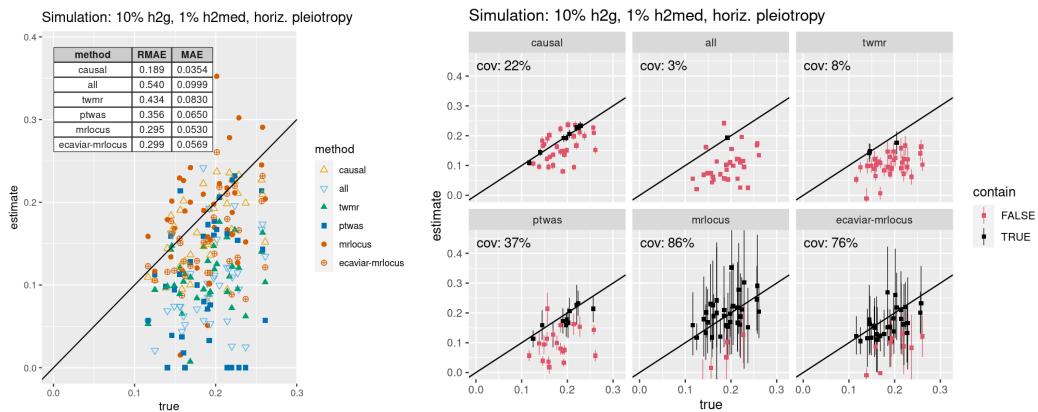
Supplementary Figure 6: Assessment of the effect of correlated instruments on the false positive rate of MRLocus slope estimation. Increasing the  $r^2$  of adjacent signal clusters raised the rate of 80% credible intervals not covering the true value of the slope  $\alpha = 0$ . The number of signal clusters was varied among  $\{4, 6, 8\}$ . Bars represent 95% binomial-based confidence intervals, and 400 simulation iterations were performed for each combination of  $r^2$  and the number of signal clusters.



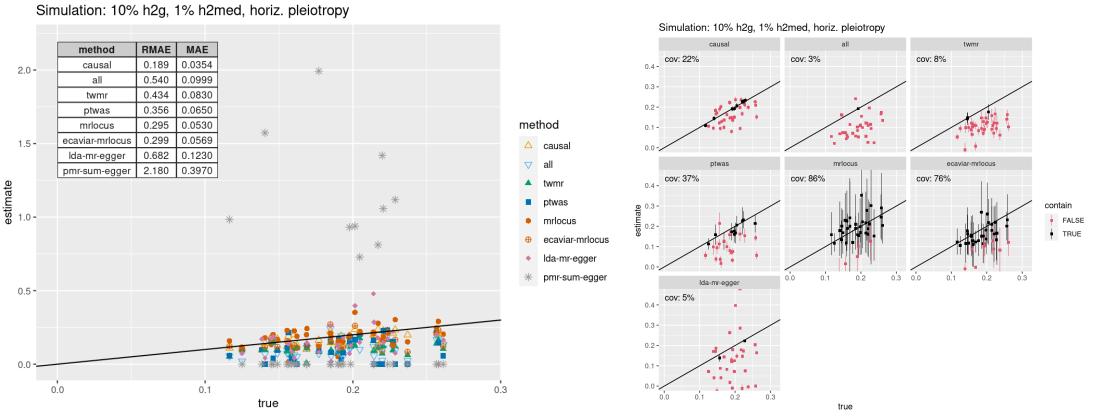
Supplementary Figure 7: Pairwise  $r^2$  across clusters for the instruments provided to MRLocus slope fitting, across all iterations and all simulated datasets. Clusters were trimmed such that  $r^2 < 0.05$ .



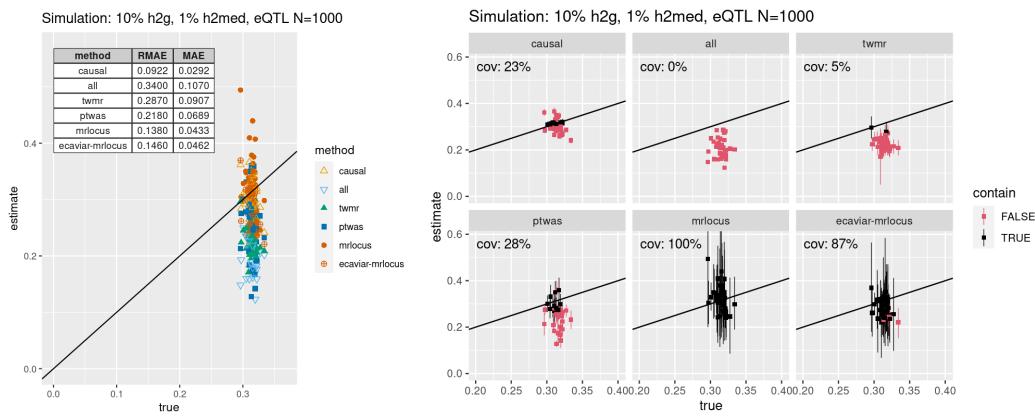
Supplementary Figure 8: Example of one of the iterations of the simulation of partial mediation with horizontal pleiotropy (HP), where three large trait-only association signals are added to a simulation with  $h^2g = 10\%$  and  $h^2med = 1\%$ . Absolute Z-scores for the GWAS population are calculated from estimated coefficients and their standard errors, and colored by the true status in the simulation.



Supplementary Figure 9: Accuracy and interval coverage of TWMR, PTWAS, MRLocus, and eCAVIAR-MRLocus on the simulation of partial mediation with horizontal pleiotropy (as in the example region in Supplementary Figure 8).

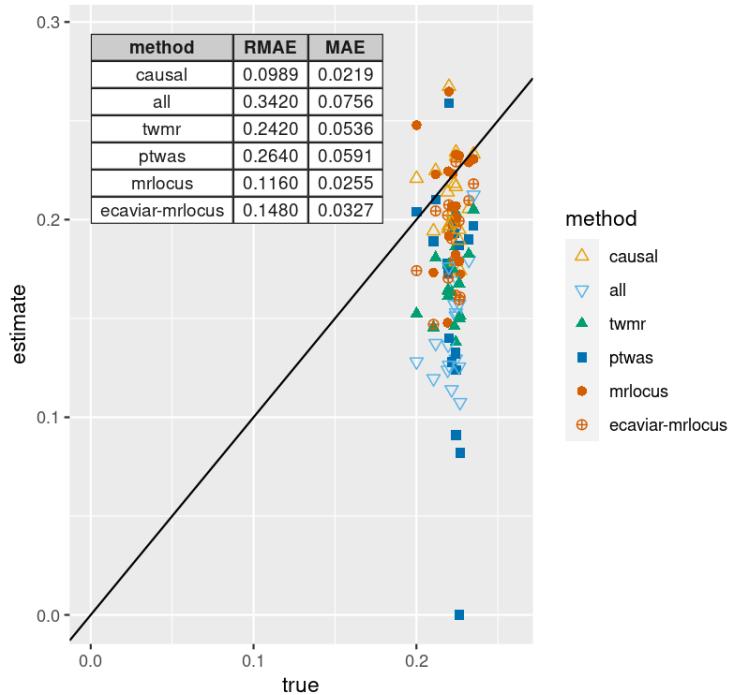


Supplementary Figure 10: Accuracy and interval coverage of methods on the simulation of partial mediation with horizontal pleiotropy, including LDA-MR-Egger and PMR-Summary-Egger. As PMR-Summary-Egger does not provide a standard error for the causal effect, interval coverage was only computed for LDA-MR-Egger.



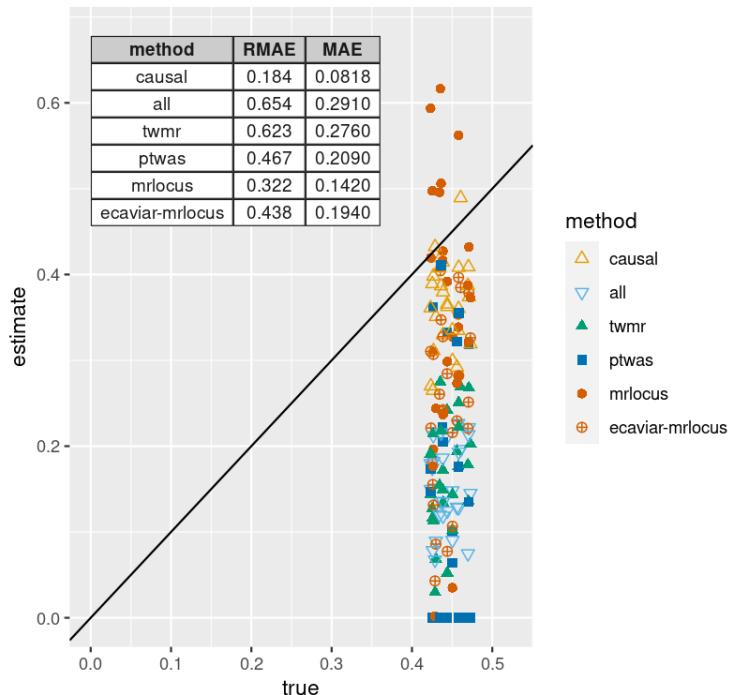
Supplementary Figure 11: Accuracy and interval coverage of TWMR, PTWAS, MRLocus, and eCAVIAR-MRLocus on the same simulation settings as simulation A but with increased  $N_{eQTL} = 1000$  (instead of the default value  $N_{eQTL} = 500$ ).

Simulation: 20% h2g, 1% h2med



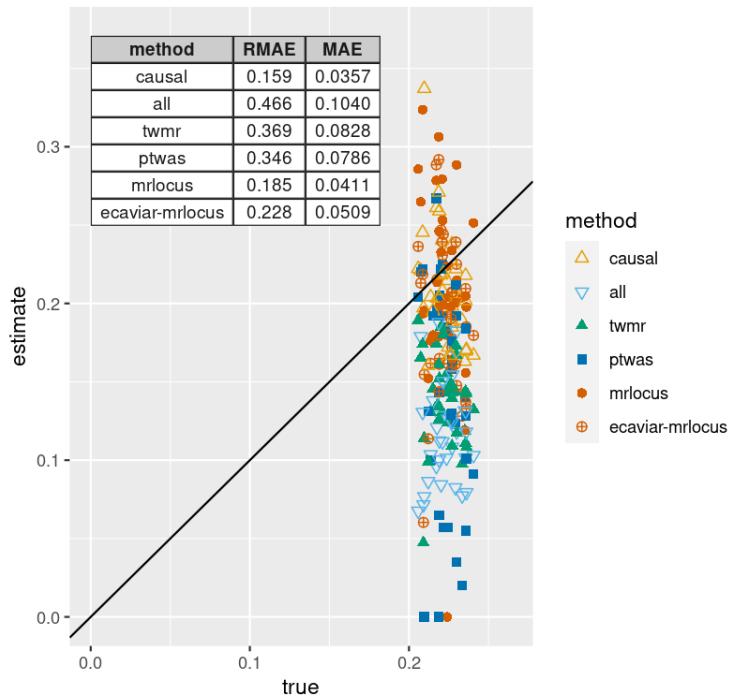
Supplementary Figure 12: Accuracy of gene-to-trait effect estimation in simulation B.

Simulation: 5% h2g, 1% h2med



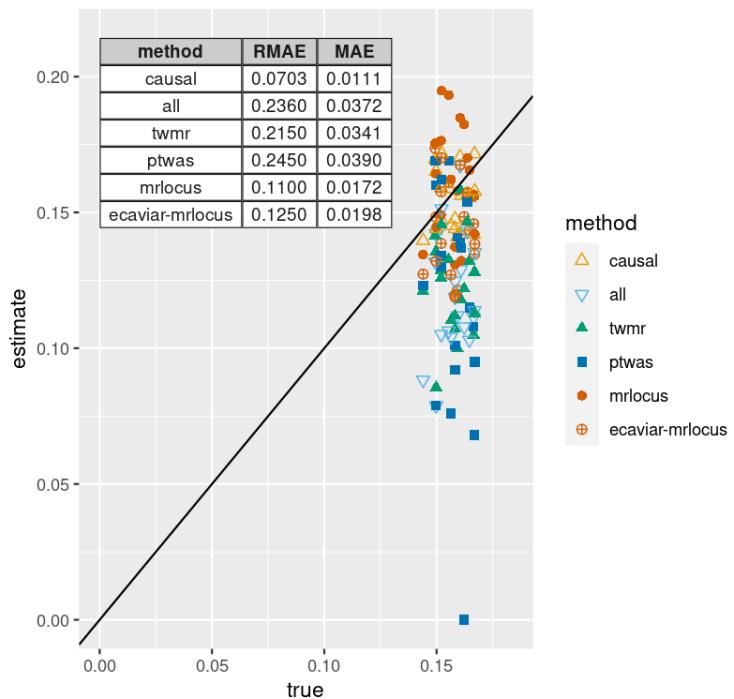
Supplementary Figure 13: Accuracy of gene-to-trait effect estimation in simulation C.

