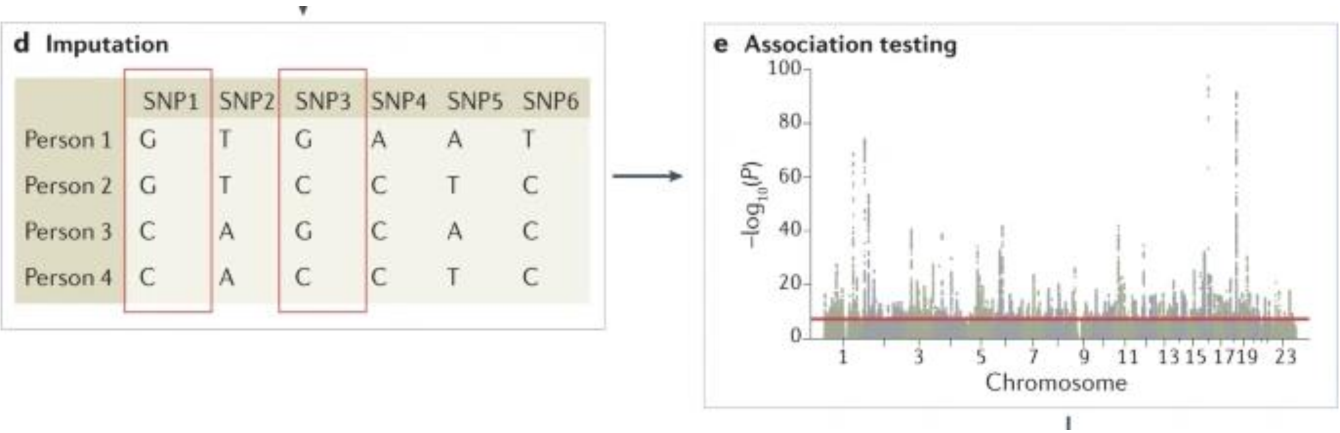
Integration of GWAS & eQTL

Sun Jianle

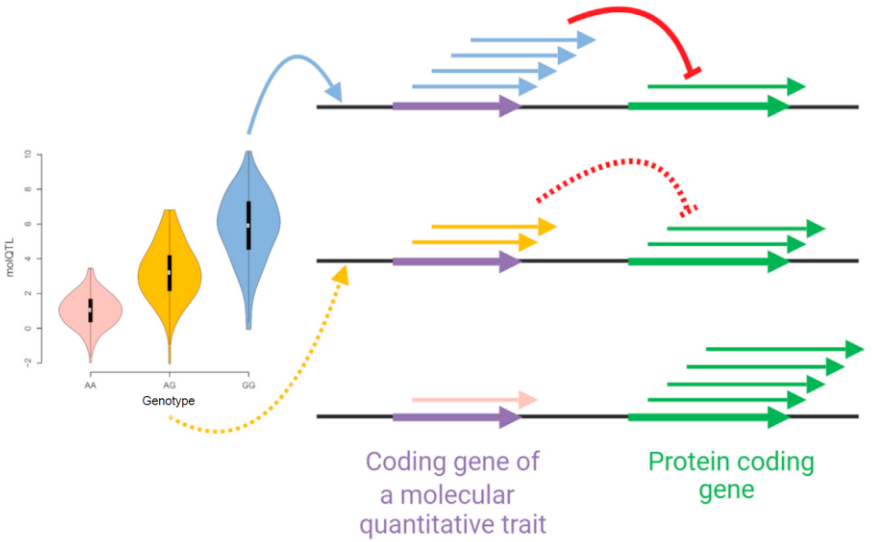
GWAS & eQTL

- GWAS



	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Person 1	G	T	G	A	A	T
Person 2	G	T	C	C	T	C
Person 3	C	A	G	C	A	C
Person 4	C	A	C	C	T	C

