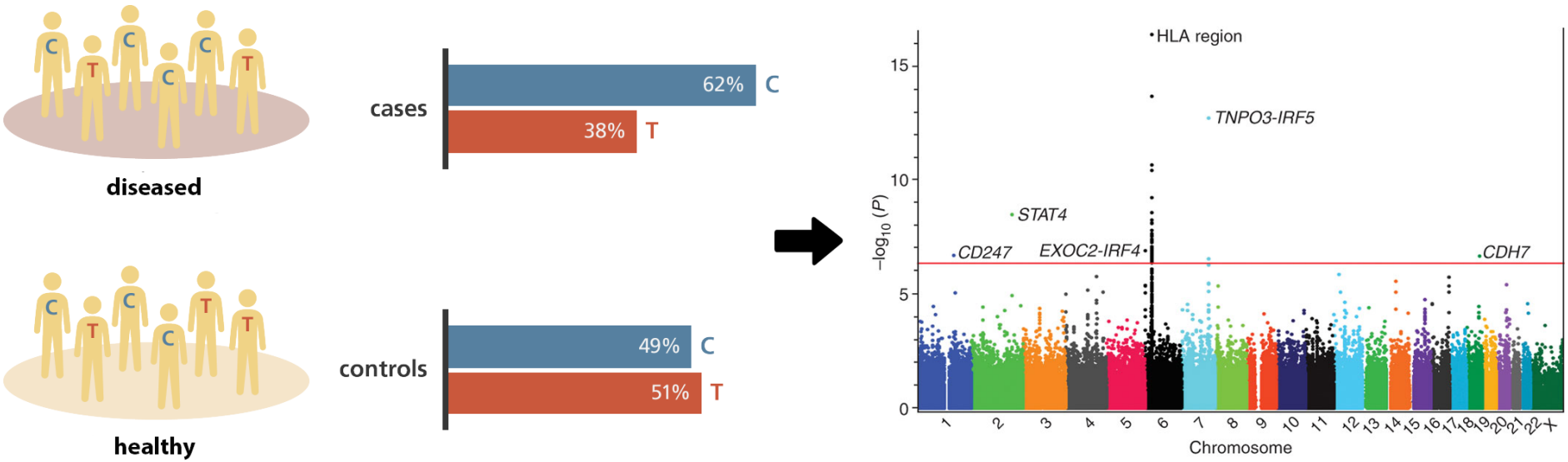
Bayesian variable selection logistic regression: multivariate metaanalysis in GWAS

Saikat Banerjee

Max Planck Institute for Biophysical Chemistry

FEBRUARY 16, 2017

Genome-wide association studies (GWAS)



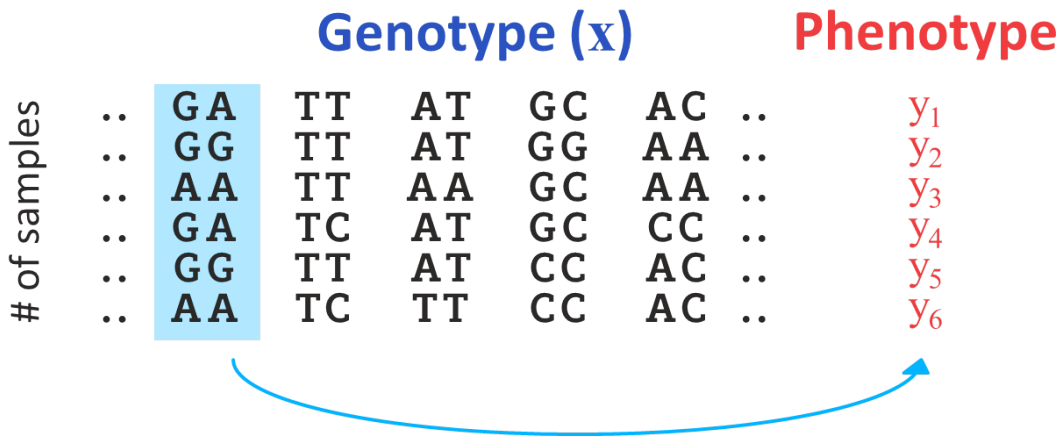
- Discovered thousands of variants associated with complex diseases

Radstake *et al.* , *Nat. Gen.* 2010

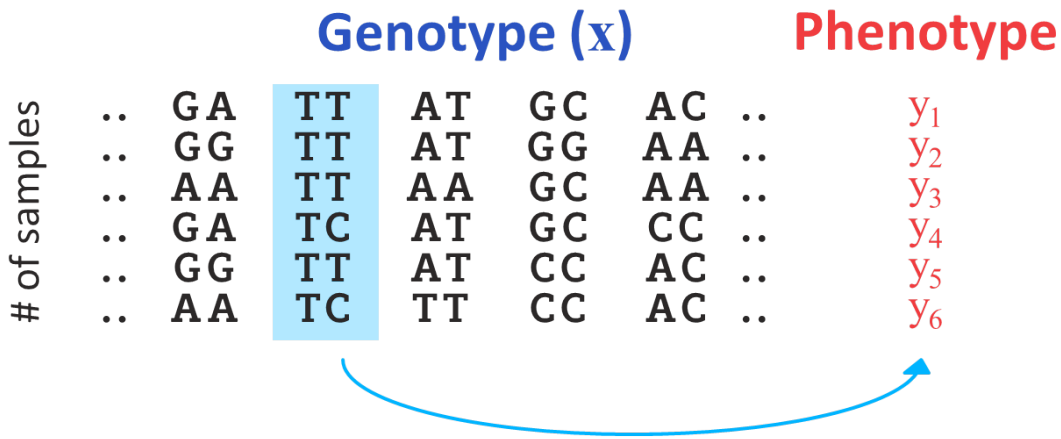
Association tests in GWAS

	Genotype (x)					Phenotype
# of samples	..	GA	TT	AT	GC	AC ..	y_1
	..	GG	TT	AT	GG	AA ..	y_2
	..	AA	TT	AA	GC	AA ..	y_3
	..	GA	TC	AT	GC	CC ..	y_4
	..	GG	TT	AT	CC	AC ..	y_5
	..	AA	TC	TT	CC	AC ..	y_6

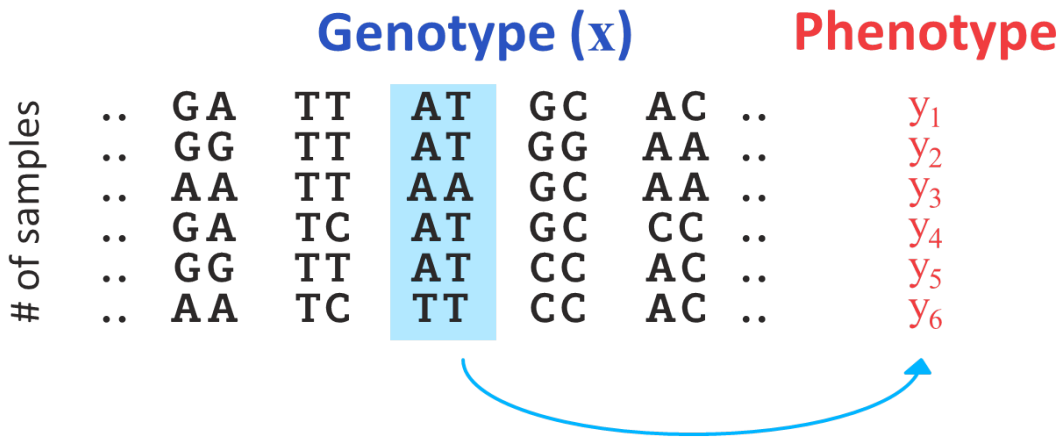
Association tests in GWAS



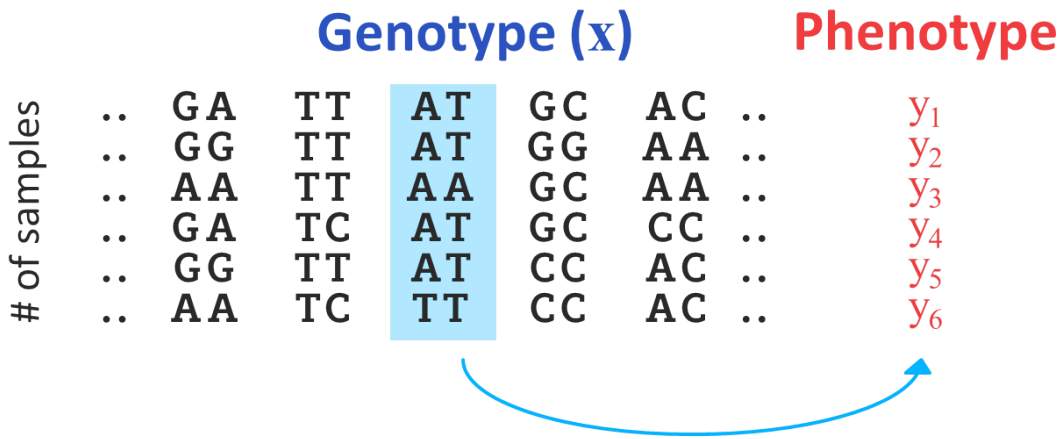
Association tests in GWAS



Association tests in GWAS

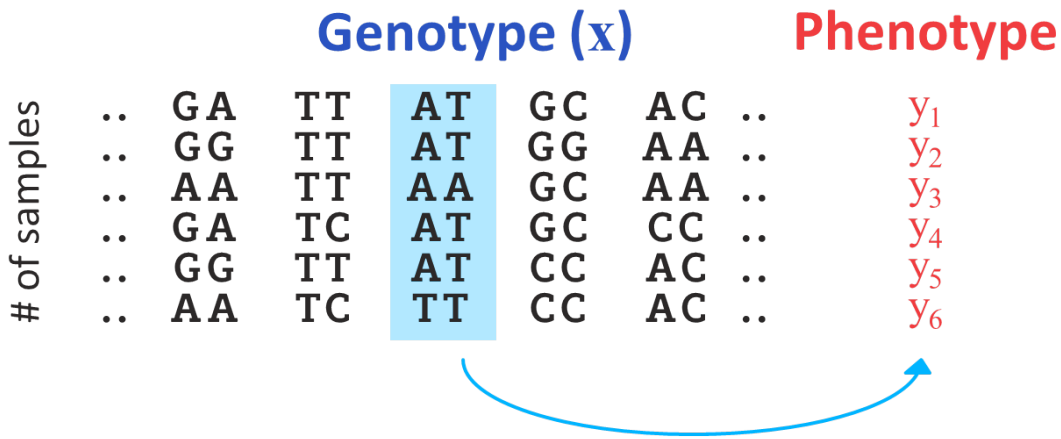


Association tests in GWAS



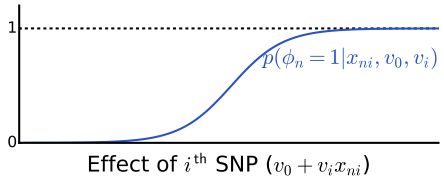
$$y_n = v_0 + v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$ Quantitative phenotype