Simulation: 10% h2g, 0.5% h2med



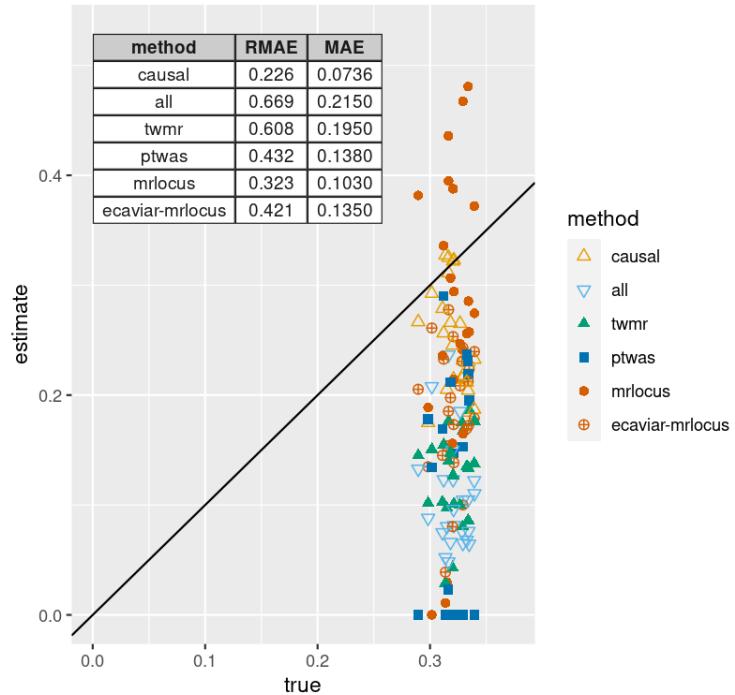
Supplementary Figure 14: Accuracy of gene-to-trait effect estimation in simulation D.

Simulation: 20% h2g, 0.5% h2med



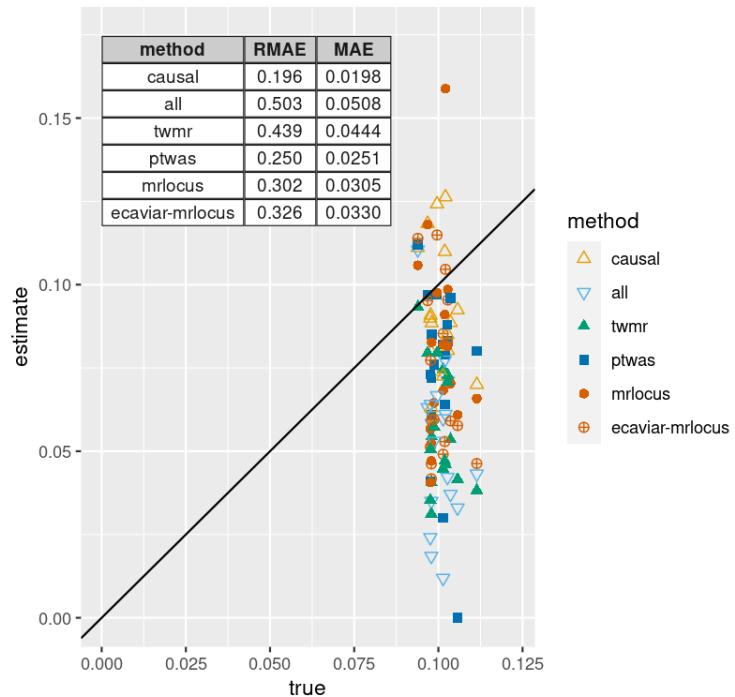
Supplementary Figure 15: Accuracy of gene-to-trait effect estimation in simulation E.

Simulation: 5% h2g, 0.5% h2med



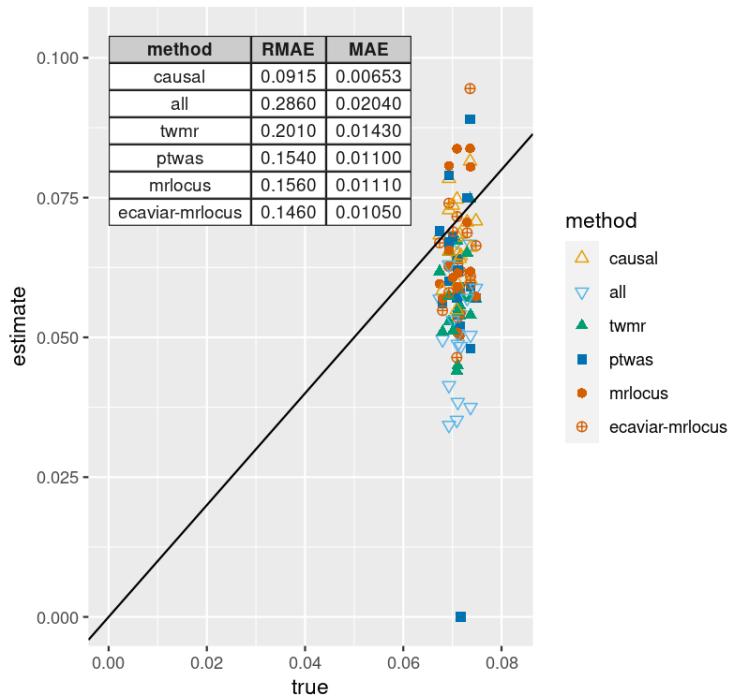
Supplementary Figure 16: Accuracy of gene-to-trait effect estimation in simulation F.

Simulation: 10% h2g, 0.1% h2med



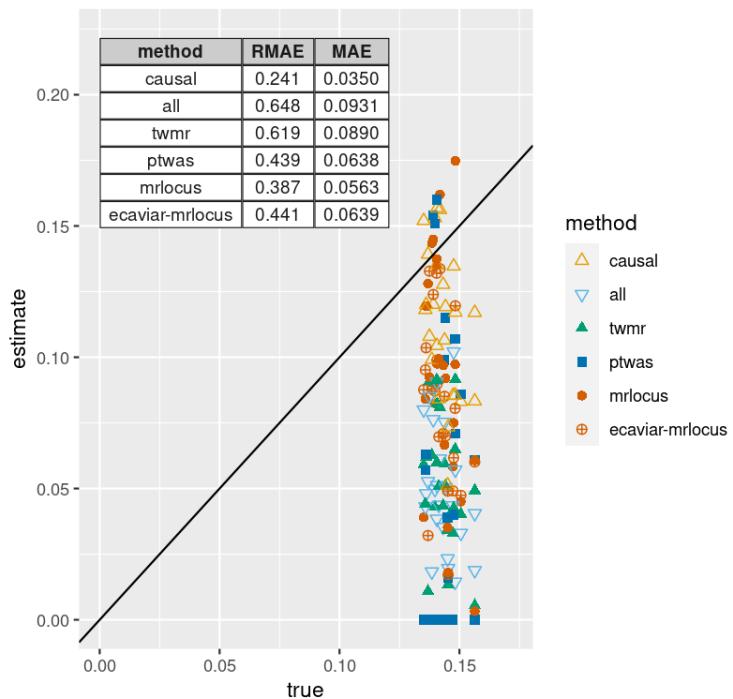
Supplementary Figure 17: Accuracy of gene-to-trait effect estimation in simulation G.

Simulation: 20% h2g, 0.1% h2med



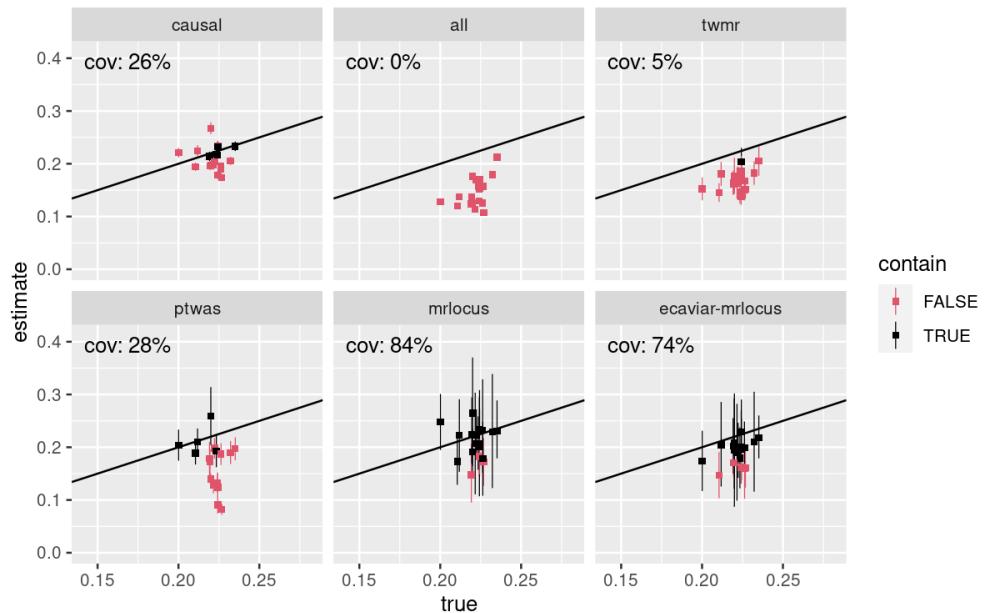
Supplementary Figure 18: Accuracy of gene-to-trait effect estimation in simulation H.

Simulation: 5% h2g, 0.1% h2med



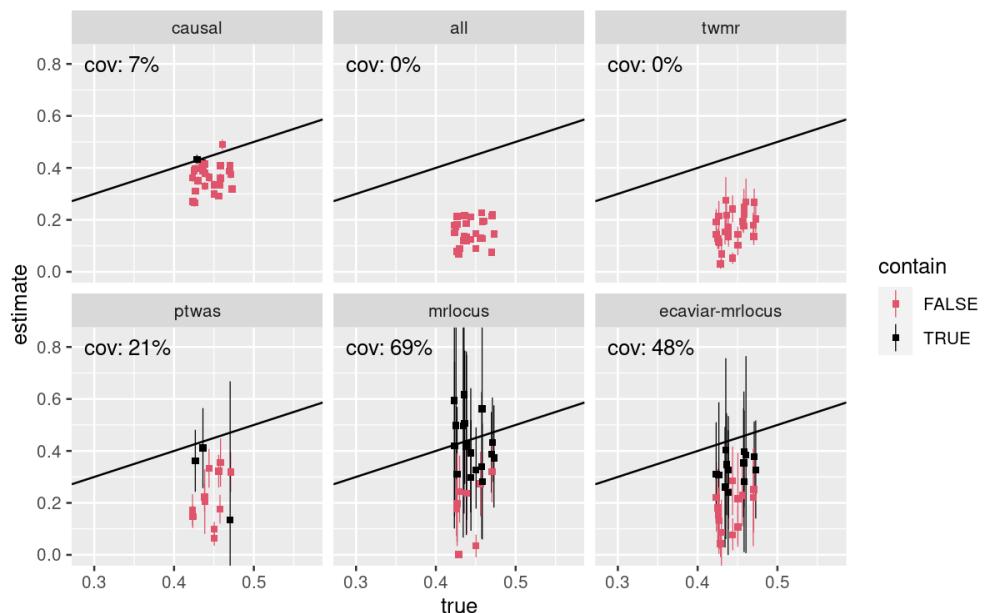
Supplementary Figure 19: Accuracy of gene-to-trait effect estimation in simulation I.

Simulation: 20% h2g, 1% h2med



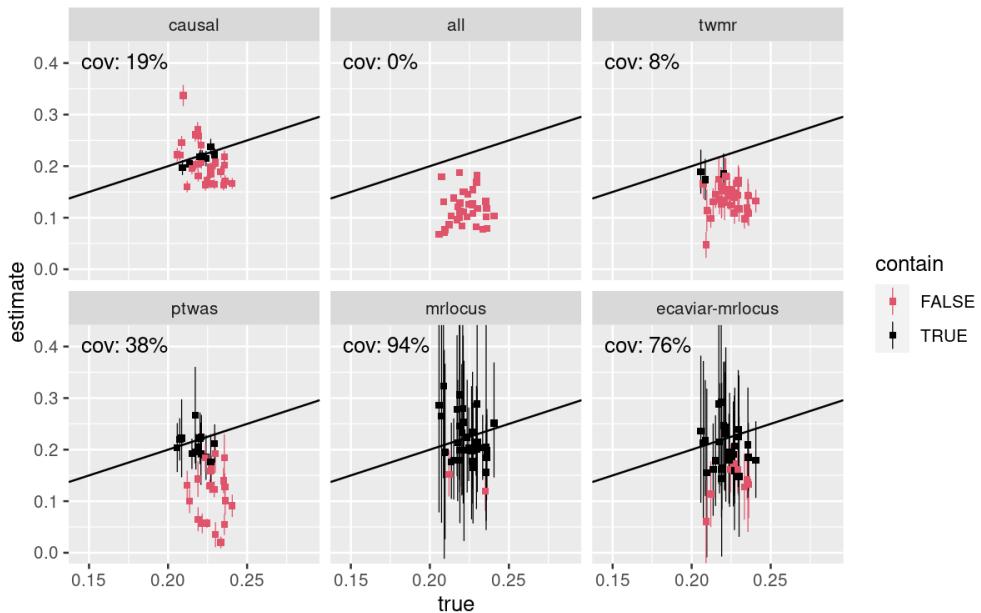
Supplementary Figure 20: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation B.