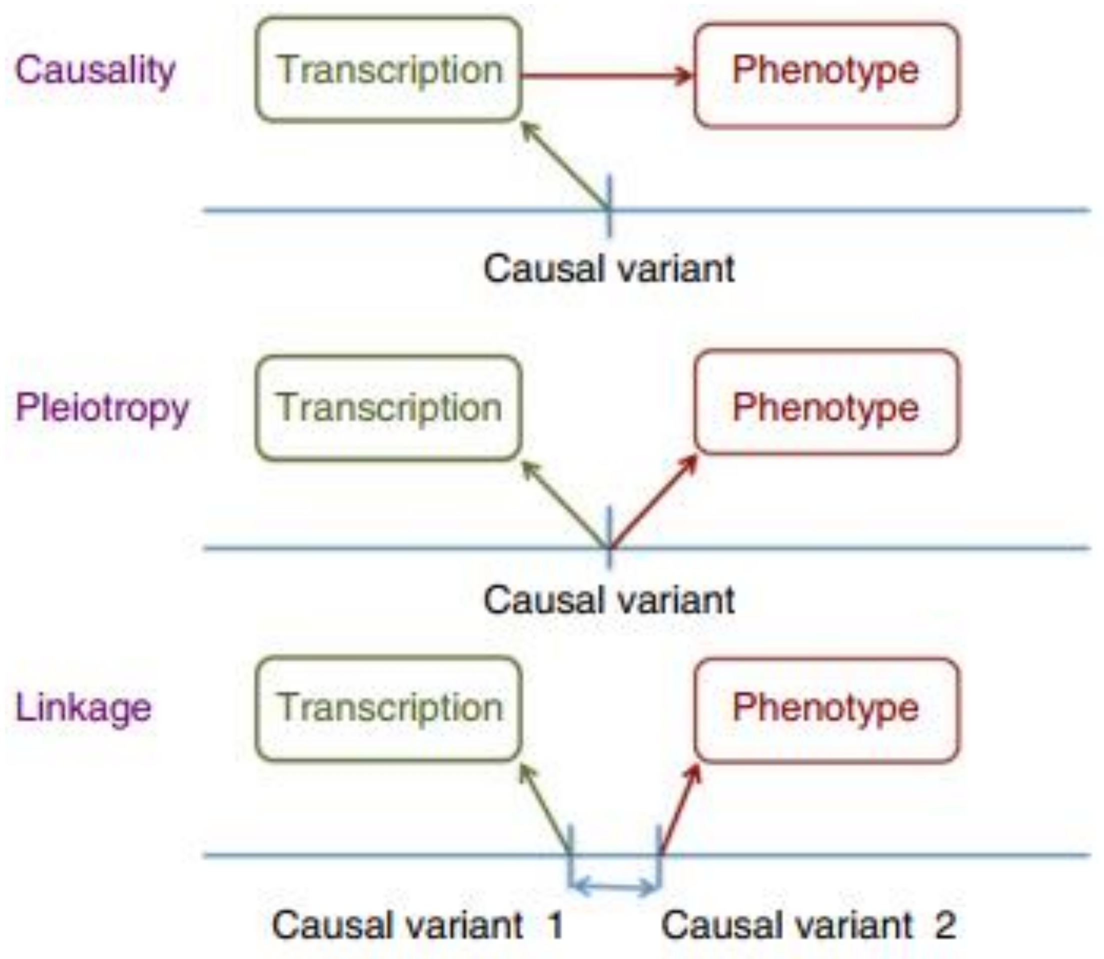
- eQTL



Trends in Molecular Medicine

Motivation

- Functional annotation of GWAS signals

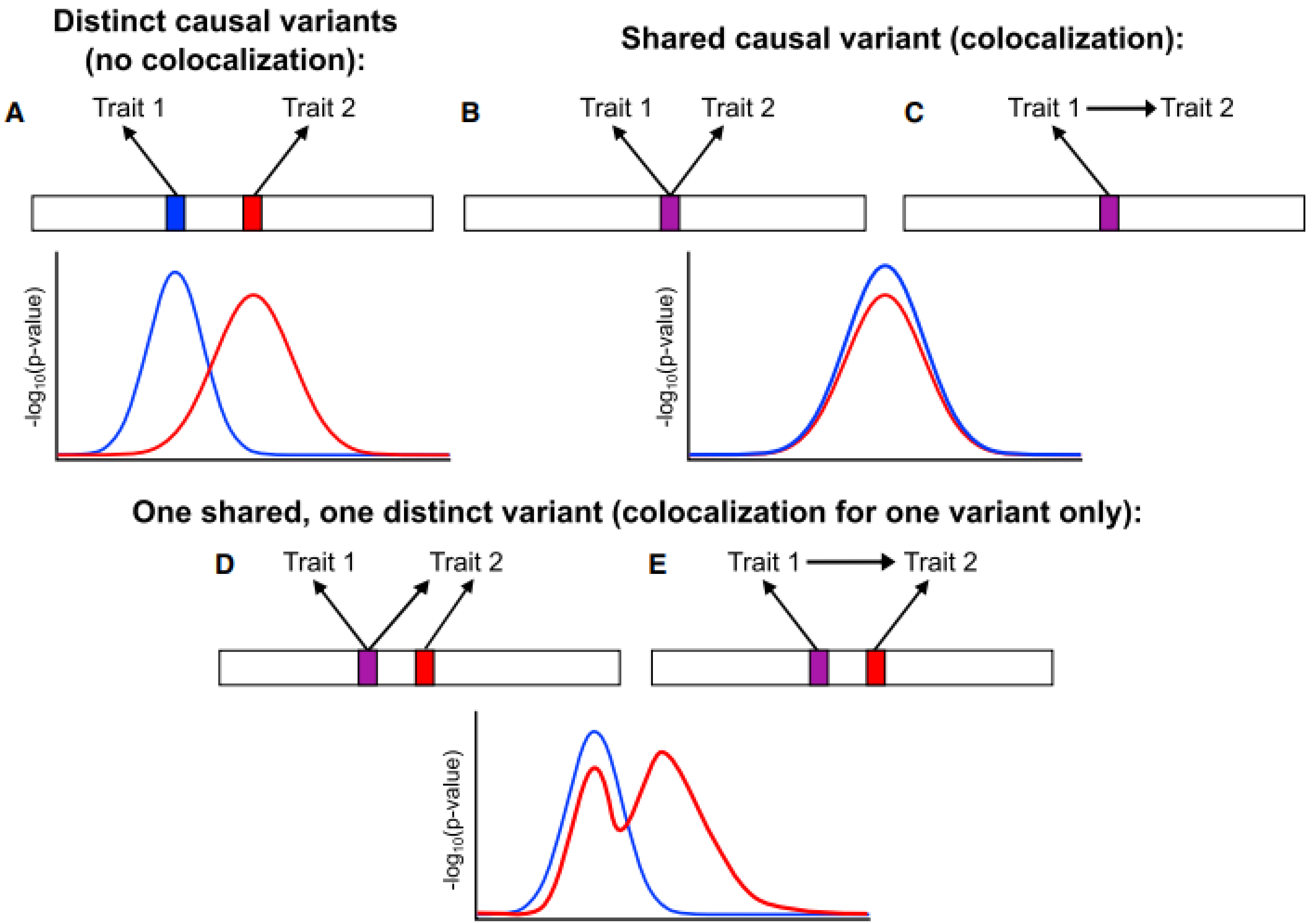


Main methods

- Colocalization
- Mendelian randomization
- Transcriptomic-wide association study (TWAS)