Association tests in GWAS

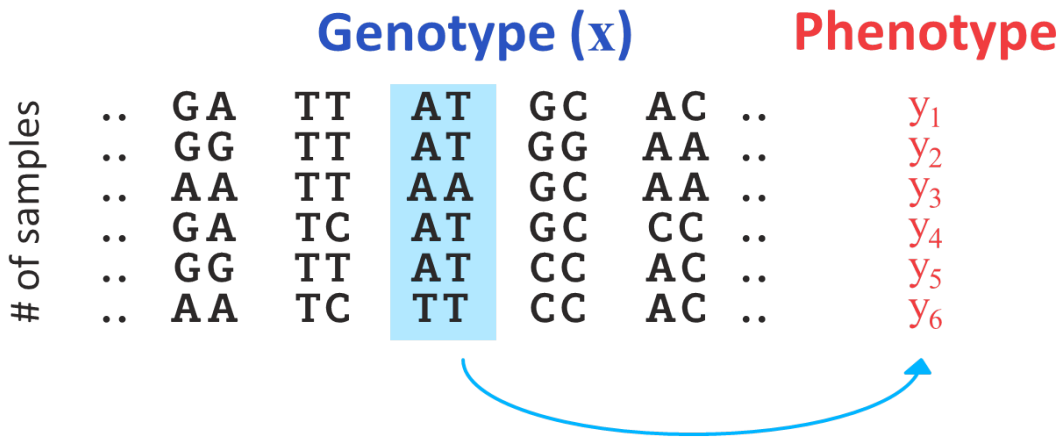


$$y_n = v_0 + v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$ Quantitative phenotype

$$p(y_n = 1 | x_{ni}, v_0, v_i) = \text{lf}(v_0 + v_i x_{ni})$$ Binary phenotype



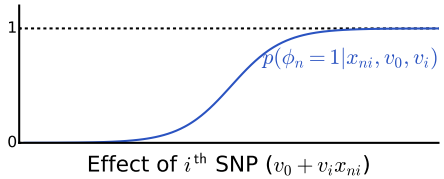
Association tests in GWAS



$$y_n = v_0 + v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$ Quantitative phenotype

$$p(y_n = 1 | x_{ni}, v_0, v_i) = \text{lf}(v_0 + v_i x_{ni})$$ Binary phenotype

- Is the coefficient v_i significantly different from 0? $\Rightarrow$ P-values



Univariate methods

Strengths

- Straightforward
- Computationally fast
- Conservative
- Easy to interpret

Univariate methods

Strengths

- Straightforward
- Computationally fast
- Conservative
- Easy to interpret

Challenges

- Linkage disequilibrium
- Genetic networks
- Low effect sizes

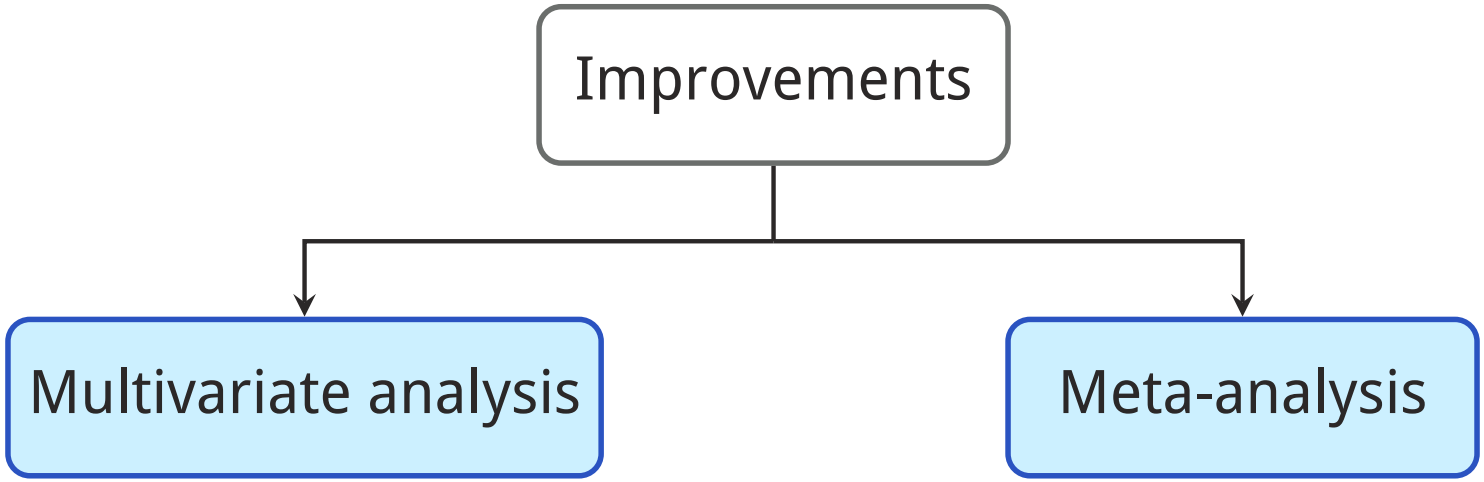
Univariate methods

Strengths

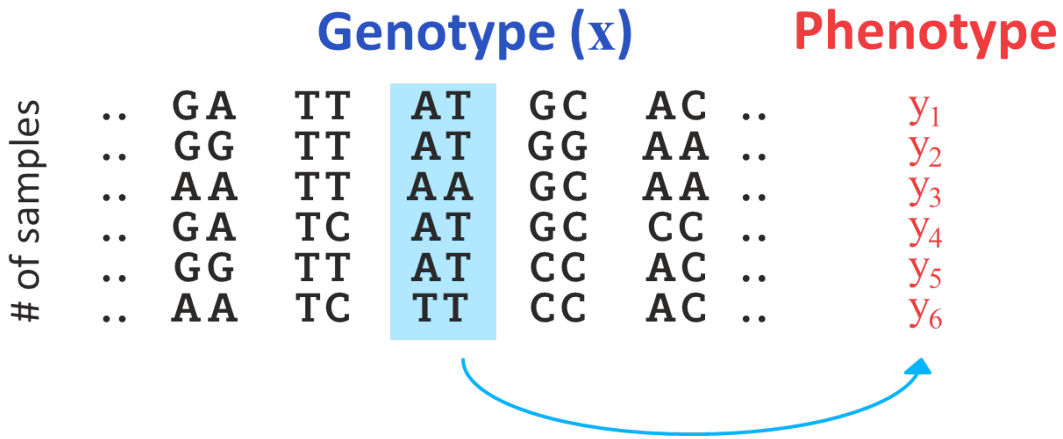
- Straightforward
- Computationally fast
- Conservative
- Easy to interpret

Challenges

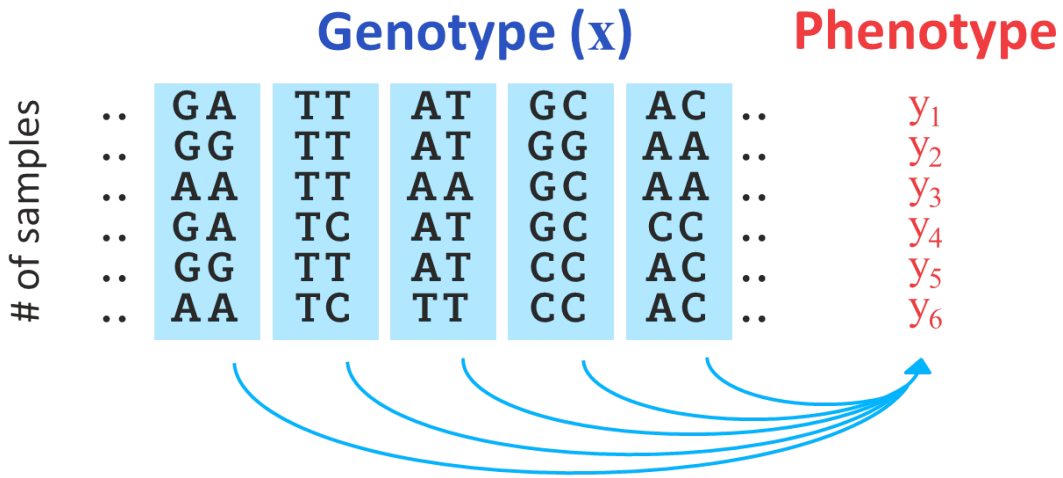
- Linkage disequilibrium
- Genetic networks
- Low effect sizes



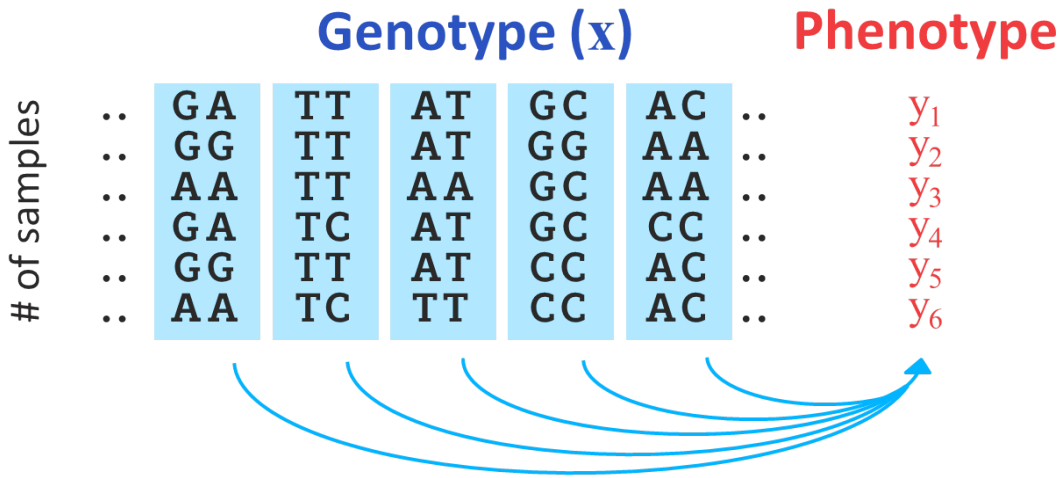
Multivariate methods



Multivariate methods



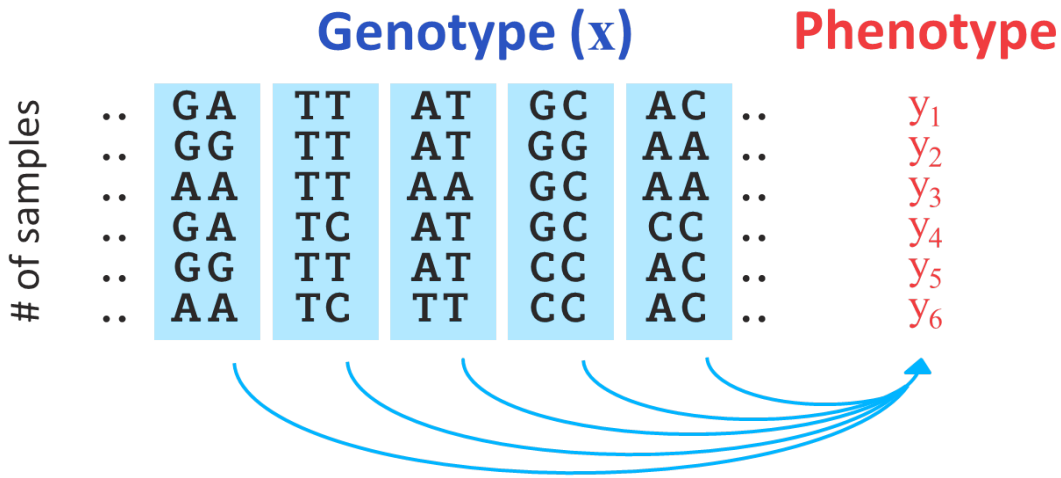
Multivariate methods



$$y_n = v_0 + \sum_i v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Quantitative phenotype

