

Supplementary Materials for

**Likelihood-based Mendelian randomization analysis with automated
instrument selection and horizontal pleiotropic modeling**

Zhongshang Yuan, Lu Liu, Ping Guo, Ran Yan, Fuzhong Xue, Xiang Zhou*

*Corresponding author. Email: xzhousph@umich.edu

Published 2 March 2022, *Sci. Adv.* **8**, eabl5744 (2022)
DOI: 10.1126/sciadv.abl5744

This PDF file includes:

Supplementary Text
Figs. S1 to S25
Tables S1 to S4

Supplementary Text

1. MRAID for individual level data

We follow the notations in the main text and rewrite the MRAID model in the following equivalent form:

$$\mathbf{x} = \mathbf{Z}_x \boldsymbol{\beta} + \boldsymbol{\varepsilon}_x, \quad (1)$$

$$\mathbf{y} = \mathbf{Z}_y \boldsymbol{\beta} \boldsymbol{\alpha} + \mathbf{Z}_y (\boldsymbol{\beta} \circ \mathbf{v}) \rho + \mathbf{Z}_y \boldsymbol{\eta}_u + \boldsymbol{\varepsilon}_y, \quad (2)$$

where \mathbf{v} is a p -vector of binary indicators that indicate whether the instrumental SNPs display correlated pleiotropy effects ($v_j = 1$) or not ($v_j = 0$); and the term $\boldsymbol{\beta} \circ \mathbf{v}$ represents the Hadamard product, also known as the element wise product, of the two vectors $\boldsymbol{\beta}$ and \mathbf{v} . For $j = 1, \dots, p$, we further have:

$$\begin{aligned} \beta_j &\sim \pi_\beta N(0, \sigma_\beta^2) + (1 - \pi_\beta) \delta_0, \\ \eta_{uj} &\sim \pi_1 N(0, \sigma_\eta^2) + (1 - \pi_1) \delta_0 \text{ if } \beta_j \neq 0, \\ \eta_{uj} &\sim \pi_0 N(0, \sigma_\eta^2) + (1 - \pi_0) \delta_0 \text{ if } \beta_j = 0, \\ p(v_j = 1) &= \pi_c, \text{ if } \beta_j \neq 0. \end{aligned}$$

Above, we assume that σ_β^2 follows an inverse gamma distribution $\sigma_\beta^2 \sim \text{InvG}(a_\beta, b_\beta)$. We set $a_\beta = \frac{p}{10} + 1$ and $b_\beta = 0.2$ to ensure a prior mean of $2/p$. Similarly, we assume $\sigma_\eta^2 \sim \text{InvG}(a_\eta, b_\eta)$, with $a_\eta = \frac{p}{5} + 1$ and $b_\eta = 0.2$ that ensure a prior mean of $1/p$. We assume the proportion parameters to follow *beta* distributions: $\pi_\beta \sim \text{beta}(\lambda_{\beta_1}, \lambda_{\beta_2})$, with $\lambda_{\beta_1} = 0.5$ and $\lambda_{\beta_2} = 4.5$ to ensure a prior mean of approximately 0.1; $\pi_c \sim \text{beta}(\lambda_{c_1}, \lambda_{c_2})$ with $\lambda_{c_1} = 0.5$ and $\lambda_{c_2} = 9.5$ to ensure a prior mean of approximately 0.05; $\pi_1 \sim \text{beta}(\lambda_{21}, \lambda_{22})$ with $\lambda_{21} = 0.5$ and $\lambda_{22} = 1.5$ to ensure a prior mean of 0.25; and $\pi_0 \sim \text{beta}(\lambda_{31}, \lambda_{32})$ with $\lambda_{31} = 0.05$ and $\lambda_{32} = 9.95$ to ensure a prior mean of 0.005. These prior means are set to represent our prior belief that a relatively small proportion of SNPs display non-zero effects on the exposure, that a relatively small proportion of SNPs display horizontal pleiotropy, that the selected instrumental SNPs are more likely to display horizontal pleiotropy than the non-selected SNPs, and that horizontal pleiotropic SNPs are more likely to display uncorrelated pleiotropic effects than correlated pleiotropic effects. For the other hyper-parameters, we use relatively non-informative priors. Specifically, for the error variance parameters, we set $p(\sigma_x^2) \propto 1$ and $p(\sigma_y^2) \propto 1$. For the parameter ρ , we set $p(\rho) \propto 1$. For the causal effect parameter α , we set $\alpha \sim N(0, \sigma_0^2)$ with $\sigma_0^2 \rightarrow \infty$.

To facilitate computation, we introduce a p -vector of binary indicators $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^T$ to indicate whether each SNP has a non-zero effect (β_j) on the exposure ($\gamma_j = 1$) or not ($\gamma_j = 0$). We also introduce another p -vector of binary indicators $\boldsymbol{\tau} = (\tau_1, \dots, \tau_p)^T$ to indicate whether each SNP has the horizontal pleiotropy effect ($\tau_j = 1$) or not ($\tau_j = 0$). Thus, we have $p(\gamma_j = 1) = \pi_\beta$, $p(\tau_j = 1 | \gamma_j = 1) = \pi_1$, $p(\tau_j = 1 | \gamma_j = 0) = \pi_0$, and $p(v_j = 1 | \gamma_j = 1) = \pi_c$.

2. Detailed sampling steps and efficient computation

Our goal is to first obtain the posterior samples for the causal effect α . To do so, we rely on Gibbs sampling to obtain the conditional distribution for one parameter at a time. Our posterior likelihood is in the form of

$$\begin{aligned}
& f(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}_u, \boldsymbol{\tau}, \mathbf{v}, \rho | \mathbf{x}, \mathbf{y}) \\
& \propto f(\mathbf{x}, \mathbf{y} | \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}_u, \boldsymbol{\tau}, \mathbf{v}, \rho) f(\boldsymbol{\beta} | \boldsymbol{\gamma}) f(\boldsymbol{\eta}_u | \boldsymbol{\tau}) f(\mathbf{v} | \boldsymbol{\gamma}) f(\boldsymbol{\tau} | \boldsymbol{\gamma}) f(\boldsymbol{\gamma} | \pi_\beta) f(\pi_\beta) f(\pi_1) f(\pi_0) f(\pi_c) \\
& \propto f(\mathbf{x} | \boldsymbol{\beta}) f(\mathbf{y} | \alpha, \boldsymbol{\beta}, \boldsymbol{\eta}_u, \mathbf{v}, \rho) f(\boldsymbol{\beta} | \boldsymbol{\gamma}) f(\boldsymbol{\eta}_u | \boldsymbol{\tau}) f(\mathbf{v} | \boldsymbol{\gamma}) f(\boldsymbol{\tau} | \boldsymbol{\gamma}) f(\boldsymbol{\gamma} | \pi_\beta) f(\pi_\beta) f(\pi_1) f(\pi_0) f(\pi_c) \propto \\
& \quad (2\pi)^{-\frac{n_2}{2}} (\sigma_y^2)^{-\frac{n_2}{2}} \exp \\
& \quad \left(- \frac{(\mathbf{y} - \alpha \mathbf{Z}_{y,-j} \boldsymbol{\beta}_{-j} - \rho \mathbf{Z}_{y,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \alpha \mathbf{Z}_{y,j} \boldsymbol{\beta}_j - \rho \mathbf{Z}_{y,j} \boldsymbol{\beta}_j v_j - \mathbf{Z}_{y,j} \boldsymbol{\eta}_{uj} - \mathbf{Z}_{y,-j} \boldsymbol{\eta}_{u,-j})^T}{2\sigma_y^2} \right. \\
& \quad \left. - \frac{(\mathbf{y} - \alpha \mathbf{Z}_{y,-j} \boldsymbol{\beta}_{-j} - \rho \mathbf{Z}_{y,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \alpha \mathbf{Z}_{y,j} \boldsymbol{\beta}_j - \rho \mathbf{Z}_{y,j} \boldsymbol{\beta}_j v_j - \mathbf{Z}_{y,j} \boldsymbol{\eta}_{uj} - \mathbf{Z}_{y,-j} \boldsymbol{\eta}_{u,-j})}{2\sigma_y^2} \right) \\
& (2\pi)^{-\frac{n_1}{2}} (\sigma_x^2)^{-\frac{n_1}{2}} \exp \left(- \frac{(\mathbf{x} - \mathbf{Z}_{x,-j} \boldsymbol{\beta}_{-j} - \mathbf{Z}_{x,j} \boldsymbol{\beta}_j)^T (\mathbf{x} - \mathbf{Z}_{x,-j} \boldsymbol{\beta}_{-j} - \mathbf{Z}_{x,j} \boldsymbol{\beta}_j)}{2\sigma_x^2} \right). \\
& \prod_{j=1}^p \left[(2\pi)^{-\frac{1}{2}} (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left(- \frac{\beta_j^2}{2\sigma_\beta^2} \right) \right]^{\gamma_j} \prod_{j=1}^p [\delta_0(\beta_j)]^{1-\gamma_j} \prod_{j=1}^p \left[(2\pi)^{-\frac{1}{2}} (\sigma_\eta^2)^{-\frac{1}{2}} \exp \left(- \frac{\eta_{uj}^2}{2\sigma_\eta^2} \right) \right]^{\tau_j} \\
& \quad \prod_{j=1}^p [\delta_0(\eta_{uj})]^{1-\tau_j} \pi_c^{\sum_{j=1}^p \gamma_j v_j} (1 - \pi_c)^{\sum_{j=1}^p \gamma_j (1-v_j)} \\
& \pi_\beta^{\sum_{j=1}^p \gamma_j} (1 - \pi_\beta)^{\sum_{j=1}^p (1-\gamma_j)} \pi_1^{\sum_{j=1}^p \gamma_j \tau_j} (1 - \pi_1)^{\sum_{j=1}^p \gamma_j (1-\tau_j)} \pi_0^{\sum_{j=1}^p (1-\gamma_j) \tau_j} (1 \\
& \quad - \pi_0)^{\sum_{j=1}^p (1-\gamma_j) (1-\tau_j)} \\
& \quad \cdot \text{beta}(\pi_\beta, \lambda_{\beta 1}, \lambda_{\beta 2}) \text{beta}(\pi_1, \lambda_{21}, \lambda_{22}) \text{beta}(\pi_0, \lambda_{31}, \lambda_{32}) \text{beta}(\pi_c, \lambda_{c1}, \lambda_{c2}) \\
& \propto C \\
& \cdot \exp \left(- \frac{(\mathbf{y} - \alpha \mathbf{Z}_{y,-j} \boldsymbol{\beta}_{-j} - \rho \mathbf{Z}_{y,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \mathbf{Z}_y \boldsymbol{\eta}_u)^T (\mathbf{y} - \alpha \mathbf{Z}_{y,-j} \boldsymbol{\beta}_{-j} - \rho \mathbf{Z}_{y,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \mathbf{Z}_y \boldsymbol{\eta}_u)}{2\sigma_y^2} \right) \\
& \quad \exp \left(- \frac{(\mathbf{x} - \mathbf{Z}_{x,-j} \boldsymbol{\beta}_{-j})^T (\mathbf{x} - \mathbf{Z}_{x,-j} \boldsymbol{\beta}_{-j})}{2\sigma_x^2} \right). \\
& \prod_{i \neq j}^p \left[\pi_\beta (2\pi)^{-\frac{1}{2}} (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left(- \frac{\beta_i^2}{2\sigma_\beta^2} \right) \right]^{\gamma_i} \prod_{i \neq j}^p [(1 - \pi_\beta) \delta_0(\beta_i)]^{1-\gamma_i} \\
& \quad \prod_{j=1}^p \left[(2\pi)^{-\frac{1}{2}} (\sigma_\eta^2)^{-\frac{1}{2}} \exp \left(- \frac{\eta_{uj}^2}{2\sigma_\eta^2} \right) \right]^{\tau_j} \\
& \quad \prod_{j=1}^p [\delta_0(\eta_{uj})]^{1-\tau_j} \pi_c^{\sum_{j=1}^p \gamma_j v_j} (1 - \pi_c)^{\sum_{j=1}^p \gamma_j (1-v_j)} \\
& \pi_1^{\sum_{j=1}^p \gamma_j \tau_j} (1 - \pi_1)^{\sum_{j=1}^p \gamma_j (1-\tau_j)} \pi_0^{\sum_{j=1}^p (1-\gamma_j) \tau_j} (1 \\
& \quad - \pi_0)^{\sum_{j=1}^p (1-\gamma_j) (1-\tau_j)} \text{beta}(\pi_\beta, \lambda_{\beta 1}, \lambda_{\beta 2}) \text{beta}(\pi_1, \lambda_{21}, \lambda_{22}) \text{beta}(\pi_0, \lambda_{31}, \lambda_{32}) \text{beta}(\pi_c, \lambda_{c1}, \lambda_{c2}) \\
& \quad \cdot \exp \left(- \frac{\beta_j^2 K - 2 \left(\frac{(\alpha + \rho v_j) (\mathbf{y} - \alpha \mathbf{Z}_{y,-j} \boldsymbol{\beta}_{-j} - \rho \mathbf{Z}_{y,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \mathbf{Z}_y \boldsymbol{\eta}_u)^T \mathbf{Z}_{y,j} + (\mathbf{x} - \mathbf{Z}_{x,-j} \boldsymbol{\beta}_{-j})^T \mathbf{Z}_{x,j}}{\sigma_y^2} \right) \beta_j}{2} \right)
\end{aligned}$$

where

$$C = (2\pi)^{-\frac{n_2}{2}} (\sigma_y^2)^{-\frac{n_2}{2}} (2\pi)^{-\frac{n_1}{2}} (\sigma_x^2)^{-\frac{n_1}{2}} (2\pi)^{-\frac{\gamma_j}{2}} \pi_\beta^{\gamma_j} (1 - \pi_\beta)^{1-\gamma_j} (\sigma_\beta^2)^{-\frac{\gamma_j}{2}} (\delta_0(\beta_j))^{1-\gamma_j};$$

K is a scalar in the form of

$$K = \frac{(\alpha + \rho v_j)^2 \mathbf{Z}_{y,j}^T \mathbf{Z}_{y,j}}{\sigma_y^2} + \frac{\mathbf{Z}_{x,j}^T \mathbf{Z}_{x,j}}{\sigma_x^2} + \frac{\gamma_j}{\sigma_\beta^2};$$

$\boldsymbol{\beta}_{-j}$ is the vector $\boldsymbol{\beta}$ without the j th element β_j ; $Z_{x,-j}$ is the $n_1 \times (p-1)$ matrix Z_x without the j th column vector $Z_{x,j}$; $Z_{y,-j}$ is the $n_2 \times (p-1)$ matrix Z_y without the j th column vector $Z_{y,j}$. Note that because Z_x, Z_y are standardized to have mean zero and unite standard deviation, we have $Z_{y,j}^T Z_{y,j} = n_2 - 1, Z_{x,j}^T Z_{x,j} = n_1 - 1$ for any $j = 1, \dots, p$.

Based on the above likelihood, we conduct Gibbs sampling on all parameters that include $\theta = (\beta_j, \gamma_j, \eta_{uj}, \tau_j, \alpha, v_j, \sigma_\beta^2, \sigma_\eta^2, \sigma_x^2, \sigma_y^2, \pi_\beta, \pi_1, \pi_0, \pi_c, \rho)$.

Specifically, given $\gamma_j=1$ and other parameters, the posterior conditional distribution of β_j given all other parameters is a normal distribution with mean U_{β_j} and variance $\sigma_{\beta_j}^2$, where

$$U_{\beta_j} = \left(\frac{(\alpha + \rho v_j)(y - \alpha Z_{y,-j} \boldsymbol{\beta}_{-j} - \rho Z_{y,-j}(\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - Z_y \boldsymbol{\eta}_u)^T \mathbf{Z}_{y,j}}{\sigma_y^2} + \frac{(x - Z_{x,-j} \boldsymbol{\beta}_{-j})^T \mathbf{Z}_{x,j}}{\sigma_x^2} \right) / K_1$$

$$\sigma_{\beta_j}^2 = 1/K_1, \quad K_1 = \frac{(\alpha + \rho v_j)^2 \mathbf{Z}_{y,j}^T \mathbf{Z}_{y,j}}{\sigma_y^2} + \frac{\mathbf{Z}_{x,j}^T \mathbf{Z}_{x,j}}{\sigma_x^2} + \frac{1}{\sigma_\beta^2}$$

Given $\tau_j=1$ and other parameters, the posterior conditional distribution of η_{uj} is a normal distribution with mean U_{η_j} and variance $\sigma_{\eta_j}^2$, where

$$U_{\eta_j} = \left(\frac{(y - Z_{y,-j} \boldsymbol{\eta}_{u,-j} - \alpha Z_y \boldsymbol{\beta} - \rho Z_y(\boldsymbol{\beta} \circ \mathbf{v}))^T \mathbf{Z}_{y,j}}{\sigma_y^2} \right) / K_2$$

$$\sigma_{\eta_j}^2 = 1/K_2, \quad K_2 = \frac{\mathbf{Z}_{y,j}^T \mathbf{Z}_{y,j}}{\sigma_y^2} + \frac{1}{\sigma_\eta^2}$$

After integrating out the parameter β_j , we obtain the posterior conditional distribution for γ_j as

$$p(\gamma_j = 1 | \boldsymbol{\beta}_{-j}, \boldsymbol{\eta}_u, \alpha, \sigma_x^2, \sigma_y^2, \sigma_\beta^2, \pi_\beta, \pi_1, \pi_0, \pi_c, \rho) \propto \exp \left(\frac{\mu_{\beta_j}^2}{2\sigma_{\beta_j}^2} + 0.5 * \log(\sigma_{\beta_j}^2) - 0.5 * \log(\sigma_\beta^2) + \log(\pi_\beta) + \tau_j \log(\pi_1) + (1 - \tau_j) \log(1 - \pi_1) + v_j \log(\pi_c) + (1 - v_j) \log(1 - \pi_c) \right),$$

$$p(\gamma_j = 0 | \boldsymbol{\beta}_{-j}, \boldsymbol{\eta}_u, \alpha, \sigma_x^2, \sigma_y^2, \sigma_\beta^2, \pi_\beta, \pi_1, \pi_0) \propto \exp(\log(1 - \pi_\beta) + \tau_j \log(\pi_0) + (1 - \tau_j) \log(1 - \pi_0)).$$

After integrating out the parameter η_j , we obtain the posterior conditional distribution for τ_j as

$$p(\tau_j = 1 | \eta_{u,-j}, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_1, \pi_0) \propto \exp \left(\frac{\mu_{\eta_j}^2}{2\sigma_{\eta_j}^2} + 0.5 * \log(\sigma_{\eta_j}^2) - 0.5 * \log(\sigma_\eta^2) + \gamma_j \log(\pi_1) + (1 - \gamma_j) \log(\pi_0) \right)$$

$$p(\tau_j = 0 | \eta_{u,-j}, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_1, \pi_0) \propto \exp(\gamma_j \log(1 - \pi_1) + (1 - \gamma_j) \log(1 - \pi_0))$$

We also obtain the posterior conditional distribution for v_j as

$$p(v_j = 1 | \eta_u, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_c, \pi_1, \pi_0) \propto \exp \left(- \frac{\rho^2 \beta_j^2 \mathbf{Z}_{y,j}^T \mathbf{Z}_{y,j} - 2(\mathbf{y} - \alpha \mathbf{Z}_y \boldsymbol{\beta} - \rho \mathbf{Z}_{y,-j}(\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \mathbf{Z}_y \boldsymbol{\eta}_u)^T \rho \mathbf{Z}_{y,j} \beta_j}{2\sigma_y^2} + \gamma_j \log(\pi_c) \right)$$

$$p(v_j = 0 | \boldsymbol{\eta}_u, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_c, \pi_1, \pi_0) \propto \exp(\gamma_j \log(1 - \pi_c))$$

The posterior conditional distribution of ρ is a normal distribution with variance $\sigma_\rho^2 =$

$$\left(\frac{(\mathbf{v} \circ \boldsymbol{\beta})^T Z_y^T Z_y (\mathbf{v} \circ \boldsymbol{\beta})}{\sigma_y^2}\right)^{-1} \text{ and mean } \mu_\rho = \frac{(\mathbf{v} \circ \boldsymbol{\beta})^T Z_y^T \mathbf{y} - (\mathbf{v} \circ \boldsymbol{\beta})^T Z_y^T Z_y \boldsymbol{\eta}_u - (\mathbf{v} \circ \boldsymbol{\beta})^T Z_y^T Z_y \boldsymbol{\beta} \alpha}{\sigma_y^2} \sigma_\rho^2.$$

The posterior conditional distribution of α is a normal distribution with variance $\sigma_\alpha^2 = \left(\frac{\boldsymbol{\beta}^T Z_y^T Z_y \boldsymbol{\beta}}{\sigma_y^2}\right)^{-1}$ and mean $\mu_\alpha = \frac{\boldsymbol{\beta}^T Z_y^T \mathbf{y} - \boldsymbol{\beta}^T Z_y^T Z_y \boldsymbol{\eta}_u - \boldsymbol{\beta}^T Z_y^T Z_y (\boldsymbol{\beta} \circ \mathbf{v}) \rho}{\sigma_y^2} \sigma_\alpha^2$.

The posterior conditional distribution of π_β is $beta(\sum_{j=1}^p \gamma_j + \lambda_{\beta 1}, \sum_{j=1}^p (1 - \gamma_j) + \lambda_{\beta 2})$.

The posterior conditional distribution of π_1 is $beta(\sum_{j=1}^p \gamma_j \tau_j + \lambda_{21}, \sum_{j=1}^p \gamma_j (1 - \tau_j) + \lambda_{22})$.

The posterior conditional distribution of π_0 is $beta(\sum_{j=1}^p (1 - \gamma_j) \tau_j + \lambda_{31}, \sum_{j=1}^p (1 - \gamma_j) (1 - \tau_j) + \lambda_{32})$.

The posterior conditional distribution of π_c is $beta(\sum_{j=1}^p \gamma_j v_j + \lambda_{c1}, \sum_{j=1}^p \gamma_j (1 - v_j) + \lambda_{c2})$.

The posterior conditional distribution of σ_x^2 is an inverse gamma distribution with the shape parameter $\sigma_{x_shape}^2 = \frac{n_1}{2} - 1$ and the scale parameter $\sigma_{x_scale}^2 = (\mathbf{x}^T \mathbf{x} + \boldsymbol{\beta}^T Z_x^T Z_x \boldsymbol{\beta} - 2\boldsymbol{\beta}^T Z_x^T \mathbf{x})/2$.