Colocalization

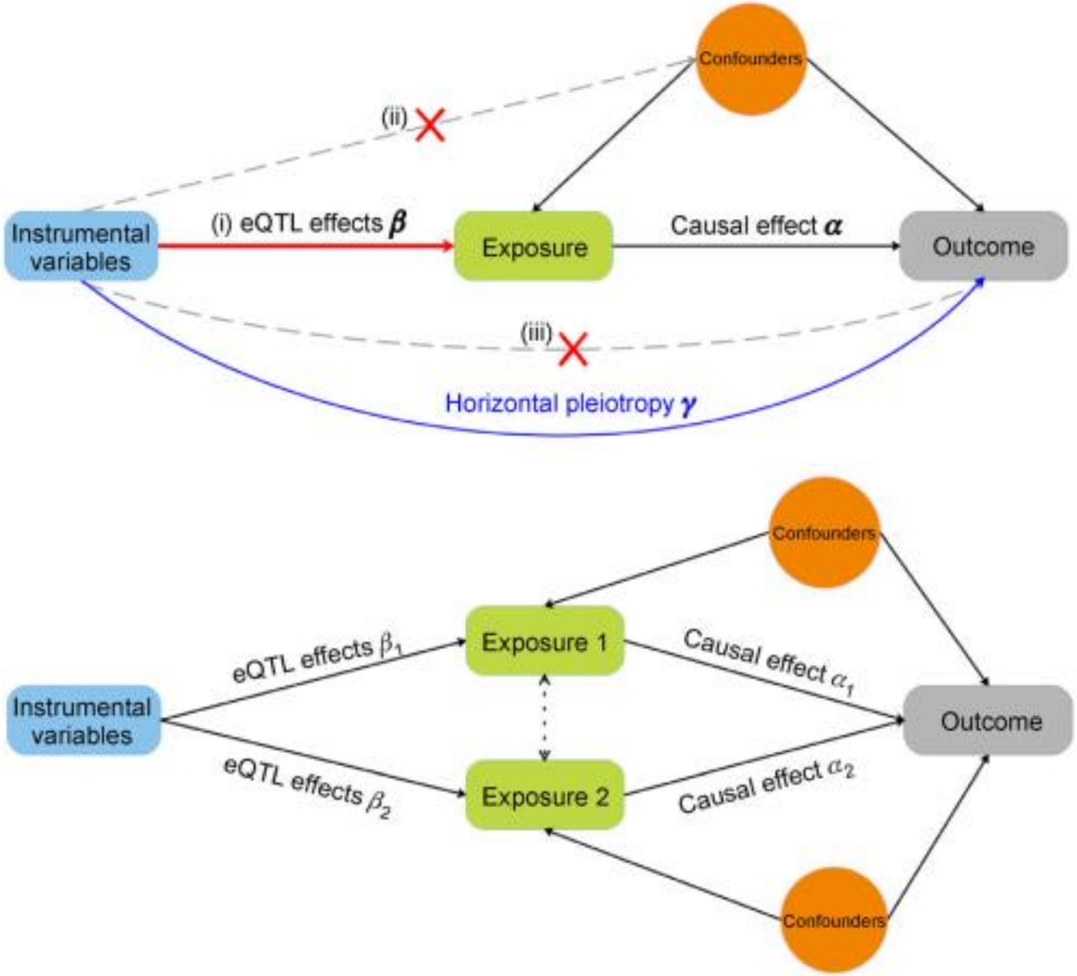
- Shared causal variants for GWAS and eQTL signals



Mendelian randomization

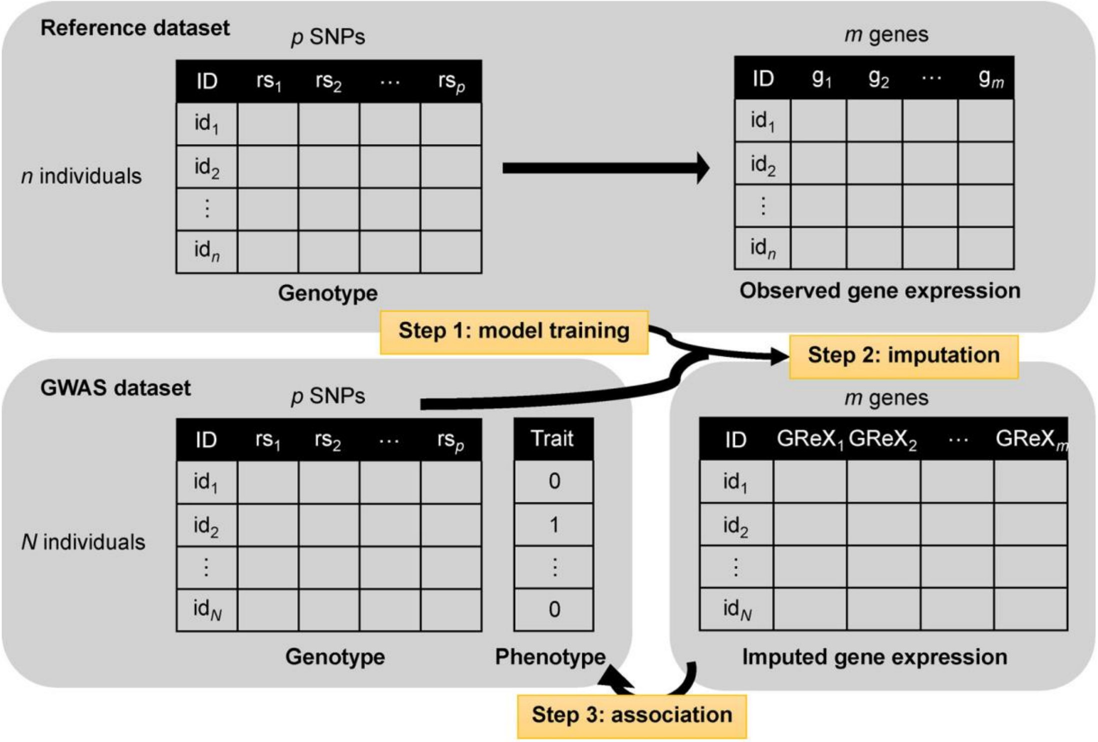
- Gene expression as a molecular trait

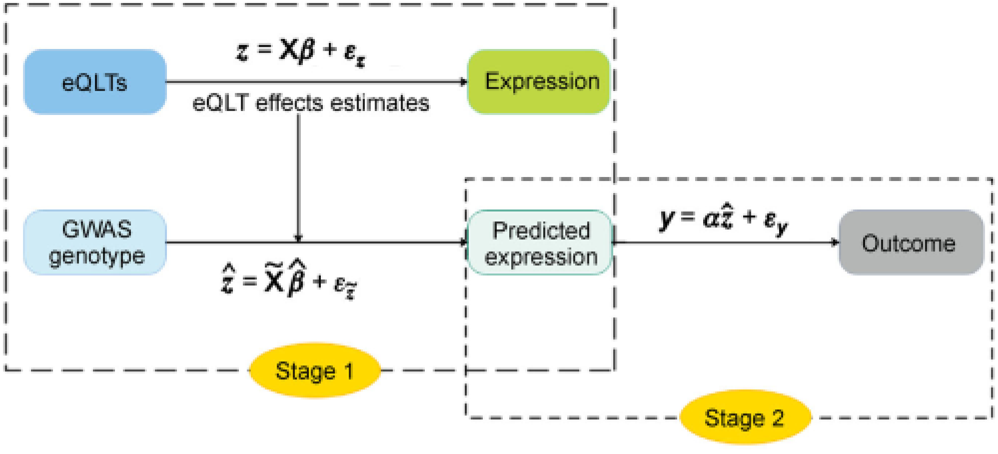
- cis-SNPs: correlation among instruments
- Multi-exposure: multiple gene expressions