Multivariate methods

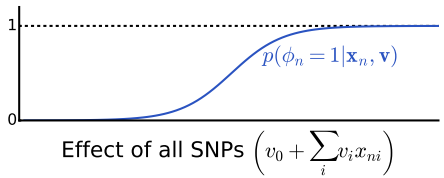


$$y_n = v_0 + \sum_i v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

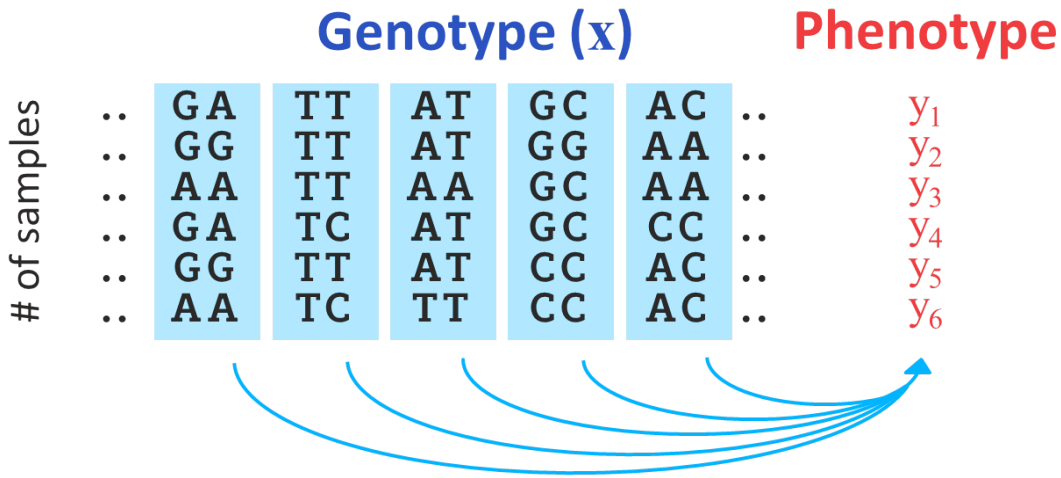
Quantitative phenotype

$$p(y_n = 1 | x_{ni}, v_0, v_i) = \mathrm{lf}\left(v_0 + \sum_i v_i x_{ni} \right)$$

Binary phenotype



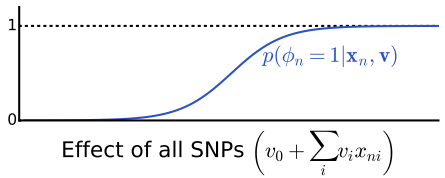
Multivariate methods



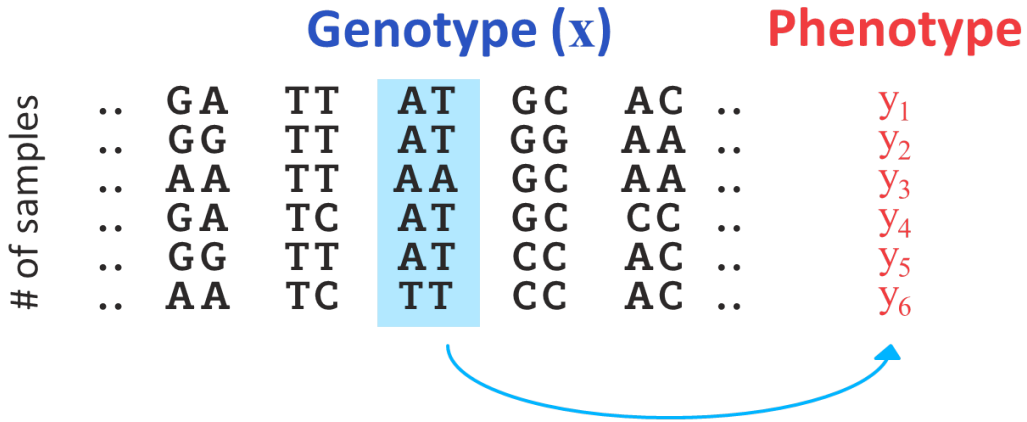
$$y_n = v_0 + \sum_i v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$ Quantitative phenotype

$$p(y_n = 1 | x_{ni}, v_0, v_i) = \mathrm{lf}\left(v_0 + \sum_i v_i x_{ni}\right)$$ Binary phenotype

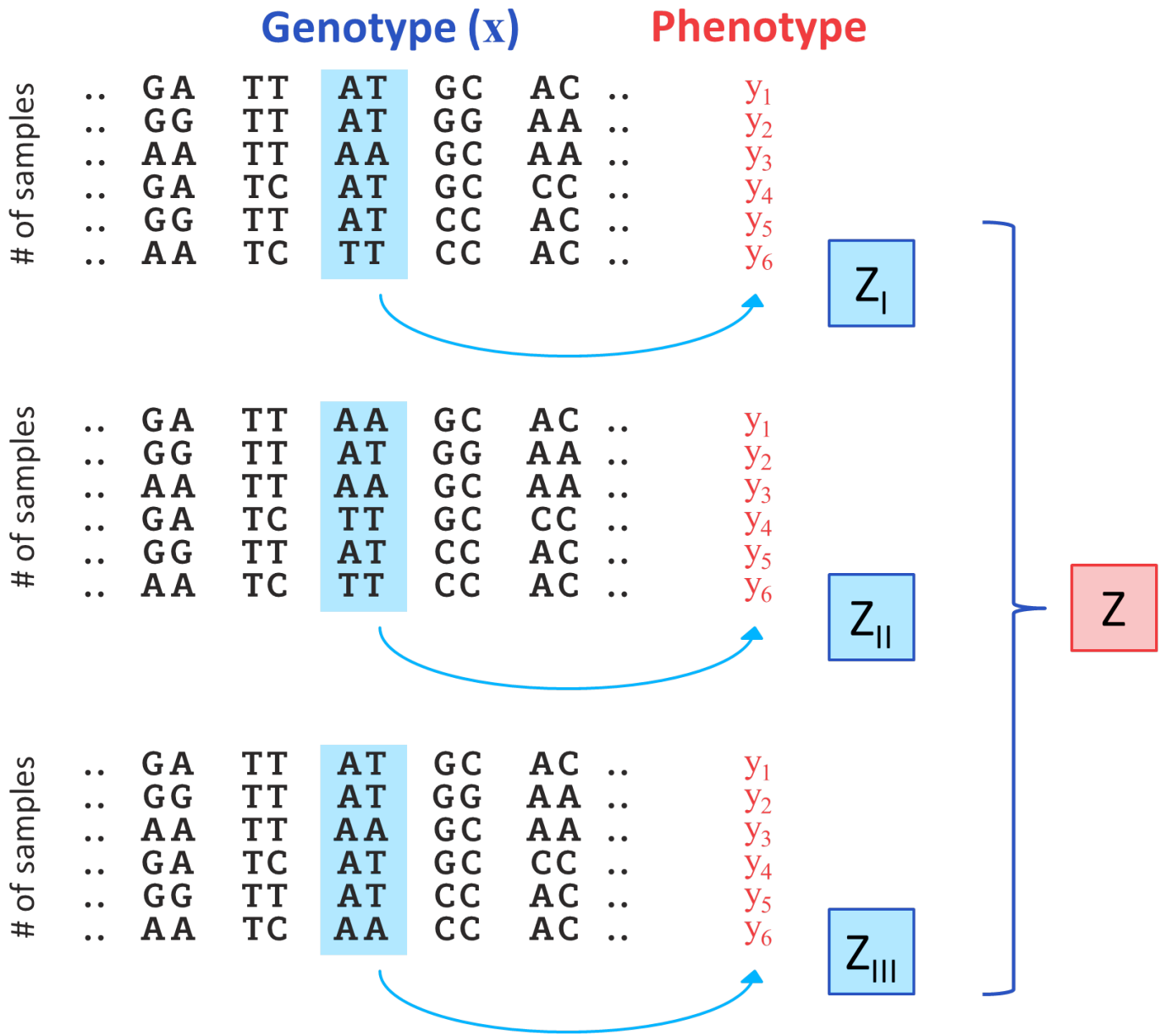
- Multivariate methods perform better than univariate methods