The posterior conditional distribution of σ_y^2 is an inverse gamma distribution with the shape parameter $\sigma_{y_shape}^2 = \frac{n_2}{2} - 1$ and the scale parameter.

$$\sigma_{y_scale}^2 = (\mathbf{y} - \alpha Z_y \boldsymbol{\beta} - \rho Z_y (\boldsymbol{\beta} \circ \mathbf{v}) - Z_y \boldsymbol{\eta}_u)^T (\mathbf{y} - \alpha Z_y \boldsymbol{\beta} - \rho Z_y (\boldsymbol{\beta} \circ \mathbf{v}) - Z_y \boldsymbol{\eta}_u) / 2$$

The posterior conditional distribution of σ_β^2 is an inverse gamma distribution with the shape parameter $\sigma_{\beta_shape}^2 = \frac{\sum_{j=1}^p \gamma_j}{2} + a_\beta$, and the scale parameter $\sigma_{\beta_scale}^2 = \frac{\sum_{j=1}^p \gamma_j \beta_j^2 + 2b_\beta}{2}$.

The posterior conditional distribution of σ_η^2 is an inverse gamma distribution with the shape parameter $\sigma_{\eta_shape}^2 = \frac{\sum_{j=1}^p \tau_j}{2} + a_\eta$, and the scale parameter $\sigma_{\eta_scale}^2 = \frac{\sum_{j=1}^p \tau_j \eta_{uj}^2 + 2b_\eta}{2}$.

For all analyses in the present study, we ran 1,000 Gibbs sampling iterations with the first 200 as burn-in and obtain the posterior samples of α for hypothesis testing, which is described in the next section.

3. Parameter estimation and p value computation

We denote μ_α and σ_α^2 as the posterior mean and the posterior variance of the causal effect parameter α . Since both the likelihood and the posterior follow normal distributions asymptotically, and because we also use a normal distribution $N(0, \sigma_0^2)$ as the prior distribution, we can easily obtain the approximate maximum likelihood estimate and its standard error by the method of moments as

$$\hat{\alpha} = \frac{\sigma_0^2 \mu_\alpha}{\sigma_0^2 - \sigma_\alpha^2},$$

$$se(\hat{\alpha}) = \frac{\sigma_0 \sigma_\alpha}{\sqrt{\sigma_0^2 - \sigma_\alpha^2}}$$

The resulting z-score $Z = \frac{\hat{\alpha}}{se(\hat{\alpha})}$ follows standard normal distribution asymptotically, which allows us to obtain a corresponding p value for testing the null hypothesis $H_0: \alpha = 0$. In present study, because we use a non-informative prior for α by set $\sigma_0^2 \rightarrow \infty$, we have $\hat{\alpha} = \mu_\alpha$ and $se(\hat{\alpha}) = \sigma_\alpha$. Note that the choice of the prior on α does not influence the inference results as we only use the likelihood, expressed as the ratio of the posterior and prior, for hypothesis test^{45,46}.

4. MRAID model for summary statistics

The MRAID model for summary statistics can be presented in the form of marginal effect size estimates and the LD matrix. We denote the LD structure of the instrumental SNPs as Σ_1 in exposure GWAS data, and Σ_2 in outcome GWAS data, both are $p \times p$ symmetric positive definite matrices. Often, Σ_1 and Σ_2 are from the same LD reference panels (e.g. the individuals with European ancestry from the 1,000 Genomes project). The proposed MRAID model for summary statistics can be constructed as the following two equations

$$\widehat{\boldsymbol{\beta}}_x = \Sigma_1 \boldsymbol{\beta} + \mathbf{e}_x \quad (3)$$

$$\widehat{\boldsymbol{\beta}}_y = \alpha \Sigma_2 \boldsymbol{\beta} + \rho \Sigma_2 (\boldsymbol{\beta} \circ \mathbf{v}) + \Sigma_2 \boldsymbol{\eta}_u + \mathbf{e}_y \quad (4)$$

where $\widehat{\boldsymbol{\beta}}_x$ is the p -vector of estimates for the marginal SNP effect sizes on the exposure; $\widehat{\boldsymbol{\beta}}_y$ is the p -vector of estimates for the marginal SNP effect sizes on the outcome; both \mathbf{e}_x and \mathbf{e}_y are p -vector of residual errors. \mathbf{e}_x is a p -vector of residual error following a multivariate normal distribution $N(0, \Sigma_1 \sigma_x^2 / (n_1 - 1))$, and \mathbf{e}_y is a p -vector of residual error following a multivariate normal distribution $N(0, \Sigma_2 \sigma_y^2 / (n_2 - 1))$.

5. Detailed sampling steps of MRAID model for summary statistics

For the summary statistics version of MRAID, we can obtain the posterior distribution for all parameters as

$$\begin{aligned} & f(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}_u, \boldsymbol{\tau}, \mathbf{v}, \rho | \widehat{\boldsymbol{\beta}}_x, \widehat{\boldsymbol{\beta}}_y) \\ & \propto f(\widehat{\boldsymbol{\beta}}_x, \widehat{\boldsymbol{\beta}}_y | \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}_u, \boldsymbol{\tau}, \mathbf{v}, \rho) f(\boldsymbol{\beta} | \boldsymbol{\gamma}) f(\boldsymbol{\eta}_u | \boldsymbol{\tau}) f(\mathbf{v} | \boldsymbol{\gamma}) f(\boldsymbol{\gamma} | \pi_\beta) f(\boldsymbol{\tau} | \boldsymbol{\gamma}) f(\pi_\beta) f(\pi_1) f(\pi_0) f(\pi_c) \\ & \propto f(\widehat{\boldsymbol{\beta}}_x | \boldsymbol{\beta}) f(\widehat{\boldsymbol{\beta}}_y | \alpha, \boldsymbol{\beta}, \boldsymbol{\eta}_u, \mathbf{v}, \rho) f(\boldsymbol{\beta} | \boldsymbol{\gamma}) f(\boldsymbol{\eta}_u | \boldsymbol{\tau}) f(\mathbf{v} | \boldsymbol{\gamma}) f(\boldsymbol{\gamma} | \pi_\beta) f(\boldsymbol{\tau} | \boldsymbol{\gamma}) f(\pi_\beta) f(\pi_1) f(\pi_0) f(\pi_c) \\ & \propto (2\pi)^{-\frac{p}{2}} \left(\frac{\sigma_y^2}{n_2 - 1} \right)^{-\frac{p}{2}} |\Sigma_2|^{-\frac{1}{2}} \\ & \exp \left(- \frac{(\widehat{\boldsymbol{\beta}}_y - \alpha \Sigma_{2,-j} \boldsymbol{\beta}_{-j} - \alpha \Sigma_{2,j} \boldsymbol{\beta}_j - \rho \Sigma_{2,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \rho \Sigma_{2,j} (\boldsymbol{\beta}_j \circ \mathbf{v}_j) - \Sigma_2 \boldsymbol{\eta}_u)^T (n_2 - 1) \Sigma_2^{-1}}{2\sigma_y^2} \right) \\ & (2\pi)^{-\frac{p}{2}} \left(\frac{\sigma_x^2}{n_1 - 1} \right)^{-\frac{p}{2}} |\Sigma_1|^{-\frac{1}{2}} \exp \left(- \frac{(\widehat{\boldsymbol{\beta}}_x - \Sigma_{1,-j} \boldsymbol{\beta}_{-j} - \Sigma_{1,j} \boldsymbol{\beta}_j)^T (n_1 - 1) \Sigma_1^{-1} (\widehat{\boldsymbol{\beta}}_x - \Sigma_{1,-j} \boldsymbol{\beta}_{-j} - \Sigma_{1,j} \boldsymbol{\beta}_j)}{2\sigma_x^2} \right) \\ & \prod_{j=1}^p \left[(2\pi)^{-\frac{1}{2}} (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left(- \frac{\beta_j^2}{2\sigma_\beta^2} \right) \right]^{\gamma_j} \prod_{j=1}^p [\delta_0(\beta_j)]^{1-\gamma_j} \prod_{j=1}^p \left[(2\pi)^{-\frac{1}{2}} (\sigma_\eta^2)^{-\frac{1}{2}} \exp \left(- \frac{\eta_{uj}^2}{2\sigma_\eta^2} \right) \right]^{\tau_j} \\ & \prod_{j=1}^p [\delta_0(\eta_{uj})]^{1-\tau_j} \pi_c^{\sum_{j=1}^p \gamma_j \nu_j} (1 - \pi_c)^{\sum_{j=1}^p \nu_j (1 - \nu_j)} \\ & \pi_\beta^{\sum_{j=1}^p \nu_j (1 - \pi_\beta)} (1 - \pi_\beta)^{\sum_{j=1}^p (1 - \nu_j)} \pi_1^{\sum_{j=1}^p \nu_j \tau_j} (1 - \pi_1)^{\sum_{j=1}^p \nu_j (1 - \tau_j)} \pi_0^{\sum_{j=1}^p (1 - \nu_j) \tau_j} (1 \\ & \quad - \pi_0)^{\sum_{j=1}^p (1 - \nu_j) (1 - \tau_j)} \\ & \text{beta}(\pi_\beta, \lambda_{\beta 1}, \lambda_{\beta 2}) \text{beta}(\pi_1, \lambda_{21}, \lambda_{22}) \text{beta}(\pi_0, \lambda_{31}, \lambda_{32}) \text{beta}(\pi_c, \lambda_{c1}, \lambda_{c2}) \end{aligned}$$

$$\begin{aligned}
& \propto C \cdot \exp \left(- \frac{(\widehat{\boldsymbol{\beta}}_y - \alpha \Sigma_{2,-j} \boldsymbol{\beta}_{-j} - \rho \Sigma_{2,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \Sigma_2 \boldsymbol{\eta}_u)^T (n_2 - 1) \Sigma_2^{-1} (\widehat{\boldsymbol{\beta}}_y - \alpha \Sigma_{2,-j} \boldsymbol{\beta}_{-j} - \rho \Sigma_{2,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \Sigma_2 \boldsymbol{\eta}_u)}{2 \sigma_y^2} \right) \\
& \exp \left(- \frac{(\widehat{\boldsymbol{\beta}}_x - \Sigma_{1,-j} \boldsymbol{\beta}_{-j})^T (n_1 - 1) \Sigma_1^{-1} (\widehat{\boldsymbol{\beta}}_x - \Sigma_{1,-j} \boldsymbol{\beta}_{-j})}{2 \sigma_x^2} \right) \\
& \prod_{i \neq j}^p \left[\pi_\beta (2\pi)^{-\frac{1}{2}} (\sigma_\beta^2)^{-\frac{1}{2}} \exp \left(- \frac{\beta_i^2}{2 \sigma_\beta^2} \right) \right]^{\gamma_i} \prod_{i \neq j}^p [(1 - \pi_\beta) \delta_0(\beta_i)]^{1 - \gamma_i} \\
& \prod_{j=1}^p \left[(2\pi)^{-\frac{1}{2}} (\sigma_\eta^2)^{-\frac{1}{2}} \exp \left(- \frac{\eta_{uj}^2}{2 \sigma_\eta^2} \right) \right]^{\tau_j} \\
& \prod_{j=1}^p [\delta_0(\eta_{uj})]^{1 - \tau_j} \pi_c^{\sum_{j=1}^p \gamma_j v_j} (1 - \pi_c)^{\sum_{j=1}^p \gamma_j (1 - v_j)} \\
& \pi_1^{\sum_{j=1}^p \gamma_j \tau_j} (1 - \pi_1)^{\sum_{j=1}^p \gamma_j (1 - \tau_j)} \pi_0^{\sum_{j=1}^p (1 - \gamma_j) \tau_j} (1 - \pi_0)^{\sum_{j=1}^p (1 - \gamma_j) (1 - \tau_j)} \\
& \text{beta}(\pi_\beta, \lambda_{\beta_1}, \lambda_{\beta_2}) \text{beta}(\pi_1, \lambda_{21}, \lambda_{22}) \text{beta}(\pi_0, \lambda_{31}, \lambda_{32}) \text{beta}(\pi_c, \lambda_{c1}, \lambda_{c2}) \\
& \cdot \exp \left(- \frac{\beta_j^2 K - 2 \left(\frac{(\alpha + \rho v_j) (\widehat{\boldsymbol{\beta}}_y - \alpha \Sigma_{2,-j} \boldsymbol{\beta}_{-j} - \rho \Sigma_{2,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \Sigma_2 \boldsymbol{\eta}_u)^T (n_2 - 1) \Sigma_2^{-1} \Sigma_{2,j}}{\sigma_y^2} + \frac{(\widehat{\boldsymbol{\beta}}_x - \Sigma_{1,-j} \boldsymbol{\beta}_{-j})^T (n_1 - 1) \Sigma_1^{-1} \Sigma_{1,j}}{\sigma_x^2} \right) \beta_j}{2} \right)
\end{aligned}$$

where

$$\begin{aligned}
C = & (2\pi)^{-\frac{p}{2}} \left(\frac{\sigma_y^2}{n_2 - 1} \right)^{-\frac{p}{2}} |\Sigma_2|^{-\frac{1}{2}} (2\pi)^{-\frac{p}{2}} \left(\frac{\sigma_x^2}{n_1 - 1} \right)^{-\frac{p}{2}} |\Sigma_1|^{-\frac{1}{2}} (2\pi)^{-\frac{\gamma_j}{2}} \pi_\beta^{\gamma_j} (1 - \\
& \pi_\beta)^{1 - \gamma_j} (\sigma_\beta^2)^{-\frac{\gamma_j}{2}} (\delta_0(\beta_j))^{1 - \gamma_j};
\end{aligned}$$

K is a scalar in the form of

$$K = \frac{(\alpha + \rho v_j)^2 (n_2 - 1)}{\sigma_y^2} + \frac{(n_1 - 1)}{\sigma_x^2} + \frac{\gamma_j}{\sigma_\beta^2}$$

$\boldsymbol{\beta}_{-j}$ is the vector $\boldsymbol{\beta}$ without the j th element β_j ;

Based on the above likelihood, we conduct Gibbs sampling on all parameters that include $\theta = (\beta_j, \gamma_j, \eta_{uj}, \tau_j, v_j, \alpha, \sigma_\beta^2, \sigma_\eta^2, \sigma_x^2, \sigma_y^2, \pi_\beta, \pi_1, \pi_0, \pi_c, \rho)$.

Given $\gamma_j=1$ and other parameters, the posterior distribution of β_j is a normal distribution with mean U_{β_j} and variance $\sigma_{\beta_j}^2$, where

$$\begin{aligned}
U_{\beta_j} = & \left(\frac{(\alpha + \rho v_j) (\widehat{\boldsymbol{\beta}}_y - \alpha \Sigma_{2,-j} \boldsymbol{\beta}_{-j} - \rho \Sigma_{2,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \Sigma_2 \boldsymbol{\eta}_u)^T (n_2 - 1) \Sigma_2^{-1} \Sigma_{2,j}}{\sigma_y^2} + \frac{(\widehat{\boldsymbol{\beta}}_x - \Sigma_{1,-j} \boldsymbol{\beta}_{-j})^T (n_1 - 1) \Sigma_1^{-1} \Sigma_{1,j}}{\sigma_x^2} \right) / K_1 \\
\sigma_{\beta_j}^2 = & 1/K_1, \quad K_1 = \frac{(\alpha + \rho v_j)^2 (n_2 - 1)}{\sigma_y^2} + \frac{(n_1 - 1)}{\sigma_x^2} + \frac{1}{\sigma_\beta^2}
\end{aligned}$$

Given $\tau_j=1$ and other parameters, the posterior conditional distribution of η_{uj} is a normal distribution with mean U_{η_j} and variance $\sigma_{\eta_j}^2$, where

$$\begin{aligned}
U_{\eta_j} = & \left(\frac{(\widehat{\boldsymbol{\beta}}_y - \Sigma_{2,-j} \boldsymbol{\eta}_u - \alpha \Sigma_{2,-j} \boldsymbol{\beta}_{-j} - \rho \Sigma_{2,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \Sigma_2 \boldsymbol{\eta}_u)^T (n_2 - 1) \Sigma_2^{-1} \Sigma_{2,j}}{\sigma_y^2} \right) / K_2 \\
\sigma_{\eta_j}^2 = & 1/K_2, \quad K_2 = \frac{n_2 - 1}{\sigma_y^2} + \frac{1}{\sigma_\eta^2}
\end{aligned}$$

After integrating out the parameter β_j , we obtain the posterior conditional distribution of γ_j

as

$$p(\gamma_j = 1 | \beta_{-j}, \eta_u, \alpha, \sigma_x^2, \sigma_y^2, \sigma_\beta^2, \pi_\beta, \pi_1, \pi_0, \pi_c, \rho) \propto \exp\left(\frac{\mu_{\beta_j}^2}{2\sigma_{\beta_j}^2} + 0.5 * \log(\sigma_{\beta_j}^2) - 0.5 * \log(\sigma_\beta^2) + \log(\pi_\beta) + \tau_j \log(\pi_1) + (1 - \tau_j) \log(1 - \pi_1) + v_j \log(\pi_c) + (1 - v_j) \log(1 - \pi_c)\right)$$

$$p(\gamma_j = 0 | \beta_{-j}, \eta_u, \alpha, \sigma_x^2, \sigma_y^2, \sigma_\beta^2, \pi_\beta, \pi_1, \pi_0) \propto \exp(\log(1 - \pi_\beta) + \tau_j \log(\pi_0) + (1 - \tau_j) \log(1 - \pi_0))$$

After integrating out the parameter η_{uj} , we obtain the posterior conditional distribution of τ_j as

$$p(\tau_j = 1 | \eta_{u,-j}, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_1, \pi_0) \propto \exp\left(\frac{\mu_{\eta_j}^2}{2\sigma_{\eta_j}^2} + 0.5 * \log(\sigma_{\eta_j}^2) - 0.5 * \log(\sigma_\eta^2) + \gamma_j \log(\pi_1) + (1 - \gamma_j) \log(\pi_0)\right)$$

$$p(\tau_j = 0 | \eta_{-j}, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_1, \pi_0) \propto \exp(\gamma_j \log(1 - \pi_1) + (1 - \gamma_j) \log(1 - \pi_0))$$

We also obtain the posterior conditional distribution of v_j as

$$p(v_j = 1 | \eta_u, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_c, \pi_1, \pi_0) \propto \exp\left(-\frac{\rho^2 \beta_j^2 (\mathbf{n}_2 - \mathbf{1}) - 2(\widehat{\boldsymbol{\beta}}_y - \alpha \Sigma_2 \boldsymbol{\beta} - \rho \Sigma_{2,-j} (\boldsymbol{\beta}_{-j} \circ \mathbf{v}_{-j}) - \Sigma_2 \boldsymbol{\eta}_u)^T (\mathbf{n}_2 - \mathbf{1}) \Sigma_2^{-1} \rho \Sigma_{2,j} \beta_j}{2\sigma_y^2} + \gamma_j \log(\pi_c)\right)$$

$$p(v_j = 0 | \eta_u, \beta, \alpha, \sigma_y^2, \sigma_\eta^2, \pi_c, \pi_1, \pi_0) \propto \exp(\gamma_j \log(1 - \pi_c))$$

The posterior conditional distribution of ρ is a normal distribution with variance $\sigma_\rho^2 = \left(\frac{(\mathbf{v} \circ \boldsymbol{\beta})^T (\mathbf{n}_2 - \mathbf{1}) \Sigma_2 (\mathbf{v} \circ \boldsymbol{\beta})}{\sigma_y^2}\right)^{-1}$ and mean $\mu_\rho = \frac{(\mathbf{v} \circ \boldsymbol{\beta})^T (\mathbf{n}_2 - \mathbf{1}) \widehat{\boldsymbol{\beta}}_y - (\mathbf{v} \circ \boldsymbol{\beta})^T (\mathbf{n}_2 - \mathbf{1}) \Sigma_2 \boldsymbol{\eta}_u - (\mathbf{v} \circ \boldsymbol{\beta})^T (\mathbf{n}_2 - \mathbf{1}) \Sigma_2 \boldsymbol{\beta} \alpha}{\sigma_y^2} \sigma_\rho^2$.

The posterior conditional distribution of α is a normal distribution with variance $\sigma_\alpha^2 = \left(\frac{\boldsymbol{\beta}^T (\mathbf{n}_2 - \mathbf{1}) \Sigma_2 \boldsymbol{\beta}}{\sigma_y^2}\right)^{-1}$ and mean $\mu_\alpha = \frac{\boldsymbol{\beta}^T (\mathbf{n}_2 - \mathbf{1}) \widehat{\boldsymbol{\beta}}_y - \boldsymbol{\beta}^T (\mathbf{n}_2 - \mathbf{1}) \Sigma_2 \boldsymbol{\eta}_u - \boldsymbol{\beta}^T (\mathbf{n}_2 - \mathbf{1}) \Sigma_2 (\boldsymbol{\beta} \circ \mathbf{v}) \rho}{\sigma_y^2} \sigma_\alpha^2$.

The posterior conditional distribution of π_β is $beta(\sum_{j=1}^p \gamma_j + \lambda_{\beta 1}, \sum_{j=1}^p (1 - \gamma_j) + \lambda_{\beta 2})$.

The posterior conditional distribution of π_1 is $beta(\sum_{j=1}^p \gamma_j \tau_j + \lambda_{21}, \sum_{j=1}^p \gamma_j (1 - \tau_j) + \lambda_{22})$

The posterior conditional distribution of π_0 is $beta(\sum_{j=1}^p (1 - \gamma_j) \tau_j + \lambda_{31}, \sum_{j=1}^p (1 - \gamma_j) (1 - \tau_j) + \lambda_{32})$

The posterior conditional distribution of π_c is $beta(\sum_{j=1}^p \gamma_j v_j + \lambda_{c1}, \sum_{j=1}^p \gamma_j (1 - v_j) + \lambda_{c2})$

The posterior conditional distribution of σ_x^2 is an inverse gamma distribution with the shape parameter $\sigma_{x_{shape}}^2 = \frac{n_1}{2} - 1$ and the scale parameter

$$\sigma_{x_{scale}}^2 = ((n_1 - 1) + \boldsymbol{\beta}^T (\mathbf{n}_1 - \mathbf{1}) \Sigma_1 \boldsymbol{\beta} - 2\boldsymbol{\beta}^T (\mathbf{n}_1 - \mathbf{1}) \widehat{\boldsymbol{\beta}}_x) / 2$$

The posterior conditional distribution of σ_y^2 is an inverse gamma distribution with the shape parameter $\sigma_{y_{shape}}^2 = \frac{n_2}{2} - 1$, and the scale parameter

$$\sigma_{y_{scale}}^2 = ((n_2 - 1) + \alpha^2 \boldsymbol{\beta}^T (n_2 - 1) \Sigma_2 \boldsymbol{\beta} + \rho^2 (\boldsymbol{\beta} \circ \mathbf{v})^T (n_2 - 1) \Sigma_2 (\boldsymbol{\beta} \circ \mathbf{v}) + 2\alpha \boldsymbol{\beta}^T (n_2 - 1) \Sigma_2 \boldsymbol{\eta}_u + \boldsymbol{\eta}_u^T (n_2 - 1) \Sigma_2 \boldsymbol{\eta}_u + 2\alpha \boldsymbol{\beta}^T (n_2 - 1) \Sigma_2 (\boldsymbol{\beta} \circ \mathbf{v}) + 2\rho (\boldsymbol{\beta} \circ \mathbf{v})^T (n_2 - 1) \Sigma_2 \boldsymbol{\eta}_u - 2\rho (\boldsymbol{\beta} \circ \mathbf{v})^T (n_2 - 1) \widehat{\boldsymbol{\beta}}_y - 2\alpha \boldsymbol{\beta}^T (n_2 - 1) \widehat{\boldsymbol{\beta}}_y - 2\boldsymbol{\eta}_u^T (n_2 - 1) \widehat{\boldsymbol{\beta}}_y) / 2$$

The posterior conditional distribution of σ_β^2 is an inverse gamma distribution with the shape parameter $\sigma_{\beta_{shape}}^2 = \frac{\sum_{j=1}^p \gamma_j}{2} + a_\beta$, and the scale parameter $\sigma_{\beta_{scale}}^2 = \frac{\sum_{j=1}^p \gamma_j \beta_j^2 + 2b_\beta}{2}$

The posterior conditional distribution of σ_η^2 is an inverse gamma distribution with the shape parameter $\sigma_{\eta_{shape}}^2 = \frac{\sum_{j=1}^p \tau_j}{2} + a_\eta$, and the scale parameter $\sigma_{\eta_{scale}}^2 = \frac{\sum_{j=1}^p \tau_j \eta_{u,j}^2 + 2b_\eta}{2}$.