TWAS

- Missing data imputation perspective
- Mendelian randomization perspective





Several methods

Integration of GWAS & eQTL

I. colocalization

$\boxed{\text{Coloc}}$ each locus i in region Q, (a_{1i}, a_{2i})
$X_1 \quad X_2$

$$p_0 + p_1 + p_2 + p_{12} = 1$$
$$(0,0) \quad (1,0) \quad (0,1) \quad (1,1)$$

<u>At most one causal locus for each trait</u>

H_0: No effect on X_1 & X_2
H_1: effect on X_1
H_2: effect on X_2
H_3: effect on X_1 & X_2, different causal loci
H_4: effect on X_1 & X_2, <u>the same locus</u>
$\Downarrow$
colocalization

$$PP_4 = P(H_4|D) = \frac{P(H_4|D)}{P(H_0|D)+P(H_1|D)+P(H_2|D)+P(H_3|D)+P(H_4|D)}$$

$$= \frac{\frac{P(H_4|D)}{P(H_0|D)}}{1 + \underline{\frac{P(H_1|D)}{P(H_0|D)}} + \frac{P(H_2|D)}{P(H_0|D)} + \frac{P(H_3|D)}{P(H_0|D)} + \frac{P(H_4|D)}{P(H_0|D)}}$$

$$\frac{P(H_h|D)}{P(H_0|D)} = \frac{\sum_{S\in S_h} P(D|S)\cdot P(S)}{\sum_{S\in S_0} P(D|S)\cdot P(S)} = \frac{P(S|S\in S_h)\sum_{S\in S_h} P(D|S)}{P(D|S_0)\cdot P(S_0)} = \sum_{S\in S_h} \underbrace{\frac{P(D|S)}{P(D|S_0)}}_{} \cdot \frac{P(S)}{P(S_0)}$$

Bayes factor (BF) $\longrightarrow$ ABF

$$S\in S_0: \quad \frac{P(S)}{P(S_0)} = \frac{p_0^Q}{p_0^Q} = 1$$

$$S\in S_1: \quad \frac{P(S)}{P(S_0)} = \frac{p_0^{Q-1}p_1}{p_0^Q} = \frac{p_1}{p_0} \approx p_1$$

$$S\in S_2: \quad \frac{P(S)}{P(S_0)} = \frac{p_0^{Q-1}p_2}{p_0^Q} = \frac{p_2}{p_0} \approx p_2$$

$$S\in S_3: \quad \frac{P(S)}{P(S_0)} = \frac{p_0^{Q-2}p_1p_2}{p_0^Q} = \frac{p_1p_2}{p_0^2} \approx p_1p_2$$

$$S\in S_4: \quad \frac{P(S)}{P(S_0)} = \frac{p_0^{Q-1}p_{12}}{p_0^Q} = \frac{p_{12}}{p_0} \approx p_{12}$$

$$\Rightarrow$$

$$\frac{P(H_0|D)}{P(H_0|D)} = 1$$

$$\frac{P(H_1|D)}{P(H_0|D)} = p_1 \sum_j ABF_j^1$$

$$\frac{P(H_2|D)}{P(H_0|D)} = p_2 \sum_j ABF_j^2$$

$$\frac{P(H_3|D)}{P(H_0|D)} = p_1 p_2 \sum_{j\neq k} ABF_j^1 ABF_k^2$$

$$\frac{P(H_4|D)}{P(H_0|D)} = p_{12} \sum_j ABF_j^1 ABF_j^2$$

<u>To calculate ABF</u>

$$Y = \mu + \beta X + e \qquad \hat{\beta} \rightarrow N(\beta, V)$$

H_0: $\beta = 0$

H_a: $\beta \sim N(0, W)$

$$\beta|y \sim N\left(\frac{\frac{1}{V}\hat{\beta} + \frac{1}{W}0}{\frac{1}{V} + \frac{1}{W}}, \frac{1}{\frac{1}{V}+\frac{1}{W}}\right)$$

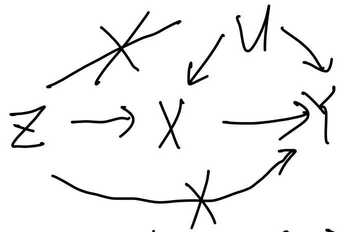
$$BF = \frac{P(Y|\beta,\mu)}{P(Y|\beta=0,\mu)} = \frac{\iint f(y|\beta,\mu)\,\pi(\beta,\mu)\,d\beta\,d\mu}{\int f(y|\beta=0,\mu)\,\pi(\mu)\,d\mu}$$

<u>Ignoring μ</u>

$$\frac{\int f(y|\beta)\,\pi(\beta)\,d\beta}{f(y|\beta=0)} = \int \frac{f(y|\beta)}{f(y|\beta=0)}\pi(\beta)\,d\beta.$$

$$\sim N\left(\frac{W}{V+W}\hat{\beta}, \frac{W}{V+W}\right) \qquad \boxed{ABF = \sqrt{1-r}\exp\left(\frac{1}{2}z^2+r\right)} \qquad r=\frac{W}{W+V}, \quad z=\hat{\beta}/V$$

II. MR



one IV: Wald estimator

$$\hat{\beta}_{XY} = \frac{\hat{\beta}_{ZY}}{\hat{\beta}_{ZX}} \xrightarrow{plim} \frac{cov(Z,Y)}{cov(Z,X)} = \beta_{XY}$$