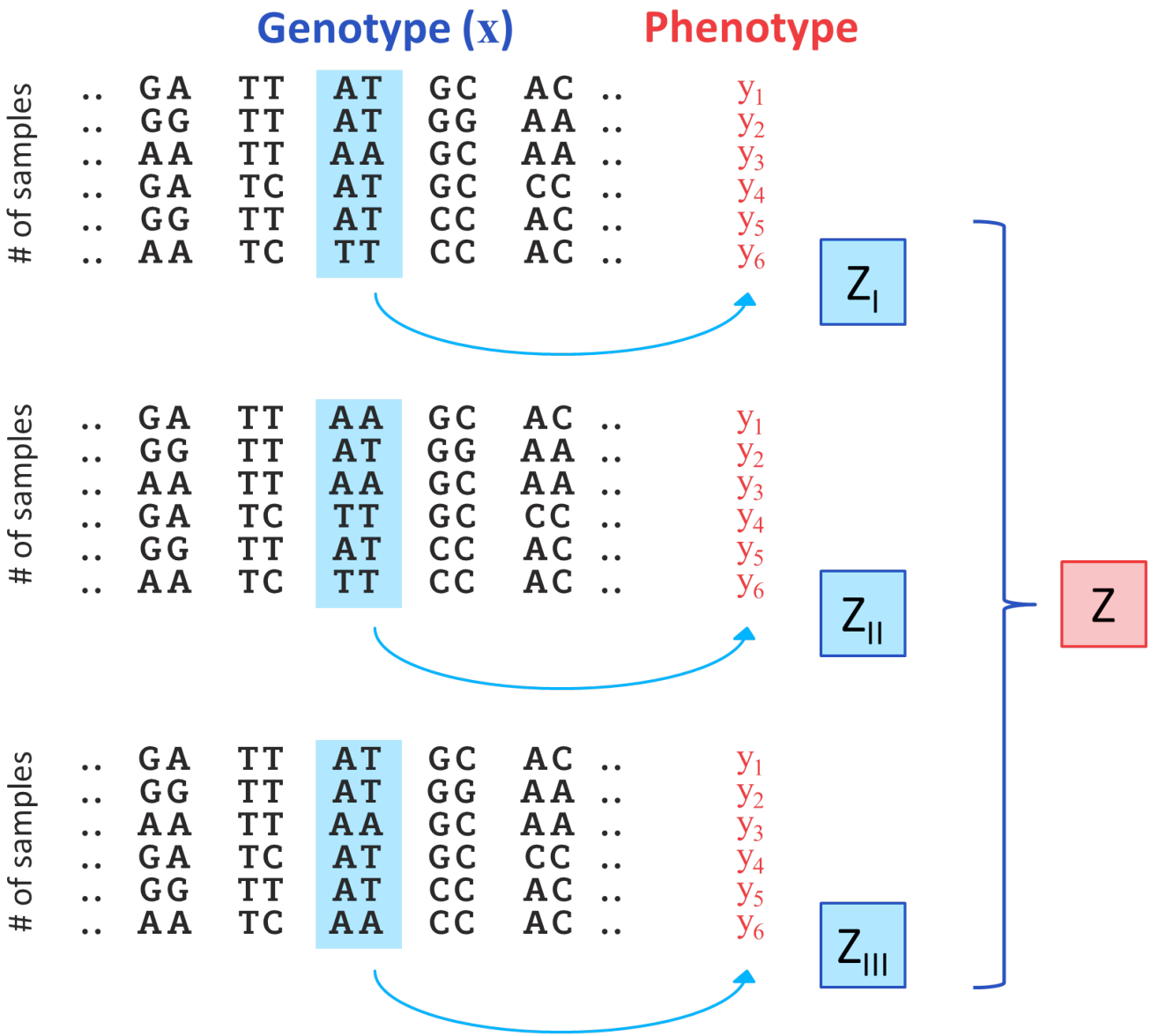
Meta-analysis



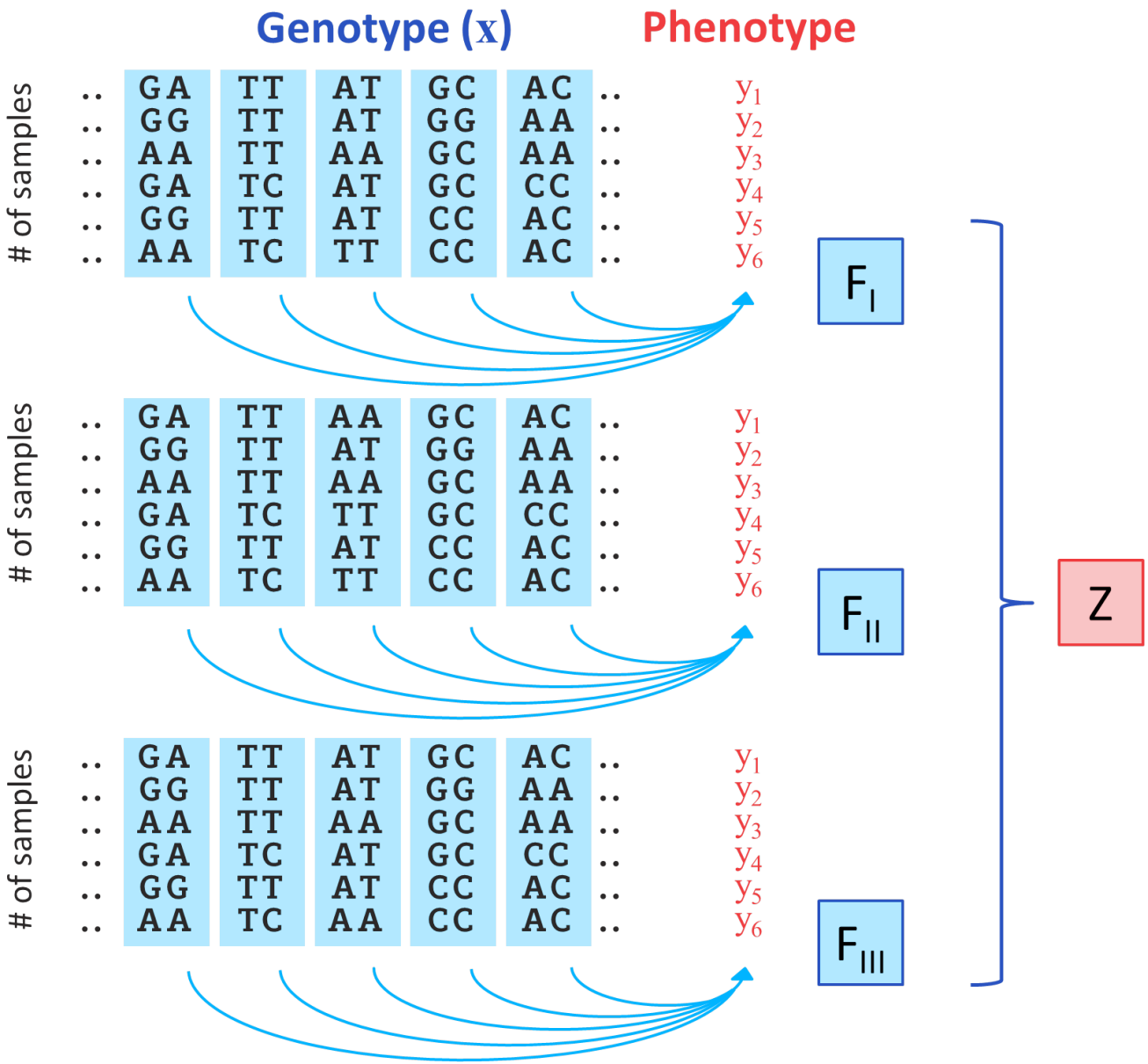
Meta-analysis



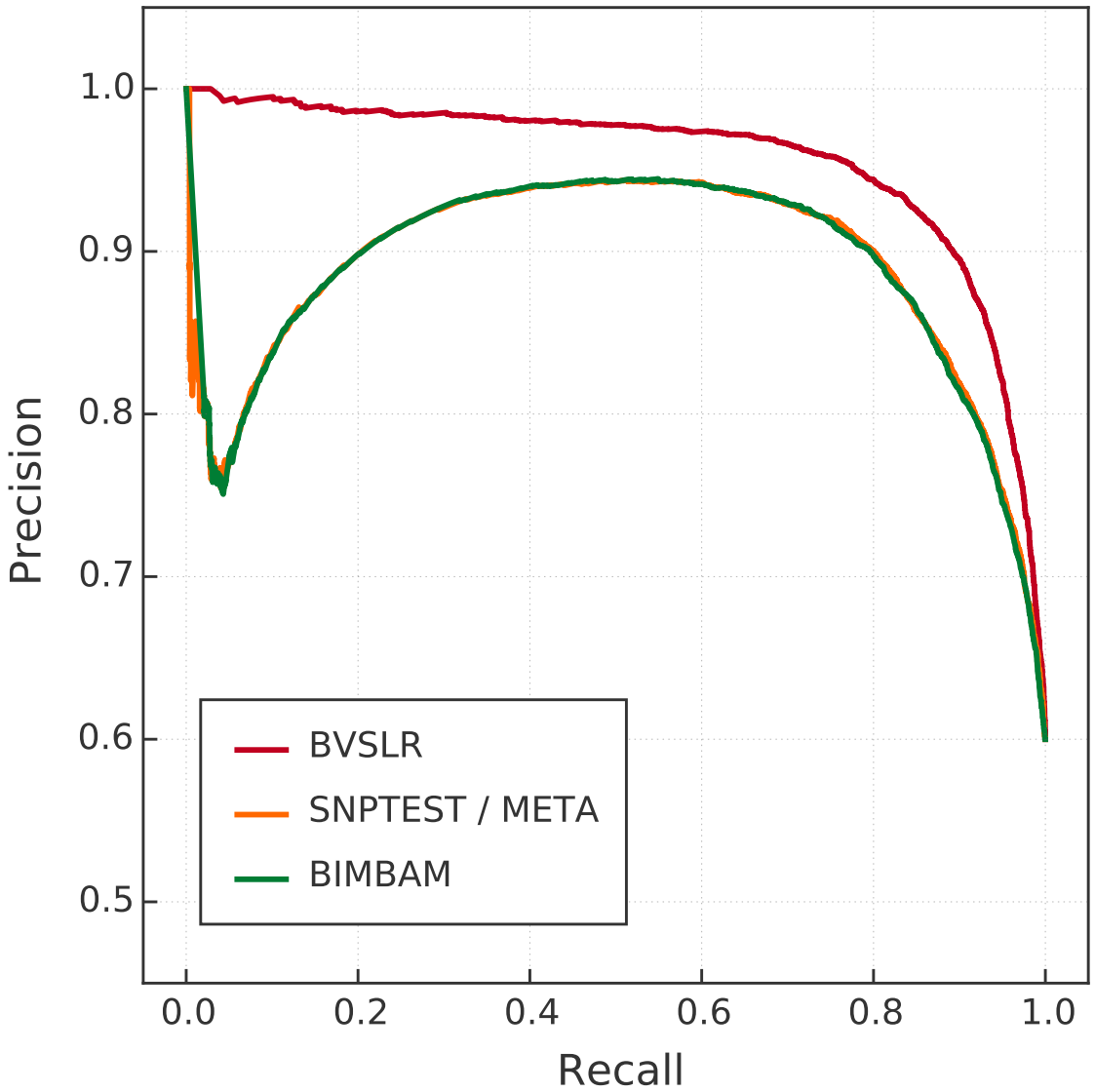
Goal of our method



Goal of our method



Preview of BVSLR



$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Bayesian variable selection regression (BVSR)

BIMBAM

Servin and Stephens, *PLoS Genetics* 2007
Guan and Stephens, *Ann. Appl. Stats.* 2011

$$y_n = v_0 + \sum_i v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \tau^{-1})$$

Quantitative phenotype

Bayesian variable selection regression (BVSR)

BIMBAM

Servin and Stephens, *PLoS Genetics* 2007
Guan and Stephens, *Ann. Appl. Stats.* 2011

$$y_n = v_0 + \sum_i v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \tau^{-1}) \qquad \text{Quantitative phenotype}$$

- Likelihood for N patients:

$$p\left(\mathbf{y} \mid \mathbf{x}, \mathbf{v}, \tau\right) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{x}^{\intercal}\mathbf{v}, \tau^{-1}\mathbb{I}\right)$$

Bayesian variable selection regression (BVSR)