Simulation: 5% h2g, 1% h2med



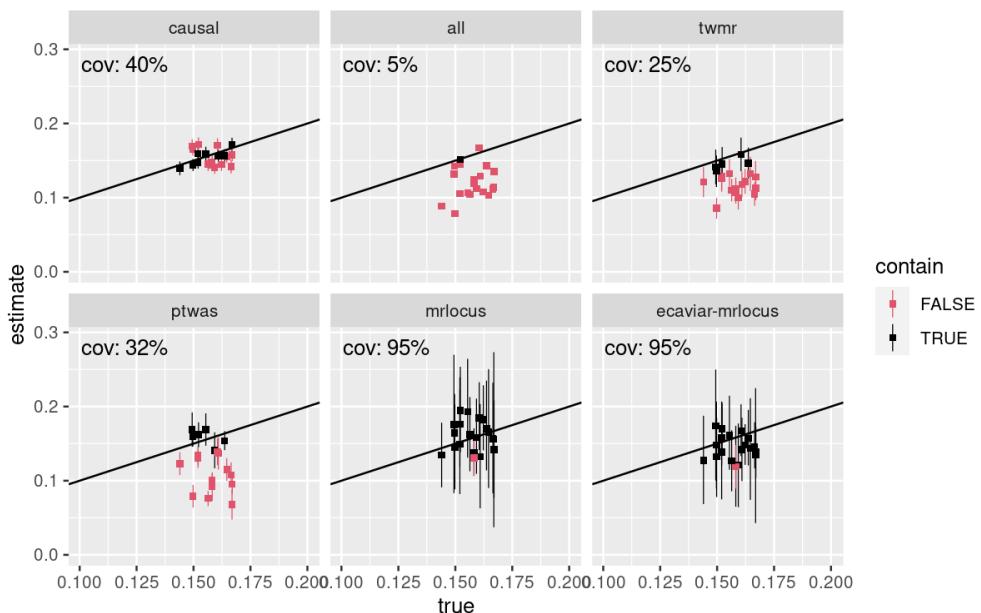
Supplementary Figure 21: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation C.

Simulation: 10% h2g, 0.5% h2med



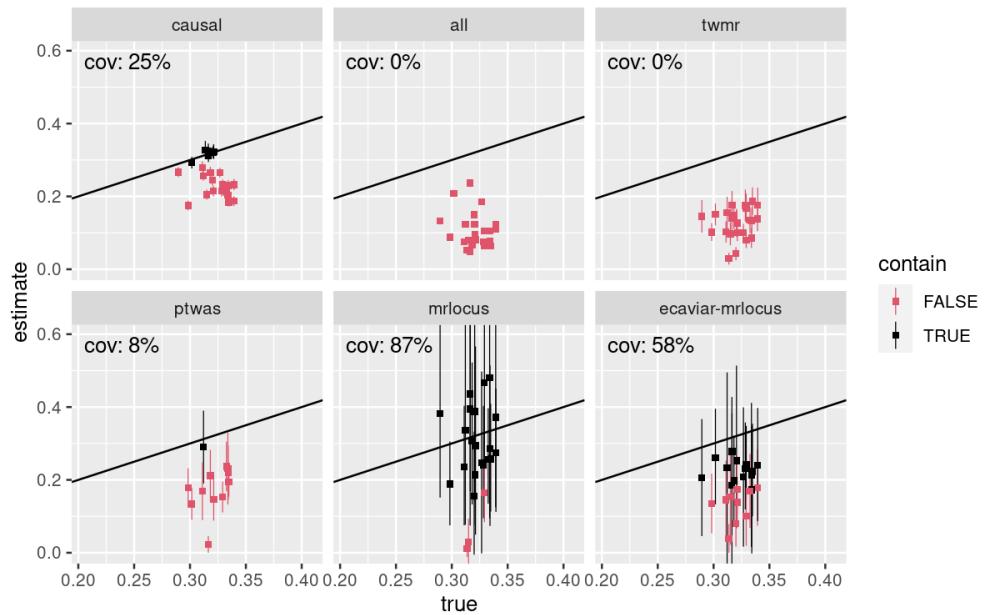
Supplementary Figure 22: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation D.

Simulation: 20% h2g, 0.5% h2med



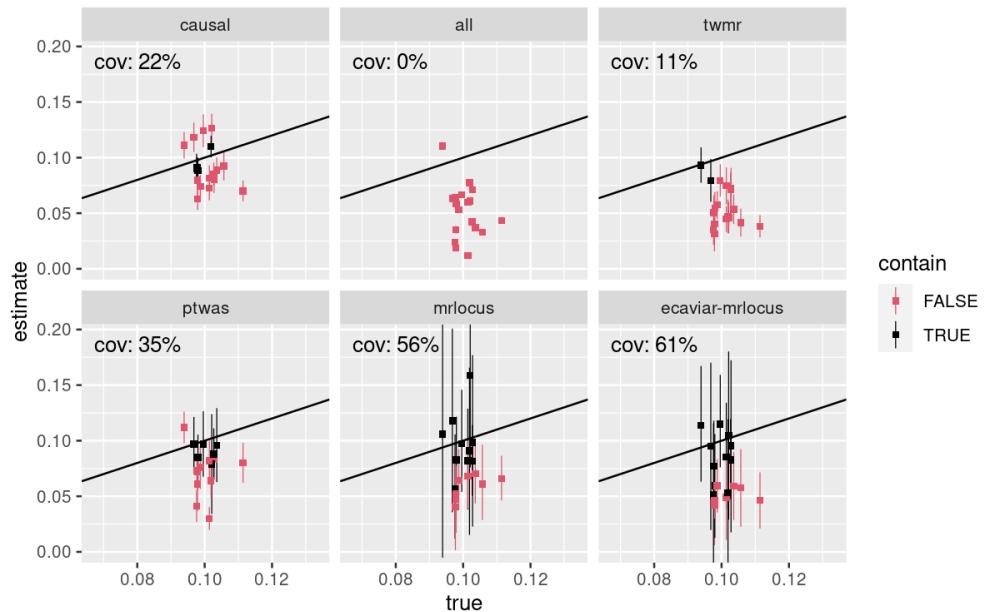
Supplementary Figure 23: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation E.

Simulation: 5% h2g, 0.5% h2med



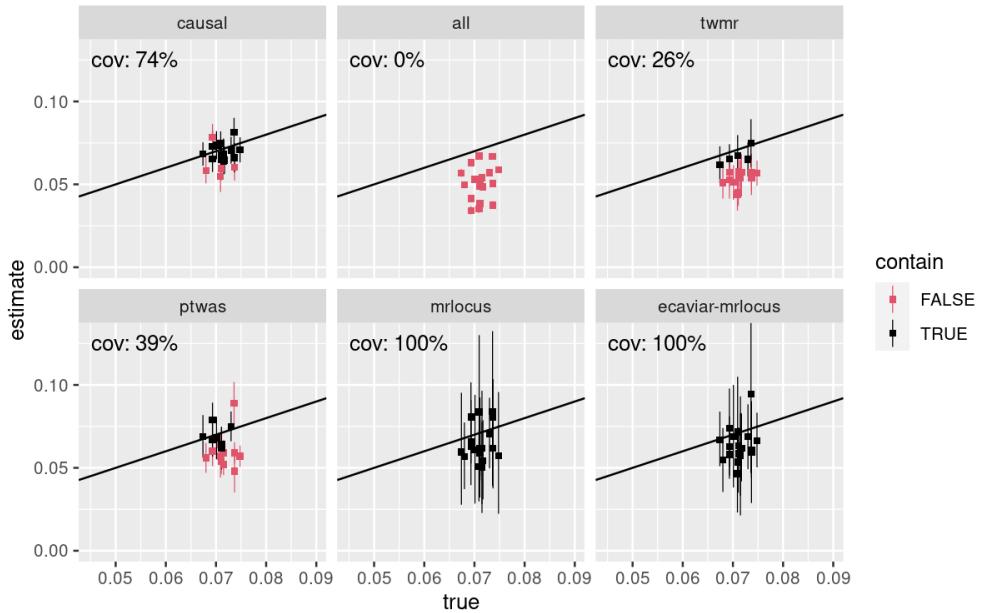
Supplementary Figure 24: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation F.

Simulation: 10% h2g, 0.1% h2med



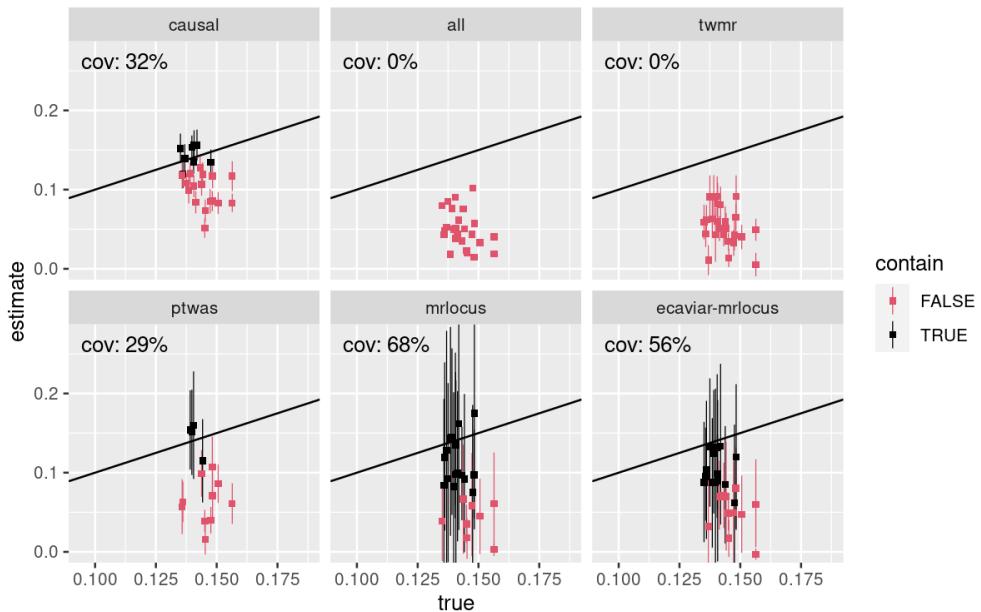
Supplementary Figure 25: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation G.

Simulation: 20% h2g, 0.1% h2med

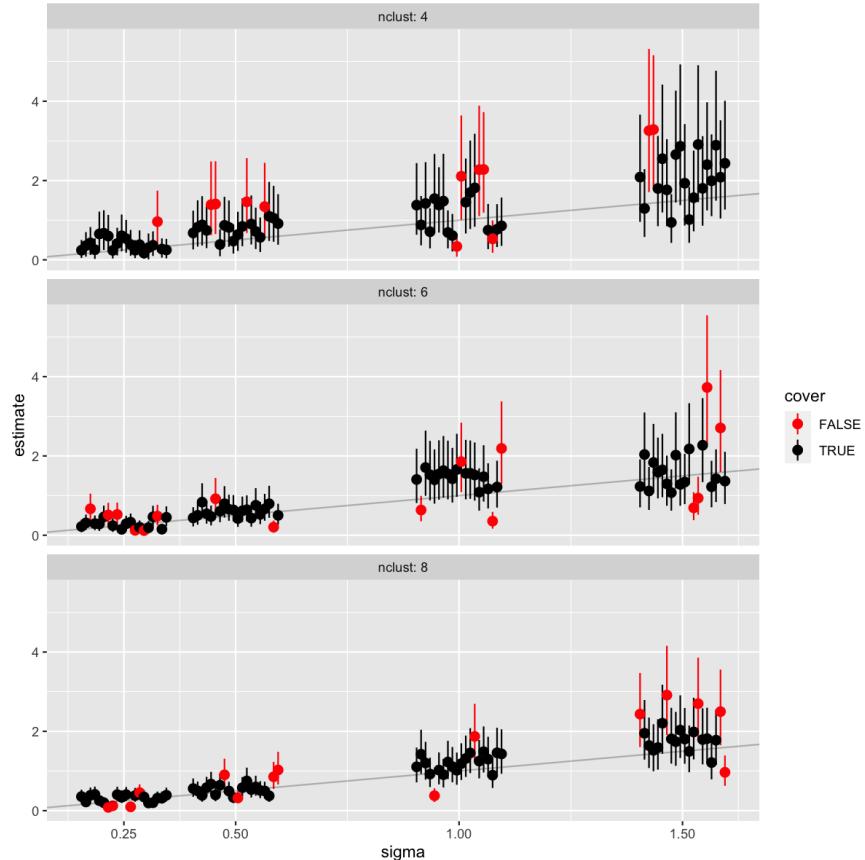


Supplementary Figure 26: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation H.