OLS:

$$Y = X\beta + \varepsilon$$

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

$$\hat{Y} = X\hat{\beta} = \underline{X(X^TX)^{-1}X^T}Y = P_X Y$$

$$P_X^T P_X = P_X \quad P_X = P_X^T$$

$$\# IV = \# \text{exogenous variables} \Rightarrow \underline{\text{exact identification}}$$

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$$

$$\# IV > \# \text{exogenous variables} \Rightarrow \text{over identification (TSLS)}$$

$$\hat{\beta}_{TSLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y \xrightarrow{\hat{X}=P_ZX} (X'P_Z'P_ZX)^{-1}X'P_Z'Y$$

$$= (X'P_ZX)^{-1}X'P_ZY$$

summary data $\boxed{IVW}$

for Z_j, $j=1,\cdots,J$, we have $\hat{\beta}_{ZX_j}$, $\hat{\beta}_{ZY_j}$ $\Rightarrow \hat{\beta}_{XY_j} = \dfrac{\hat{\beta}_{ZY_j}}{\hat{\beta}_{ZX_j}}$

Δ method:

$$Var(\hat{\beta}_{XY_j}) = Var\left(\frac{\hat{\beta}_{ZY_j}}{\hat{\beta}_{ZX_j}}\right) \approx \underbrace{\frac{Var(\hat{\beta}_{ZY_j})}{\beta_{ZX_j}^2}} + \frac{\beta_{ZY_j}^2}{\beta_{ZX_j}^4}Var(\hat{\beta}_{ZX_j}) - 2\frac{\beta_{ZY_j}}{\hat{\beta}_{ZX_j}^3}cov(\hat{\beta}_{ZX_j}, \hat{\beta}_{ZY_j})$$

$$\hat{Var}(\hat{\beta}_{XY_j}) \approx \frac{Var(\hat{\beta}_{ZY_j})}{\hat{\beta}_{ZX_j}^2}$$

let $\hat{\gamma}_j = \hat{\beta}_{ZX_j}$, $\hat{\Gamma}_j = \hat{\beta}_{ZY_j}$, $\sigma_{X_j}^2 = Var(\hat{\beta}_{ZX_j})$. $\sigma_{Y_j}^2 = Var(\hat{\beta}_{ZY_j})$ $\Rightarrow \hat{\beta}_j = \dfrac{\hat{\Gamma}_j}{\hat{\gamma}_j}$, $Var(\hat{\beta}_j) \approx \dfrac{\sigma_{Y_j}^2}{\hat{\gamma}_j^2}$

meta-analysis

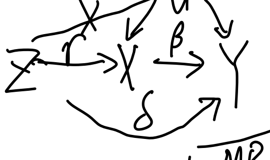
$$\hat{\beta}_{IVW} = \frac{\sum_j \hat{\gamma}_j \hat{\Gamma}_j \sigma_{Y_j}^{-2}}{\sum_j \hat{\gamma}_j^2 \sigma_{Y_j}^{-2}} \rightarrow \underline{(\gamma'\Sigma_Y^{-1}\gamma)^{-1}\gamma'\Sigma_Y^{-1}\Gamma}$$

$\boxed{\text{independent SNPs}}$

$\Downarrow$

Another perspective: $\Gamma_j = \beta_{IVW}\gamma_j + \varepsilon_j$ $\quad \varepsilon_j \sim N(0, \sigma_{Y_j}^2)$, $\underline{cov(\varepsilon_i, \varepsilon_j) = 0}$.

Horizontal pleiotropy



$$Y = \beta X + \delta Z + \varepsilon = \alpha\beta Z + \delta Z + \varepsilon.$$

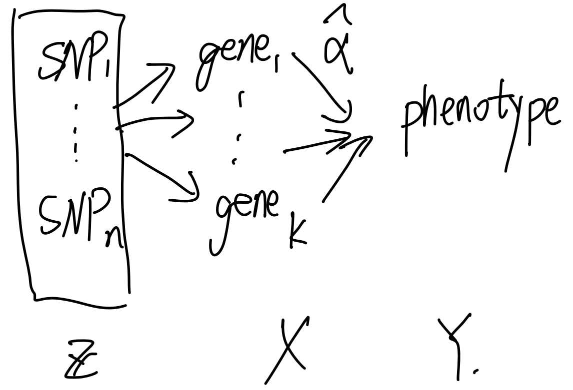
$$\Gamma_j = \beta\gamma_j + \delta_j + \varepsilon_j'$$

$$\begin{cases} \delta_j = \delta & \text{fixed-effect} \\ \alpha_j \perp\!\!\!\perp \delta & \text{InSIDE} \end{cases}$$

$$\Longleftarrow$$

$\boxed{\text{MR-Egger}}$ $\Gamma_j = \beta\gamma_j + \delta + \varepsilon_j$, $\varepsilon_j \sim N(0, \sigma_{Y_j}^2)$

TWMR : cis-eQTL → SNPs in LD



$\hat{\beta}_{ZX} \rightarrow E_{n\times k}$.

$\hat{\beta}_{ZY} \rightarrow G_{n\times 1}$

$\hat{\beta}_{XY} \rightarrow \vec{\alpha}_{k\times 1}$

LD matrix of Z C

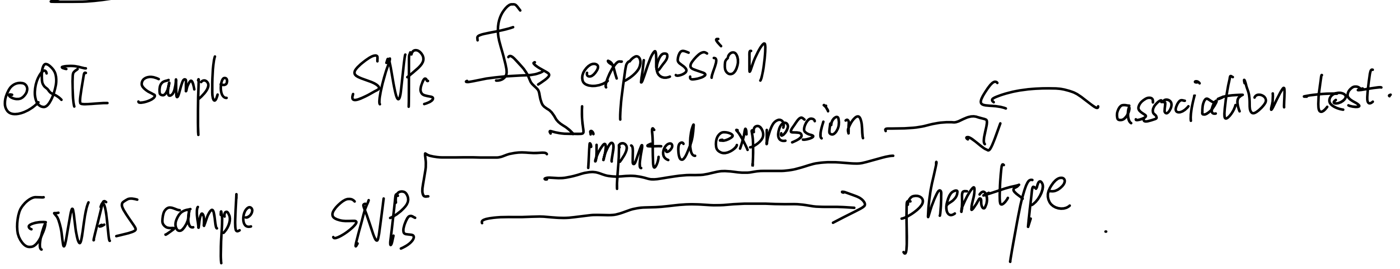
$$\Rightarrow \hat{\alpha} = (E'\underline{\underline{C^{-1}}}E)^{-1}(E'\underline{\underline{C^{-1}}}G)$$