BIMBAM

Servin and Stephens, *PLoS Genetics* 2007
Guan and Stephens, *Ann. Appl. Stats.* 2011

$$y_n = v_0 + \sum_i v_i x_{ni} + \epsilon, \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \tau^{-1}) \qquad \text{Quantitative phenotype}$$

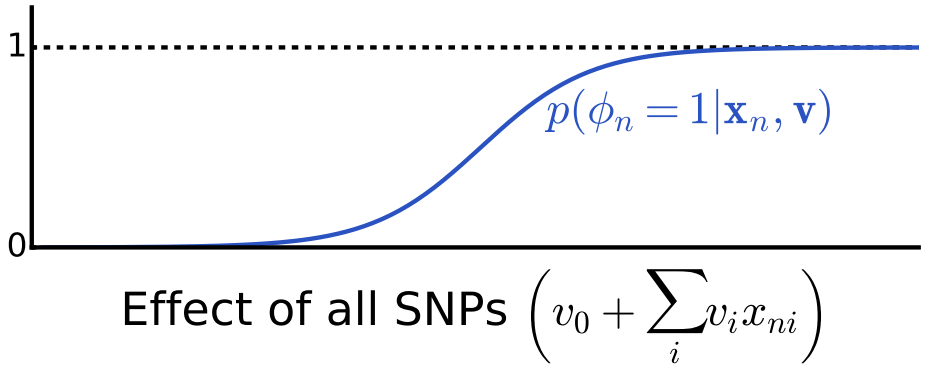
- Likelihood for N patients:

$$p\left(\mathbf{y} \mid \mathbf{x}, \mathbf{v}, \tau\right) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{x}^{\intercal}\mathbf{v}, \tau^{-1}\mathbb{I}\right)$$

- Number of SNPs $\gg$ samples $\rightarrow$ Overfitting
- Effective priors on $\mathbf{v}$ for sparsity

Bayesian variable selection logistic regression (BVSLR)

$$p\left(\phi_n = 1 | \mathbf{x}_n, \mathbf{v}\right) = \mathrm{lf}\left(v_0 + \sum_i v_i x_{ni}\right)$$

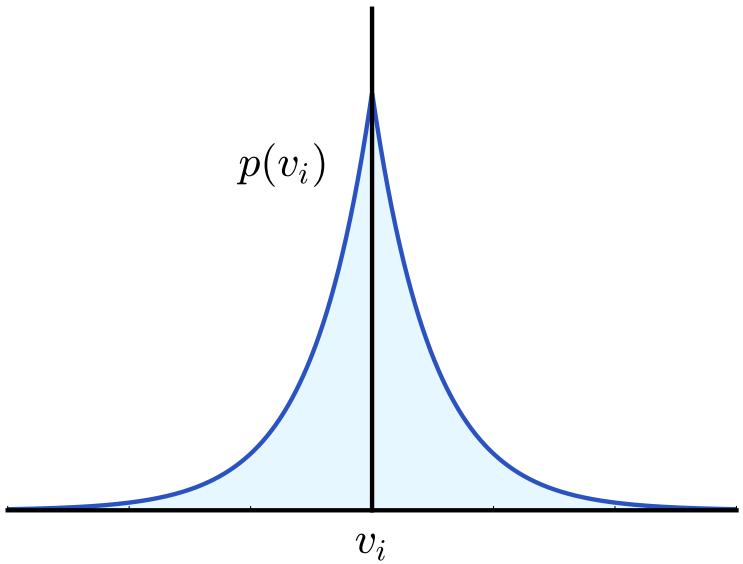


- Likelihood for N patients:

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) = \prod_{n=1}^{N} p\left(\phi_n | \mathbf{x}_n, \mathbf{v}\right) = \prod_{n=1}^{N} \frac{\exp(\phi_n \mathbf{v}^\intercal \mathbf{x}_n)}{1 + \exp(\mathbf{v}^\intercal \mathbf{x}_n)}$$

LASSO penalisation

- Constraint: $\sum_i |v_i| \leq t$, where $t(>0)$ is a *tuning parameter*.
- Applied as a Lagrangian penalty to the joint log-likelihood.



LASSO penalty assumes implicit Laplace prior → more zero-valued coefficients.

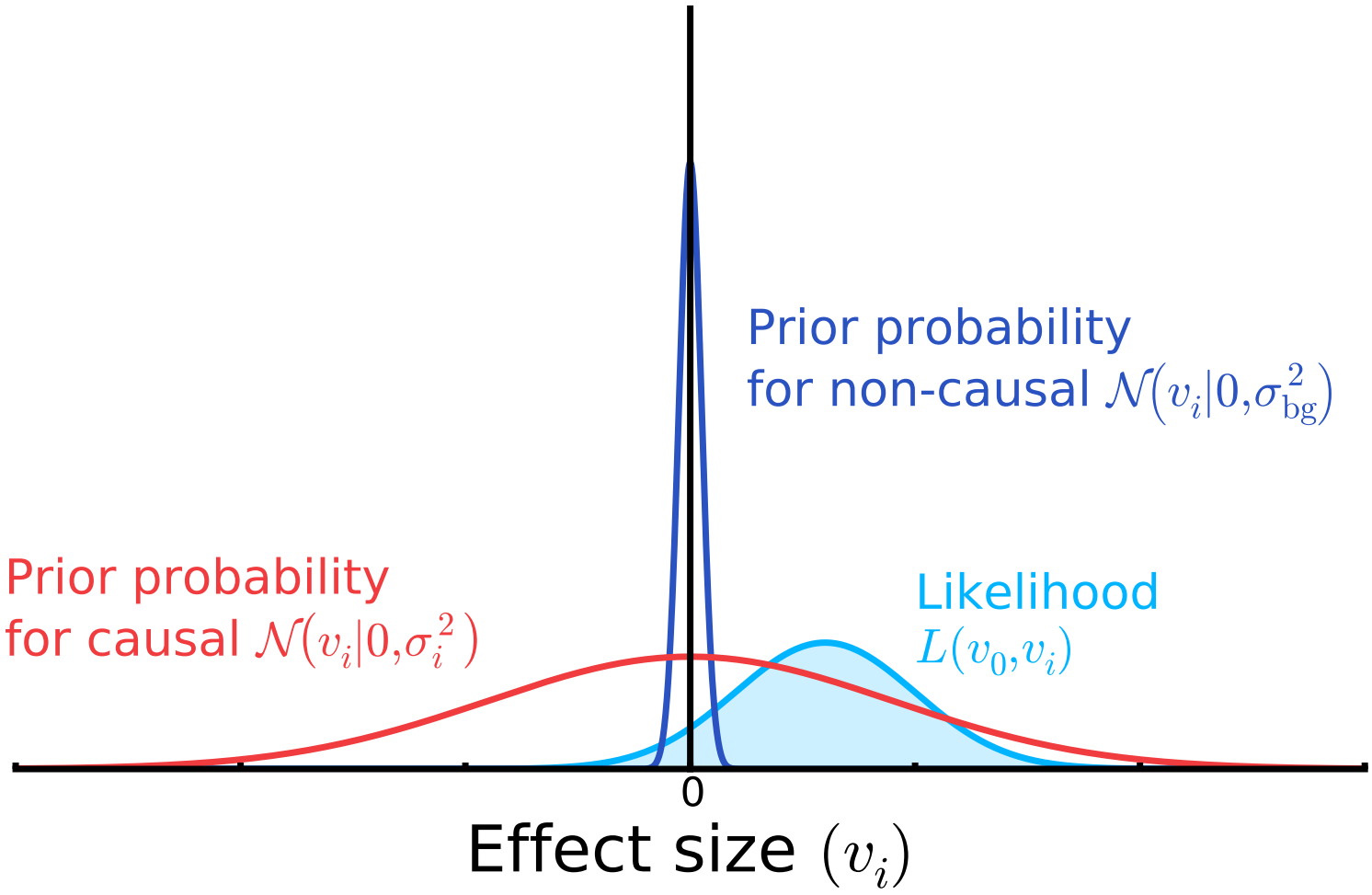
Problems:

- Multicollinearity
- Variability in penalty parameter

Tibshirani, 1996

Sparsity in BVSR / BVSLR

Looks at both null hypothesis and alternate hypothesis



Priors in BVSR

$$p\left(\mathbf{y} \mid \mathbf{x}, \mathbf{v}, \tau\right) = \mathcal{N}\left(\mathbf{y} \mid \mathbf{x}^{\intercal}\mathbf{v}, \tau^{-1}\mathbb{I}\right)$$

$$p(\tau) = \text{Gamma}\left(\frac{\lambda}{2}, \frac{\kappa}{2}\right)$$

$$p(v_0 \mid \tau) = \mathcal{N}\left(v_0 \mid 0, \sigma_\mu^2/\tau\right)$$

$$p(v_i \mid \tau) = \overbrace{(1-\pi)\delta_0}^{\text{Non-causal}} + \underbrace{\pi\,\mathcal{N}\left(v_i \middle| 0, \sigma_a^2/\tau\right)}_{\text{Causal}}$$

Hyperparameters $\Rightarrow\ \lambda, \kappa, \pi, \sigma_\mu, \sigma_a$

Priors in BVSLR

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) = \prod_{n=1}^{N} \frac{\exp(\phi_n \mathbf{v}^\top \mathbf{x}_n)}{1 + \exp(\mathbf{v}^\top \mathbf{x}_n)}$$

$$p\left(v_i \mid \boldsymbol{\theta}\right) = \overbrace{(1-\pi)\mathcal{N}\left(v_i \mid 0, \sigma^2_{\text{bg}}\right)}^{\text{Non-causal}} + \underbrace{\pi\mathcal{N}\left(v_i \mid \mu, \sigma^2\right)}_{\text{Causal}}$$

Hyperparameters $\boldsymbol{\theta} \Rightarrow \pi, \mu, \sigma_{bg}, \sigma$

Causality configurations

$$p(v_i \mid \boldsymbol{\theta}) = \overbrace{(1-\pi)\mathcal{N}\left(v_i \mid 0, \sigma^2_{\text{bg}}\right)}^{\text{Non-causal}} + \underbrace{\pi\mathcal{N}\left(v_i \mid \mu, \sigma^2\right)}_{\text{Causal}}$$

$$= \sum_{z_i=0,1} \pi^{z_i}(1-\pi)^{(1-z_i)}\mathcal{N}\left(v_i \mid \mu_{\mathbf{z},i}, \sigma^2_{\mathbf{z},i}\right)$$

$$\mu_{\mathbf{z},i} = z_i\mu \quad \text{and} \quad \sigma^2_{\mathbf{z},i} = \sigma^2_{\text{bg}} + z_i\left[\sigma^2 - \sigma^2_{\text{bg}}\right]$$