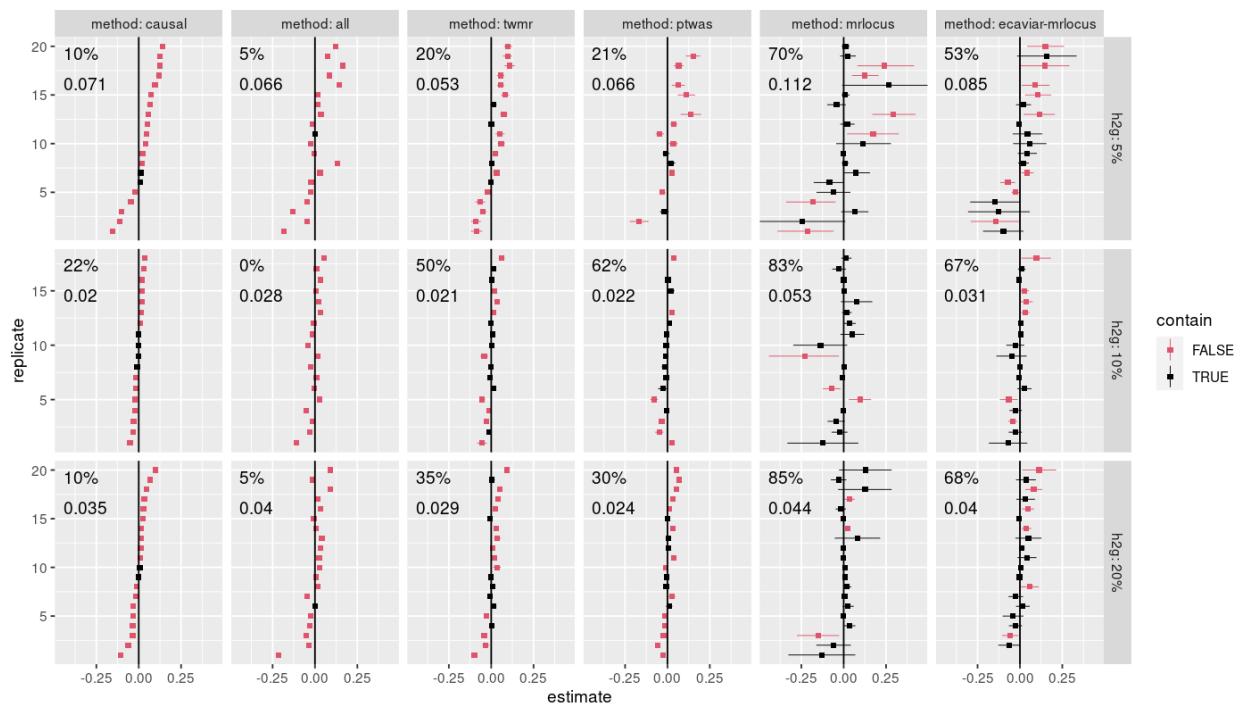
Simulation: 5% h2g, 0.1% h2med



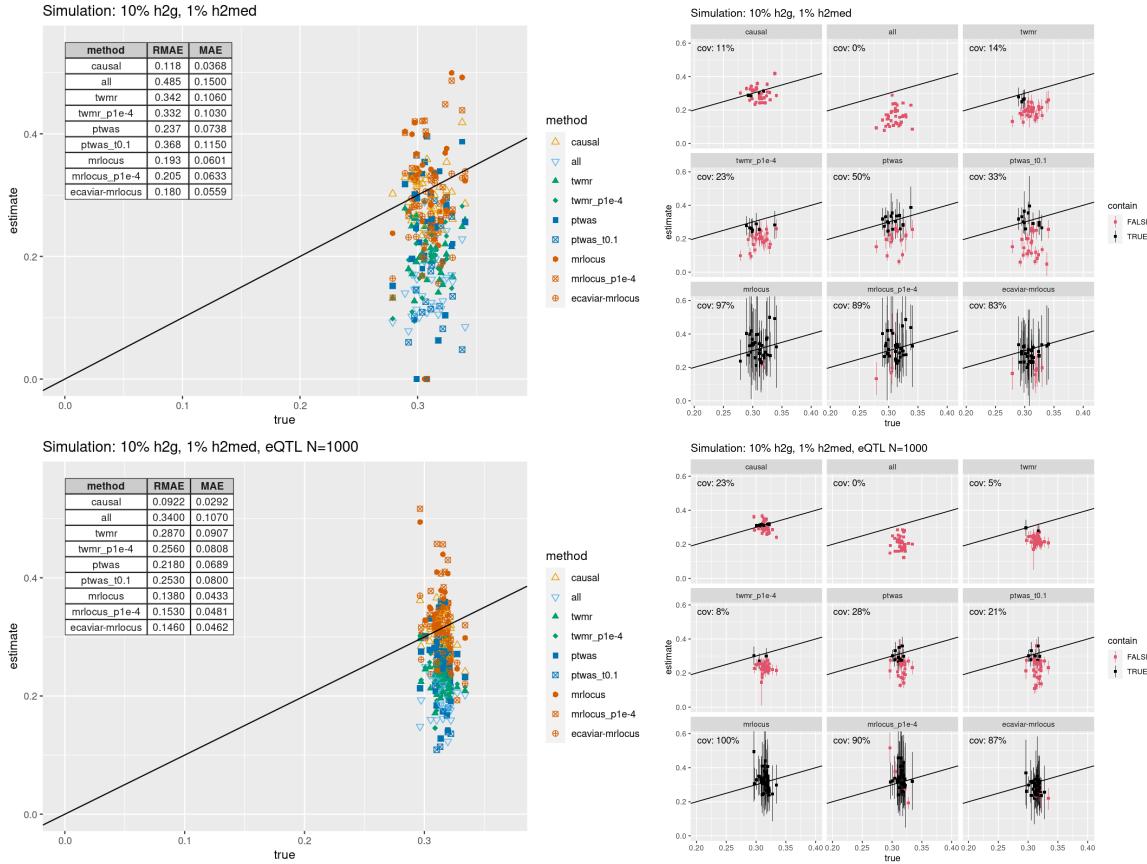
Supplementary Figure 27: Coverage of confidence or credible intervals for the gene-to-trait effect in simulation I.



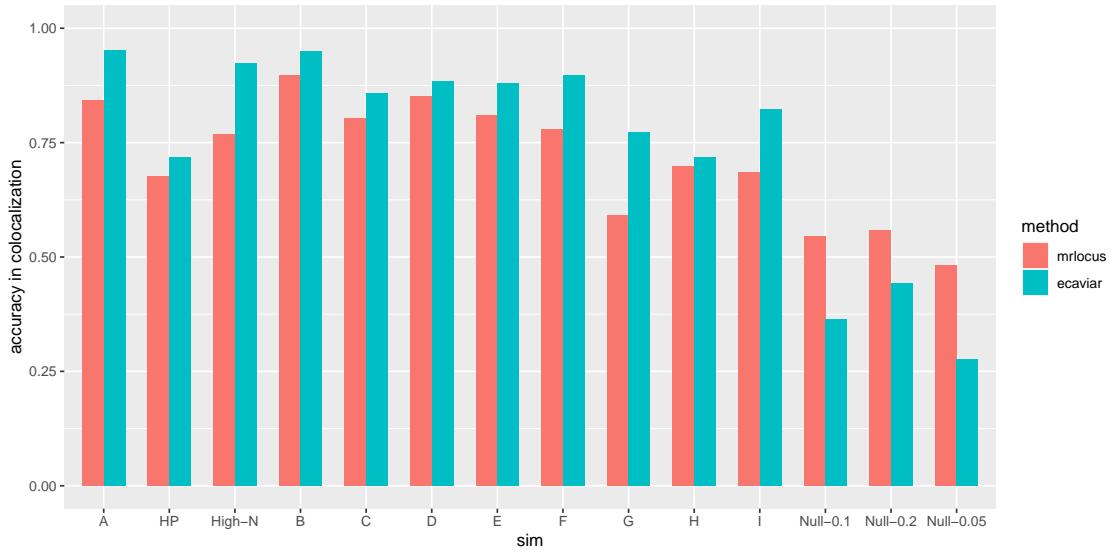
Supplementary Figure 28: Simulation assessing MRLocus' estimation of the scale of dispersion ( $\sigma$ ) across independent signal clusters. The summary statistics for eQTL and GWAS were generated from a multivariate normal distribution as in the eCAVIAR model, using a simulated LD matrix. The true slope ( $\alpha$ ) was set to 1, the true  $\sigma$  varied among the values  $\{0.25, 0.5, 1, 1.5\}$  (x-axis), and the number of LD-independent clusters varied between 4, 6, and 8 (top, middle, bottom panels), with 20 iterations per setting (plotted with horizontal spacing to avoid overplotting). The posterior mean is indicated with a dot, while 80% quantile-based credible intervals and their coverage of the true value are indicated with the line and its color. The simulation script is included within the MRLocus package test directory, as an R script "test\_sigma.R".



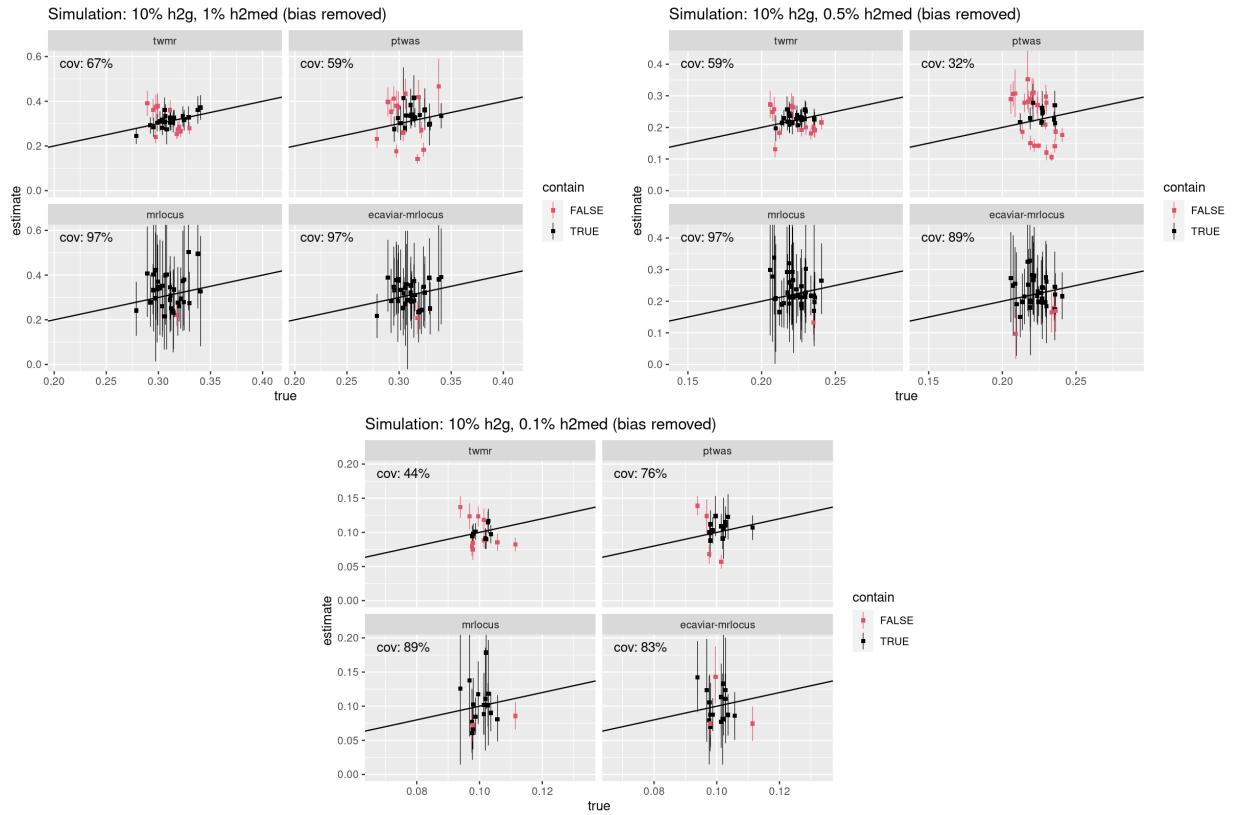
Supplementary Figure 29: Coverage of confidence or credible intervals for the 3 null simulation settings.



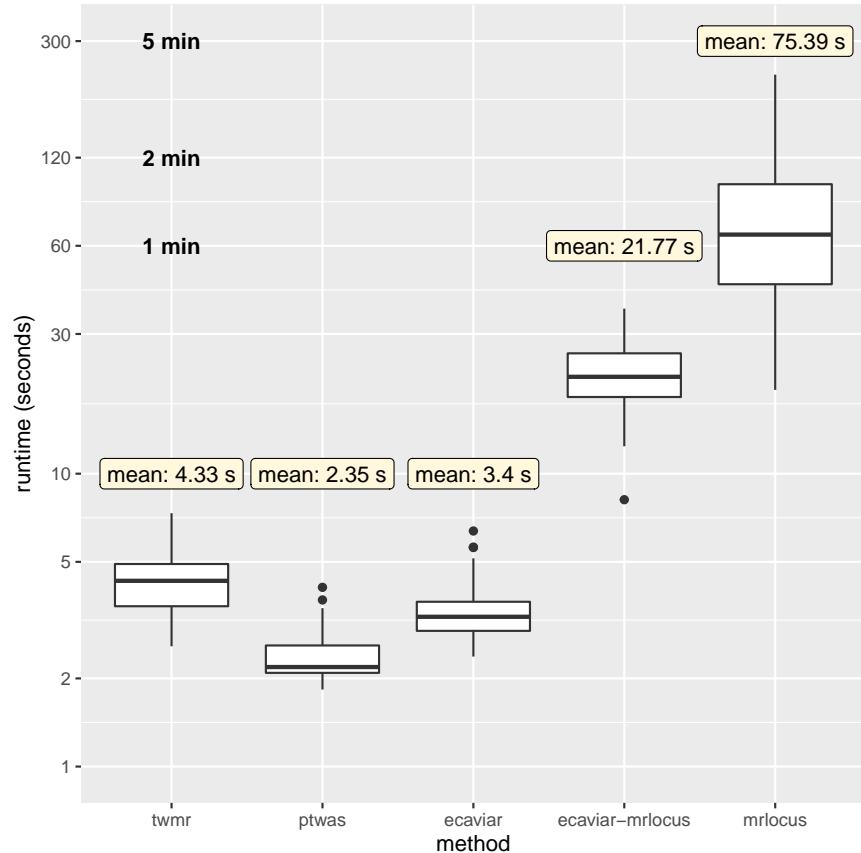
Supplementary Figure 30: Assessment of methods at eQTL  $N = 500$  and  $N = 1000$  at alternative thresholds: p-value 0.0001 for PLINK clumping (labeled  $_p1e-4$ ) and using a more lenient PIP threshold of 0.1 (labeled  $_t0.1$ ).