Supplementary Figures

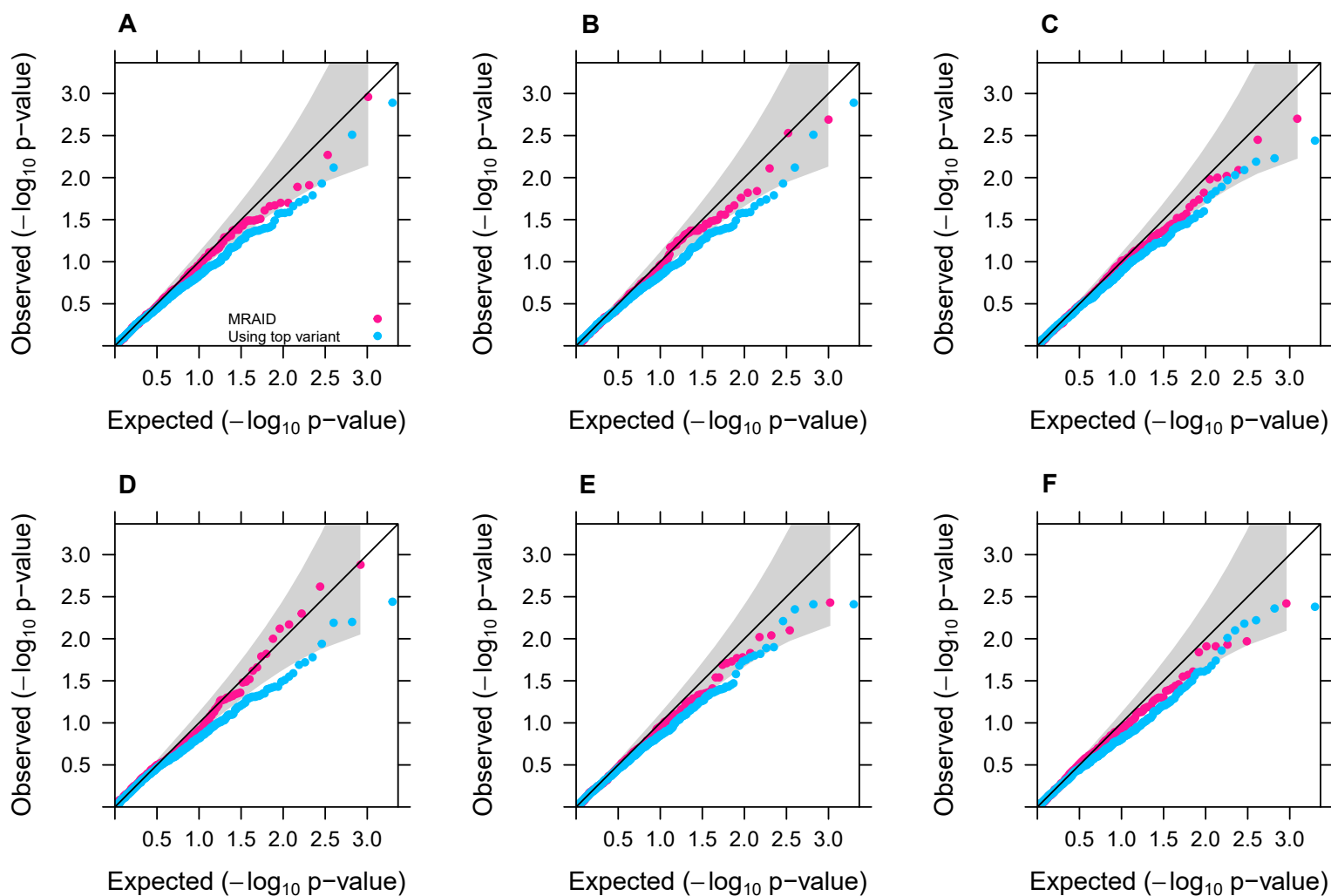


Fig. S1 Type I error control of different MR methods in simple simulations. Type I error control is evaluated by quantile-quantile plot of $-\log_{10}$ p-values from different MR methods on testing the causal effect under the null simulations. Compared methods include MRAID (magenta) and the MR method using only the top variant (blue). We randomly selected one LD block and randomly selected 10 SNPs in the block (A-D) for simulations. Among the 10 SNPs, we randomly selected either one SNP (A, B) or two SNPs (C, D) to be causal and applied different methods to perform MR analysis either with the causal SNP (A, C) or without the causal SNPs (B, D). In addition, we randomly selected two neighborhood LD blocks on chromosome 1 and randomly selected 10 SNPs from each block for another set of simulations (E, F) with one SNP in each LD block randomly selected to be causal and with the MR analysis performed either with (E) or without the two causal SNPs (F). We set $n_1=30000$, $n_2=30000$, $PVE_{zx}=0.25\%$, $PVE_{\alpha}=0$ to examine the performance of MRAID and the MR method that uses only the lead variant.

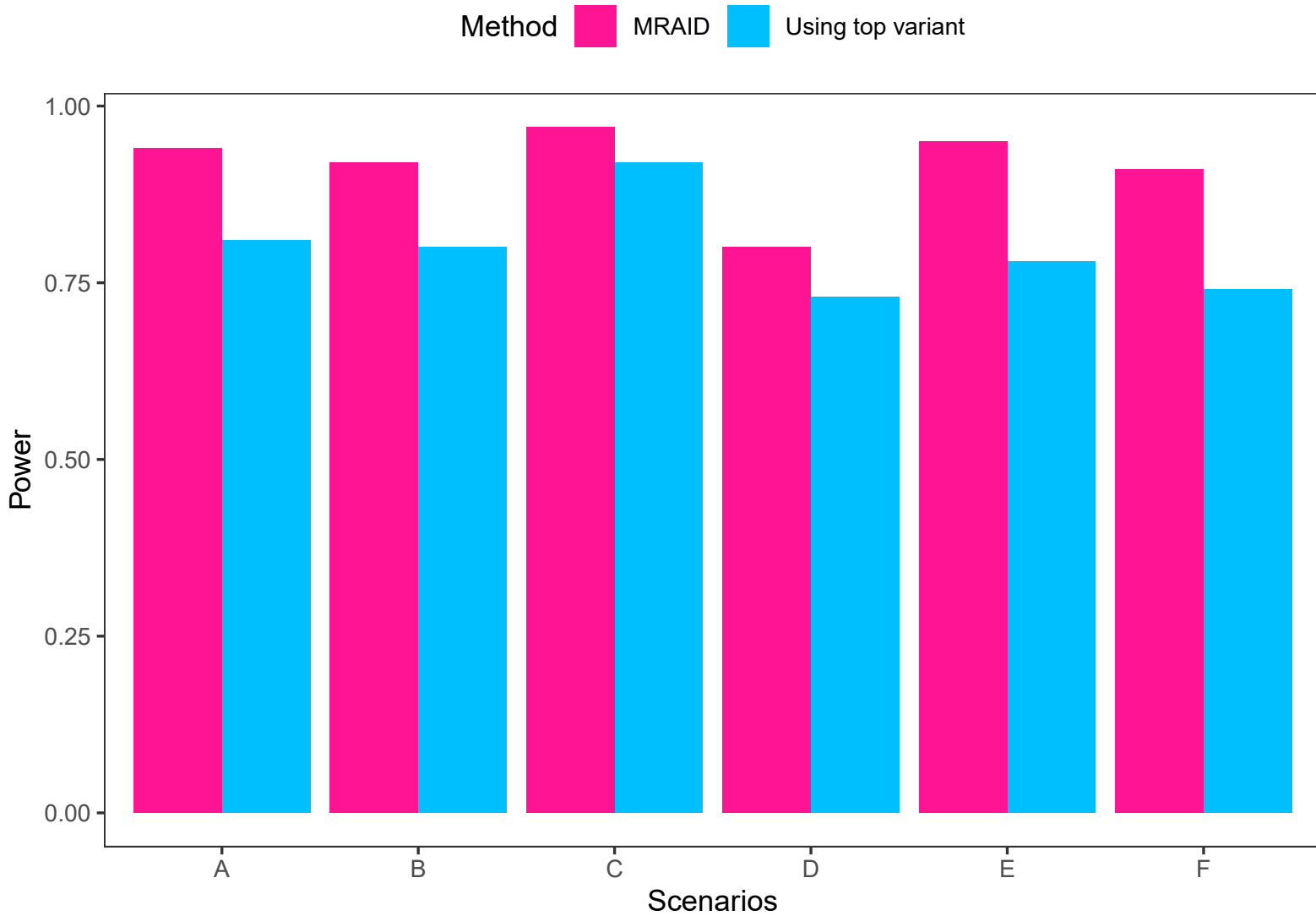


Fig. S2 Power of different MR methods in simple simulations. Power (y-axis) at a false discovery rate of 0.05 to detect the causal effect is plotted against different scenarios (x-axis). Compared methods include MRAID (magenta) and the MR method using only the top variant (blue). Six alternative simulation scenarios are examined. We randomly selected one LD block and randomly selected 10 SNPs in the block (scenarios: A-D). Among the 10 SNPs, we randomly selected either one SNP (scenarios: A, B) or two SNPs (scenarios: C, D) to be causal and applied different methods to perform MR analysis either with the causal SNP (scenarios: A, C) or without the causal SNPs (scenarios: B, D). In addition, we randomly selected two neighborhood LD blocks on chromosome 1 and randomly selected 10 SNPs from each block for another set of simulations (scenarios: E, F) with one SNP in each LD block randomly selected to be causal and with the MR analysis performed either with (scenarios E) or without the two causal SNPs (scenario F). We set $n_1=30000$, $n_2=30000$, $PVE_{zx}=0.25\%$, $PVE_{\alpha}=1\%$ to examine the performance of MRAID and the MR method that uses only the lead variant.

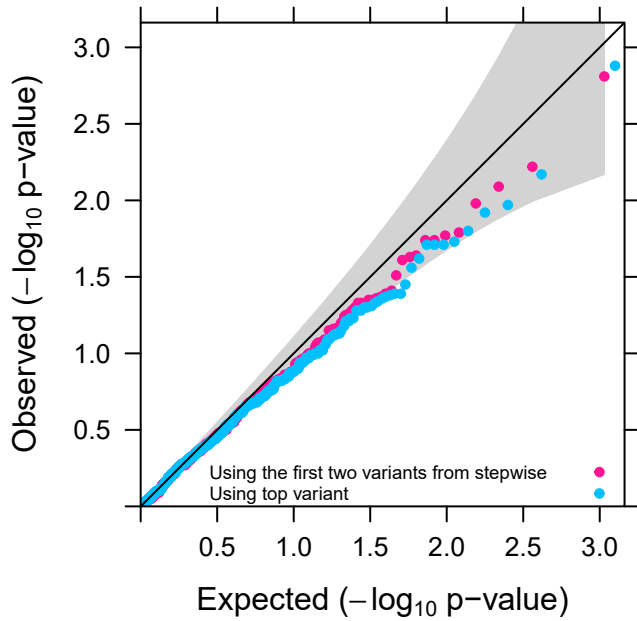
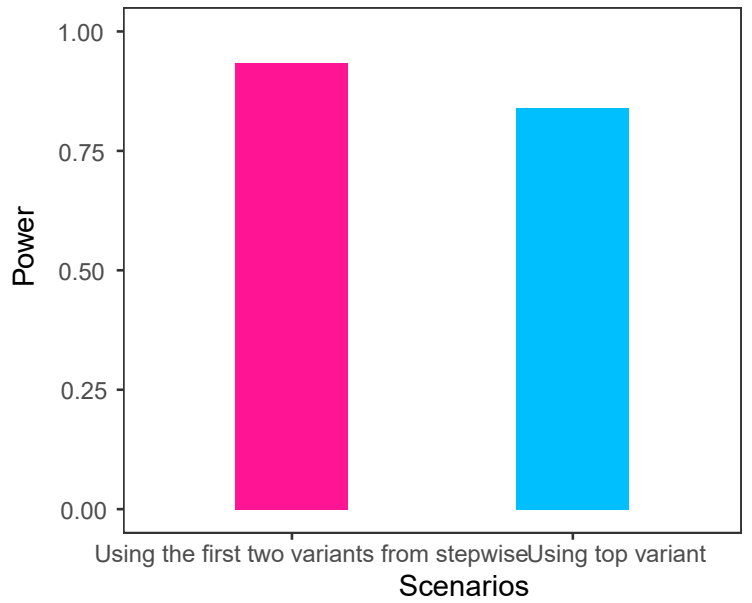
A**B**

Fig. S3 Comparison of different MR methods in simple simulations. Compared methods include MRAID using the lead SNP (deep pink) and MRAID using the top two SNPs obtained from a stepwise regression (blue). We randomly selected one LD block on chromosome 1 and randomly selected 10 SNPs in the block as the candidate SNP set. We then randomly selected one causal SNP from these 10 SNPs to generate data in the absence of both correlated and uncorrelated horizontal pleiotropic effects. We set $n_1=30000$, $n_2=30000$, $PVE_{ZX}=0.25\%$, $PVE_{\alpha}=0$ (null simulation, **A**) or $PVE_{\alpha}=1\%$ (Power simulation, **B**) to examine the performance of these two methods.

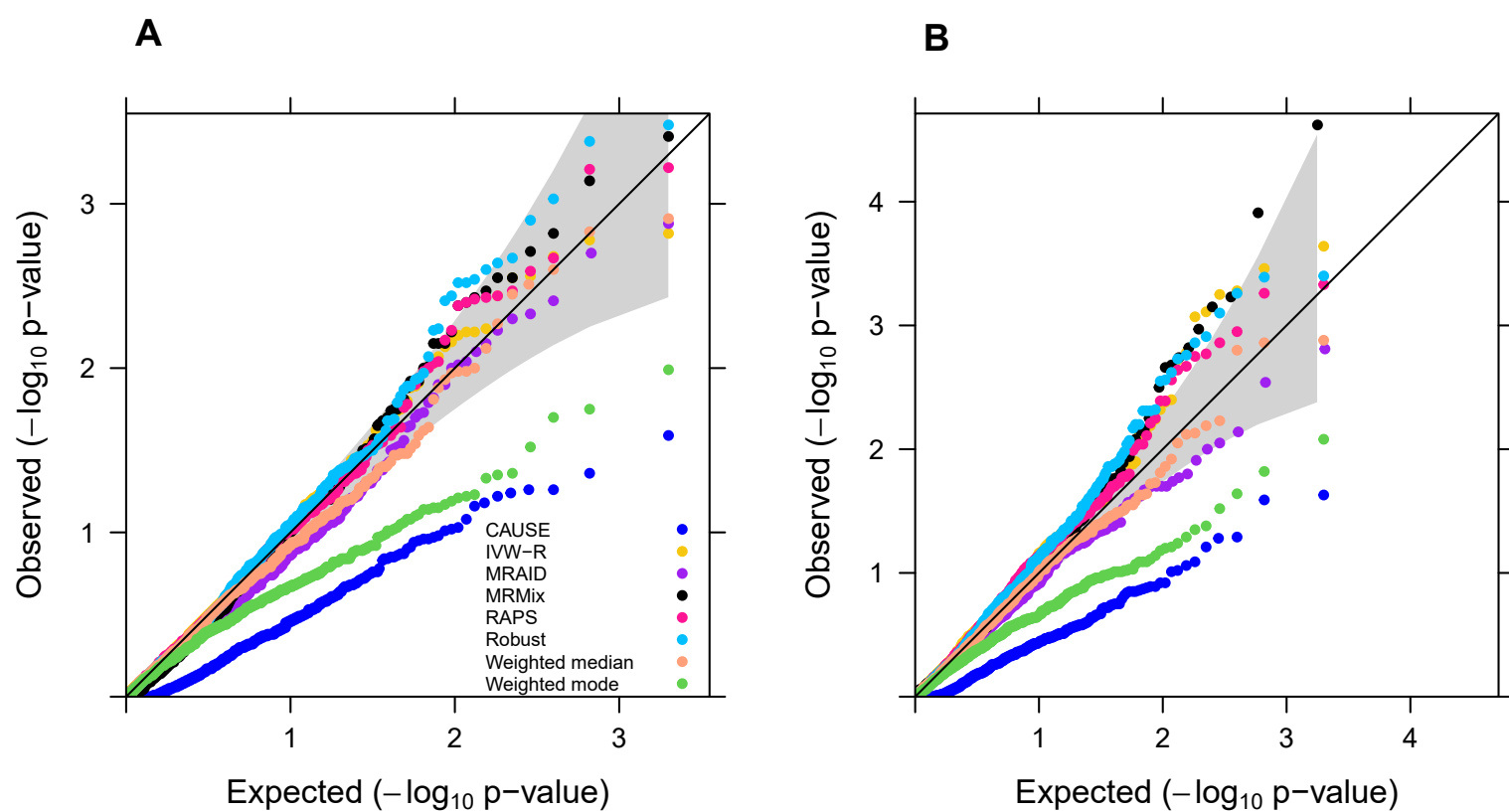


Fig. S4 Type I error control of different MR methods in simulations. Type I error control is evaluated by quantile-quantile plots of $-\log_{10}$ p-values from different MR methods on testing the causal effect in the absence of both correlated and uncorrelated horizontal pleiotropic effects under the null simulations. Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). (A) We simulated 1,000 instrumental SNPs with $PVE_{Z_x} = 10\%$; (B) We simulated 100 instrumental SNPs with $PVE_{Z_x} = 5\%$.

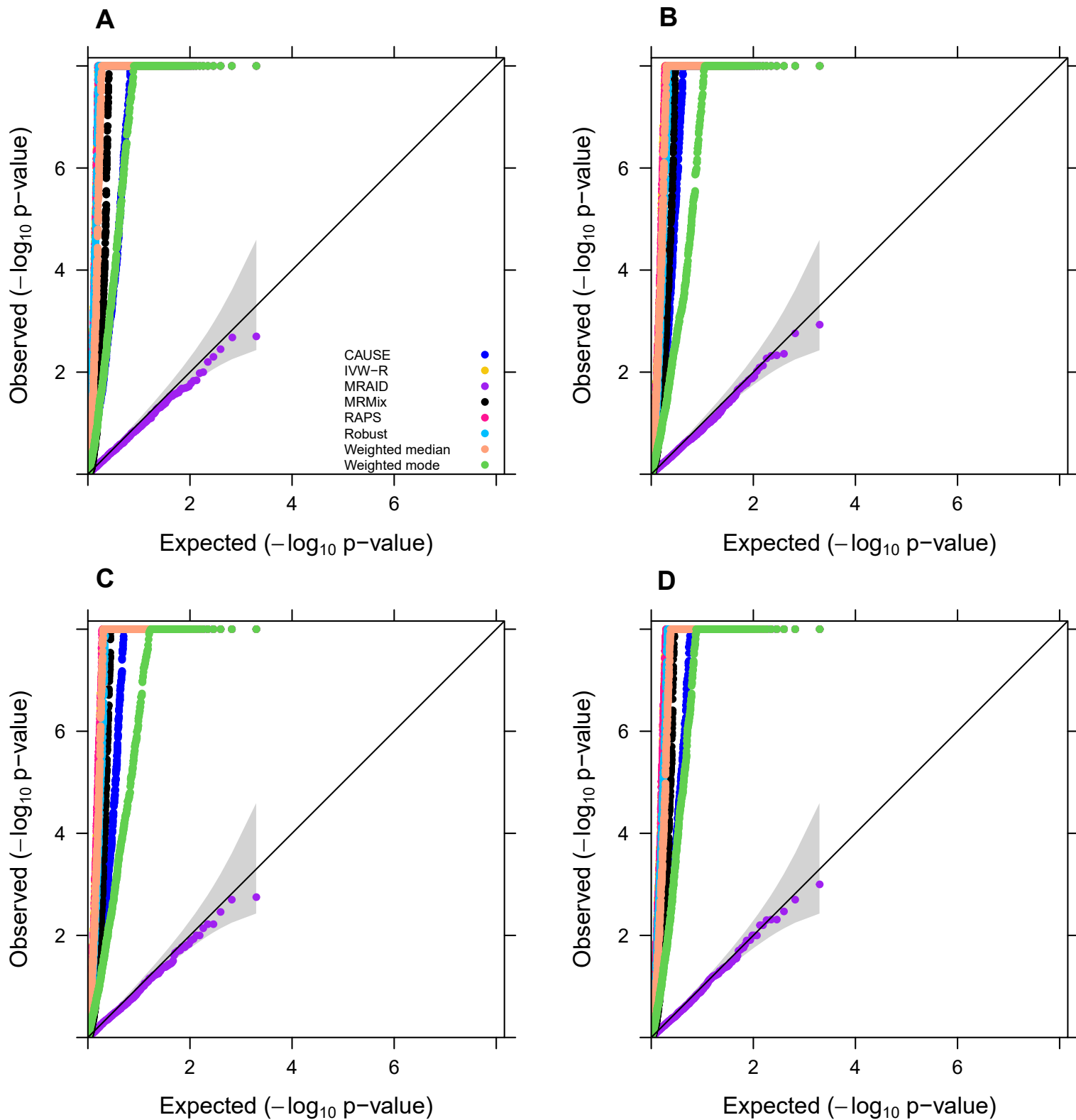


Fig. S5 Type I error control of different MR methods in simulations. Type I error control is evaluated by quantile-quantile plot of $-\log_{10}$ p-values from different MR methods on testing the causal effect in the absence of both correlated and uncorrelated horizontal pleiotropic effects under null simulations. All methods used all exposure-associated SNPs that achieve genome-wide significance level ($p < 5 \times 10^{-8}$) with high LD. Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). (A) We simulated 100 instrumental SNPs with their effect sizes drawing from a normal distribution; (B) We simulated 1,000 instrumental SNPs with their effects size drawing from a normal distribution; (C) We simulated 1,000 instrumental SNPs with their effects size drawing from a BSLMM distribution with 10% SNPs having large effects and 90% SNPs having small effects; (D) We simulated 1,000 instrumental SNPs with their effects size drawing from a BSLMM distribution with 1% SNPs having large effects and 99% SNPs having small effects.