$\mathbf{z} \in \{0,1\}^I \Rightarrow$ *Causality configurations*

- $z_i = 1$ *SNP i is causal*
- $z_i = 0$ *SNP i is non-causal*

Optimising the hyperparameters

$$p(v_i \mid \boldsymbol{\theta}) = \sum_{z_i=0,1} \pi^{z_i} (1-\pi)^{(1-z_i)} \mathcal{N}\left(v_i \mid \mu_{\mathbf{z},i}, \sigma^2_{\mathbf{z},i}\right)$$

Optimising the hyperparameters

$$p(v_i \mid \boldsymbol{\theta}) = \sum_{z_i=0,1} \pi^{z_i} (1-\pi)^{(1-z_i)} \mathcal{N}\left(v_i \mid \mu_{\mathbf{z},i}, \sigma^2_{\mathbf{z},i}\right)$$

$$p(\mathbf{v} \mid \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \mathrm{diag}(\boldsymbol{\sigma}^2_{\mathbf{z}})\right)$$

Optimising the hyperparameters

$$p(\mathbf{v} \mid \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \mathrm{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)$$

Optimising the hyperparameters

$$p(\mathbf{v} \mid \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)$$

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) = \prod_{n=1}^{N} \frac{\exp(\phi_n \mathbf{v}^\top \mathbf{x}_n)}{1 + \exp(\mathbf{v}^\top \mathbf{x}_n)}$$

Optimising the hyperparameters

$$p(\mathbf{v} \mid \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)$$

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) = \prod_{n=1}^{N} \frac{\exp(\phi_n \mathbf{v}^\top \mathbf{x}_n)}{1 + \exp(\mathbf{v}^\top \mathbf{x}_n)}$$

Evidence approximation: Maximising the marginal likelihood

$$\begin{aligned} mL(\boldsymbol{\theta}) &= p\left(\boldsymbol{\phi} \mid \mathbf{x}, \boldsymbol{\theta}\right) \\ &= \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) p(\mathbf{v} \mid \boldsymbol{\theta}) d\mathbf{v} \rightarrow \max \end{aligned}$$

$$mL(\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right) d\mathbf{v}$$

Optimising the hyperparameters

$$p(\mathbf{v} \mid \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)$$

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) = \prod_{n=1}^{N} \frac{\exp(\phi_n \mathbf{v}^\top \mathbf{x}_n)}{1 + \exp(\mathbf{v}^\top \mathbf{x}_n)}$$

Evidence approximation: Maximising the marginal likelihood

$$\begin{aligned} mL(\boldsymbol{\theta}) &= p\left(\boldsymbol{\phi} \mid \mathbf{x}, \boldsymbol{\theta}\right) \\ &= \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) p(\mathbf{v} \mid \boldsymbol{\theta}) d\mathbf{v} \rightarrow \max \end{aligned}$$

$$mL(\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right) d\mathbf{v}$$

Laplace approximation (?)

Optimising the hyperparameters

$$p(\mathbf{v} \mid \boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)$$

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) = \prod_{n=1}^{N} \frac{\exp(\phi_n \mathbf{v}^\top \mathbf{x}_n)}{1 + \exp(\mathbf{v}^\top \mathbf{x}_n)}$$

Evidence approximation: Maximising the marginal likelihood

$$\begin{aligned} mL(\boldsymbol{\theta}) &= p\left(\boldsymbol{\phi} \mid \mathbf{x}, \boldsymbol{\theta}\right) \\ &= \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) p(\mathbf{v} \mid \boldsymbol{\theta}) d\mathbf{v} \rightarrow \max \end{aligned}$$

$$mL(\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right) d\mathbf{v}$$

~~Laplace approximation (?)~~

Quasi-Laplace approximation

$$mL(\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right) d\mathbf{v}$$

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)$$

$$= \underbrace{p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}, \operatorname{diag}(\tilde{\boldsymbol{\sigma}}^2)\right)}_{\propto \mathcal{N}\left(\mathbf{v} \mid \hat{\mathbf{v}}, \hat{\boldsymbol{\Lambda}}^{-1}\right)} \times \frac{\mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)}{\mathcal{N}\left(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}, \operatorname{diag}(\tilde{\boldsymbol{\sigma}}^2)\right)}$$

Quasi-Laplace approximation

$$mL(\boldsymbol{\theta}) = \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \int p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right) d\mathbf{v}$$

$$p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)$$

$$= \underbrace{p\left(\boldsymbol{\phi} \mid \mathbf{x}, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}, \operatorname{diag}(\tilde{\boldsymbol{\sigma}}^2)\right)}_{\propto \mathcal{N}\left(\mathbf{v} \mid \tilde{\mathbf{v}}, \tilde{\boldsymbol{\Lambda}}^{-1}\right)} \times \frac{\mathcal{N}\left(\mathbf{v} \mid \boldsymbol{\mu}_{\mathbf{z}}, \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{z}}^2)\right)}{\mathcal{N}\left(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}, \operatorname{diag}(\tilde{\boldsymbol{\sigma}}^2)\right)}$$

Analytical solution

$$mL(\boldsymbol{\theta}) = p\left(\boldsymbol{\phi} \mid \mathbf{x}, \boldsymbol{\theta}\right)$$

$$\approx D' \sum_{\mathbf{z}} p(\mathbf{z} \mid \boldsymbol{\theta}) \frac{|\mathrm{diag}(\boldsymbol{\lambda}_{\mathbf{z}})|^{\frac{1}{2}}}{|\boldsymbol{\Lambda}_{\mathbf{z}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{\mu}_{\mathbf{z}}^{\intercal}\,\mathrm{diag}(\boldsymbol{\lambda}_{\mathbf{z}})\boldsymbol{\mu}_{\mathbf{z}} + \frac{1}{2}\mathbf{v}_{\mathbf{z}}^{\intercal}\boldsymbol{\Lambda}_{\mathbf{z}}\mathbf{v}_{\mathbf{z}}\right)$$

where

$$\boldsymbol{\Lambda}_{\mathbf{z}} := \tilde{\boldsymbol{\Lambda}} - \mathrm{diag}(\tilde{\boldsymbol{\lambda}}) + \mathrm{diag}(\boldsymbol{\lambda}_{\mathbf{z}})$$

$$\mathbf{v}_{\mathbf{z}} := \boldsymbol{\Lambda}_{\mathbf{z}}^{-1}\left[\tilde{\boldsymbol{\Lambda}}\tilde{\mathbf{v}} + \mathrm{diag}(\boldsymbol{\lambda}_{\mathbf{z}})\boldsymbol{\mu}_{\mathbf{z}} - \mathrm{diag}(\tilde{\boldsymbol{\lambda}})\tilde{\boldsymbol{\mu}}\right]$$

- Optimization can be done by gradient descent methods (*e.g.* L-BFGS).

Inference of causality in BVSLR

Using the definition of conditional probability,

$$p(\mathbf{z} \mid \boldsymbol{\phi}, \mathbf{x}, \boldsymbol{\theta}) = \frac{p(\boldsymbol{\phi}, \mathbf{z} \mid \mathbf{x}, \boldsymbol{\theta})}{p(\boldsymbol{\phi} \mid \mathbf{x}, \boldsymbol{\theta})}$$

The posterior probability for SNP i to be causal is

$$p(z_i = 1 \mid \boldsymbol{\phi}, \mathbf{x}, \theta) = \sum_{\mathbf{z}: z_i = 1} p(\mathbf{z} \mid \boldsymbol{\phi}, \mathbf{x}, \boldsymbol{\theta})$$

Inference of causality in BVSLR

Using the definition of conditional probability,

$$p(\mathbf{z} \mid \boldsymbol{\phi}, \mathbf{x}, \boldsymbol{\theta}) = \frac{p(\boldsymbol{\phi}, \mathbf{z} \mid \mathbf{x}, \boldsymbol{\theta})}{p(\boldsymbol{\phi} \mid \mathbf{x}, \boldsymbol{\theta})}$$

The posterior probability for SNP i to be causal is

$$p(z_i = 1 \mid \boldsymbol{\phi}, \mathbf{x}, \theta) = \sum_{\mathbf{z}: z_i = 1} p(\mathbf{z} \mid \boldsymbol{\phi}, \mathbf{x}, \boldsymbol{\theta})$$

The probability for a locus to be causal $= 1 -$ probability of *not* having a single causal SNP

$$p\left(\text{locus is causal} \mid \boldsymbol{\phi}, \mathbf{x}, \boldsymbol{\theta}\right) = 1 - p\left(\mathbf{z} = \mathbf{0} \mid \boldsymbol{\phi}, \mathbf{x}, \boldsymbol{\theta}\right)$$

BVSLR can be extended to multiple studies

$$
\begin{aligned}
mL(\boldsymbol{\theta}) &= p\left(\boldsymbol{\phi} \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_S, \boldsymbol{\theta}\right) \\
&= \int p\left(\boldsymbol{\phi} \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_S, \mathbf{v}\right) p\left(\mathbf{v} \mid \boldsymbol{\theta}\right) d\mathbf{v} \rightarrow \max
\end{aligned}
$$

Assuming the quasi-Laplace approximation holds for each study,

$$
\begin{aligned}
\prod_{s=1}^{S} \left[p\left(\boldsymbol{\phi} \mid \mathbf{x}_s, \mathbf{v}\right) \mathcal{N}\left(\mathbf{v} \mid \tilde{\boldsymbol{\mu}}_{\mathbf{z},s}, \operatorname{diag}(\tilde{\boldsymbol{\sigma}}^2_{\mathbf{z},s})\right)\right] &\propto \prod_{s=1}^{S} \mathcal{N}\left(\mathbf{v} \mid \tilde{\mathbf{v}}_s, \tilde{\boldsymbol{\Lambda}}_s^{-1}\right) \\
&= \mathcal{N}\left(\mathbf{v} \mid \tilde{\mathbf{v}}, \tilde{\boldsymbol{\Lambda}}^{-1}\right)
\end{aligned}
$$

where $\tilde{\boldsymbol{\Lambda}} = \sum_{s=1}^{S} \tilde{\boldsymbol{\Lambda}}_s$ and $\tilde{\mathbf{v}} = \tilde{\boldsymbol{\Lambda}}^{-1} \sum_{s=1}^{S} \tilde{\boldsymbol{\Lambda}}_s \tilde{\mathbf{v}}_s$