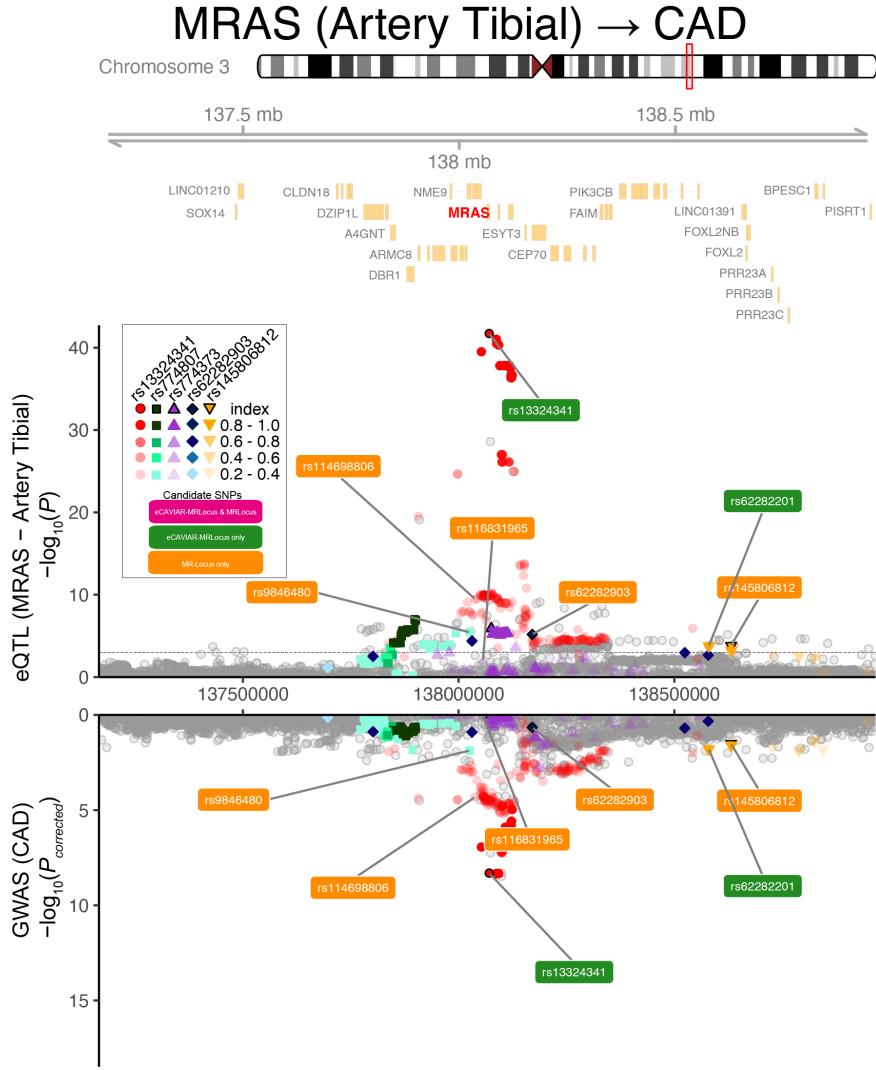
Supplementary Figure 31: Assessment of colocalization accuracy across simulation settings. Shown is the ratio of correctly identified causal eSNPs for all signal clusters containing a true causal eSNP, aggregating across all iterations. Additionally, if the method identified a SNP with correlation of  $> 0.95$  to the true causal eSNP, it was counted as a correct identification.



Supplementary Figure 32: Coverage of confidence and credible intervals after removing estimator bias using the sample average for the estimator and information about the true population-level gene-to-trait effect. Bias was removed and coverage recalculated for simulations with  $h^2g = 10\%$ : simulations A, D and G. For TWMR, PTWAS, MRLocus, and eCAVIAR-MRLocus, the bias in simulation A was  $\{-0.112, -0.079, -0.004, -0.052\}$  respectively. For simulation D the bias was  $\{-0.083, -0.086, -0.014, -0.037\}$ , and for simulation G the bias was  $\{-0.044, -0.027, -0.020, -0.028\}$ .

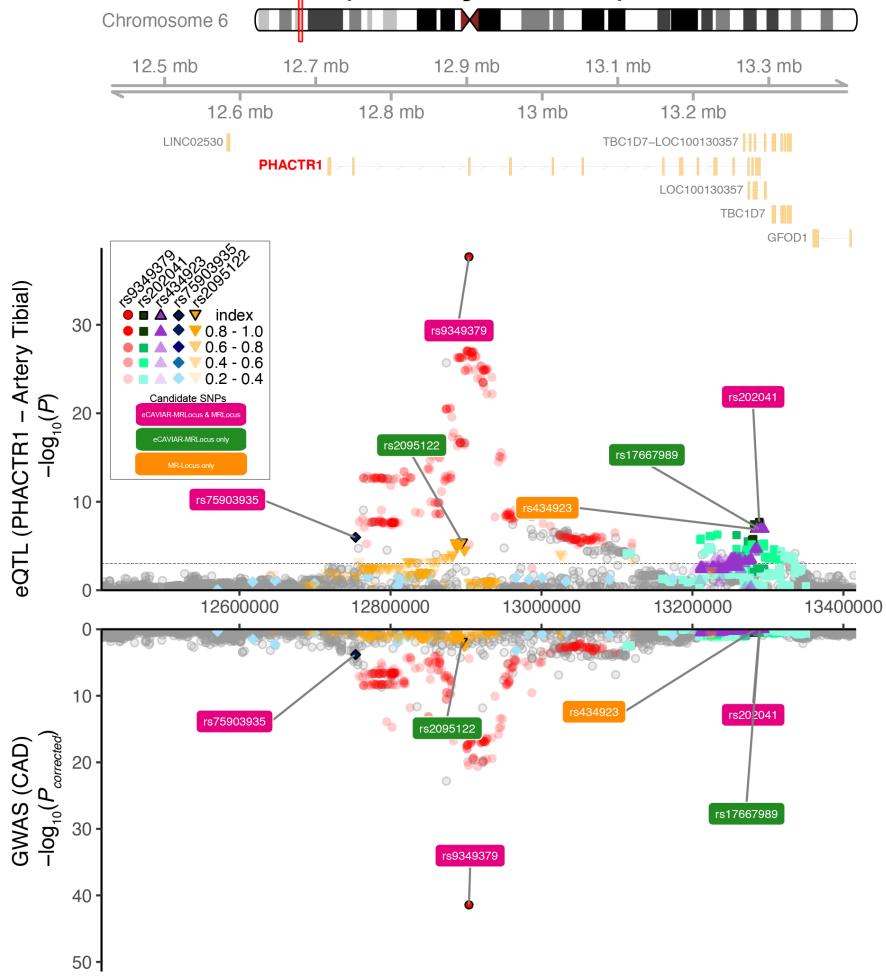


Supplementary Figure 33: Runtime for TWMR, PTWAS, and MRLocus on the 80 iterations from simulation A and eQTL  $N = 1000$  simulation. The runtime for a single locus is shown on the y-axis (log scale), with MRLocus colocalization using 4 cores. “eCAVIAR-MRLocus” indicates the time for MRLocus slope fitting following eCAVIAR colocalization, while “MRLocus” indicates the time for both colocalization and slope fitting.

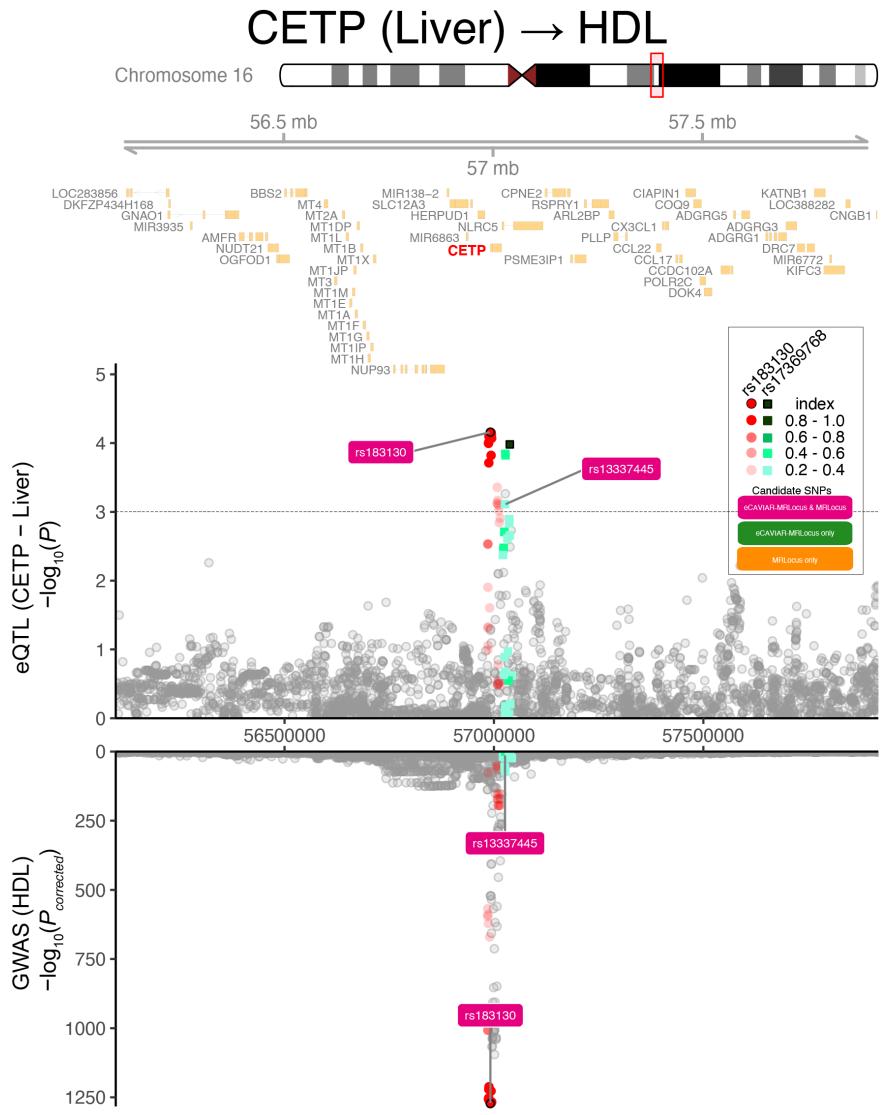


Supplementary Figure 34: Colocalized signals in the *MRAS* region. From top panel to bottom, gene model (NCBI Refseq), eQTL for *MRAS* in artery tibial (GTEx;  $N = 663$ ) and CAD association within CARDIoGRAMplusC4D ( $N_{\text{case}} = 60,801$  and  $N_{\text{control}} = 123,504$ ) (M. Nikpay et al., 2015). LD was calculated to nearly-LD-independent index SNPs within 1KG EUR and colored accordingly. Dashed line indicates a significance threshold at  $p = 0.001$  or  $p = 5 \times 10^{-8}$  for eQTL and GWAS respectively. Colored labels indicate eSNPs used for slope fitting with both methods, eCAVIAR-MRLocus, or MRlocus.