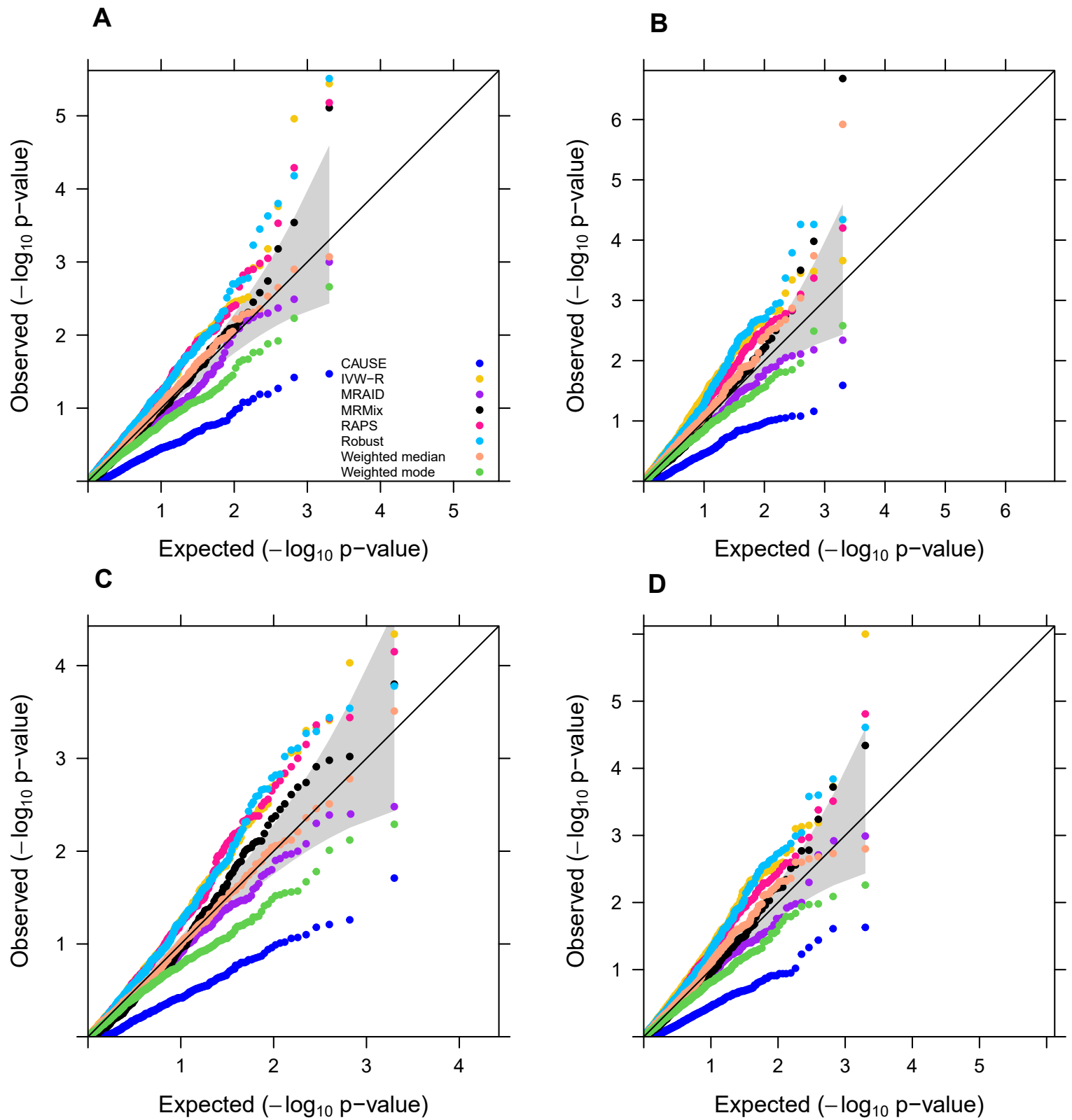


Fig. S6 Type I error control of different MR methods in simulations. Type I error control is evaluated by quantile-quantile plot of $-\log_{10}$ p-values from different MR methods on testing the causal effect under the null simulations in the absence of correlated horizontal pleiotropic effects but in the presence of uncorrelated horizontal pleiotropic effects. Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). We simulated 100 instrumental SNPs with different uncorrelated horizontal pleiotropy effects ($PVE_u=5\%$, **A**, **B**; $PVE_u=2.5\%$, **C**, **D**) and set the proportions of instrumental SNPs having the uncorrelated horizontal pleiotropy effect to be either no overlap (**A**, **C**) or 10% (**B**, **D**).

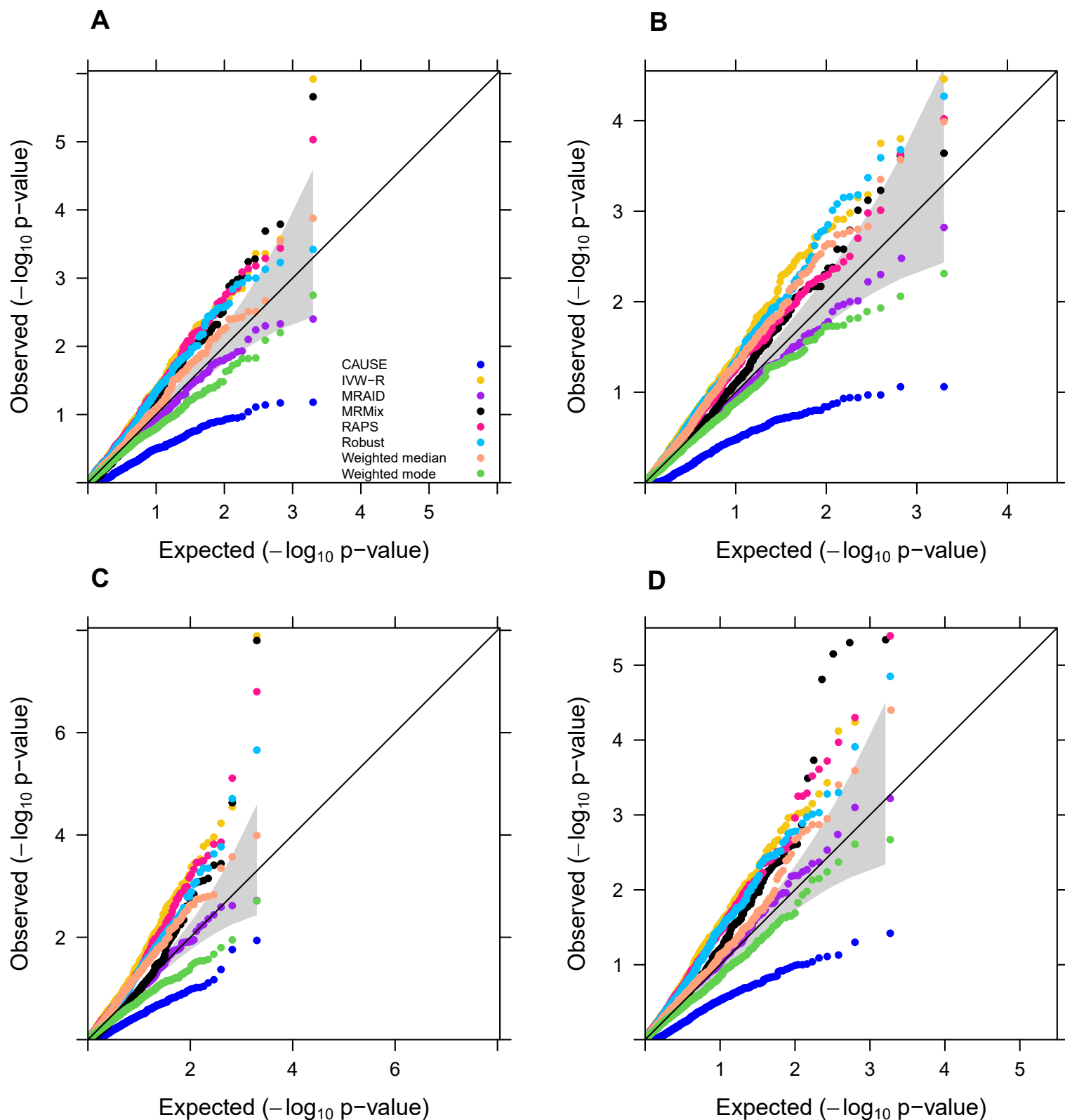


Fig. S7 Type I error control of different MR methods in simulations. Type I error control is evaluated by quantile-quantile plot of $-\log_{10}$ p-values from different MR methods on testing the causal effect under the null simulations in the presence of both correlated and uncorrelated horizontal pleiotropic effects ($PVE_u=5\%$). Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). We simulated 100 instrumental SNPs with different proportions of instrumental SNPs having uncorrelated horizontal pleiotropy effects (no overlap; **A**, **C**, **D**; 20%; **B**), different probability of having the correlated pleiotropy effect from an unobserved confounder ($\pi_c=5\%$, **A**, **B**, **C**; $\pi_c=10\%$, **D**) and different correlated pleiotropy effect sizes $\rho=\sqrt{0.02}$ (**A**, **B**, **D**) or $\rho=\sqrt{0.05}$ (**C**).

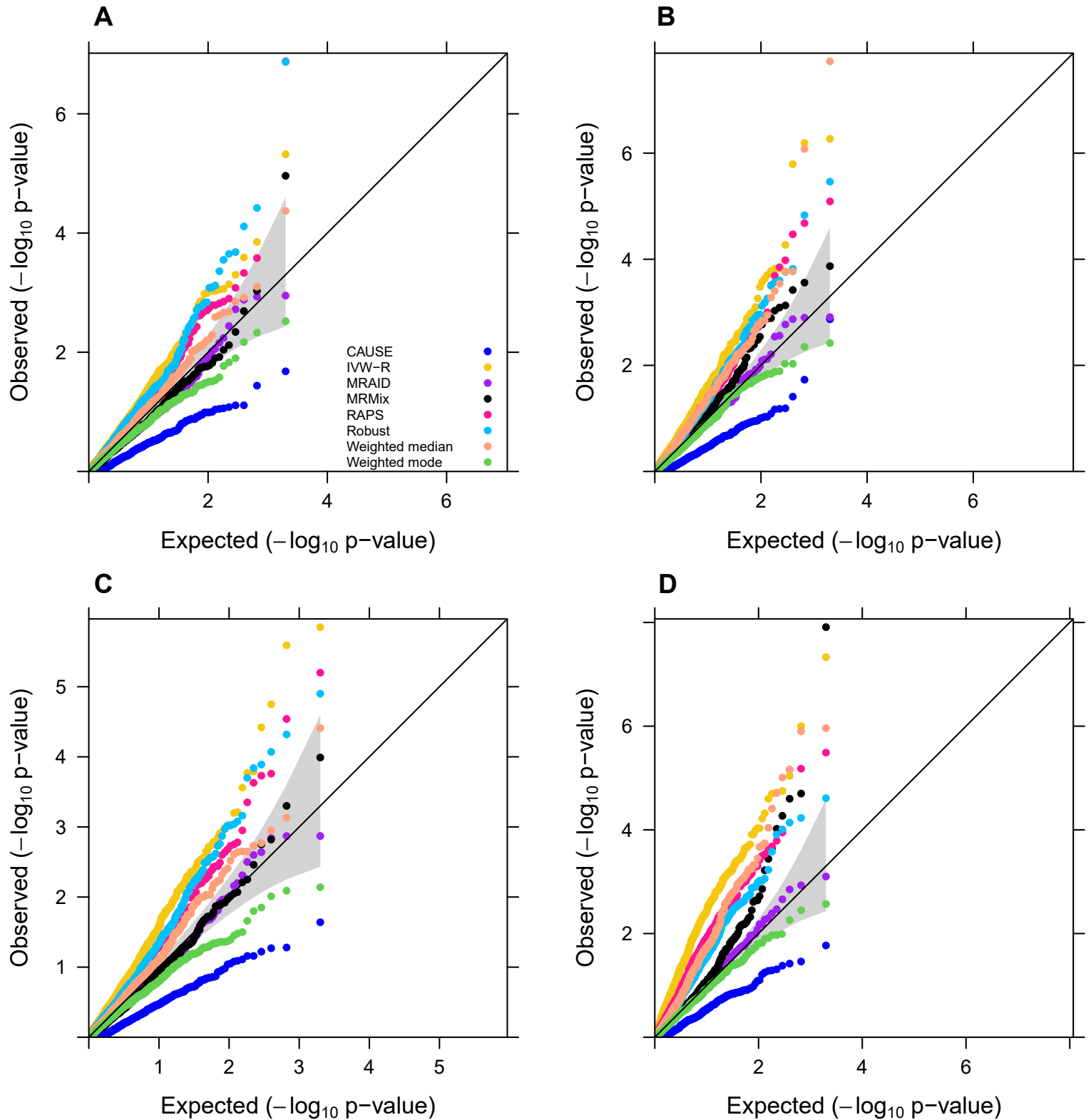


Fig. S8 Type I error control of different MR methods in simulations. Type I error control is evaluated by quantile-quantile plot of $-\log_{10}$ p-values from different MR methods on testing the causal effect under the null simulations in the presence of both correlated and uncorrelated horizontal pleiotropic effect ($PVE_{ii}=5\%$). Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). We simulated 100 instrumental SNPs with different overlapped proportion among instrumental SNPs and the uncorrelated pleiotropic SNPs (10%, **A**, **C**; 30%, **B**, **D**). The correlated horizontal pleiotropy is, rather than simulated from an unobserved confounder, generated through assigning correlation either 0.1 (**A**, **B**) or 0.3 (**C**, **D**) to the effect size of the overlapped SNPs from bivariate normal distribution.

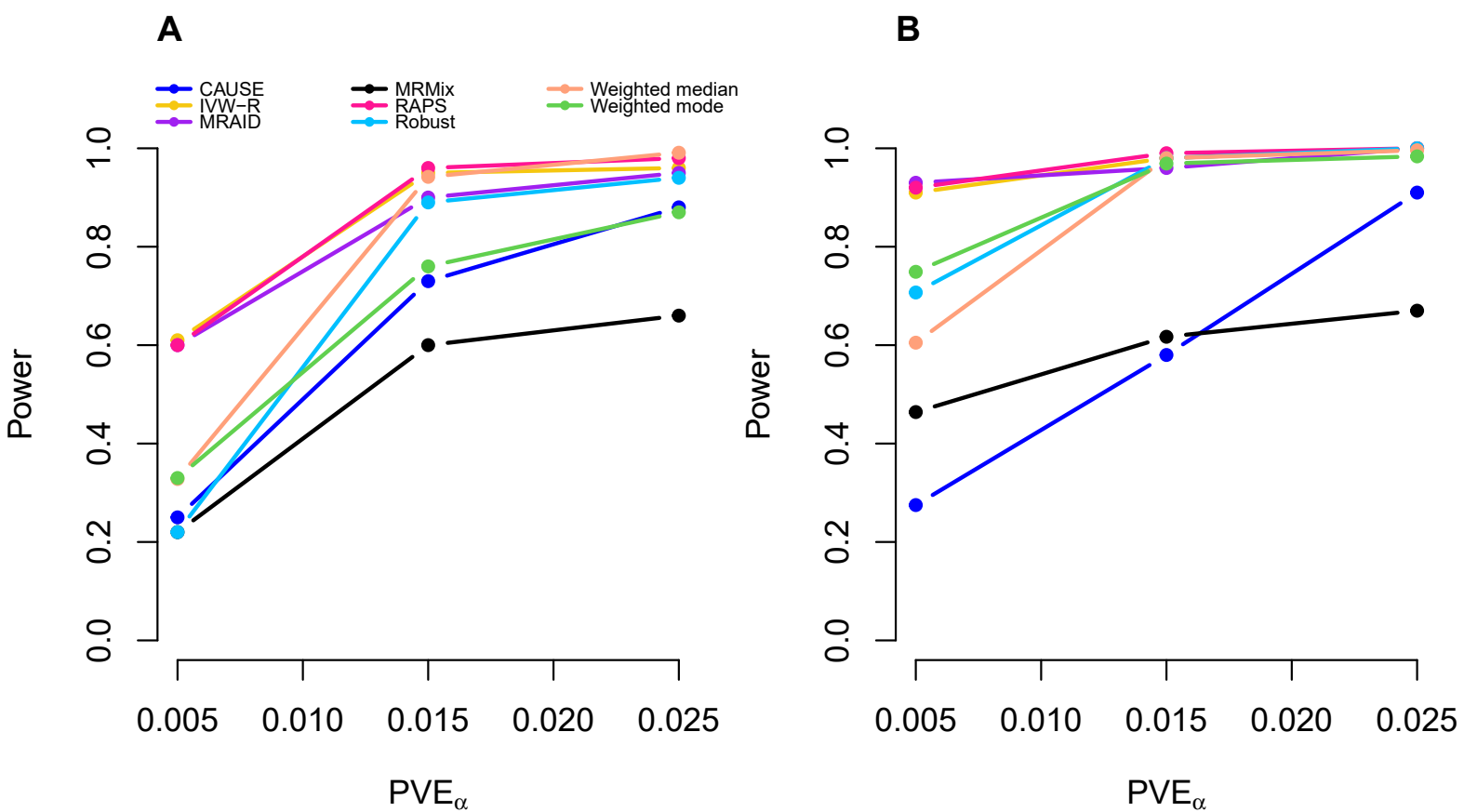


Fig. S9 Power of different MR methods in the absence of both uncorrelated and correlated pleiotropic effects. Power (y-axis) at a false discovery rate of 0.05 to detect the causal effect is plotted against different causal effect size characterized by PVE_α (x-axis). Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). **(A)** We simulated 1,000 instrumental SNPs with their effect size drawing from a normal distribution; **(B)** We simulated 1,000 instrumental SNPs with their effect size drawing from a BSLMM distribution with 10% SNPs having large effects and 90% SNPs having small effects.

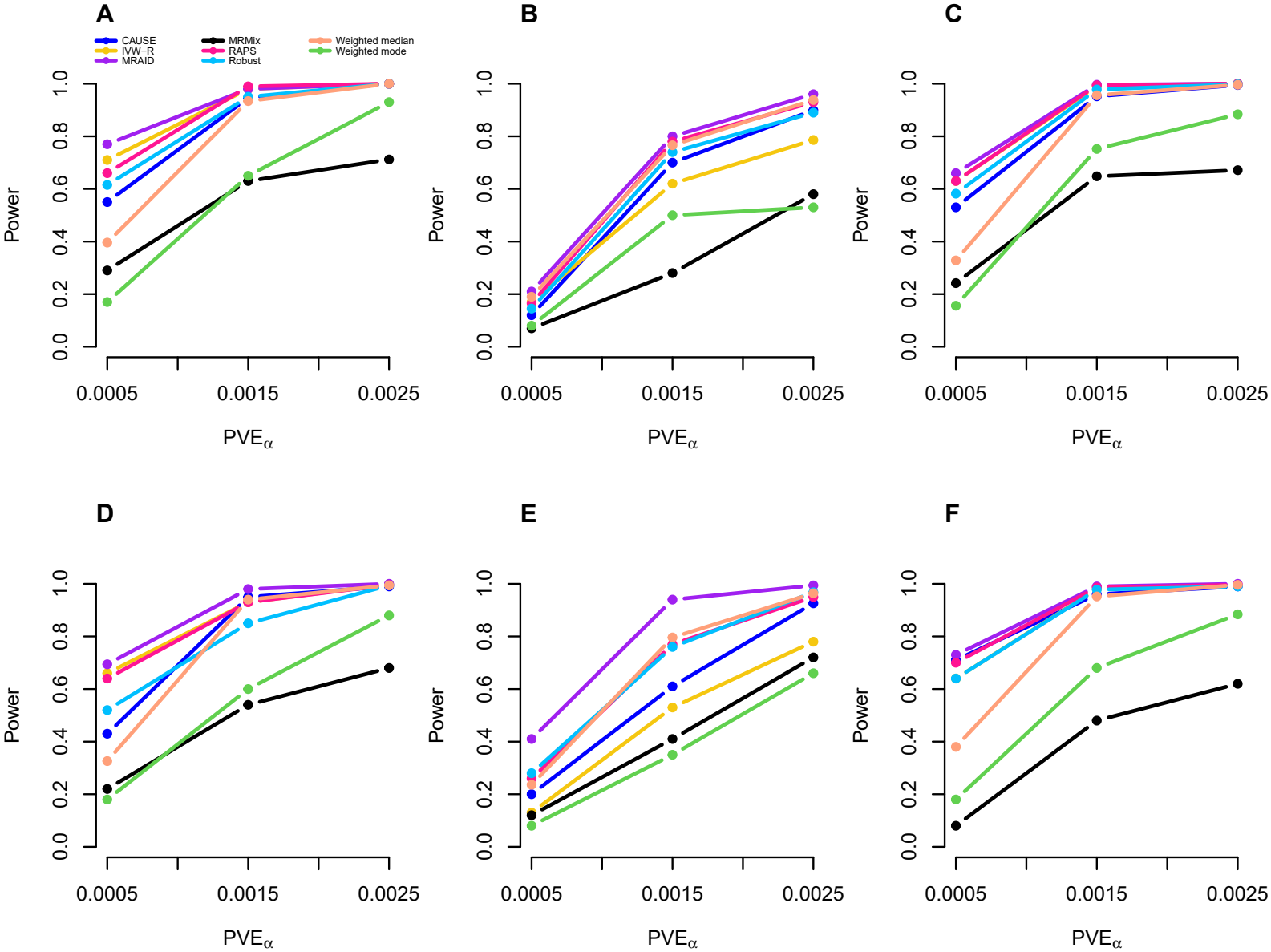


Fig. S10 Power of different MR methods in simulations. Power (y-axis) at a false discovery rate of 0.05 to detect the causal effect is plotted against different causal effect size characterized by PVE_α (x-axis). Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). We simulated 100 instrumental SNPs with different uncorrelated horizontal pleiotropy effects ($PVE_u=2.5\%$; **A, B**; $PVE_u=5\%$; **C, D, E, F**), different proportions of instrumental SNPs having the uncorrelated horizontal pleiotropy effects (no overlap, **A, C, D, F**; 20%, **E**; 30%, **B**), and different correlated horizontal pleiotropy effect sizes ($\pi_c=0, \rho=0$, **A, B, C**; $\pi_c=5\%, \rho=\sqrt{0.02}$, **D, E**; $\pi_c=5\%, \rho=\sqrt{0.05}$, **F**).