Simulation details

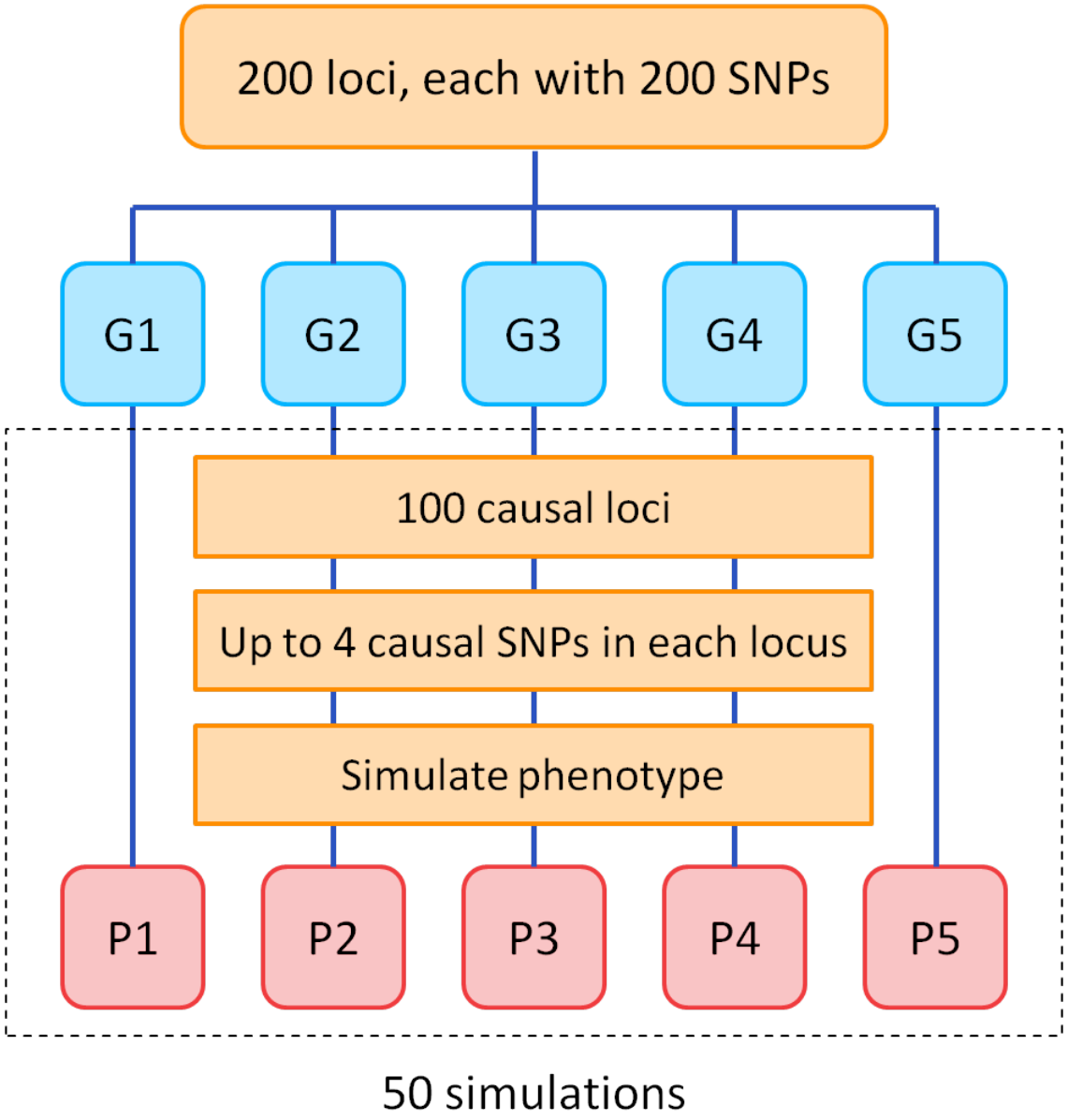
- Genotype: German Myocardial Infarction Family Study (GERMIFS)
- Five cohorts : G1, G2, G3, G4, G5
- Phenotype simulation:

$$\text{Disease liability} \quad Y_n = \sum_i v_i x_{ni} + \varepsilon_n$$

$$\text{Var}\left(\sum_i v_i x_{ni}\right) = h_g^2 = 0.4 \text{ and } \text{Var}(\varepsilon) = 0.6$$

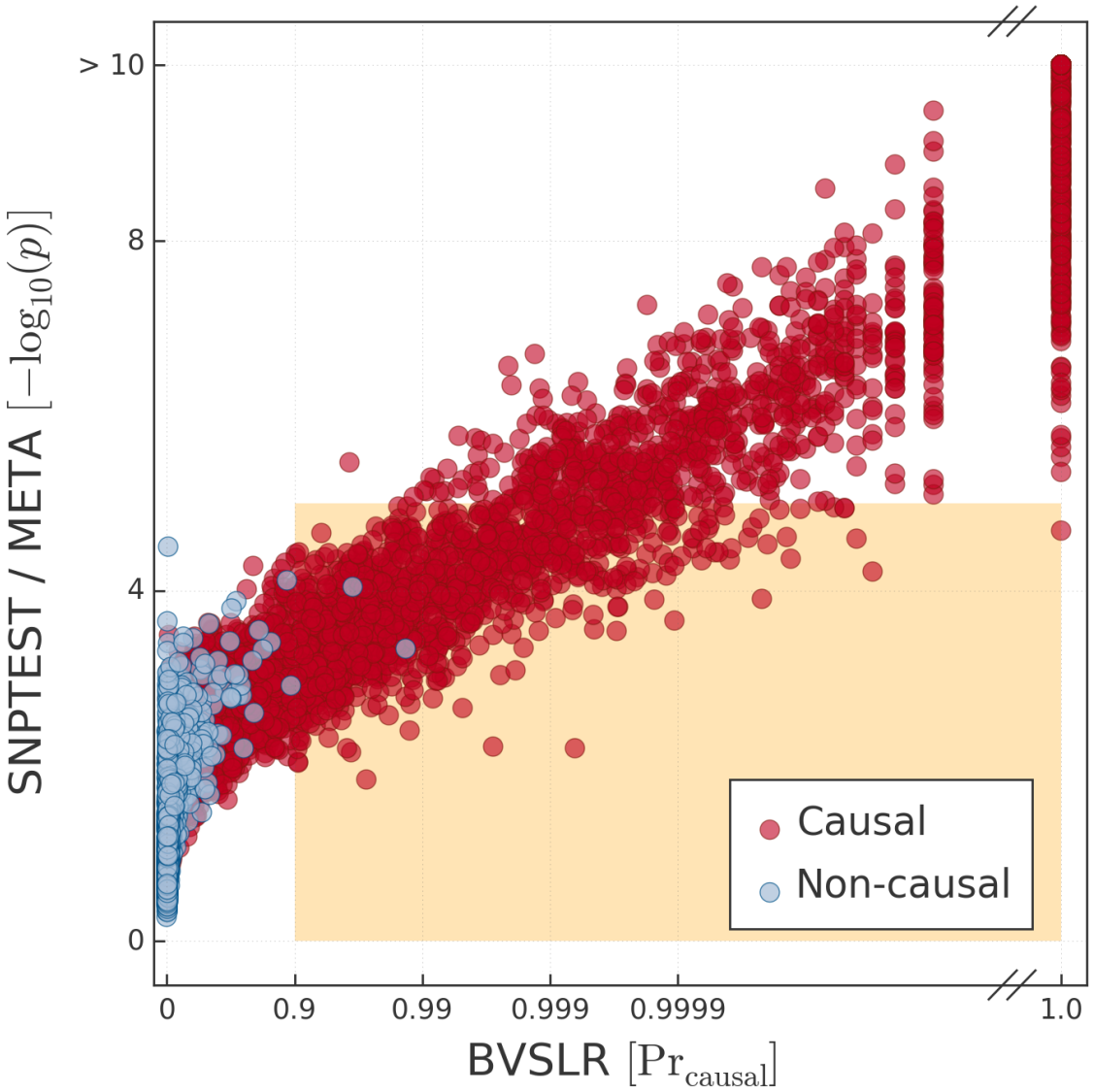
Cases are sampled from Y_n exceeding the threshold of normal distribution truncating the proportion of k (disease prevalence)

Simulation details



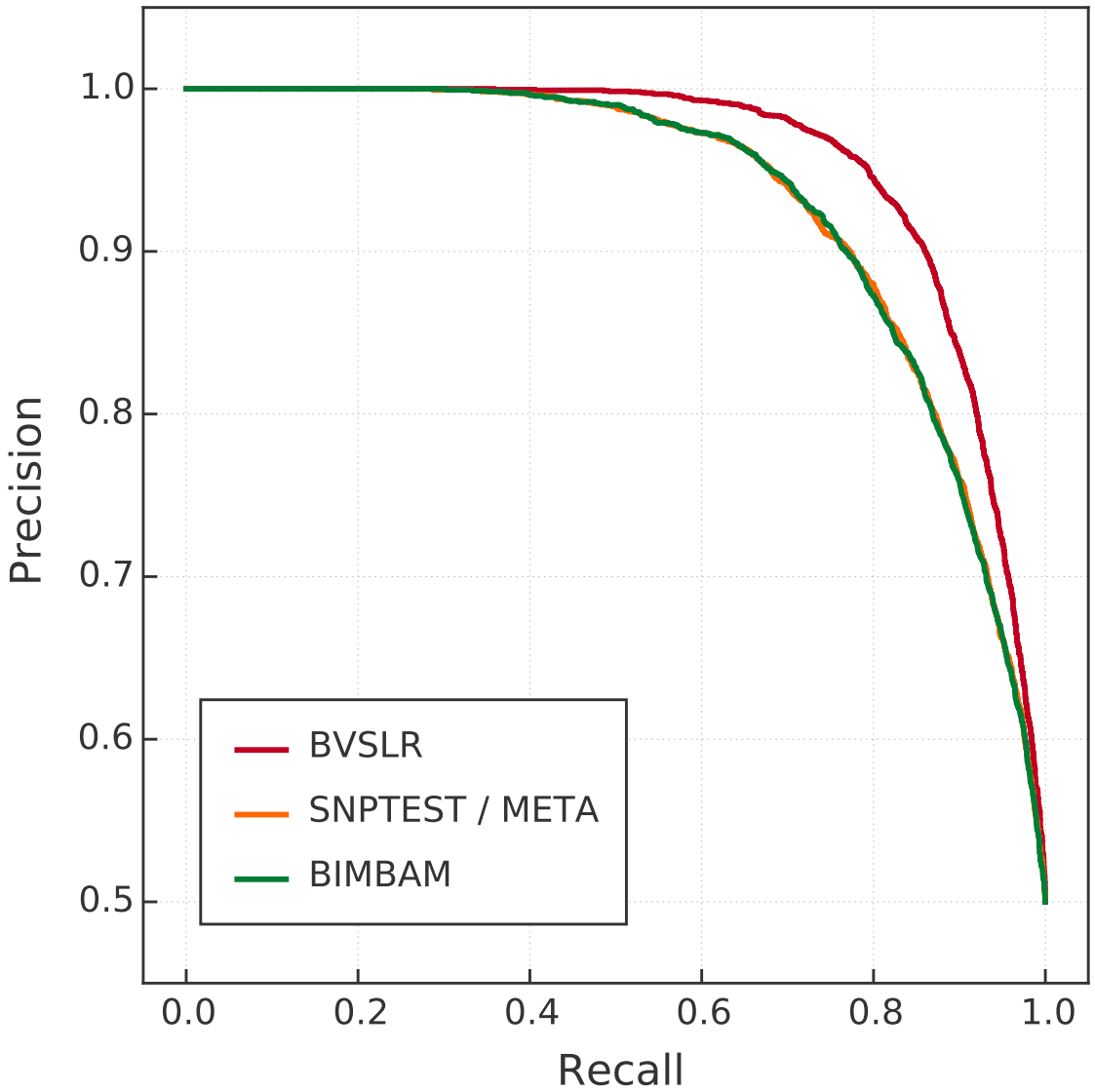
- BVSLR
- BIMBAM
- SNPTEST / META
- PAINTOR