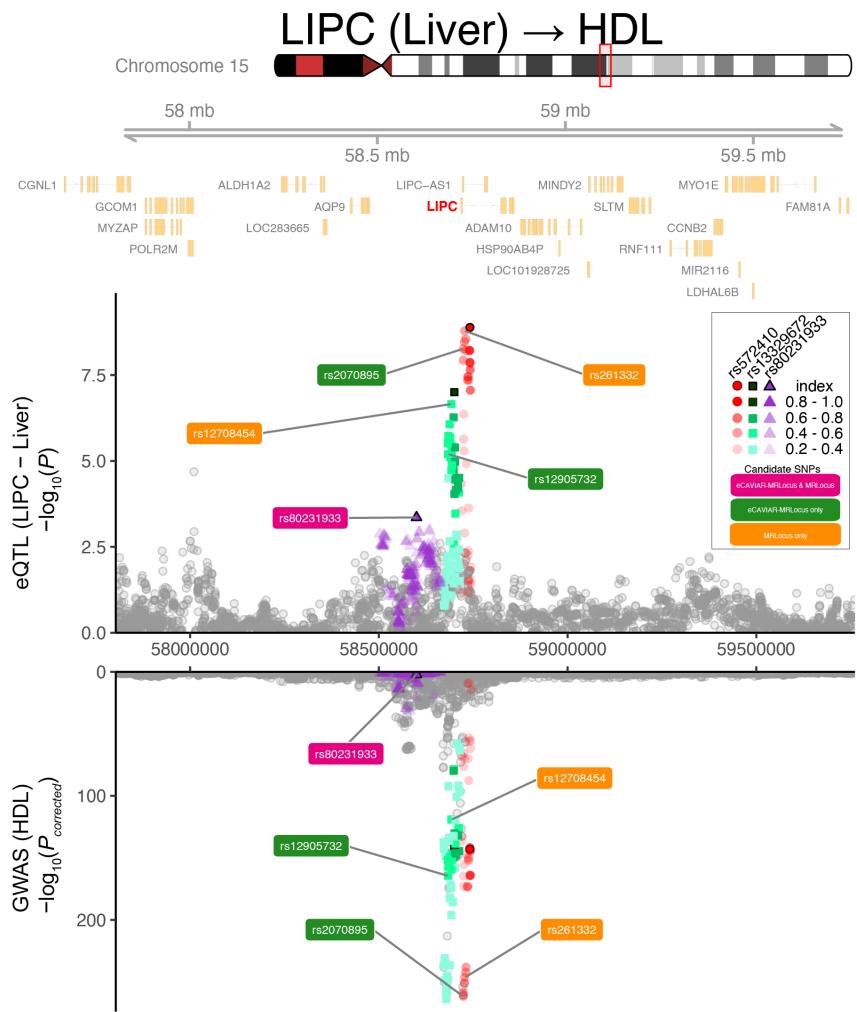
## PHACTR1 (Artery Tibial) → CAD



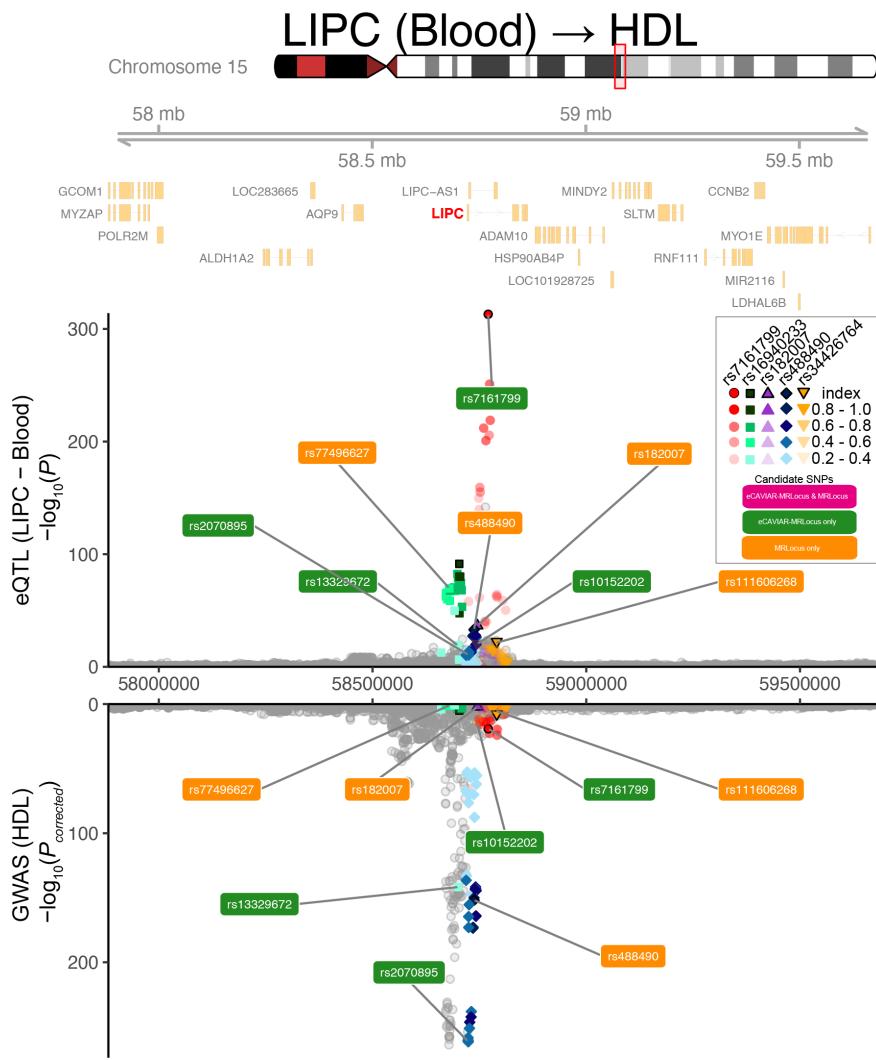
Supplementary Figure 35: Colocalized signals in the *PHACTR1* region. From top panel to bottom, gene model (NCBI Refseq), eQTL for *PHACTR1* in artery tibial (GTEX;  $N = 663$ ) and CAD association within CARDIoGRAMplusC4D ( $N_{\text{case}} = 60,801$  and  $N_{\text{control}} = 123,504$ ) (M. Nikpay et al., 2015). LD was calculated to nearly-LD-independent index SNPs within 1KG EUR and colored accordingly. Dashed line indicates a significance threshold at  $p = 0.001$  or  $p = 5 \times 10^{-8}$  for eQTL and GWAS respectively. Colored labels indicate eSNPs used for slope fitting with both methods, eCAVIAR-MRLocus, or MRLocus.



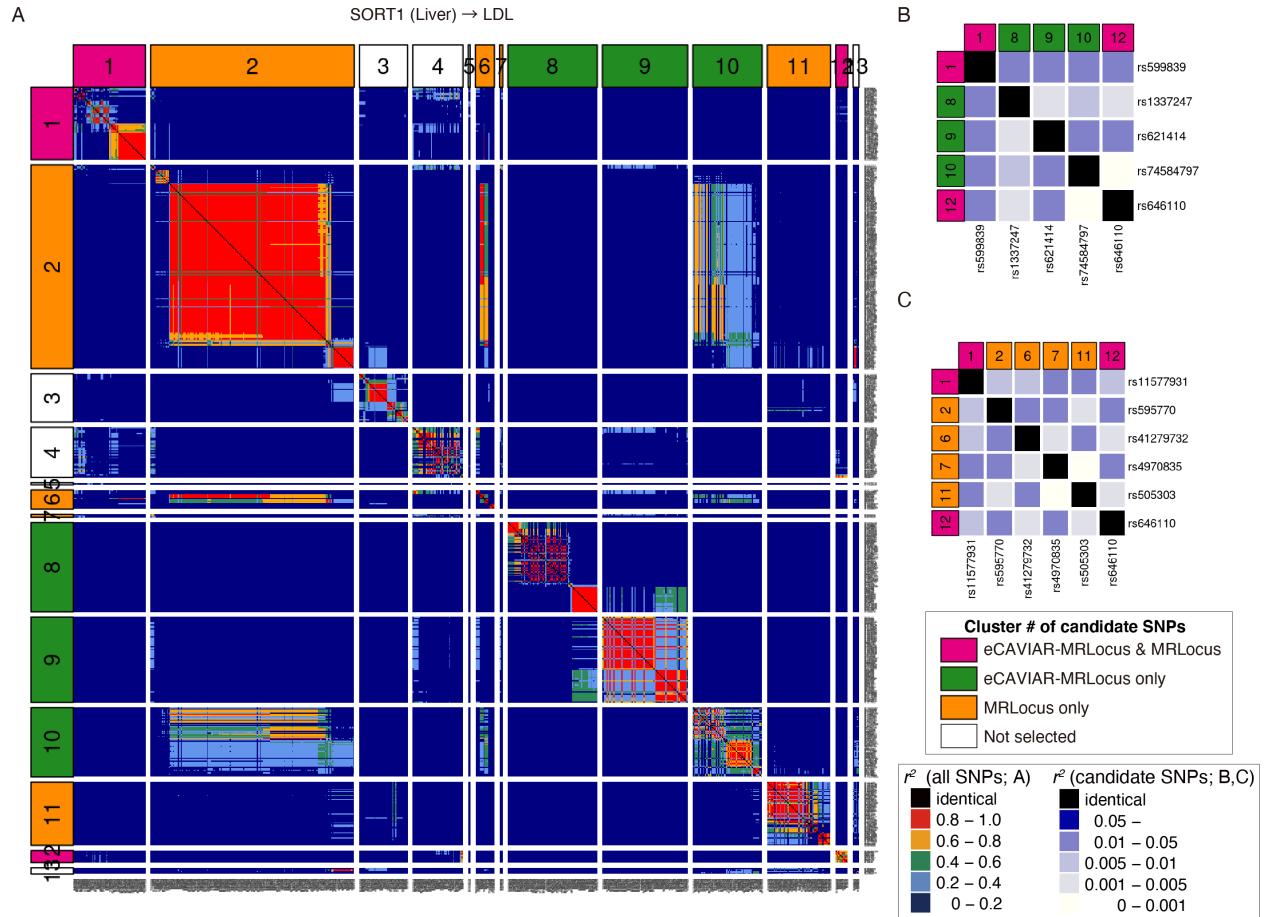
Supplementary Figure 36: Colocalized signals in the *CETP* region. From top panel to bottom, gene model (NCBI Refseq), eQTL for *CETP* in liver ( $N = 588$ ) (Strunz et al., 2018) and HDL association within UKBB ( $N = 315,133$ ). LD was calculated to nearly-LD-independent index SNPs within 1KG EUR and colored accordingly. Dashed line indicates a significance threshold at  $p = 0.001$  or  $p = 5 \times 10^{-8}$  for eQTL and GWAS respectively. Colored labels indicate eSNPs used for slope fitting with both methods, eCAVIAR-MRLocus, or MRLocus.



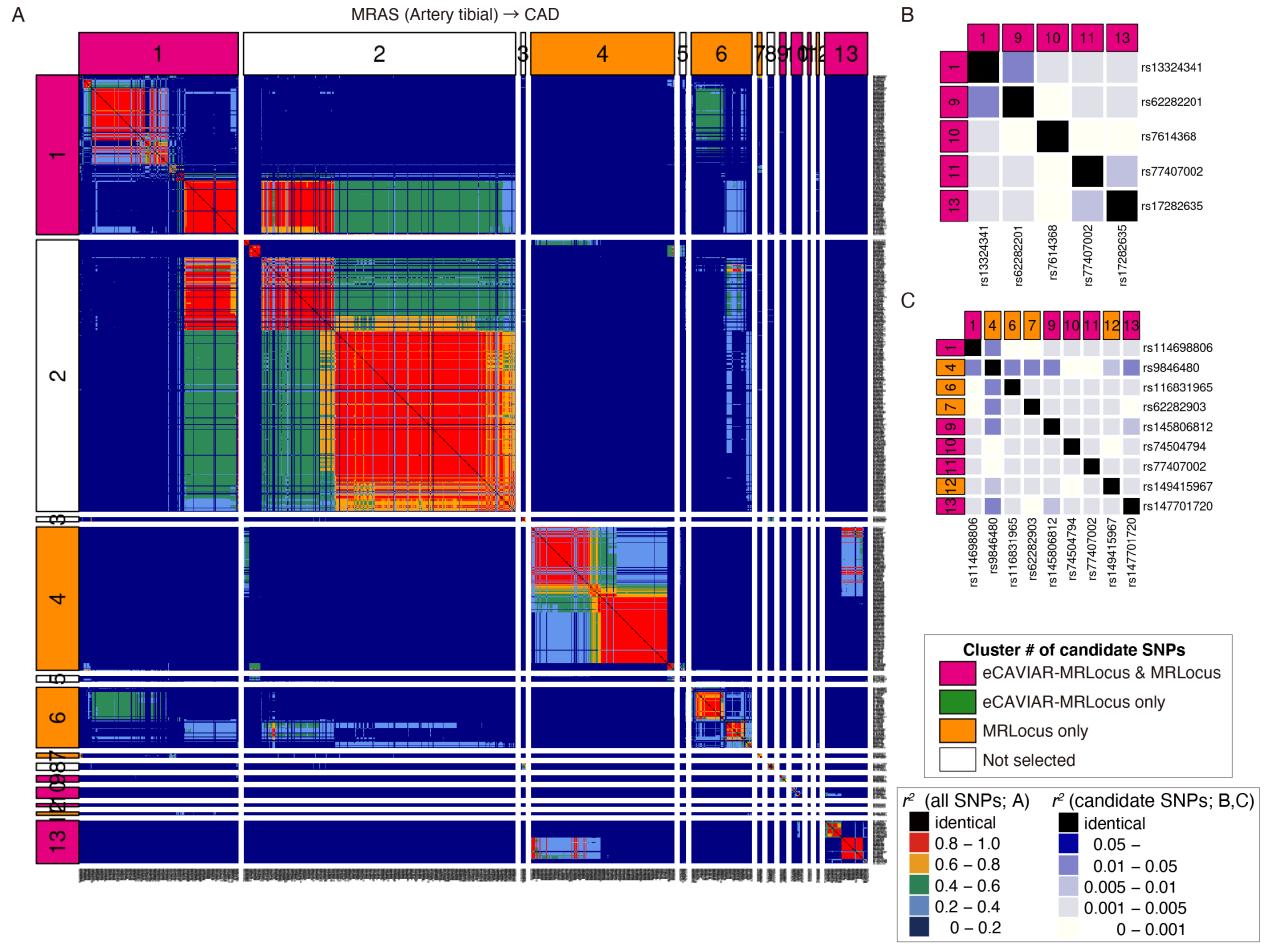
Supplementary Figure 37: Colocalized signals in the *LIPC* region for liver eQTL. From top panel to bottom, gene model (NCBI Refseq), eQTL for *LIPC* in liver ( $N = 588$ ) (Strunz et al., 2018) and HDL association within UKBB ( $N = 315,133$ ). LD was calculated to nearly-LD-independent index SNPs within 1KG EUR and colored accordingly. Dashed line indicates a significance threshold at  $p = 0.001$  or  $p = 5 \times 10^{-8}$  for eQTL and GWAS respectively. Colored labels indicate eSNPs used for slope fitting with both methods, eCAVIAR-MRLocus, or MRlocus.