seemingly more reasonable approach

$$\hat{\alpha} = (E'(\sqrt{\Sigma}'C\sqrt{\Sigma})^{-1}E)(E'(\sqrt{\Sigma}'C\sqrt{\Sigma})^{-1}E)$$

where $\Sigma = diag(var(\hat{\beta}_{ZY_1}), \cdots, Var(\hat{\beta}_{ZY_n}))$

III. TWAS.



PrediXcan

$X_g = \sum_k w_{k,g} Z_g + \varepsilon$ Elastic Network $w \propto \exp(\lambda_1 \|w\|_1 + \lambda_2 \|w_2\|_2)$

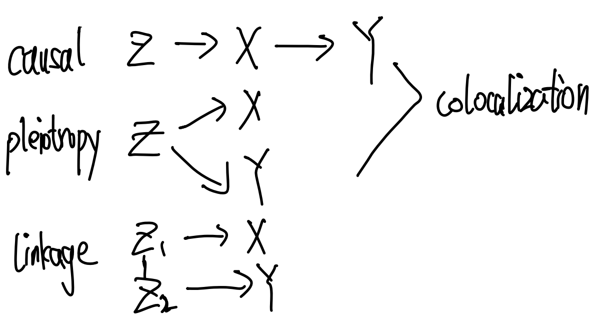
FUSION Bayesian sparse linear mixed model (BSLMM) $w \sim \pi N(0, \sigma_a^2 + \sigma_b^2) + (1-\pi) N(0, \sigma_b^2)$ omnigenic model

> spike & slab

BGW-TWAS Bayesian variable selection regression (BVSR) $w \sim \pi(0, \sigma_w^2) + (1-\pi)\underline{\delta_0}$

↪ point mass at zero

SMR



$$\hat{\beta}_{xy} = \frac{\hat{\beta}_{yz}}{\hat{\beta}_{xz}}$$

$$Var(\hat{\beta}_{xy}) = Var\left(\frac{\hat{\beta}_{yz}}{\hat{\beta}_{xz}}\right) \approx \frac{Var(\hat{\beta}_{yz})}{\beta_{xz}^2} + \frac{\beta_{zy}^2}{\beta_{zx}^4} Var(\hat{\beta}_{zy}) - \frac{2\beta_{zy}}{\beta_{zx}^3} cov(\hat{\beta}_{zx}, \hat{\beta}_{zy})$$

$$\hat{Var}(\hat{\beta}_{xy}) = \left(\frac{\hat{\beta}_{zy}}{\hat{\beta}_{zx}}\right)^2 \left[\frac{Var(\hat{\beta}_{zy})}{\hat{\beta}_{zy}^2} + \frac{Var(\hat{\beta}_{zx})}{\hat{\beta}_{zx}^2} - \frac{2Cov(\hat{\beta}_{zx}, \hat{\beta}_{zy})}{\hat{\beta}_{zx}\hat{\beta}_{z}}\right]$$

$$= \hat{\beta}_{xy}\left(\frac{1}{Z_{zy}^2} + \frac{1}{Z_{zx}^2}\right)$$

No sample overlap

$\hookrightarrow \text{cov}(\hat{\beta}_{zx}, \hat{\beta}_{zy}) = 0$

$z\text{-score} = \frac{\hat{\beta}}{\text{Var}(\hat{\beta})}$

$$\chi^2_{SMR} = \frac{\hat{\beta}_{xy}^2}{\text{Var}(\hat{\beta}_{xy})} \approx \frac{Z_{zy}^2 Z_{zx}^2}{Z_{zy}^2 + Z_{zx}^2}$$

If only z-scores available:

$$y = bx + b_0 + \varepsilon \qquad \hat{b} = \frac{S_{xy}}{S_{xx}} = \sum_i \frac{(x_i - \bar{x})}{S_{xx}} y_i$$

$X \sim B(2, p) \mapsto \text{MAF}$

Y standardized

$$\text{Var}(\hat{b}) = \sigma_e^2 \cdot \frac{1}{S_{xx}} = \frac{\sigma_e^2}{n\text{Var}(x)}$$

$$R^2 = r^2 = \frac{\text{Var}(x)}{\text{Var}(y)} \hat{b}^2 = 1 - \frac{SSE}{n\text{Var}(y)} \Rightarrow SSE = n\text{Var}(y) - n\text{Var}(x)\hat{b}^2$$

$$= n[1 - \hat{b}^2 \cdot 2p(1-p)]$$

$$\sigma_e^2 \approx 1 - 2\hat{b}^2 p(1-p)$$

$$\Rightarrow SE_{\hat{b}} = \sqrt{\frac{\sigma_e^2}{n\text{Var}(x)}} = \sqrt{\frac{1-2p(1-p)\hat{b}^2}{2p(1-p)n}}$$

$$Z = \frac{\hat{b}}{SE_{\hat{b}}} = \frac{\hat{b}\sqrt{2p(1-p)n}}{\sqrt{1-2p(1-p)\hat{b}^2}} \Rightarrow \hat{b} = \frac{z}{\sqrt{2p(1-p)(n+z^2)}}$$

Distinguish colocalization and linkage: HEIDI test

if Z_0 is a causal locus for both X and Y, Z_i is linked to Z_0

then:

$$\hat{\beta}_{xy_i} = \frac{\hat{\beta}_{zy_i}}{\hat{\beta}_{zx_i}} = \frac{\hat{\beta}_{zy_0} r_{0i}\sqrt{h_0/h_i}}{\hat{\beta}_{zx_0} r_{0i}\sqrt{h_0/h_i}} = \frac{\hat{\beta}_{zy_0}}{\hat{\beta}_{zx_0}} = \hat{\beta}_{xy_0}$$

$$\begin{tikzcd} Z_i \arrow[r] & Z_0 \arrow[r, "\hat{\beta}_{zx_0}"] & X \\ r_{0i}\sqrt{\frac{h_0}{h_i}} \arrow[u] & & \\ & \Downarrow & \\ Z_i \arrow[rr, "r_{0i}\sqrt{\frac{h_0}{h_i}} \cdot \hat{\beta}_{zx_0}"'] & & X \end{tikzcd}$$

here: $h_i = 2p_i(1-p_i) = \text{var}(Z_i)$ under HWE

Testing colocalization → Testing whether there is a difference between $\hat{\beta}_{xy}$ estimated by loci in LD.