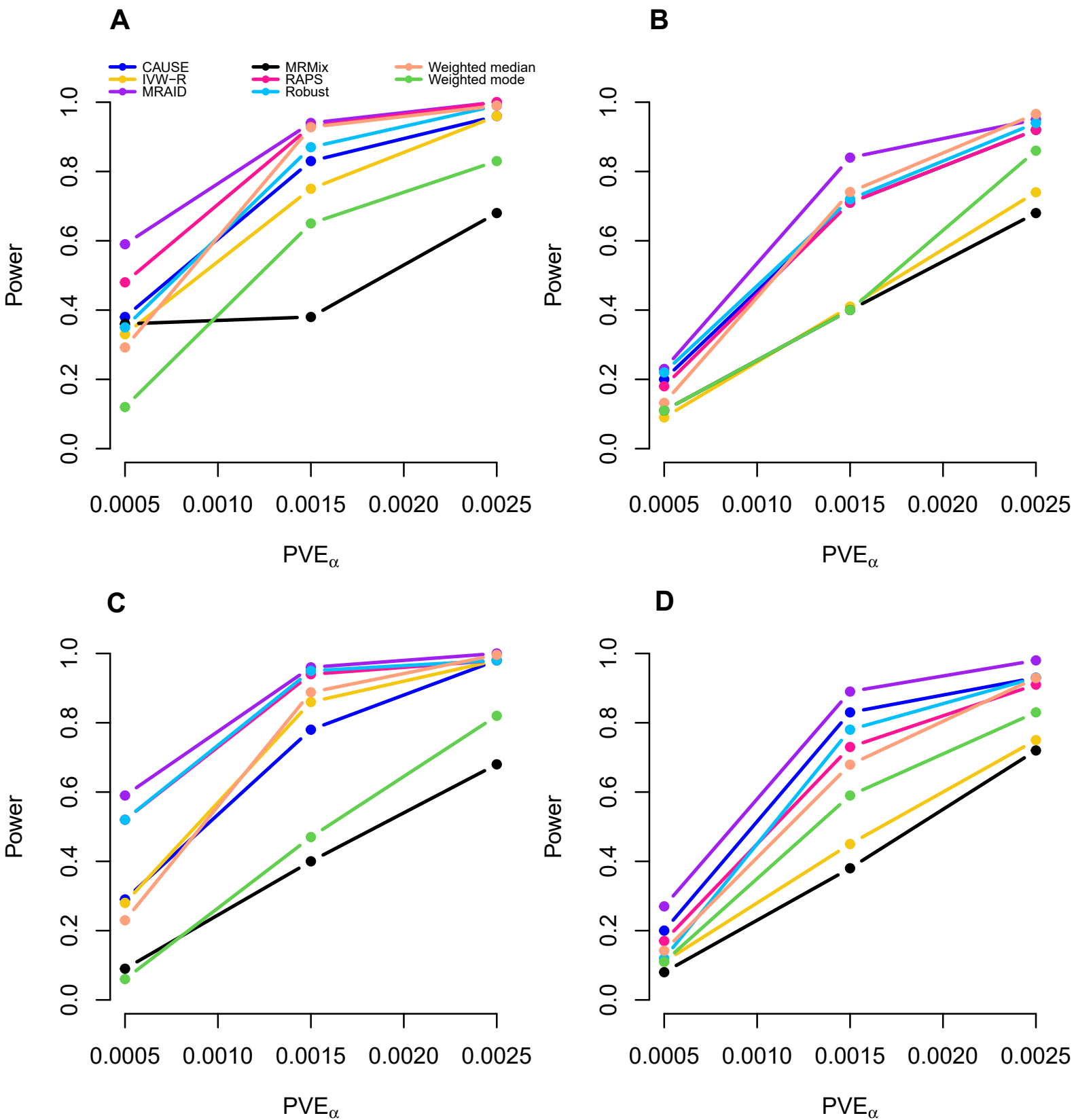


Fig. S11 Power of different methods in the presence of both correlated and uncorrelated horizontal pleiotropic effect ($PVE_u=5\%$). Power (y-axis) at a false discovery rate of 0.05 to detect the causal effect is plotted against different causal effect size characterized by PVE_{α} (x-axis). Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). We simulated 100 instrumental SNPs with different overlapped proportion among instrumental SNPs and the pleiotropic SNPs (10%, **A**, **C**; 30%, **B**, **D**). The correlated horizontal pleiotropy is, rather than simulated from an unobserved confounder, generated through assigning correlation either 0.1 (**A**, **B**) or 0.3 (**C**, **D**) to the effect size of the overlapped SNPs from bivariate normal distribution.

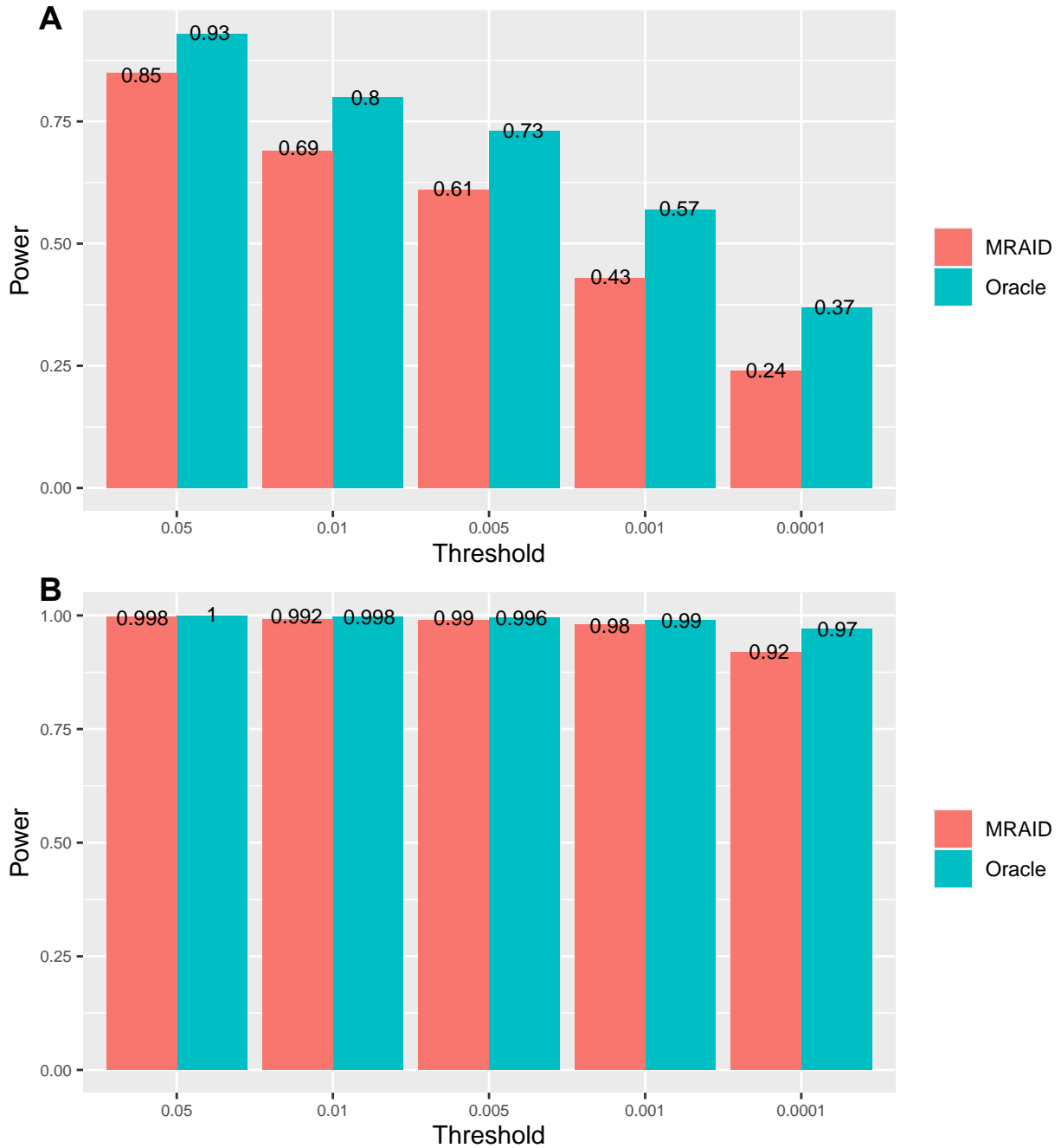


Fig. S12 Power of MRAID and Oracle method in the absence of both uncorrelated and correlated pleiotropic effects. Power (y-axis) to detect the causal effect is plotted against different p value threshold (x-axis). Compared methods include MRAID (red) and Oracle (blue). We simulated 100 instrumental SNPs with 500 replicates. The power of the oracle method is calculated from the IVW-R method using the independent SNPs screened from the 100 instruments, which is the gold standard method in the absence of any type of horizontal pleiotropy effect. (A) simulations with $PVE_{\alpha}=0.05\%$; (B) simulations with $PVE_{\alpha}=0.15\%$.

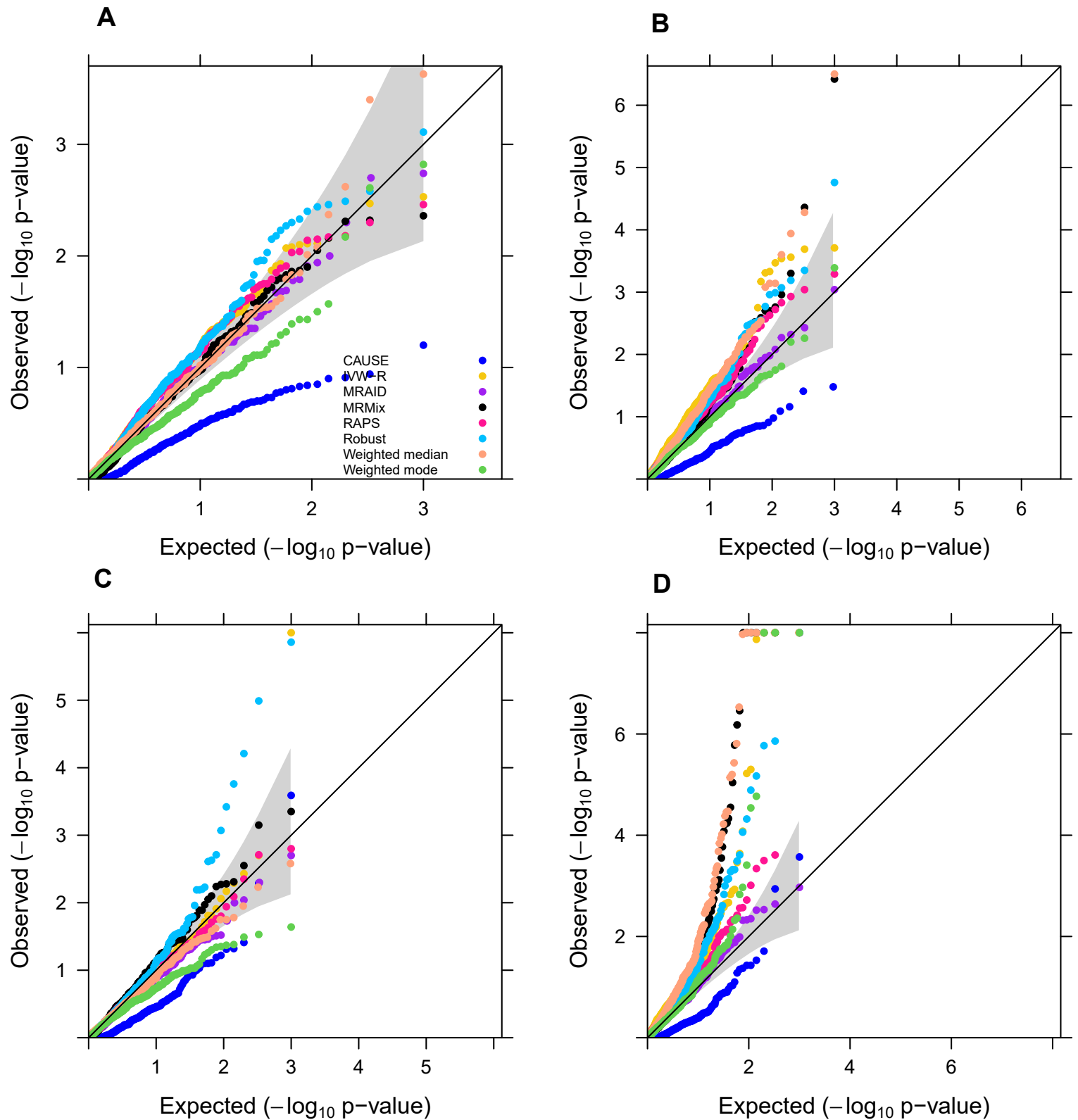


Fig. S13 Quantile-quantile plot of $-\log_{10}$ p-values from different MR methods in distinguishing the causal effect direction through reverse causality analysis in the absence of correlated horizontal pleiotropic effects but in the presence of uncorrelated horizontal pleiotropic effects. We tested the causal effect of the outcome on the exposure in the alternative simulations where the exposure has causal effect on the outcome ($PVE_{\alpha}=0.15\%$) but not vice versa. Compared methods include CAUSE (blue), IVW-R (gold), MRaid (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). We simulated 100 instrumental SNPs with different uncorrelated horizontal pleiotropy effects ($PVE_u=5\%$, **A**, **B**; $PVE_u=2.5\%$, **C**, **D**) and set the proportions of instrumental SNPs having the uncorrelated horizontal pleiotropy effect to be either no overlap (**A**, **C**) or 30% (**B**, **D**). MRaid is the only method showing calibrated p values across a range of simulation scenarios.

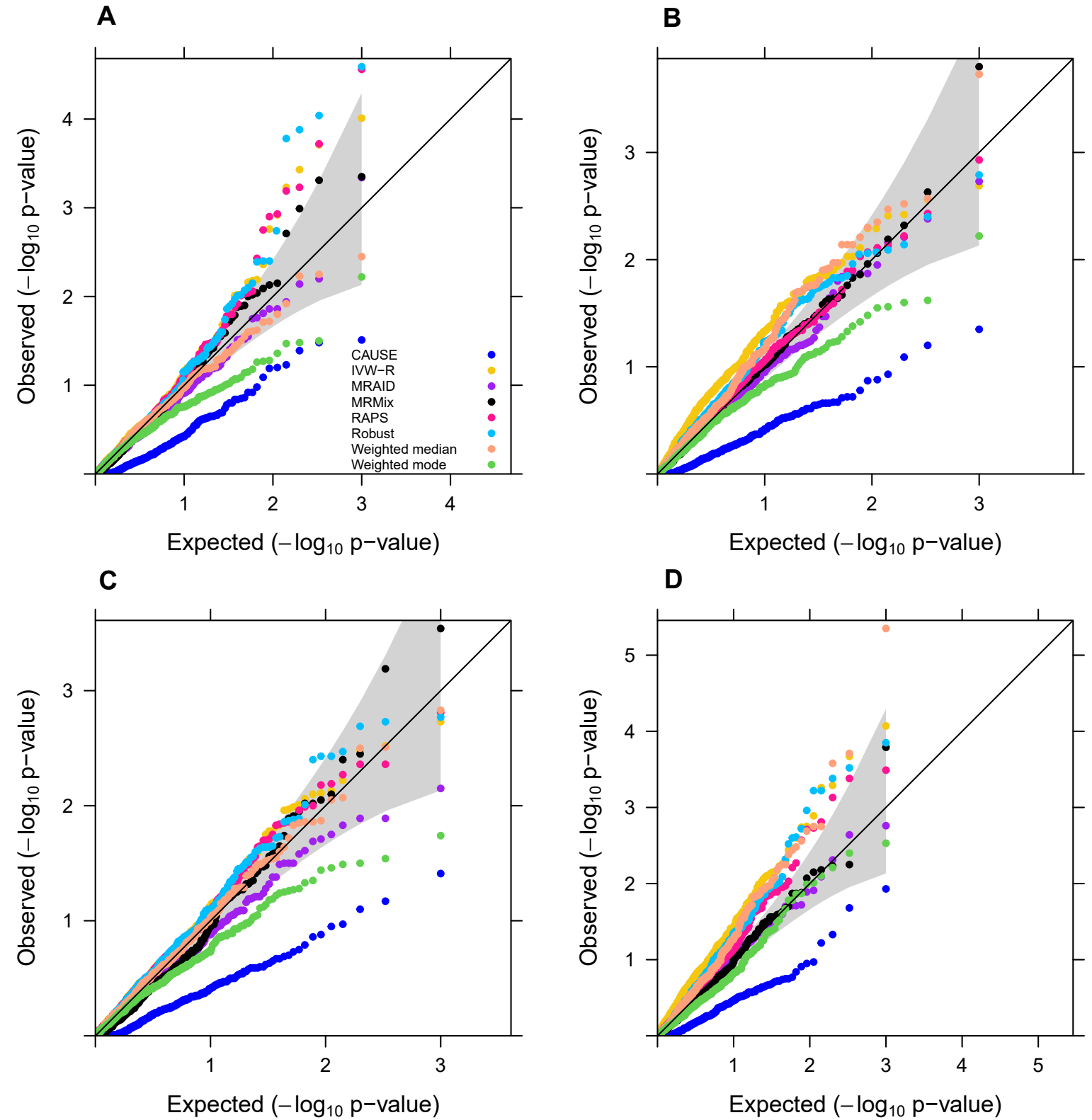


Fig. S14 Quantile-quantile plot of $-\log_{10}$ p-values from different MR methods in distinguishing the causal effect direction through reverse causality analysis in the presence of both correlated ($\pi_c=5\%$) and uncorrelated horizontal pleiotropic effect ($PVE_u=5\%$). We tested the causal effect of the outcome on the exposure in the alternative simulations where the exposure has causal effect on the outcome ($PVE_\alpha=0.15\%$) but not vice versa. Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). We simulated 100 instrumental SNPs with different proportions of instrumental SNPs having the uncorrelated horizontal pleiotropy effect being either no overlap (A, C) or 20% (B, D) and different correlated horizontal pleiotropy effect sizes ($\rho=\sqrt{0.05}$, A, B; $\rho=\sqrt{0.02}$, C, D). MRAID is the only method showing calibrated p values across a range of simulation scenarios.

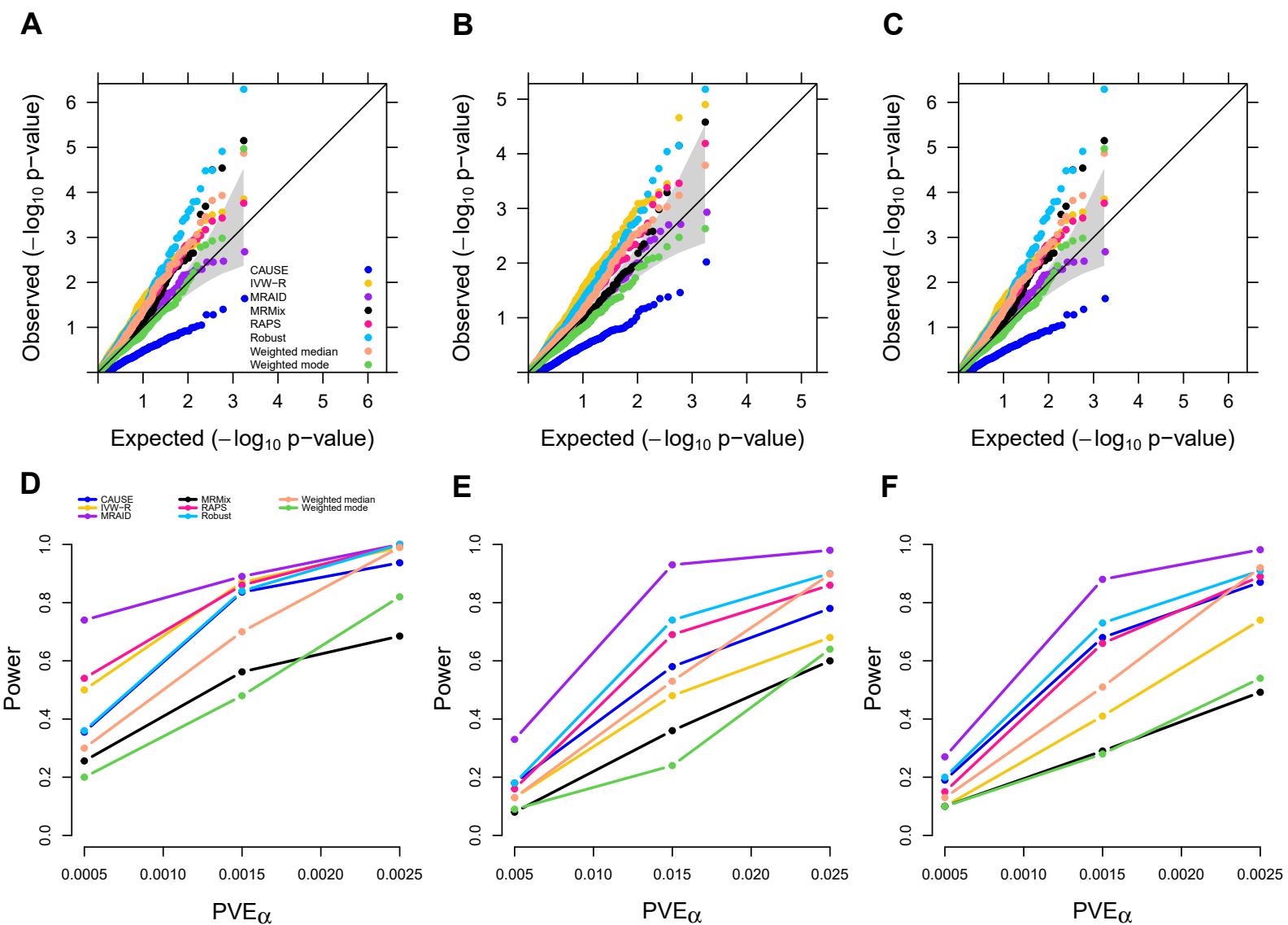


Fig. S15 Simulation results for different MR methods when multiple causal SNPs are selected to be in the same LD block. We randomly selected 50 among the 133 LD blocks on chromosome 1. In each of the 50 LD blocks, we randomly selected two SNPs to be causal, resulting in a total of 100 causal SNPs. Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). Six simulation scenarios are examined, including null (A) or alternative (D) simulations in the absence of both correlated and uncorrelated horizontal pleiotropic effects, null (B) or alternative (E) simulations in the absence of correlated horizontal pleiotropic effect but in the presence of uncorrelated horizontal pleiotropic effect ($PVE_u=5\%$), with the proportion of instrumental SNPs having uncorrelated horizontal pleiotropy to being 20%, and null (C) or alternative (F) simulations in the presence of both correlated ($\pi_c=5\%$, $\rho=\sqrt{0.05}$) and uncorrelated horizontal pleiotropic effects ($PVE_u=5\%$), with the proportion of instrumental SNPs having the uncorrelated horizontal pleiotropy effect being 20%.