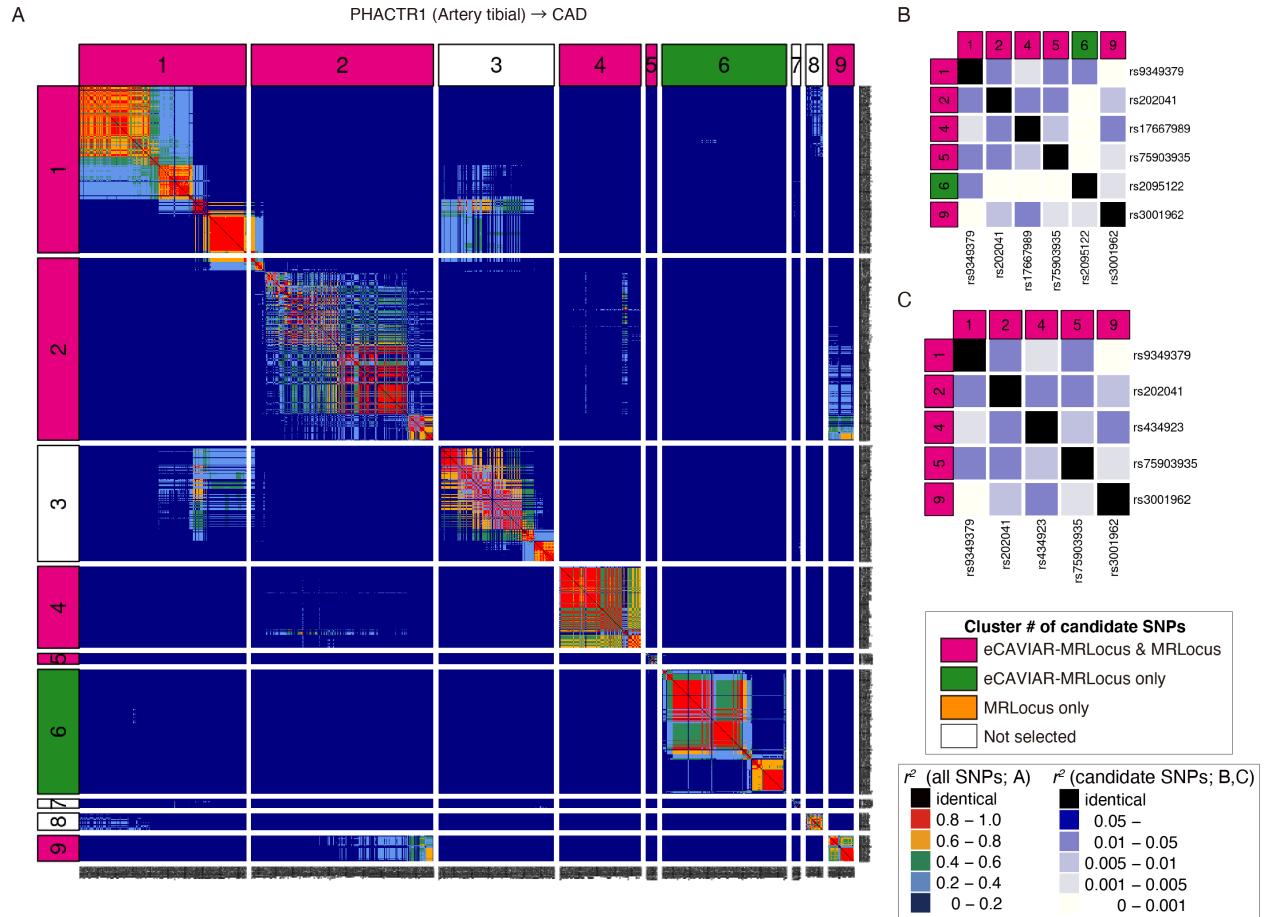
Supplementary Figure 38: Colocalized signals in the *LIPC* region for blood eQTL. From top panel to bottom, gene model (NCBI Refseq), eQTL for *LIPC* in blood ( $N = 31,684$ ) (Vosa et al. 2018) and HDL association within UKBB ( $N = 315,133$ ). LD was calculated to nearly-LD-independent index SNPs within 1KG EUR and colored accordingly. Dashed line indicates a significance threshold at  $p = 0.001$  or  $p = 5 \times 10^{-8}$  for eQTL and GWAS respectively. Colored labels indicate eSNPs used for slope fitting with both methods, eCAVIAR-MRLocus, or MRlocus.



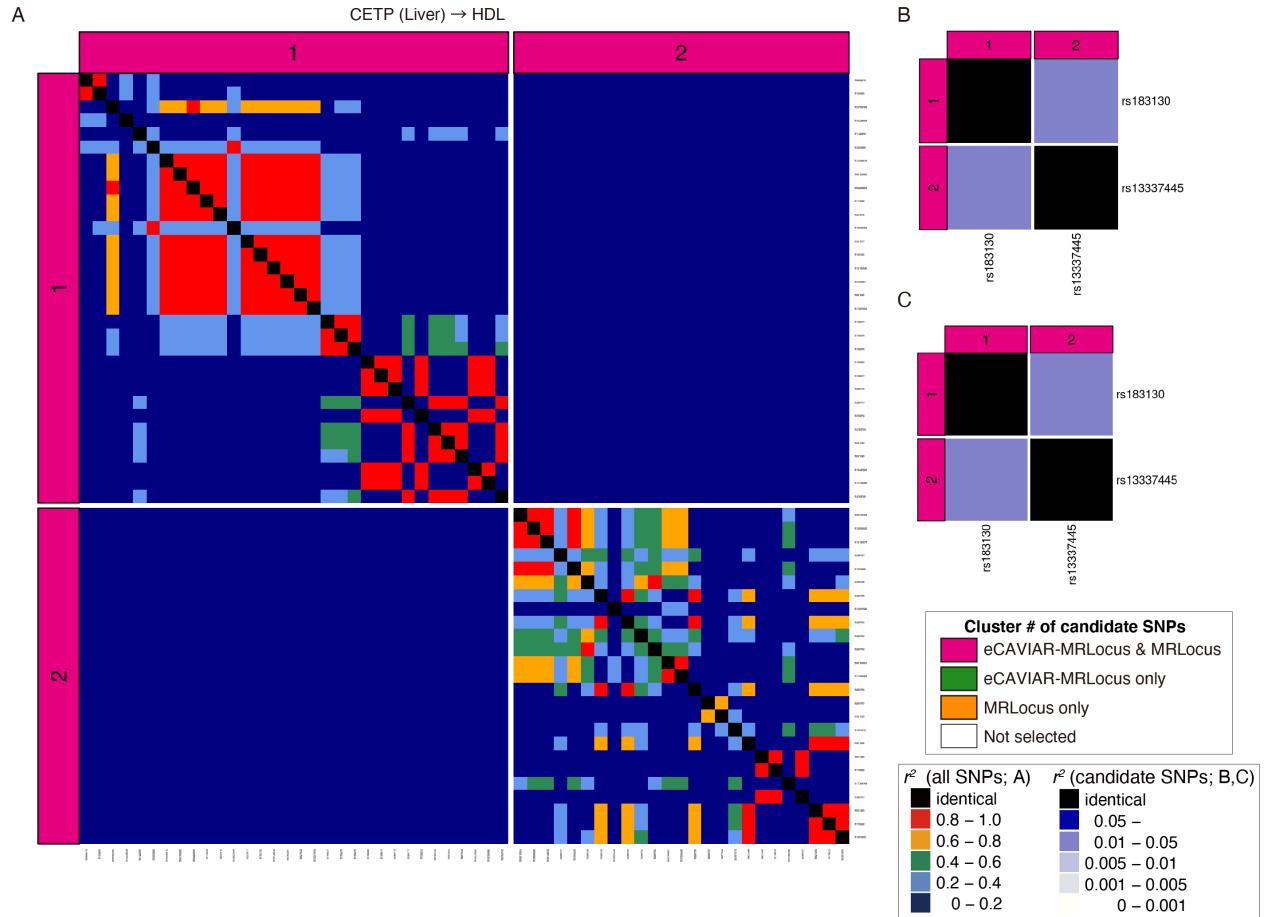
Supplementary Figure 39: LD pattern across nearly-LD-independent signal clusters from *SORT1* eQTL (liver). (A)  $r^2$  between all SNPs in signal clusters. Color bars at the top and left represent whether the cluster was used for estimation of gene-to-trait effect with both methods, only eCAVIAR-MRLocus, only MRLocus, or neither (trimmed due to pairwise  $r^2$  with other signal clusters). (B) Pairwise  $r^2$  for instruments selected for slope fitting for eCAVIAR-MRLocus. (C) Pairwise  $r^2$  for instruments selected for slope fitting for MRLocus.



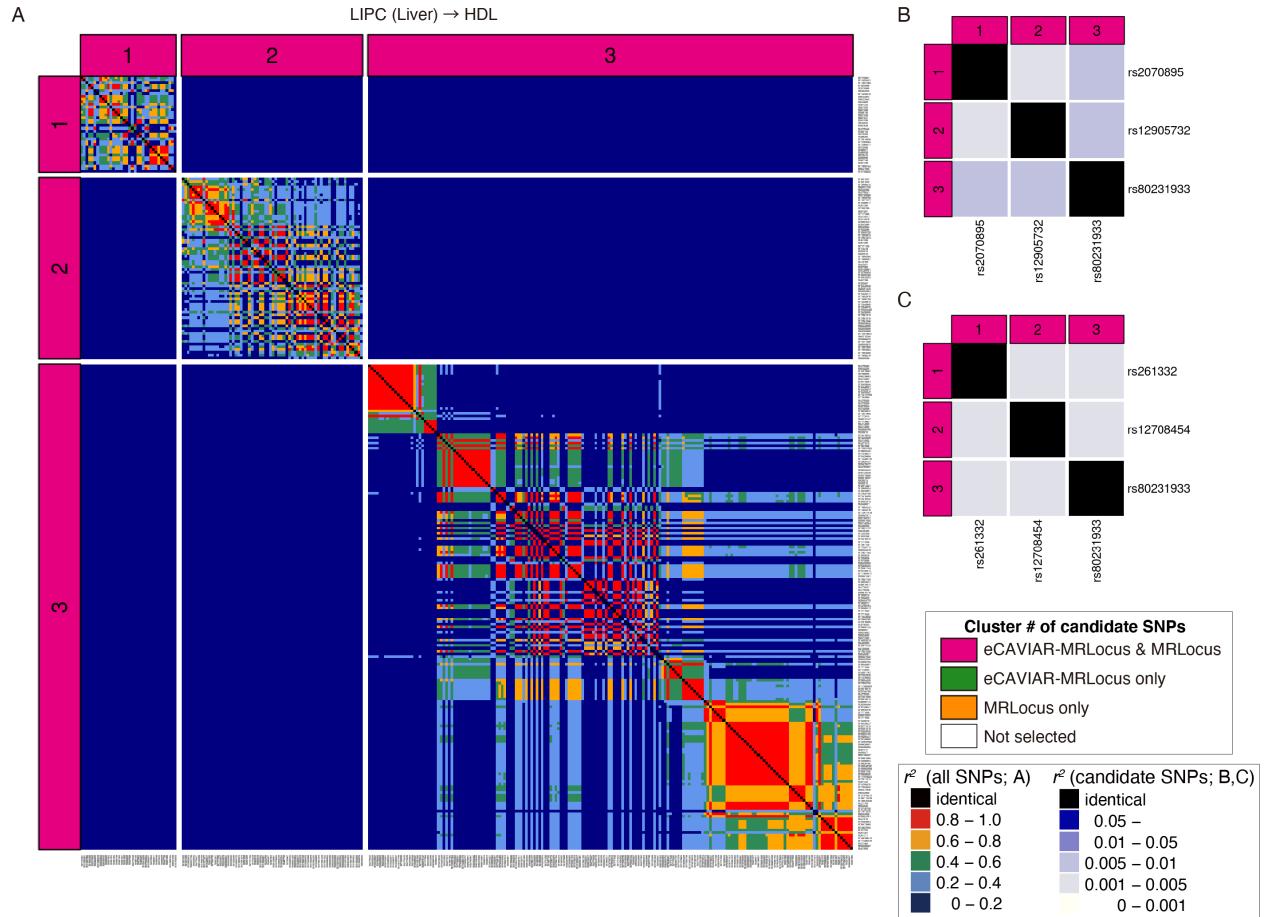
Supplementary Figure 40: LD pattern across nearly-LD-independent signal clusters from *MRAS* eQTL (artery). (A)  $r^2$  between all SNPs in signal clusters. Color bars at the top and left represent whether the cluster was used for estimation of gene-to-trait effect with both methods, only eCAVIAR-MRLocus, or neither (trimmed due to pairwise  $r^2$  with other signal clusters). (B) Pairwise  $r^2$  for instruments selected for slope fitting for eCAVIAR-MRLocus. (C) Pairwise  $r^2$  for instruments selected for slope fitting for MRLocus.