$$d_i = \hat{\beta}_{xy_i} - \hat{\beta}_{xy_0} \qquad \vec{\hat{d}_i} \sim MVN(\vec{d}, \mathbb{V})$$

$$\text{cov}(\hat{d}_i, \hat{d}_j) = \text{cov}(\hat{\beta}_{xy_i} - \hat{\beta}_{xy_0}, \hat{\beta}_{xy_j} - \hat{\beta}_{xy_0}) = \text{cov}(\hat{\beta}_{xy_i}, \hat{\beta}_{xy_j}) - \text{cov}(\hat{\beta}_{xy_0}, \hat{\beta}_{xy_j}) - \text{cov}(\hat{\beta}_{xy_j}, \hat{\beta}_{xy_0})$$

$$+ \text{var}(\hat{\beta}_{xy_0})$$

solve $\text{cov}(\hat{d}_i, \hat{d}_j)$:

$$\text{cov}(\hat{\beta}_{xy_i}, \hat{\beta}_{xy_j}) = E\left(\frac{\hat{\beta}_{zy_i}, \hat{\beta}_{zy_j}}{\hat{\beta}_{zx_i}, \hat{\beta}_{zx_j}}\right) - E\left(\frac{\hat{\beta}_{zy_i}}{\hat{\beta}_{zx_i}}\right)E\left(\frac{\hat{\beta}_{zy_j}}{\hat{\beta}_{zx_j}}\right)$$

second-order Δ method

$$E_\theta g(T) \approx g(\theta) + \underbrace{\sum_i g_i'(\theta) E_\theta(T_i - \theta_i)}_{} \overset{=0}{} + \sum_i \frac{1}{2} g_i''(\theta) E_\theta (T_i - \theta_i)^2$$

$$E\left(\frac{\hat{\beta}_{zy_i}\hat{\beta}_{zy_j}}{\hat{\beta}_{zx_i}\hat{\beta}_{zx_j}}\right) \approx \frac{\beta_{zy_i}\beta_{zy_j}}{\beta_{zx_i}\beta_{zx_j}} + 0 \cdot \text{Var}(\hat{\beta}_{zy_i}) + 0 \cdot \text{Var}(\hat{\beta}_{zy_j}) + \frac{\beta_{zy_i}\beta_{zy_j}}{\beta_{zx_i}^3 \beta_{zy_j}} \text{Var}(\hat{\beta}_{zx_i})$$

$$+\frac{\beta_{zy_i}\beta_{zy_j}}{\beta_{zx_i}\beta_{zx_j}^3}Var(\hat{\beta}_{zx_j})+\frac{1}{\beta_{zx_i}\beta_{zx_j}}cov(\hat{\beta}_{zy_i},\hat{\beta}_{zy_j})+\frac{\beta_{zy_i}\beta_{zy_j}}{\beta_{zx_i}^2\beta_{zx_j}^2}cov(\hat{\beta}_{zx_i},\hat{\beta}_{zx_j})$$

$$+\square\cdot\underbrace{cov(\hat{\beta}_{zx_i},\hat{\beta}_{zy_i})}_{\substack{\shortparallel\\0}}+\square\cdot\underbrace{cov(\hat{\beta}_{zx_j},\hat{\beta}_{zy_j})}_{\substack{\shortparallel\\0}}+\square\cdot\underbrace{cov(\hat{\beta}_{zx_i},\hat{\beta}_{zy_j})}_{\substack{\shortparallel\\0}}+\square\cdot\underbrace{cov(\hat{\beta}_{zx_j},\hat{\beta}_{zy_i})}_{\substack{\shortparallel\\0}}$$

$\boxed{\text{two-sample (independent)}}$

$$E\left(\frac{\hat{\beta}_{zy_i}}{\hat{\beta}_{zx_i}}\right)=\frac{\beta_{zy_i}}{\beta_{zx_i}}\left(1+\frac{Var(\hat{\beta}_{zx_i})}{\beta_{zx_i}^2}-\underbrace{\frac{cov(\hat{\beta}_{zx_i},\hat{\beta}_{zy_i})}{\beta_{zy_i}\beta_{zx_i}}^{=0}}\right)$$

$$\Rightarrow cov(\hat{\beta}_{xy_i},\hat{\beta}_{xy_j})\approx\frac{1}{\beta_{zx_i}\beta_{zx_j}}\underbrace{cov(\hat{\beta}_{zy_i},\hat{\beta}_{zy_j})}_{\substack{\Downarrow\\ r_{ij}\sqrt{Var(\hat{\beta}_{zy_i})Var(\hat{\beta}_{zy_j})}}}+\frac{\beta_{zy_i}\beta_{zy_j}}{\beta_{zx_i}^2\beta_{zx_j}^2}\underbrace{cov(\hat{\beta}_{zx_i},\hat{\beta}_{zx_j})}_{\substack{\shortparallel\\ r_{ij}\sqrt{Var(\hat{\beta}_{zx_i})Var(\hat{\beta}_{zx_j})}}}\underbrace{-\frac{\beta_{zy_i}\beta_{zy_j}}{\beta_{zx_i}^3\beta_{zx_j}^3}Var(\hat{\beta}_{zx_i})Var(\hat{\beta}_{zx_j})}_{\substack{\shortparallel\\ -\frac{\beta_{zy_i}\beta_{zy_j}}{\beta_{zx_i}\beta_{zx_j}}\cdot\frac{1}{z_{zx_i}^2z_{zx_j}^2}}}$$

$$=\frac{r_{ij}}{\beta_{zx_i}\beta_{zx_j}}\sqrt{Var(\hat{\beta}_{zy_i})Var(\hat{\beta}_{zy_j})}+\underline{\hat{\beta}_{xy_i}\hat{\beta}_{xy_j}}\left(\frac{r}{z_{zx_i}z_{zx_j}}-\frac{1}{z_{zx_i}^2z_{zx_j}^2}\right)$$

$$z_{d_i}=\frac{\hat{d}_i}{\sqrt{var(d_i)}}\qquad T_{HEIDI}=z_dIz_d^T=\sum_i z_{d_i}^2\qquad H_0:\vec{d}=0$$

quadric form of standard normal variables. Satterthwaite method.