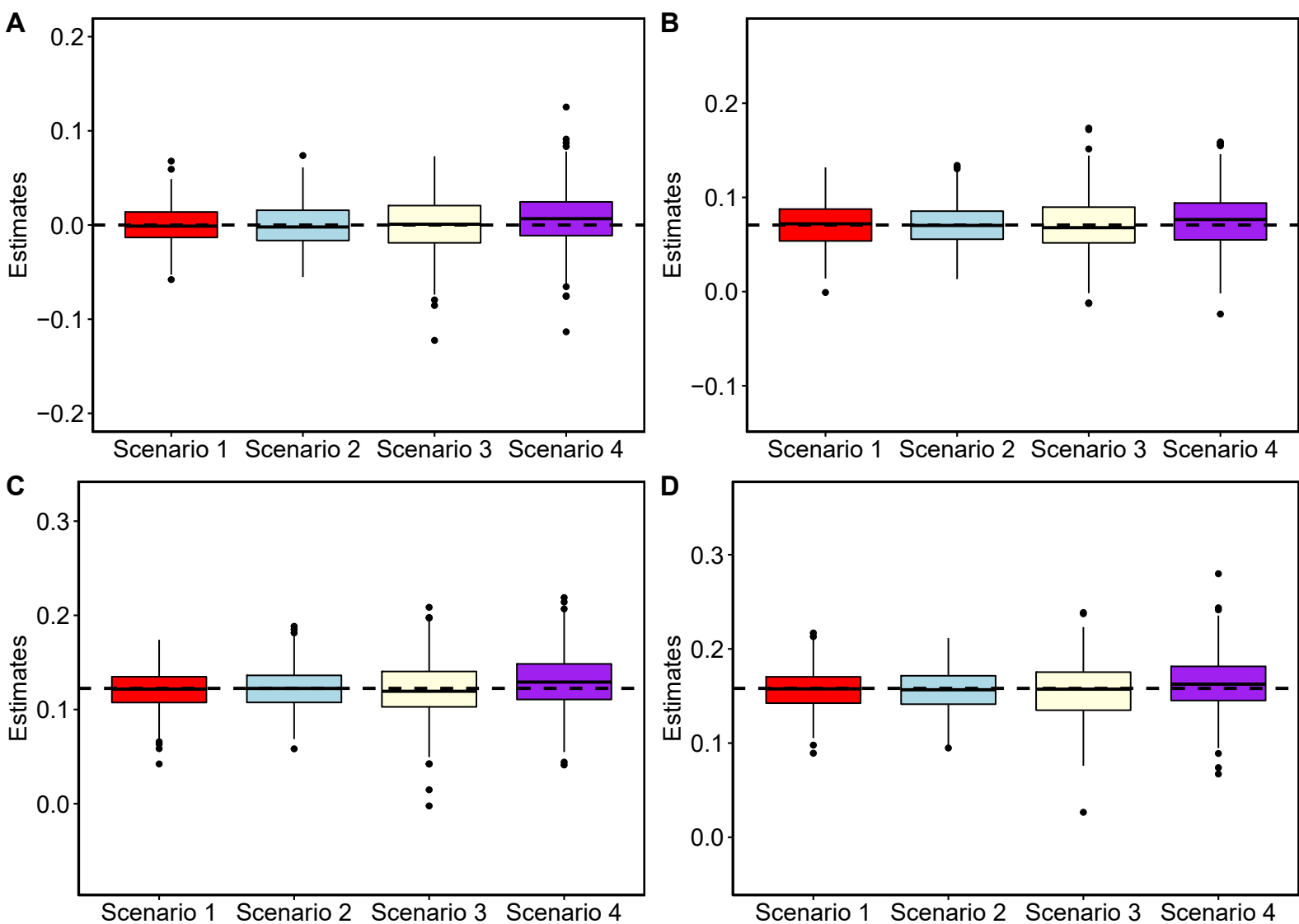


Fig. S16 Boxplot displays causal effect estimates in various simulations scenarios by MRAID with 100 instrumental SNPs. **Scenario 1** (red): simulations in the absence of both correlated and uncorrelated horizontal pleiotropic effects; **Scenario 2** (light blue): simulations in the absence of correlated horizontal pleiotropic effect but in the presence of uncorrelated horizontal pleiotropic effect ($PVE_u=5\%$); **Scenario 3** (light yellow): simulations in the absence of correlated horizontal pleiotropic effect but in the presence of uncorrelated horizontal pleiotropic effect ($PVE_u=5\%$) with the proportion of instrumental SNPs having the uncorrelated horizontal pleiotropy effect being 30%; **Scenario 4** (purple): simulations in the presence of both correlated ($\pi_c=5\%$, $\rho=\sqrt{0.05}$) and uncorrelated horizontal pleiotropic effects ($PVE_u=5\%$) with the proportion of instrumental SNPs having the horizontal pleiotropy effect being 20%. The horizontal dashed lines represent the four true values of $\alpha=0$ (A), $\alpha=\sqrt{0.005}$ (B), $\alpha=\sqrt{0.015}$ (C) and $\alpha=\sqrt{0.025}$ (D). MRAID produces approximately unbiased causal effect size estimates across different scenarios. 500 replicates are performed for each simulation scenario.

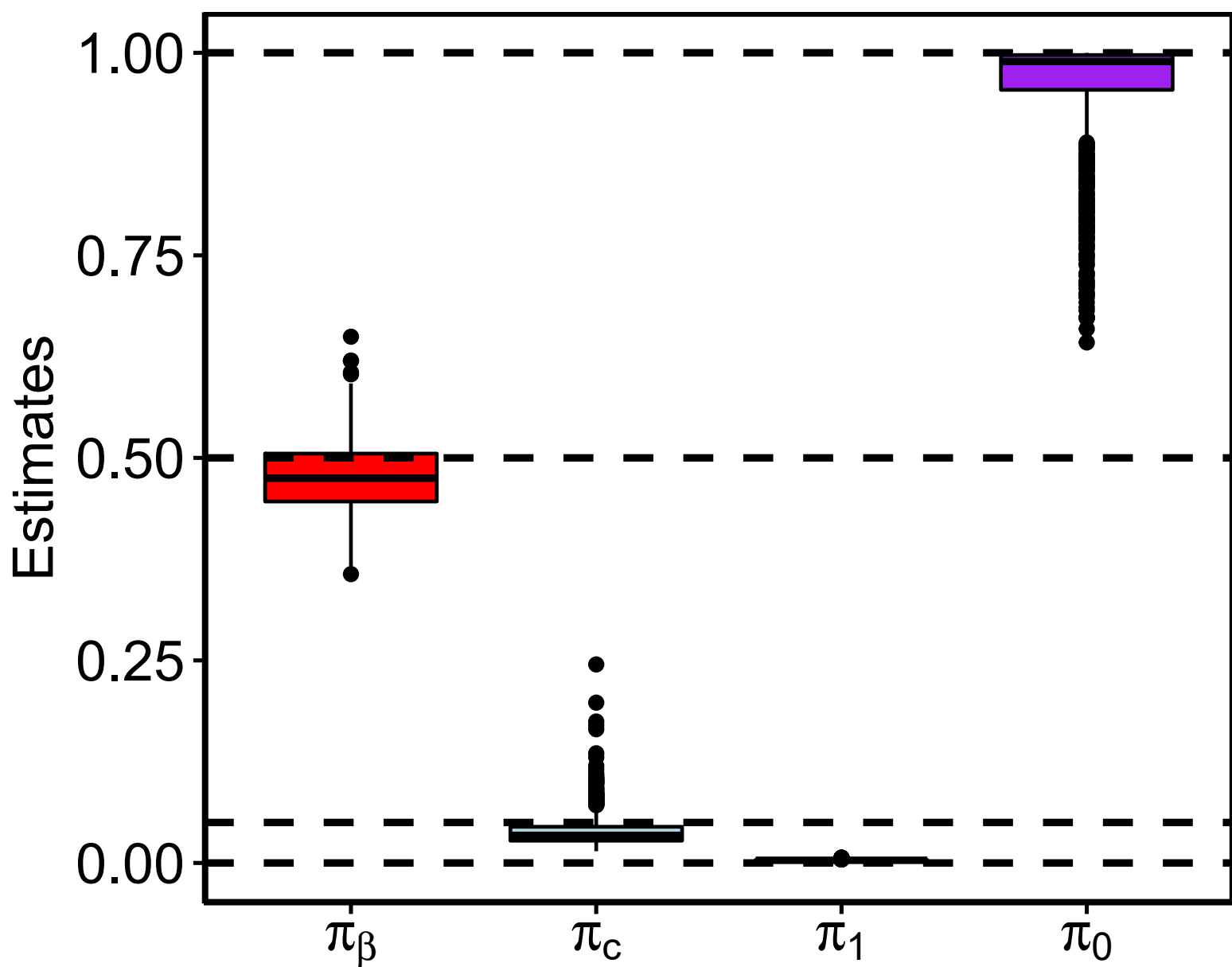


Fig. S17 Boxplot displays the estimates of the proportion parameters in the simulations. The horizontal dashed lines represent the true values of π_β (red), π_c (green), π_1 (blue) and π_0 (purple). We simulated the null simulations in the presence of both correlated ($\pi_c=5\%$, $\rho=\sqrt{0.05}$) and uncorrelated horizontal pleiotropic effects ($PVE_u=5\%$). We set the number of instrumental SNPs and the number of uncorrelated pleiotropic SNPs to be 100, with no overlap between the two sets. Thus, a total of 200 SNPs are used and the corresponding true values of π_β , π_c , π_1 and π_0 are 0.5, 0.05, 0, 1 respectively.

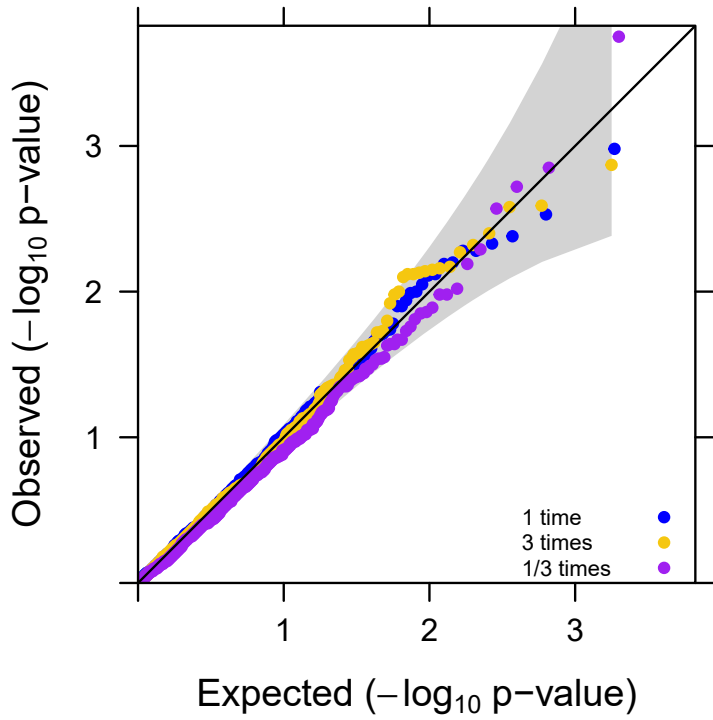
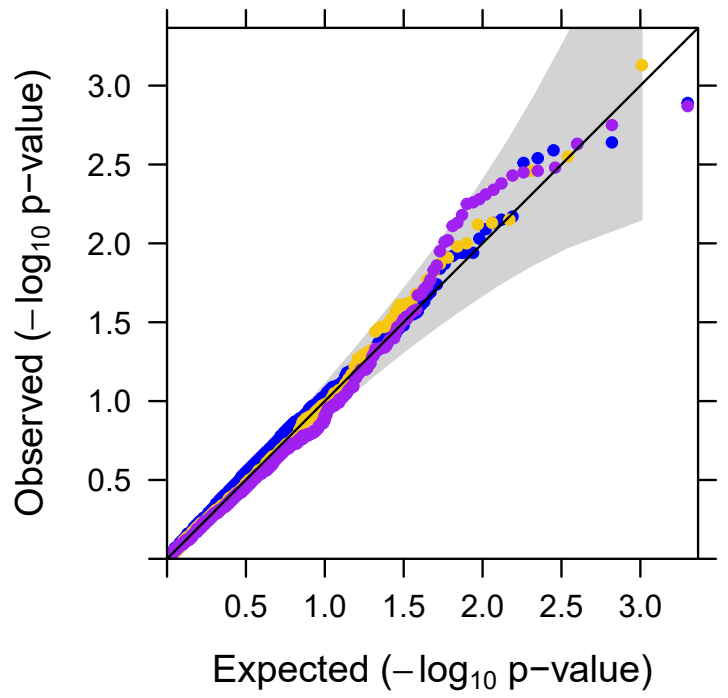
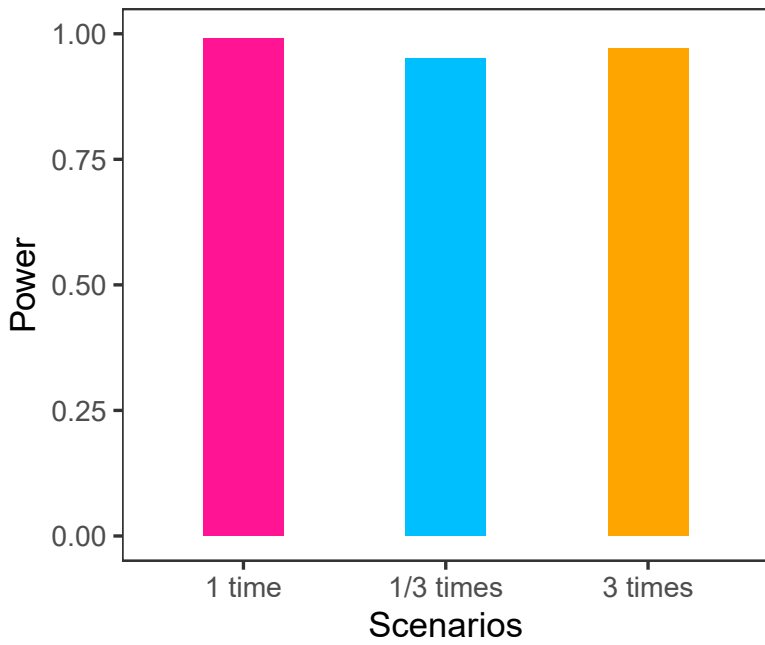
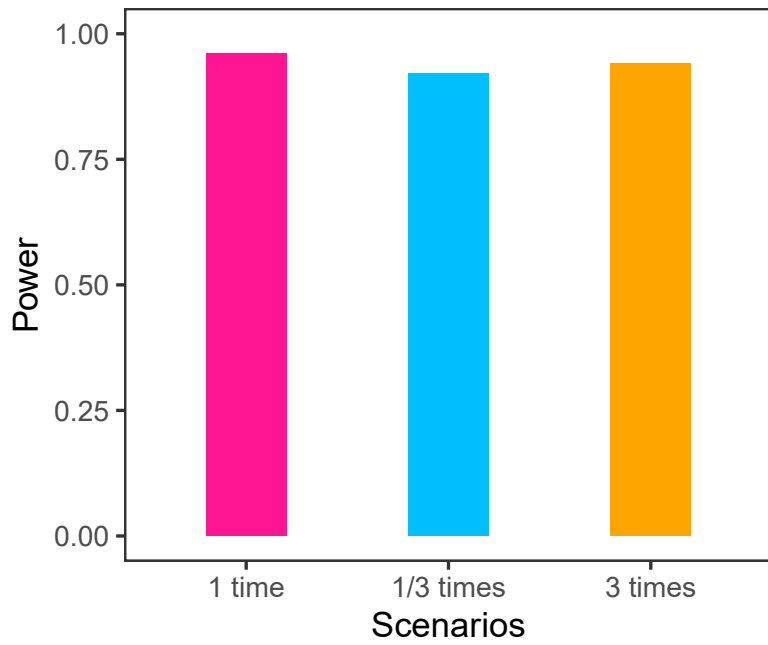
A**B****C****D**

Fig. S18 Simulation results for MRAID when the variance parameter for simulating the uncorrelated horizontal pleiotropic effects is different between instrumental and non-instrumental SNPs. Simulations are performed in the absence of correlated horizontal pleiotropic effect but in the presence of uncorrelated horizontal pleiotropic effect ($PVE_{it}=5\%$). We set the variance parameter for modeling the uncorrelated horizontal pleiotropic effects from the non-instrumental SNPs to be either 3 times or 1/3 times that of the instrumental SNPs. We simulated 100 instrumental SNPs and set the proportion of instrumental SNPs having uncorrelated horizontal pleiotropy to be 20% (A, C), 30% (B, D), respectively. The null ($PVE_{\alpha}=0$) and power ($PVE_{\alpha}=0.15\%$) simulations are displayed in (A, B) and (C, D), respectively. The results of MRAID under simulations with the same variance parameter are also included for comparison.

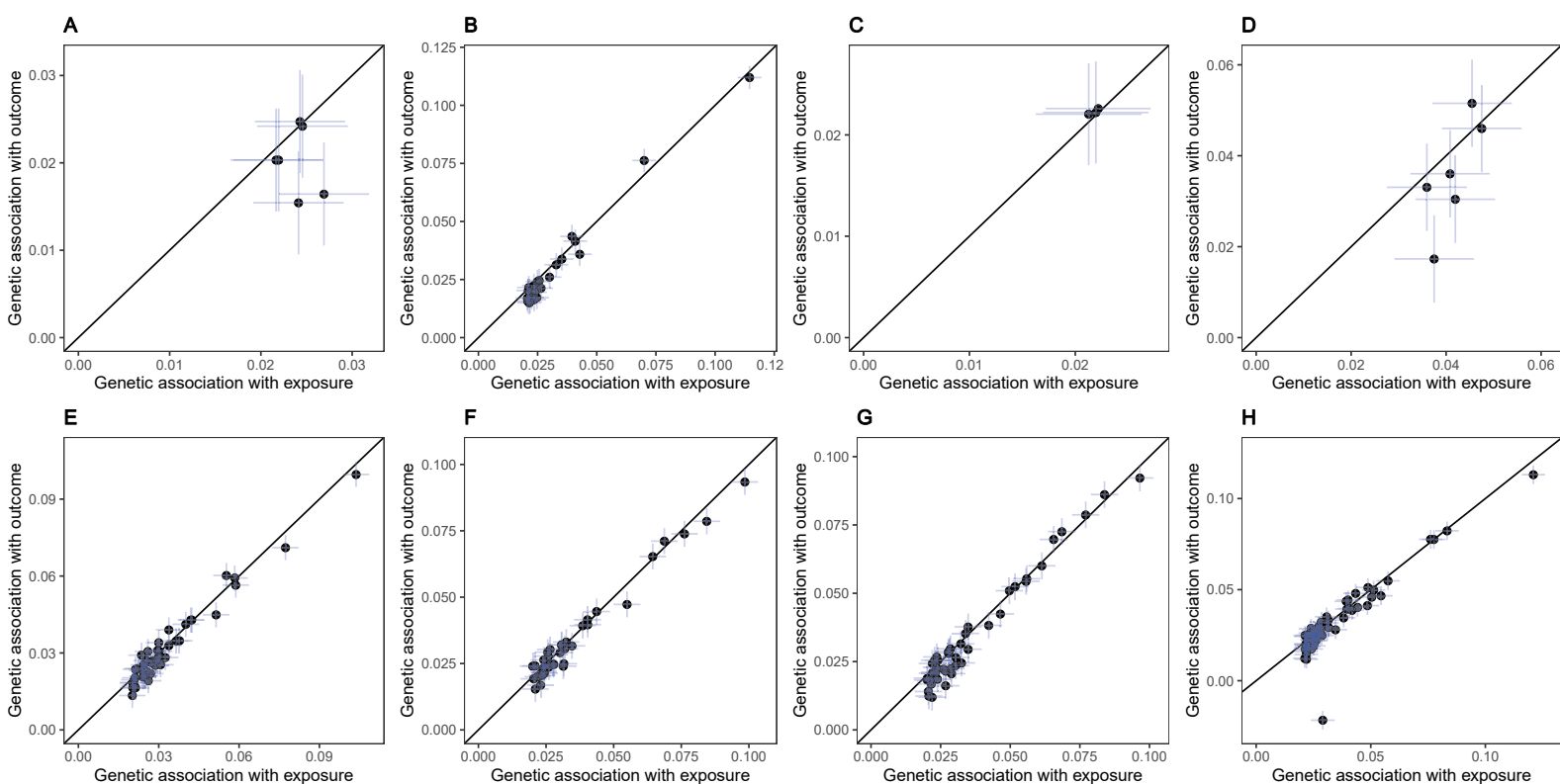


Fig. S19 Scatter plots of genetic associations in the trait on itself analysis in UKBB. Trait pairs include SBP-SBP (A), BMI-BMI (B), DBP-DBP (C), Pulse rate-Pulse rate (D), TC-TC (E), LDL-LDL (F), TG-TG (G), and HDL-HDL (H). The SNP effect on the outcome is plotted against the SNP effect on the exposure. We show these plots for all MR methods that use a pre-selected set of independent instrumental SNPs. For MR-IVW, we are unable to show such plot because there is no formal set of instruments, as MR-IVW uses a probabilistic model to select instruments from a large set of candidate instruments, none of which has 100% chance of being selected.