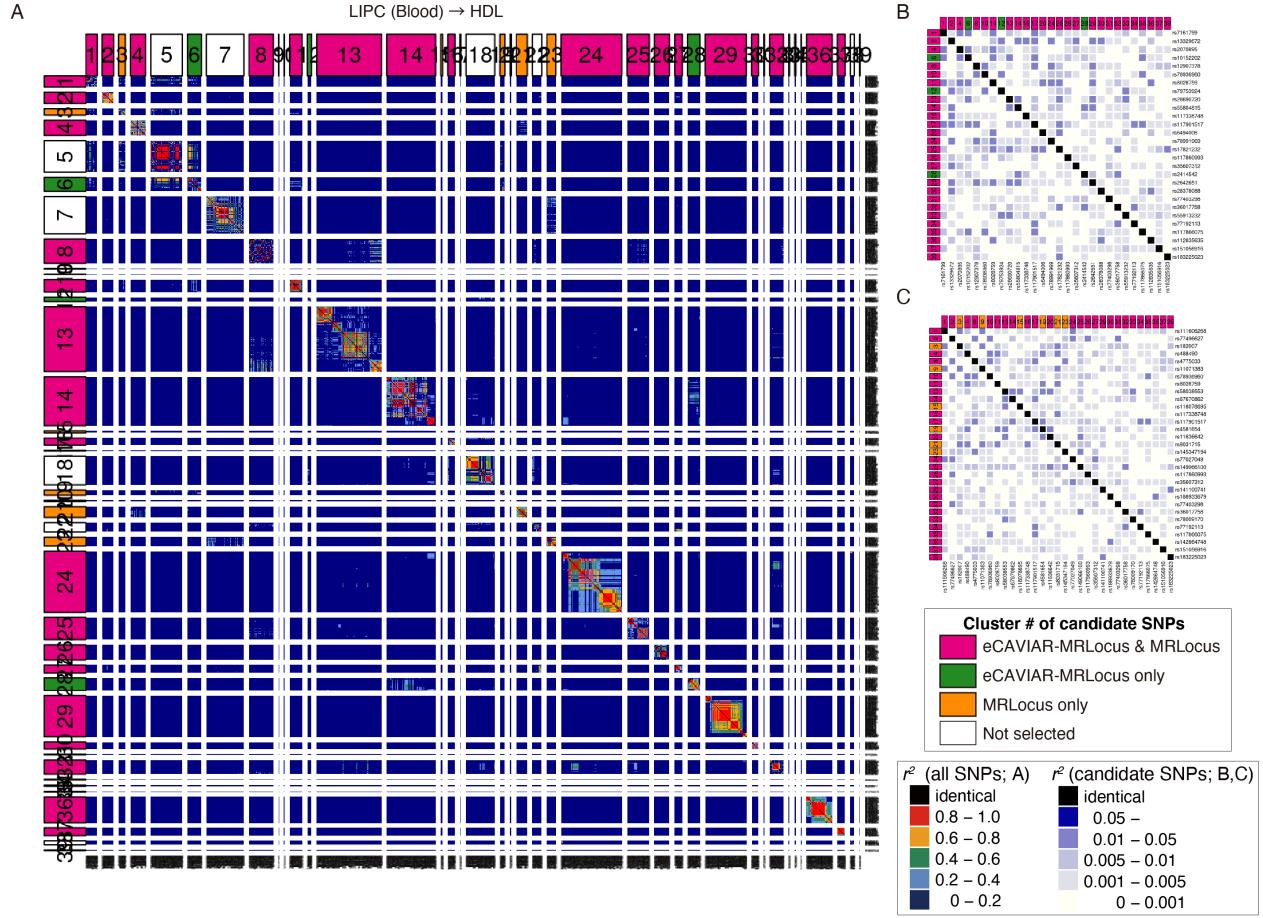
Supplementary Figure 41: LD pattern across nearly-LD-independent signal clusters from *PHACTR1* eQTL (artery). (A)  $r^2$  between all SNPs in signal clusters. Color bars at the top and left represent whether the cluster was used for estimation of gene-to-trait effect with both methods, only eCAVIAR-MRLocus, only MRLocus, or neither (trimmed due to pairwise  $r^2$  with other signal clusters). (B) Pairwise  $r^2$  for instruments selected for slope fitting for eCAVIAR-MRLocus. (C) Pairwise  $r^2$  for instruments selected for slope fitting for MRLocus.



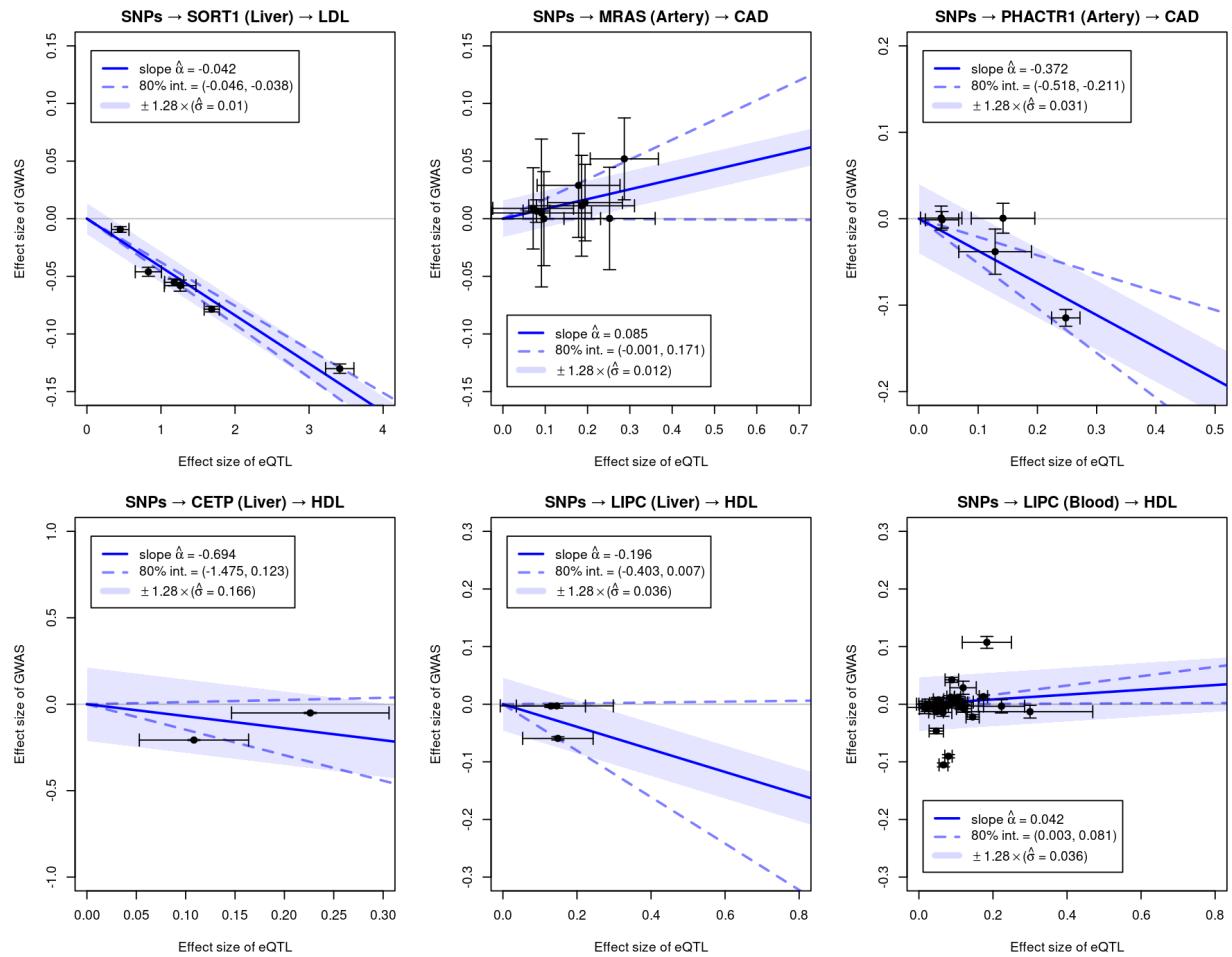
Supplementary Figure 42: LD pattern across nearly-LD-independent signal clusters from *CETP* eQTL (liver). (A)  $r^2$  between all SNPs in signal clusters. Color bars at the top and left represent whether the cluster was used for estimation of gene-to-trait effect with both methods, only eCAVIAR-MRLocus, only MRLocus, or neither (trimmed due to pairwise  $r^2$  with other signal clusters). (B) Pairwise  $r^2$  for instruments selected for slope fitting for eCAVIAR-MRLocus. (C) Pairwise  $r^2$  for instruments selected for slope fitting for MRLocus.



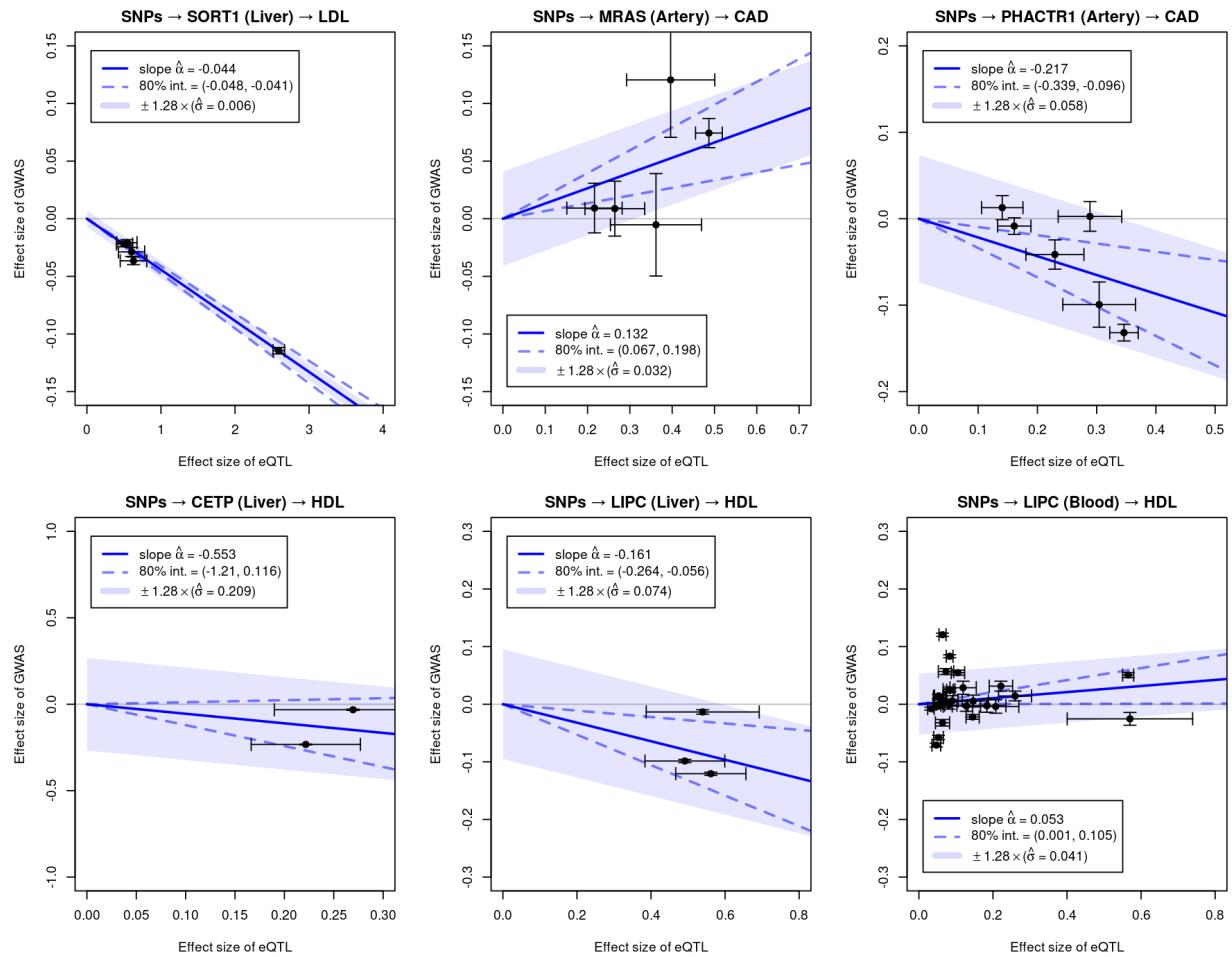
Supplementary Figure 43: LD pattern across nearly-LD-independent signal clusters from *LIPC* eQTL (liver). (A)  $r^2$  between all SNPs in signal clusters. Color bars at the top and left represent whether the cluster was used for estimation of gene-to-trait effect with both methods, only eCAVIAR-MRLocus, only MRLocus, or neither (trimmed due to pairwise  $r^2$  with other signal clusters). (B) Pairwise  $r^2$  for instruments selected for slope fitting for eCAVIAR-MRLocus. (C) Pairwise  $r^2$  for instruments selected for slope fitting for MRLocus.



Supplementary Figure 44: LD pattern across nearly-LD-independent signal clusters from *LIPC* eQTL (blood). (A)  $r^2$  between all SNPs in signal clusters. Color bars at the top and left represent whether the cluster was used for estimation of gene-to-trait effect with both methods, only eCAVIAR-MRLocus, only MRLocus, or neither (trimmed due to pairwise  $r^2$  with other signal clusters). (B) Pairwise  $r^2$  for instruments selected for slope fitting for eCAVIAR-MRLocus. (C) Pairwise  $r^2$  for instruments selected for slope fitting for MRLocus.



Supplementary Figure 45: MRLocus plots for six eQTL-GWAS dataset pairs, using MRLocus for colocalization.



Supplementary Figure 46: MRLocus plots for six eQTL-GWAS dataset pairs, using eCAVIAR for colocalization.