GSMR

only one SNP used in SMR. cannot distinguish causality & pleiotropy

causality $Z\to X\to Y$

pleiotropy $Z\to X$, $Z\to Y$

} GSMR (independent SNPs)

linkage $Z_1\to X$, $Z_2\to Y$ (Z_1 — Z_2) $\Rightarrow$ HEIDI test using variants in LD. (SMR)

m SNPs (near independent)

$$\vec{\hat{\beta}}_{xy}\sim MNV(\underline{\mathbb{1}\beta_{xy}},\mathbb{V})$$

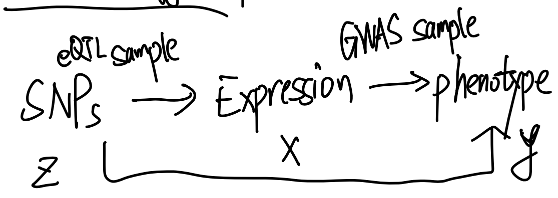
$\mathbb{V}$: $cov(\hat{\beta}_{xy_i},\hat{\beta}_{xy_j})$, $Var(\hat{\beta}_{xy_i})$

$$\Downarrow$$

$$\hat{\beta}_{xy}=(\mathbb{1}'V^{-1}\mathbb{1})^{-1}\mathbb{1}'V^{-1}\vec{\hat{\beta}}_{xy}$$

weighted least square

PMR-Egger



$$x=\mu_x+Z_x\beta+\varepsilon_x$$
$$\tilde{x}=\mu_x+Z_y\beta+\varepsilon_{\tilde{x}}$$
$$y=\mu_y+\tilde{x}\alpha+Z_y\gamma+\varepsilon$$
$$>y=\tilde{\mu}_y+\underline{Z_y\beta\alpha}+\underline{Z_y\gamma}+\varepsilon_y.$$

unidentifiable

$\Rightarrow$ Assumption: $\underbrace{\vec{\beta} \sim N(0, \sigma_z^2 I_p)}$; $\underbrace{\gamma_j = \gamma, \text{for } j = 1, \cdots, p}$ $\Rightarrow$ fixed-effect ; InSIDE

Estimation (EM) $\hookrightarrow x \sim N(\vec{\mu}_x + Z_x \vec{\beta}, \sigma_x^2 \mathbb{I}_{n_1})$, $y \sim N(\vec{\mu}_y + Z_y \vec{\beta} \alpha + Z_y \gamma \vec{1}, \sigma_y^2 \mathbb{I}_{n_2})$

observed likelihood $f(x.y) = \int f(x, y, \beta) d\beta = \int f(x, y|\beta) f(\beta) d\beta \underset{\substack{\uparrow \\ \text{two-sample : independent}}}{=} f(y|\beta) f(x|\beta) f(\beta) d\beta$

$\beta \rightarrow$ latent variables

$\alpha, \gamma, \sigma_i^2, \mu_i \longrightarrow$ parameters

parameter-expanded EM (λ)

$x = \mu_x + \lambda Z_x \beta + \varepsilon_x$

$y = \mu_y + Z_y \beta \alpha + Z_y \gamma + \varepsilon_y$

$R(\lambda, \alpha, \gamma, \sigma_y^2, \sigma_x^2, \mu_x, \mu_y) = (\frac{\alpha}{\lambda}, \gamma, \sigma_y^2, \sigma_x^2, \mu_x, \mu_y)$

$\hookrightarrow \beta | x, y, Z_x, Z_y, \underline{\theta}^t \sim N(\mu_\beta . \Sigma_B)$

$$\boxed{\begin{aligned} \text{EM}: Q(\theta) &= E_{z \sim P(z|X, \theta^t)} \ln P(X z|\theta) \\ &= \sum_z P(z|X, \theta^t) \ln P(X z|\theta) \end{aligned}}$$

E-step: $\mathbb{E}_{\beta|x, y, Z_x, Z_y, \theta^t} \log f(x, y, \beta|\theta)$

$= \mathbb{E}_{\beta \sim N(\mu_B, \Sigma_B)} \log \underline{f(x|\beta) f(y|\beta) f(\beta)}$

$= Q(\theta, \theta^{(t)})$ $\quad \curvearrowleft \vec{\beta} \sim N(0, \sigma_z^2 I_p)$ $\quad x|\beta \sim N(\mu_x + Z_x \beta, \sigma_x^2 \mathbb{I}_{n_1})$

$y|\beta \sim N(\mu_y + Z_y \beta \alpha + Z_y \gamma \mathbb{1}, \sigma_y^2 \mathbb{I}_{n_2})$

$E_{\beta \sim N(\mu_B, \Sigma_B)}(\beta^T A \beta) = \mu_B^T A \mu_B + Tr(A \Sigma_B)$

M-step: Maximum $Q(\theta|\theta^{(t)}) \Rightarrow \theta^{(t+1)} = ?$

Reduction-step: $R(\lambda = 1)$

Testing likelihood $\int f(x, y, \beta) d\beta = \int f(y|\beta) f(x|\beta) f(\beta) d\beta$

LRT $H_0: \alpha = 0 \quad \Lambda_\alpha = 2[\log f(x, y|Z_x, Z_y, \hat{\theta}) - \log f(x, y|Z_x, Z_y, \hat{\theta}_{\alpha=0})]$

$H_0: \gamma = 0 \quad \Lambda_\gamma = 2[\log f(x, y|Z_x, Z_y, \hat{\theta}) - \log f(x, y|Z_x, Z_y, \hat{\theta}_{\gamma=0})]$