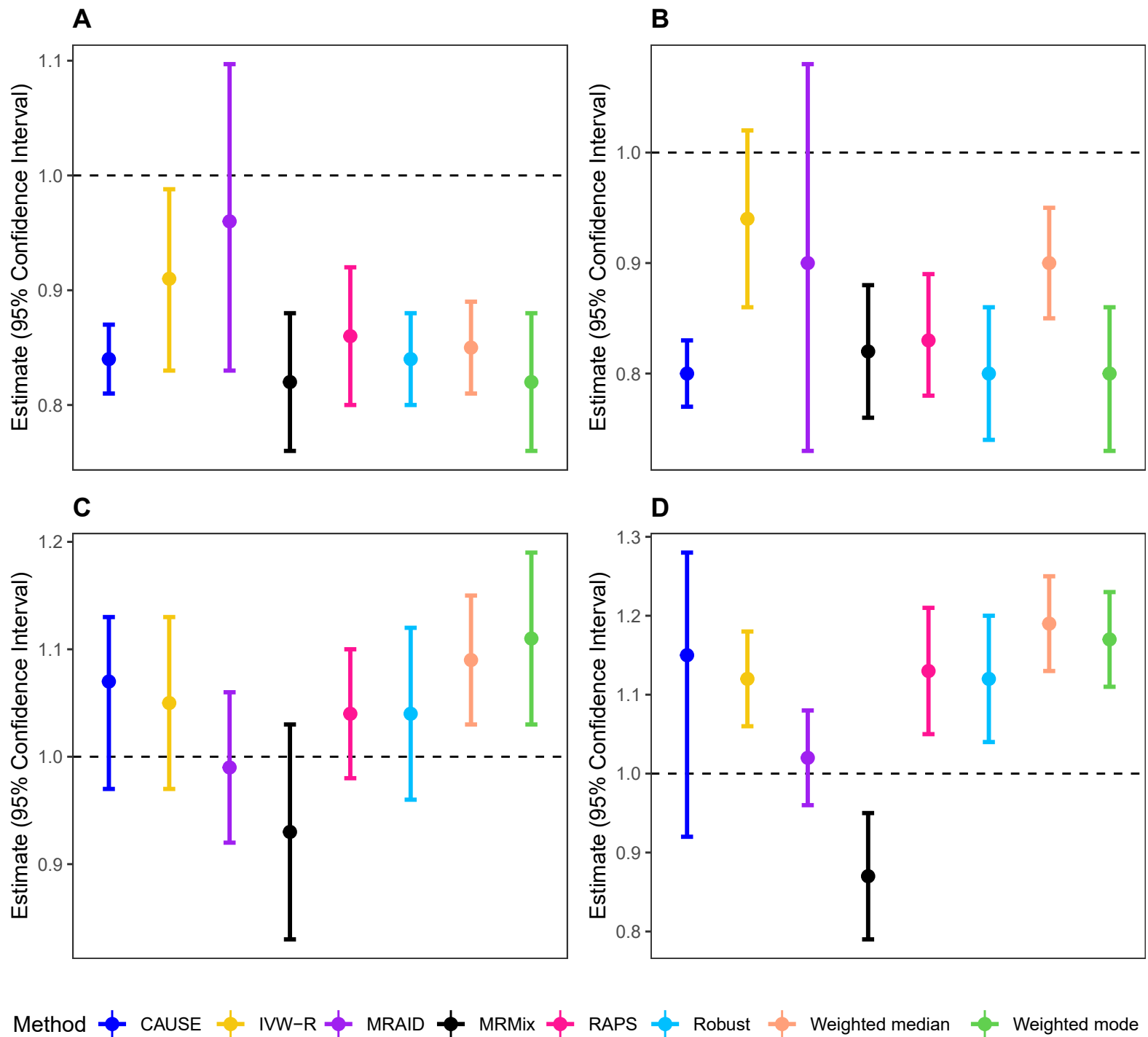


Fig. S20 Point estimates and 95% confidence intervals in the trait on itself analysis in GLGC and UKBB. Each of the four blood lipid traits in GLGC is used as the exposure and is examined on its causal effect on the same trait in UKBB. Because only summary statistics are available GLGC, we constructed the LD matrix in the exposure GWAS using individuals of European ancestry from the 1,000 Genomes project. We constructed the LD matrix in the outcome GWAS using the complete individual level data in the UKBB. Compared methods include CAUSE (blue), IVW-R (gold), MRAID (purple), MRMix (black), RAPS (deep pink), Robust (deep sky blue), Weighted median (light salmon), Weighted mode (green). Analyzed trait pairs include TC-TC (A), HDL-HDL (B), LDL-LDL (C), and TG-TG (D). The horizontal black dashed line in each panel represents the true causal effect size of $\alpha=1$. Only MRAID is able to produce 95% confidence intervals that cover the true causal effects for all four trait pairs.

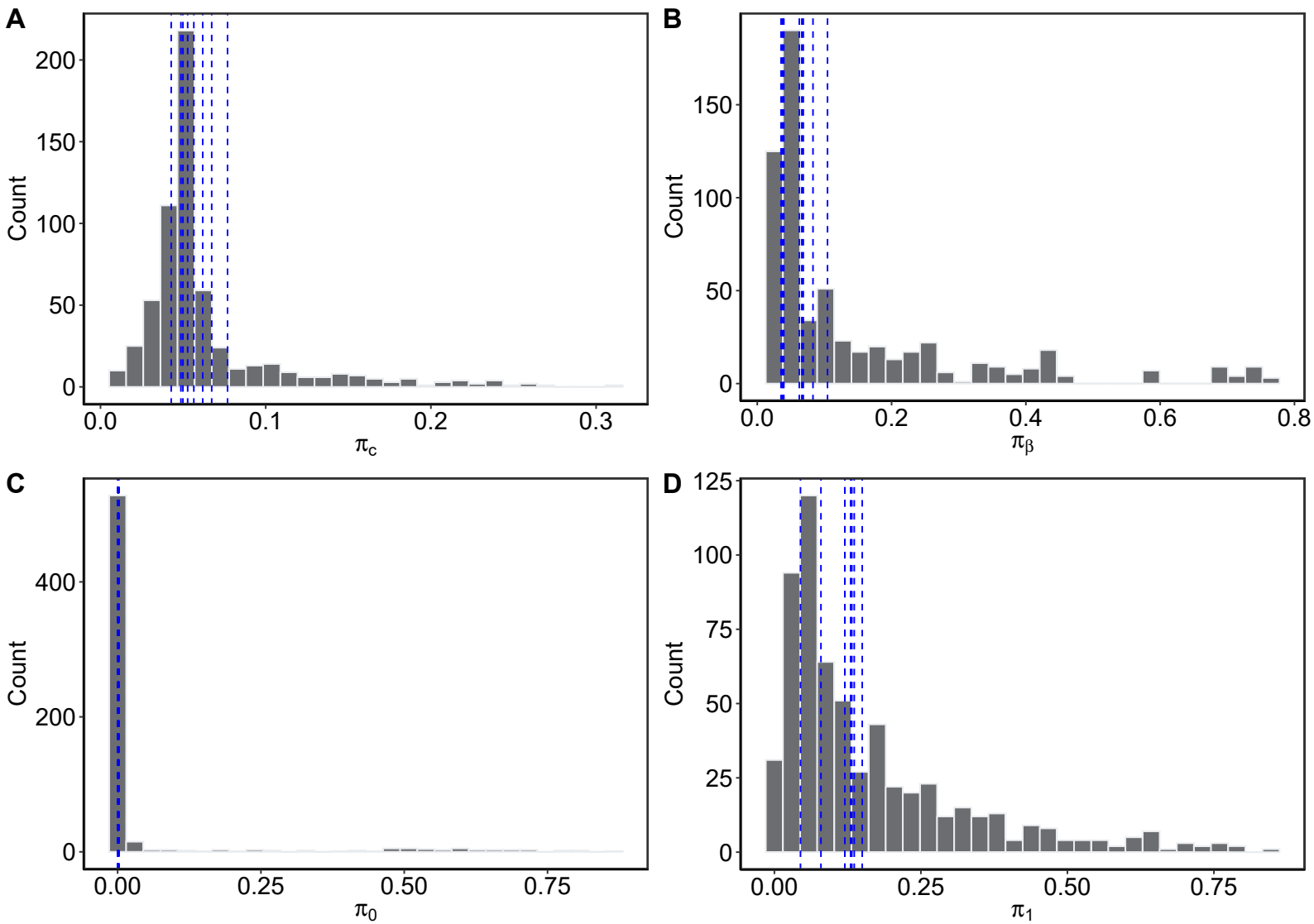


Fig. S21 Histograms of the four proportion parameters in MARID across the 645 analyzed trait pairs from UK Biobank, with blue dashed line in each panel representing the value from the 8 significant trait pairs detected by MRAID. Histograms are for π_c (A), π_β (B), π_0 (C) and π_1 (D) respectively.

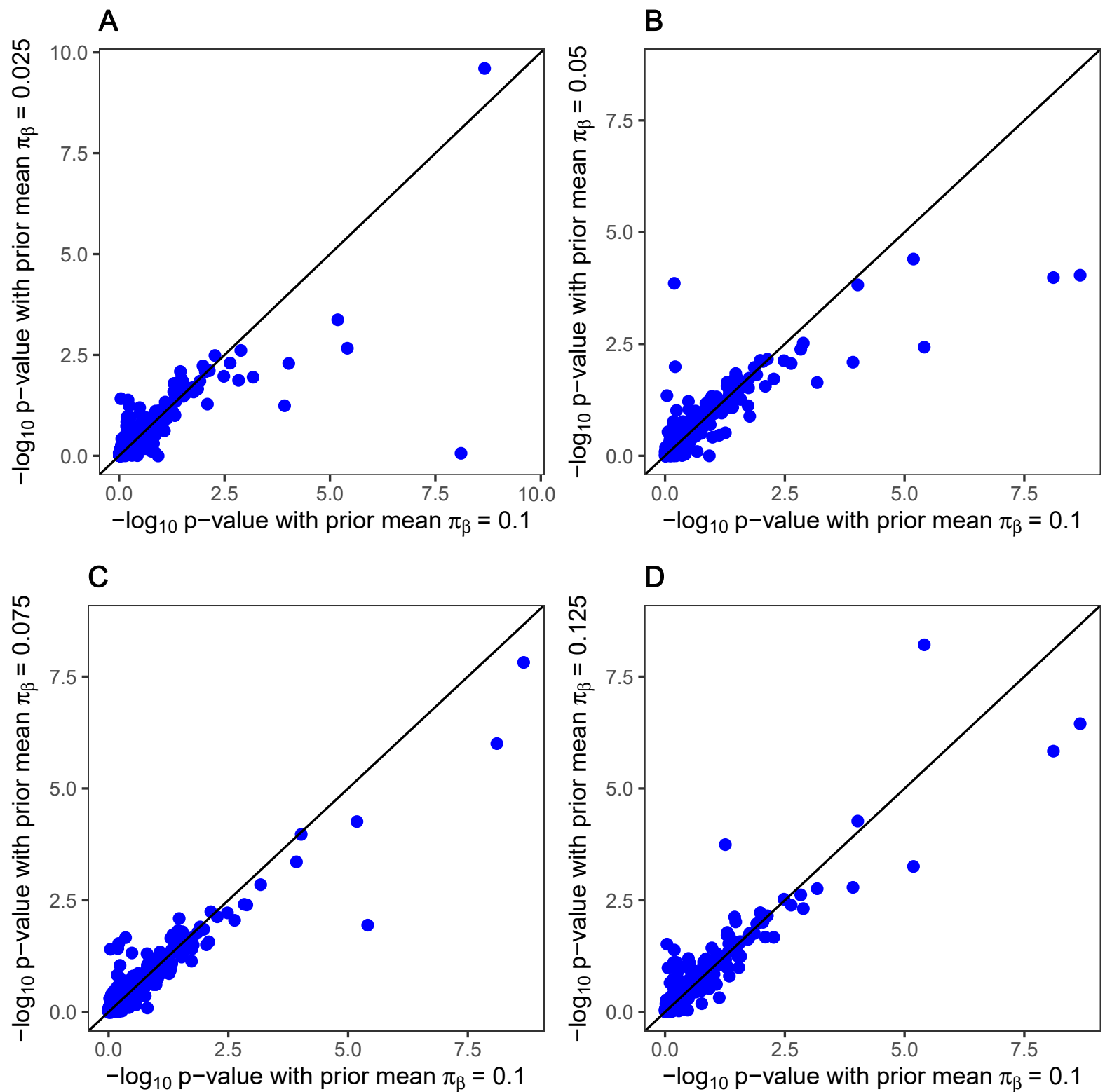


Fig. S22 Comparison of p values from MRAID under different prior distribution of π_β . We used different parameters in the Beta distribution to set the prior mean of π_β to be either 0.025, 0.05, 0.075, 0.1, and 0.125. The $-\log_{10}$ p values of MRAID from these different priors are plotted against the $-\log_{10}$ p values from MRAID with the default prior, where the prior mean of π_β is 0.1. The p values results from MRAID are robust with respect to the prior distribution of π_β .

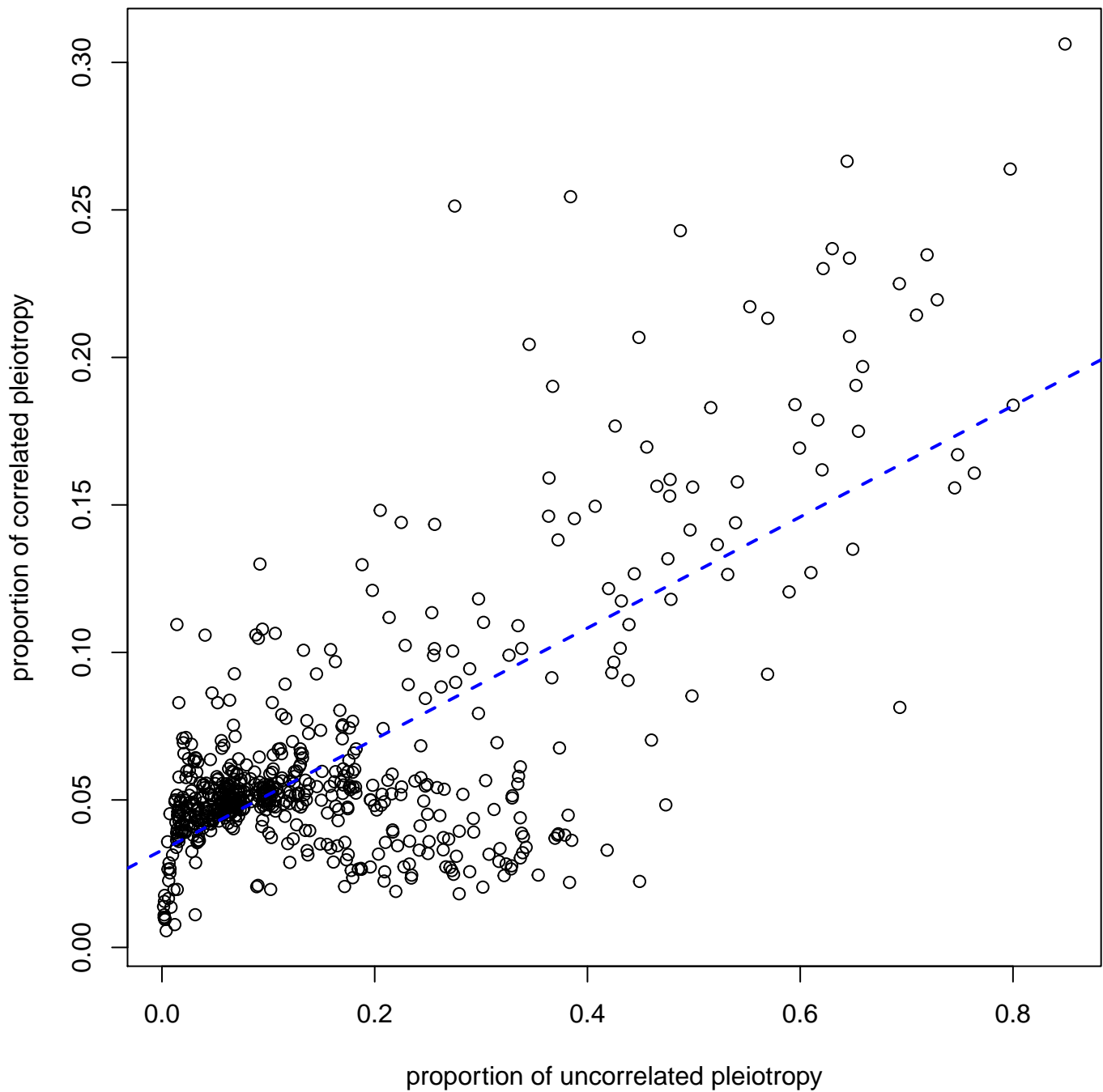


Fig. S23 Scatterplot displays the relationship between the proportion of SNPs exhibiting uncorrelated horizontal pleiotropy among the selected instruments (π_1) and the proportions of SNPs displaying correlated horizontal pleiotropy among the selected instruments (π_c) across the 645 analyzed trait pairs from UK Biobank. The proportion of SNPs displaying correlated pleiotropy is correlated with the proportions of SNPs displaying uncorrelated pleiotropy, with the latter generally being larger than the former.

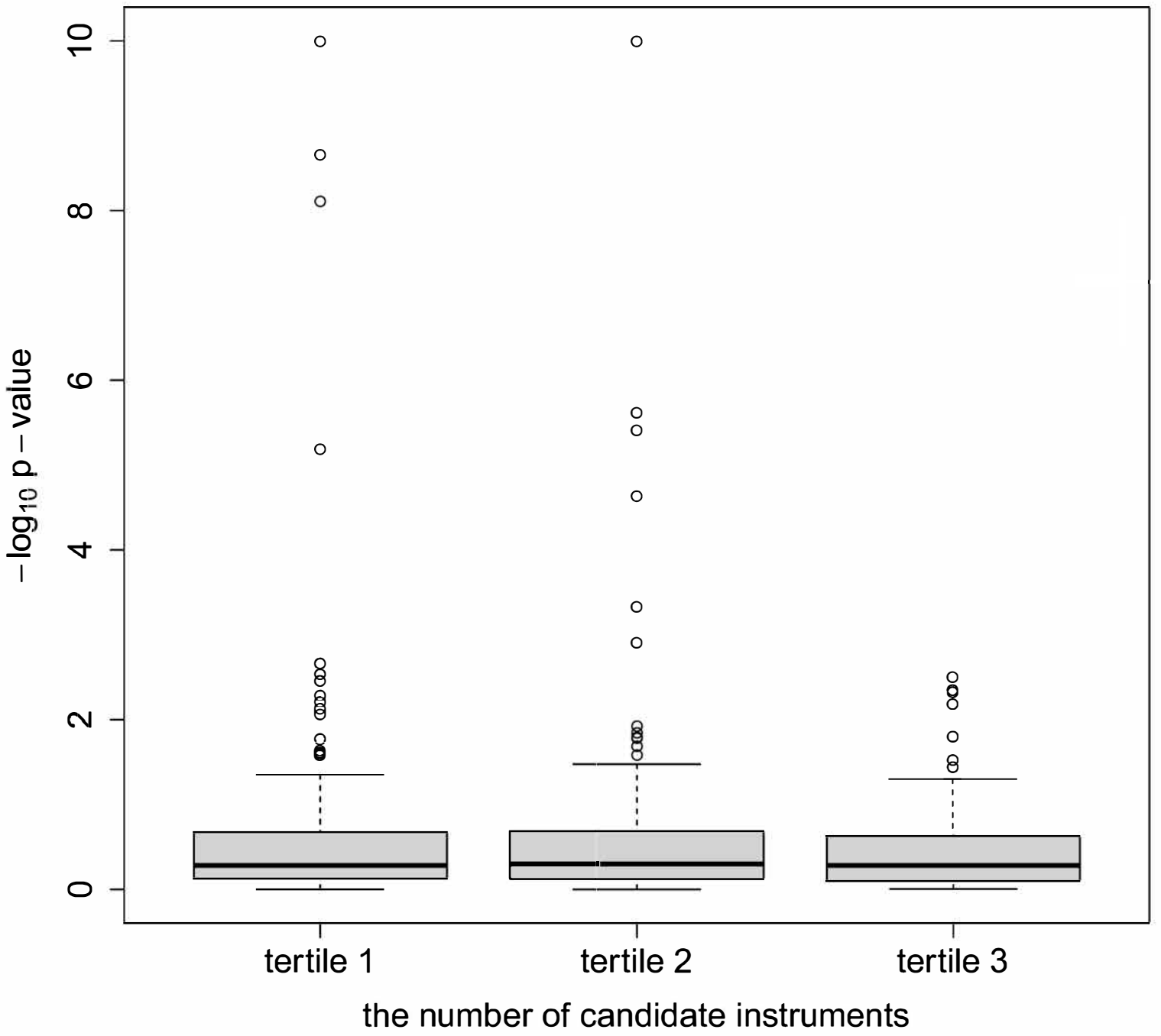


Fig. S24 Boxplot displays the $-\log_{10}$ p-values of MRAID stratified by the tertiles of the number of candidate instruments across the 645 analyzed trait pairs from UK Biobank.

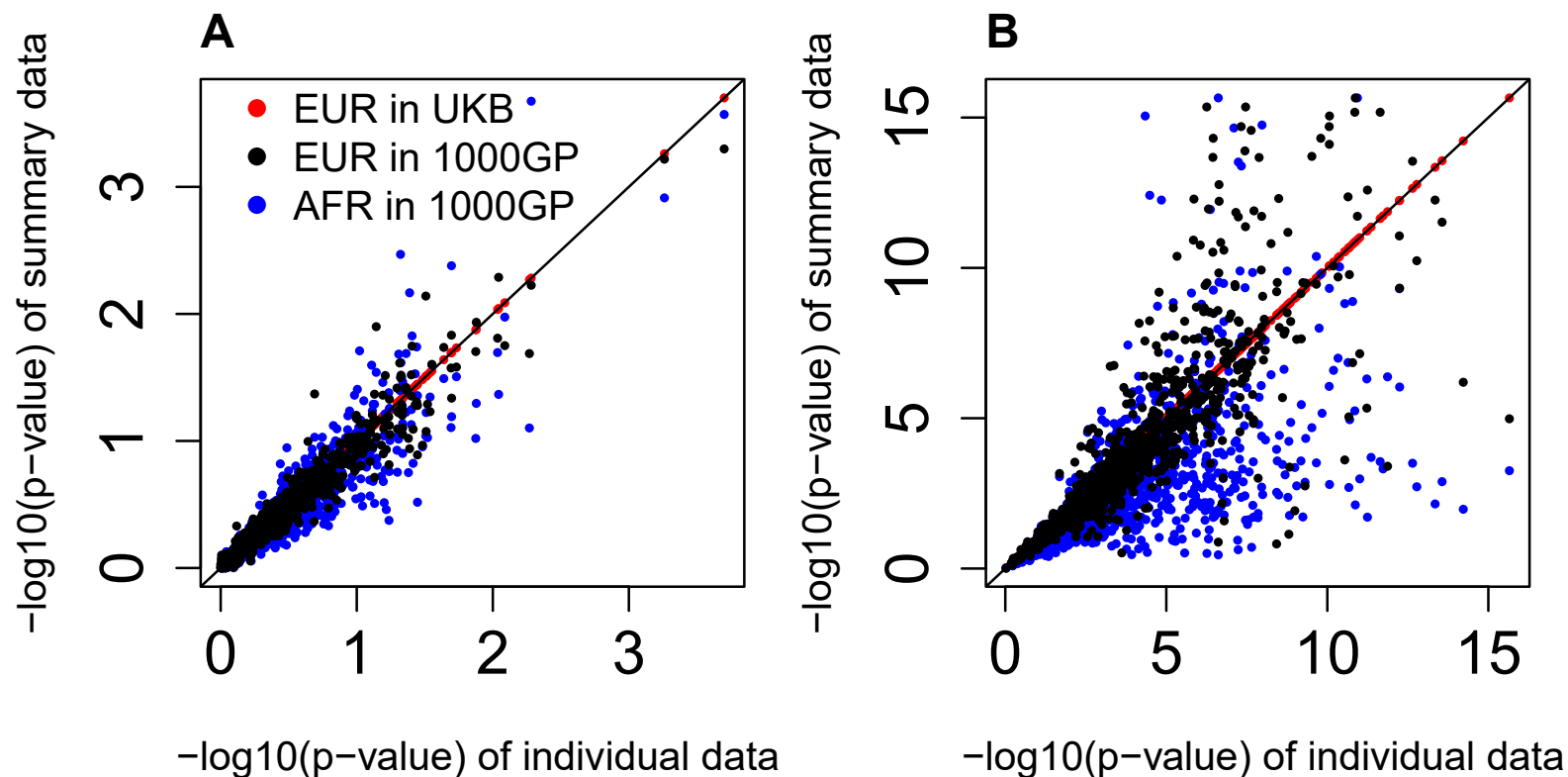


Fig. S25 Comparison of the summary-statistics version of MRAID with the individual-data version of MRAID. The LD matrix in the exposure data is directly computed using the individual level from the exposure GWAS study. The LD matrix in the outcome GWAS data is computed through three different ways: all individuals from the outcome GWAS data ($n=337,129$; red); individuals of European ancestry from the 1,000 Genomes project phase 3 ($n=503$; black); or individuals with African ancestry from the 1,000 Genomes project phase 3 ($n=611$; blue). We plotted $-\log_{10}$ p-values from the summary-statistics version of MRAID (y-axis) versus $-\log_{10}$ p-values from the individual-level data version of MRAID (x-axis). Simulations are performed with 100 instrumental SNPs in the absence of both correlated and uncorrelated horizontal pleiotropic effects under different causal effect $PVE_{\alpha}=0$ (A) or $PVE_{\alpha}=0.05\%$ (B). As expected, except for the African reference panel, the results from the summary-statistics version of MRAID are largely consistent with that from the individual-level data version of MRAID.

Supplementary Tables

Table S1. Simulation scenarios in this study

Type	Scenario	Figure
simple simulations	randomly select one SNP to be causal from ten SNPs in one region	Fig. S1A, Fig. S2
	randomly select two SNPs to be causal from ten SNPs in one region	Fig. S1C, Fig. S2
	randomly select one SNP to be causal from ten SNPs in one region and analyze without the causal SNP	Fig. S1B, Fig. S2
	randomly select two SNPs to be causal from ten SNPs in one region and analyze without the causal SNPs	Fig. S1D, Fig. S2
	randomly select two SNPs to be causal from two regions with each region containing ten SNPs	Fig. S1E, Fig. S2
	randomly select two SNPs to be causal from two regions with each region containing ten SNPs and analyze without the causal SNPs using the top two SNPs obtained from a stepwise regression	Fig. S1F, Fig. S2 Fig. S3
main simulations	two different instrumental SNPs settings in the absence of both correlated and uncorrelated horizontal pleiotropy	Fig. 2A,B, Fig. 3A,B, Fig.S9
	absence of correlated horizontal pleiotropy but in the presence of uncorrelated horizontal pleiotropy	Fig. 2C, Fig. 3C, Fig. S6 Fig. 2D, Fig. 3D,
	presence of both correlated and uncorrelated horizontal pleiotropy	Fig.S7, Fig.S8, Fig.S11
	different variance of the exposure explained by instrumental SNPs	Fig. S4
	different genetics architectures using all correlated SNPs	Fig. S5
	different uncorrelated or correlated horizontal pleiotropy effects, different proportions of instrumental SNPs having uncorrelated horizontal pleiotropy	Fig. S10
	Oracle method in the absence of both correlated and uncorrelated horizontal pleiotropy	Fig. S12
	reverse causality analysis in the absence of correlated horizontal pleiotropy but in the presence of uncorrelated horizontal pleiotropy	Fig. S13
	reverse causality analysis in the presence of both correlated and uncorrelated horizontal pleiotropy	Fig. S14
	multiple casual SNPs selected to be in the same LD block	Fig. S15
	estimates of the causal effects in various simulations	Fig. S16
	estimates of the proportion parameters in the presence of both correlated and uncorrelated horizontal pleiotropy	Fig. S17
different variance parameter for the uncorrelated horizontal pleiotropy between instrumental and non-instrumental SNP	Fig. S18	

Table S2. All causal trait associations detected by different MR methods.

Traits	MRAID	CAUSE	IVW-R	MRMix	RAPS	Robust	Weighted median	Weighted mode
Age started smoking in former smokers								
PR	0.67 (0.65)	-0.30 (1.3×10⁻¹¹)	-0.40 (2.2×10 ⁻⁴)	-0.40 (3.0×10 ⁻⁴)	-0.40 (5.28×10 ⁻⁴)	-0.40 (3.1×10⁻¹⁵)	-0.38 (3.6×10⁻⁵)	-0.39 (1.3×10 ⁻³)
PWRIx	0.46 (0.62)	0.28 (1.5×10 ⁻³)	0.35 (1.1×10 ⁻³)	0.36 (1.2×10 ⁻³)	0.35 (2.0×10 ⁻³)	0.35 (2.7×10⁻¹²)	0.23 (1.7×10 ⁻²)	0.37 (2.4×10 ⁻³)
PWP2PT	0.24 (0.41)	0.03 (7.2×10⁻⁵⁷)	0.14 (0.19)	0.14 (0.19)	0.14 (0.21)	0.14 (5.5×10 ⁻³)	0.15 (0.10)	0.14 (0.25)
BMI	2.80 (0)	4.57 (0)	2.64 (0)	0.00 (1)	2.65 (0)	2.64 (8.9×10⁻²⁴)	2.30 (2.1×10⁻⁴⁶)	2.68 (2.9×10⁻⁴⁸)
SBP	-0.42 (0.06)	-0.33 (1.3×10 ⁻³)	-0.34 (7.8×10⁻⁸)	-0.35 (4.1×10⁻⁷)	-0.34 (8.9×10⁻⁷)	-0.34 (1.5×10⁻³⁰)	-0.28 (1.0×10⁻⁶)	-0.35 (1.7×10⁻⁶)
DBP	-0.14 (0.14)	-0.04 (0.02)	-0.14 (0.03)	-0.14 (0.03)	-0.14 (0.04)	-0.14 (3.3×10⁻⁶)	-0.14 (0.01)	-0.15 (0.03)
TC	-0.19 (0.75)	0.11 (4.0×10 ⁻³)	0.18 (3.7×10 ⁻³)	0.19 (4.0×10 ⁻³)	0.18 (5.9×10 ⁻³)	0.18 (5.6×10⁻¹⁹)	0.18 (8.0×10 ⁻⁴)	0.19 (5.6×10 ⁻³)
HDL	0.70 (2.2×10⁻⁹)	0.13 (0.11)	0.70 (8.0×10⁻²⁶)	0.70 (7.1×10⁻¹⁸)	0.70 (0)	0.70 (1.1×10⁻¹¹)	0.70 (2.9×10⁻²⁶)	0.70 (5.4×10⁻¹⁶)
LDL	0.06 (0.81)	0.05 (3.6×10 ⁻³)	0.13 (0.04)	0.14 (0.04)	0.13 (0.05)	0.13 (1.0×10⁻⁵)	0.13 (1.3×10 ⁻²)	0.14 (0.05)
TG	-0.64 (7.8×10⁻⁹)	-0.92 (4.9×10⁻²⁶⁷)	-0.65 (6.0×10⁻²⁵)	-0.66 (3.9×10⁻¹⁷)	-0.65 (0)	-0.65 (1.0×10⁻¹⁰⁷)	-0.57 (4.1×10⁻¹⁹)	-0.65 (1.4×10⁻¹⁵)
Number of unsuccessful stop-smoking attempts								
PR	0.20 (0.31)	0.20 (0.03)	0.23 (3.8×10 ⁻³)	0.23 (4.2×10 ⁻³)	0.23 (5.5×10 ⁻³)	0.23 (6.0×10⁻¹⁴)	0.24 (7.8×10 ⁻³)	0.23 (0.01)
PWRIx	-0.56 (0.59)	-0.14 (0.06)	-0.20 (0.01)	-0.21 (0.01)	-0.20 (0.02)	-0.20 (4.7×10⁻⁵)	-0.21 (1.8×10 ⁻²)	-0.23 (0.02)
BMI	-1.94 (0)	-1.93 (1.8×10 ⁻³)	-1.82 (1.8×10⁻³⁸)	-1.00 (1.9×10 ⁻⁴)	-1.84 (0)	-1.82 (7.0×10⁻⁵⁹)	-1.90 (6.9×10⁻¹⁷)	-1.91 (2.7×10⁻⁶²)
SBP	0.15 (0.44)	0.19 (0.01)	0.23 (1.4×10⁻⁶)	0.23 (2.8×10⁻⁶)	0.23 (5.5×10⁻⁶)	0.23 (4.1×10⁻⁴)	0.23 (2.4×10⁻⁵)	0.22 (1.3×10 ⁻⁴)
TC	-0.15 (0.07)	-0.11 (0.08)	-0.13 (5.3×10 ⁻³)	-0.13 (5.5×10 ⁻³)	-0.13 (7.6×10 ⁻³)	-0.13 (6.7×10⁻⁶)	-0.13 (1.7×10 ⁻²)	-0.14 (0.02)
HDL	-0.46 (2.4×10⁻⁶)	-0.24 (1.38×10 ⁻³)	-0.44 (1.1×10⁻¹⁹)	-0.45 (2.2×10⁻¹⁶)	-0.45 (4.4×10⁻¹⁶)	-0.44 (1.3×10⁻³⁸)	-0.48 (8.4×10⁻¹⁶)	-0.48 (4.8×10⁻¹⁴)
LDL	-0.07 (0.66)	-0.08 (0.07)	-0.11 (0.02)	-0.11 (0.02)	-0.11 (0.03)	-0.11 (2.6×10⁻⁵)	-0.11 (4.0×10 ⁻²)	-0.11 (0.06)
TG	0.48 (3.9×10⁻⁶)	0.40 (1.9×10 ⁻³)	0.48 (3.0×10⁻¹²)	0.00 (1)	0.49 (2.1×10⁻¹⁴)	0.48 (9.6×10⁻²⁷)	0.50 (1.5×10⁻¹⁷)	0.48 (5.0×10⁻¹⁴)
Number of cigarettes previously smoked daily								
PWRIx	-0.07 (0.08)	-0.16 (0.05)	-0.09 (3.2×10 ⁻³)	-0.09 (6.8×10 ⁻³)	-0.09 (3.4×10 ⁻³)	-0.09 (1.2×10⁻¹¹)	-0.09 (1.3×10 ⁻²)	-0.08 (0.01)
BMI	-0.05 (0.31)	-0.48 (0.60)	-0.26 (5.4×10⁻¹⁷)	-0.29 (5.6×10⁻⁴⁵)	-0.27 (0)	-0.27 (3.3×10⁻²¹)	-0.28 (1.1×10⁻³⁵)	-0.29 (7.6×10⁻³⁹)
TC	-0.03 (0.32)	-0.05 (0.93)	-0.07 (0.01)	-0.07 (0.01)	-0.07 (3.8×10 ⁻³)	-0.07 (3.8×10⁻¹²)	-0.07 (1.0×10 ⁻³)	-0.07 (1.4×10 ⁻³)
HDL	-0.07 (3×10 ⁻³)	-0.12 (0.09)	-0.08 (2.8×10⁻⁵)	-0.08 (5.6×10 ⁻³)	-0.08 (2.7×10⁻⁵)	-0.08 (6.8×10⁻¹⁸)	-0.09 (6.5×10⁻⁵)	-0.08 (7.4×10⁻⁵)
LDL	-0.02 (0.46)	-0.06 (0.32)	-0.06 (0.03)	-0.06 (0.11)	-0.06 (0.01)	-0.06 (6.9×10⁻⁷)	-0.06 (3.6×10 ⁻³)	-0.06 (4.4×10 ⁻³)
Age stopped smoking								
BMI	-0.16 (8.1×10 ⁻³)	-1.08 (0.01)	-0.41 (3.4×10⁻⁵⁶)	-0.34 (4.7×10⁻³³)	-0.42 (0)	-0.42 (1.1×10⁻⁹⁴)	-0.42 (6.3×10⁻⁷⁵)	-0.43 (3.8×10⁻⁸²)
SBP	0.03 (0.55)	0.06 (0.14)	0.04 (0.01)	0.05 (0.01)	0.04 (0.01)	0.05 (9.5×10⁻⁹)	0.05 (1.2×10 ⁻²)	0.05 (0.01)
HDL	-0.10 (2.3×10⁻⁵)	-0.10 (0.61)	-0.09 (9.2×10 ⁻⁴)	0.07 (0.06)	-0.09 (6.9×10 ⁻⁴)	-0.09 (4.0×10⁻⁸)	-0.11 (7.0×10⁻⁸)	-0.11 (7.0×10⁻⁸)
TG	0.09 (6.7×10 ⁻⁴)	0.20 (0.12)	0.08 (6.1×10 ⁻⁴)	0.09 (3.1×10⁻⁶)	0.09 (6.4×10⁻⁷)	0.09 (1.1×10⁻²⁶)	0.09 (2.04×10⁻⁶)	0.09 (8.7×10⁻⁷)
Age started smoking in current smokers								
BMI	2.53 (0.07)	-0.12 (0.99)	3.03 (0.03)	-0.15 (0.26)	3.09 (0.02)	3.04 (0.02)	4.3 (7.3×10⁻¹⁷)	-0.12 (0.38)
SBP	0.23 (0.14)	-0.23 (0.45)	-0.26 (0.18)	0.24 (0.07)	-0.27 (0.18)	-0.26 (0.15)	-0.45 (2.0×10 ⁻⁴)	-0.49 (7.9×10⁻⁵)
HDL	-0.14 (0.34)	0.05 (0.88)	0.78 (0.05)	-0.16 (0.24)	0.80 (0.05)	0.79 (0.02)	0.97 (2.4×10⁻⁷)	-0.03 (0.86)
TG	-0.07 (0.91)	0.12 (0.79)	-0.60 (0.12)	0.32 (0.02)	-0.61 (0.12)	-0.60 (0.08)	-0.79 (7.0×10⁻⁶)	0.20 (0.48)
Number of cigarettes currently smoked daily (current cigarette smokers)								
PWRIx	0.07 (0.36)	0.11 (0.14)	0.10 (0.05)	0.11 (0.07)	0.00 (0.06)	0.13 (4.1×10⁻⁹)	-0.04 (0.52)	0.13 (0.06)
BMI	-0.06 (0.38)	-0.58 (0.39)	-0.57 (2.5×10⁻⁹)	-0.70 (6.2×10⁻⁵⁰)	-0.58 (1.6×10⁻⁹)	-0.57 (5.8×10⁻⁷)	-0.68 (5.0×10⁻³⁴)	-0.69 (6.4×10⁻⁴¹)
TC	-0.08 (0.05)	-0.18 (0.08)	-0.13 (1.9×10 ⁻³)	-0.28 (8.0×10⁻¹⁰)	-0.13 (3.2×10 ⁻³)	-0.13 (2.0×10 ⁻³)	-0.13 (2.5×10 ⁻³)	-0.10 (0.04)
HDL	-0.13 (1.5×10 ⁻³)	-0.02 (0.98)	-0.11 (1.1×10 ⁻³)	-0.11 (1.4×10 ⁻³)	-0.11 (1.5×10 ⁻³)	-0.11 (6.0×10⁻⁶)	-0.11 (5.6×10 ⁻³)	-0.12 (3.9×10 ⁻³)
Average weekly red wine intake								
TG	-0.50 (6.5×10⁻⁶)	-0.73 (0.02)	-0.52 (4.0×10⁻⁶)	-0.33 (3.2×10⁻⁵)	-0.51 (2.8×10⁻⁶)	-0.50 (4.2×10 ⁻³)	-0.46 (3.5×10⁻⁵)	-0.40 (0.06)
Average monthly intake of other alcoholic drinks								
TG	0.15 (0.58)	0.12 (0.47)	0.18 (0.25)	0.37 (4.4×10 ⁻³)	0.23 (0.09)	0.34 (2.7×10⁻¹⁶)	0.33 (5.3×10 ⁻³)	0.36 (0.01)
Duration of walks								
PWAIx	-0.23 (0.35)	-0.05 (2.1×10⁻⁵)	-0.29 (0.17)	-0.29 (0.16)	-0.29 (0.21)	-0.29 (4.4×10 ⁻³)	-0.27 (1.4×10 ⁻³)	-0.30 (0.18)
Duration of moderate activity								
PWP2PT	-0.21 (0.37)	-0.10 (0.46)	-0.23 (0.21)	-0.23 (0.23)	-0.24 (0.23)	-0.41 (0)	-0.41 (0.09)	-0.41 (0.18)
Number of days/week of vigorous physical activity 10+ minutes								
PWP2PT	0.40 (0.03)	0.41 (0.16)	0.42 (0.01)	0.71 (2.9×10⁻⁵)	0.43 (0.01)	0.43 (0.01)	0.59 (6.7×10 ⁻⁴)	0.65 (0.01)
BMI	2.48 (0.02)	0.15 (0.98)	2.35 (0.07)	-0.08 (0.53)	2.22 (0.08)	2.15 (0.31)	1.53 (1.4×10⁻⁸)	0.01 (0.94)
SBP	-0.33 (0.03)	-0.35 (0.21)	-0.31 (0.35)	-0.46 (4.8×10⁻⁷)	-0.36 (5.6×10 ⁻³)	-0.35 (0.01)	-0.35 (8.8×10 ⁻⁴)	-0.35 (6.6×10 ⁻³)
HDL	0.26 (0.05)	0.92 (0.49)	0.74 (0.01)	0.26 (0.26)	0.72 (0.01)	0.71 (0.03)	0.68 (5.4×10⁻⁶)	0.31 (0.08)
Time spent watching television (TV)								
HDL	-0.06 (0.16)	-0.15 (0.24)	-0.13 (3.4×10⁻⁵)	-0.07 (0.67)	-0.12 (1.0×10 ⁻⁴)	-0.09 (0.04)	-0.10 (0.02)	-0.08 (0.53)
BMI								
TSDriv	0.65 (0.01)	0.71 (0.02)	0.82 (2.5×10⁻⁶)	0.77 (0.12)	0.71 (3.1×10⁻⁵)	0.67 (1.5×10 ⁻⁴)	0.50 (0.04)	0.52 (0.04)
AveWRWI	0.01 (0.99)	-0.08 (0.88)	-0.10 (0.63)	-0.96 (4.9×10⁻⁵)	-0.06 (0.78)	-0.08 (0.75)	0.20 (0.49)	0.37 (0.21)
TC								
AveWFWI	-1.08 (0.32)	-1.46 (0.01)	-1.47 (2.2×10⁻⁵)	-0.56 (0.42)	-1.43 (6.1×10⁻⁵)	-1.42 (2.9×10⁻⁷)	-1.32 (0.01)	-1.35 (0.03)

Values are bold if $p < 8 \times 10^{-5}$. PR, pulse rate; PWRIx, pulse wave reflection index; PWP2PT, pulse wave peak to peak time; PWAIx, pulse wave Arterial Stiffness index; BMI, body mass index; SBP, systolic blood pressure, automated reading; DBP, diastolic blood pressure, automated reading; TC, cholesterol; HDL, high density lipoprotein; LDL, low density lipoprotein; TG, triglycerides; TSDriv, time spent driving; AveWRWI, Average weekly red wine intake; AveWFWI, Average weekly fortified wine intake.

Table S3. The UK Biobank traits analyzed in this study.

	Category	Trait	Field ID
lifestyle risk factors	Physical activity	Number of days/week walked 10+ minutes	864
		Duration of walks	874
		Number of days/week of moderate physical activity 10+ minutes	884
		Duration of moderate activity	894
		Number of days/week of vigorous physical activity 10+ minutes	904
		Time spent watching television (TV)	1070
		Time spent using computer	1080
		Time spent driving	1090
	Alcohol	Average weekly red wine intake	1568
		Average weekly champagne plus white wine intake	1578
		Average weekly beer plus cider intake	1588
		Average weekly spirits intake	1598
		Average weekly fortified wine intake	1608
		Average monthly red wine intake	4407
		Average monthly champagne plus white wine intake	4418
		Average monthly beer plus cider intake	4429
		Average monthly spirits intake	4440
		Average monthly fortified wine intake	4451
		Average monthly intake of other alcoholic drinks	4462
		Average weekly intake of other alcoholic drinks	5364
	Diet	Cooked vegetable intake	1289
		Salad / raw vegetable intake	1299
		Fresh fruit intake	1309
		Dried fruit intake	1319
		Bread intake	1438
		Cereal intake	1458
		Tea intake	1488
		Coffee intake	1498
		Water intake	1528
	Age when last ate meat	3680	
	Smoking	Exposure to tobacco smoke at home	1269
		Exposure to tobacco smoke outside home	1279
		Age started smoking in former smokers	2867
Number of cigarettes previously smoked daily		2887	
Age stopped smoking		2897	
Number of unsuccessful stop-smoking attempts		2926	
Age started smoking in current smokers		3436	
Number of cigarettes currently smoked daily (current cigarette smokers)		3456	
CVD-related traits	Arterial stiffness	Pulse rate	4194
		Pulse wave reflection index	4195
		Pulse wave peak to peak time	4196
		Pulse wave Arterial Stiffness index	21021
	Body size measures	Body mass index (BMI)	21001
	Blood pressure	Diastolic blood pressure, automated reading	4079
		Systolic blood pressure, automated reading	4080
	Blood biochemistry	Cholesterol	30690
HDL cholesterol		30760	
LDL direct		30780	
Triglycerides		30870	

Table S4. Eight MR methods in this study and the corresponding software.

Method	R package	Version	Function	Tuning parameters
CAUSE ⁸	cause	NA	cause	default settings
IVW-R ⁷³	MendelianRandomization	0.4.0	mr_ivw	default settings
MRAID	MRAID	NA	MRAID	default settings
MRMix ⁷	MRMix	NA	MRMix	default settings
Weighted median ⁴	MendelianRandomization	0.4.0	mr_median	default settings
Weighted mode ⁷⁴	MendelianRandomization	0.4.0	mr_mbe	default settings
RAPS ⁶	mr.raps	NA	mr.raps.overdispersed.robust	(b_exp,b_out,se_exp,se_out,loss.function="huber",k=1.345,initialization=c("I2"),suppress.warning=FALSE,diagnostics=FALSE,niter=20,tol=.Machine\$double.eps^0.5)
Robust ⁷⁶	MendelianRandomization	0.4.0	mr_ivw	(model="random",robust=TRUE), other settings are default