Prediction of causal loci



- Shaded region shows improvement with BVSLR
- 5000 causal loci (100 from each of the 50 simulations)
- 5000 non-causal loci (100 from each of the 50 simulations)

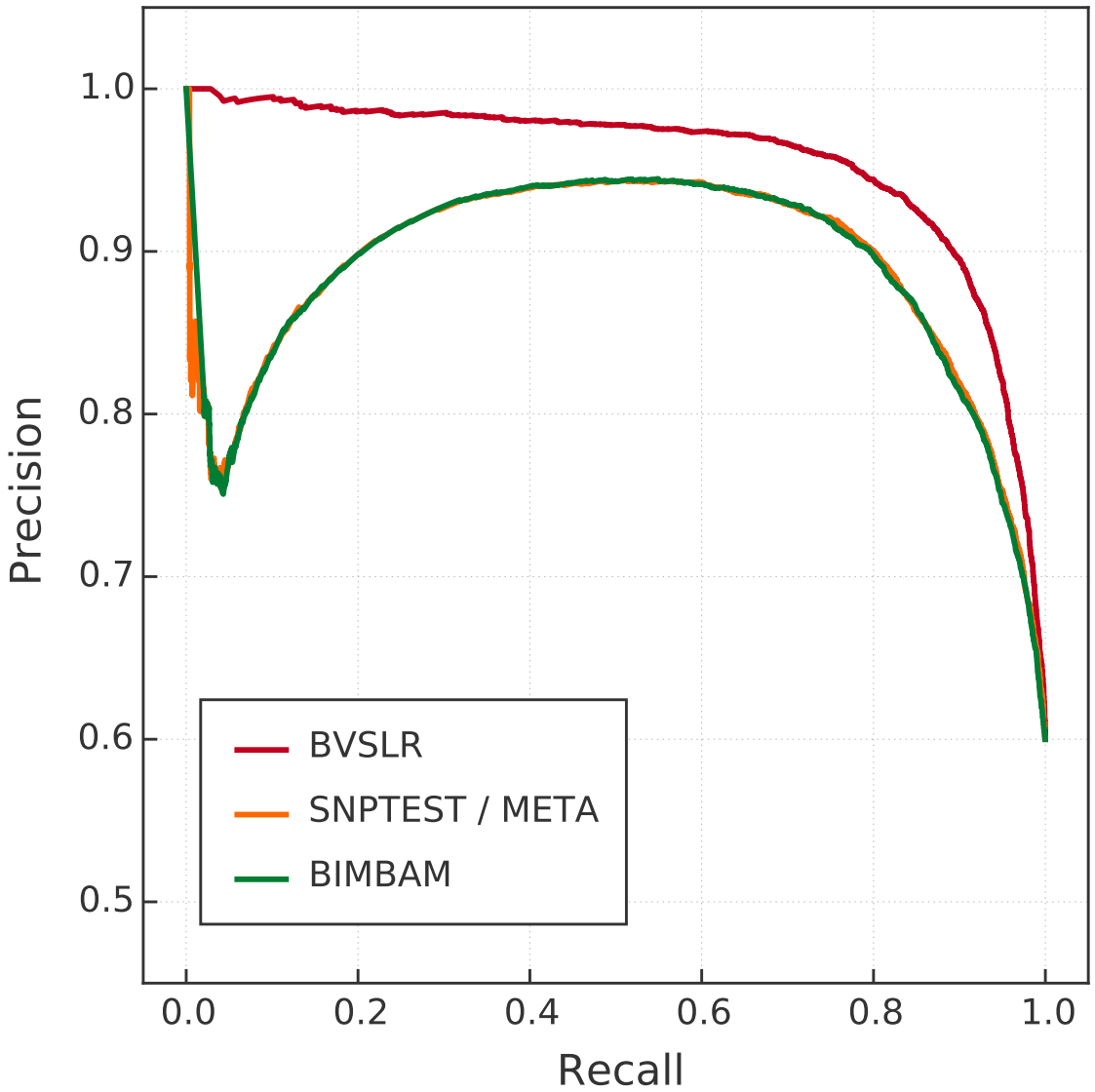
Prediction of causal loci



$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

If there are non-causal loci in LD with causal regions

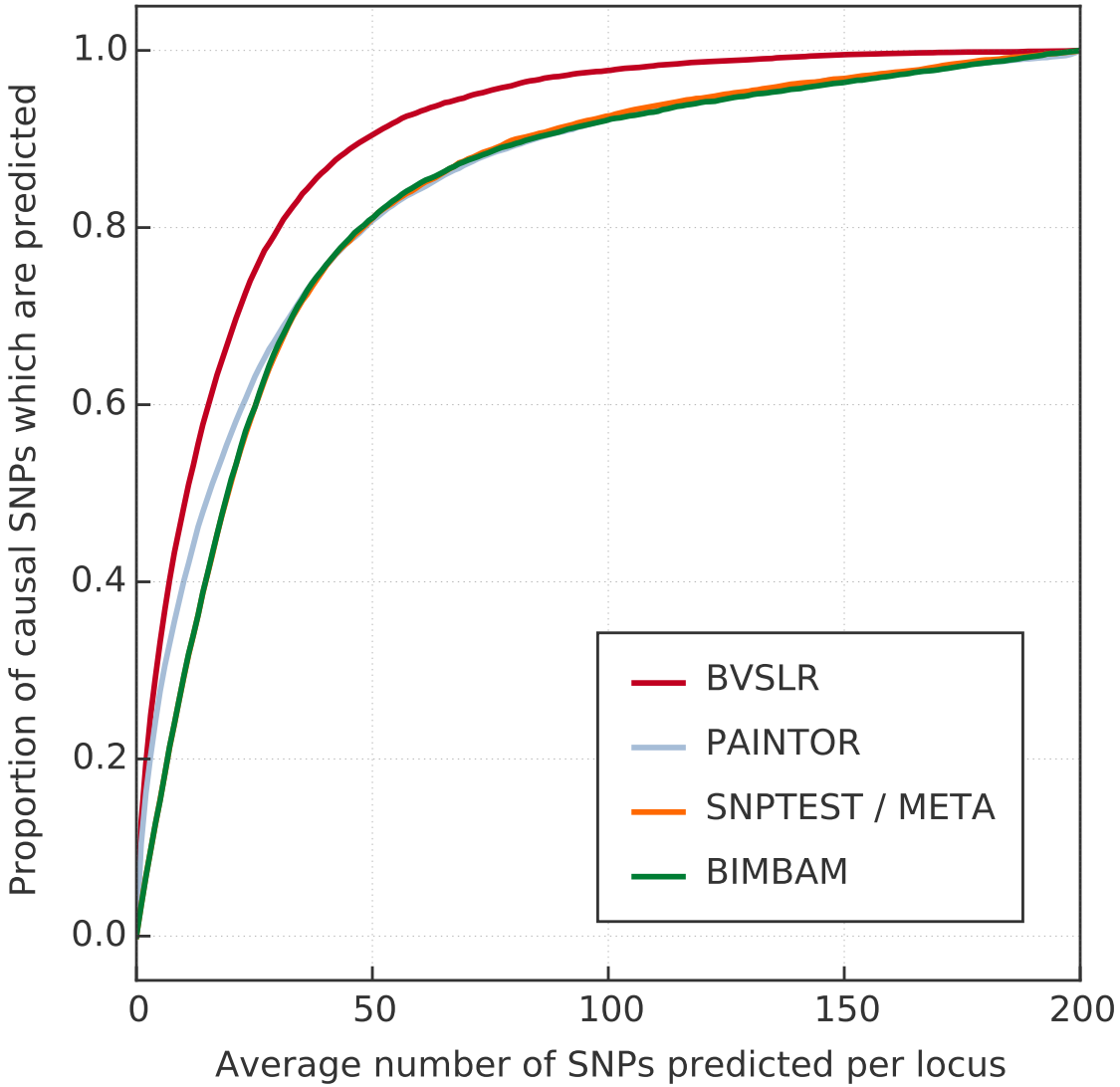


$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

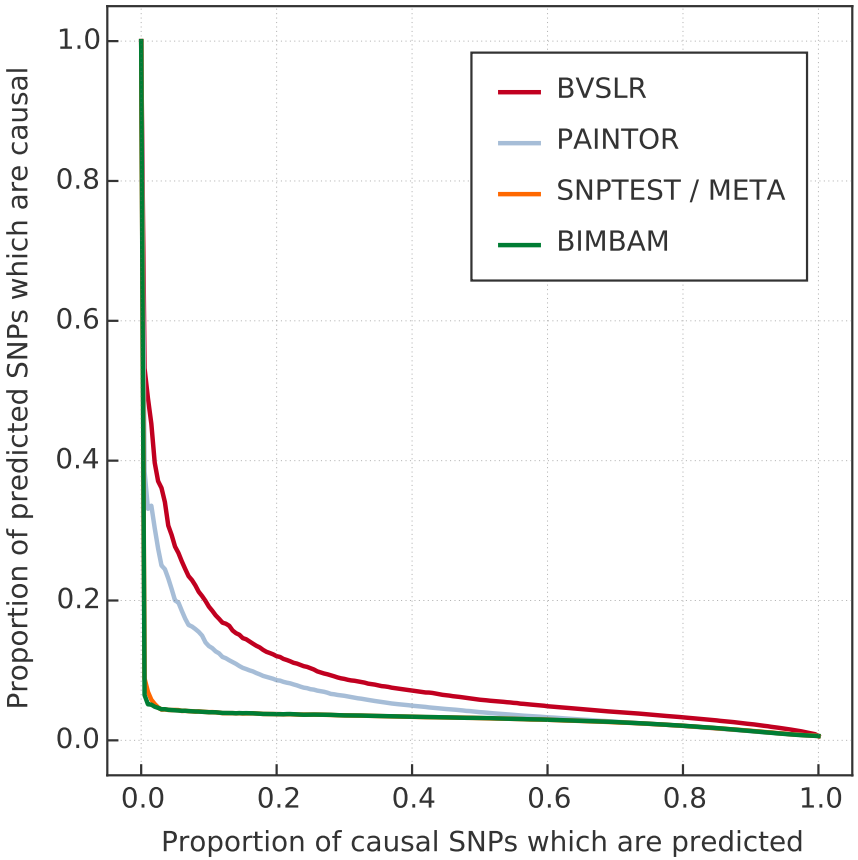
- 8 loci (out of 200) in LD with each other were introduced in the simulation

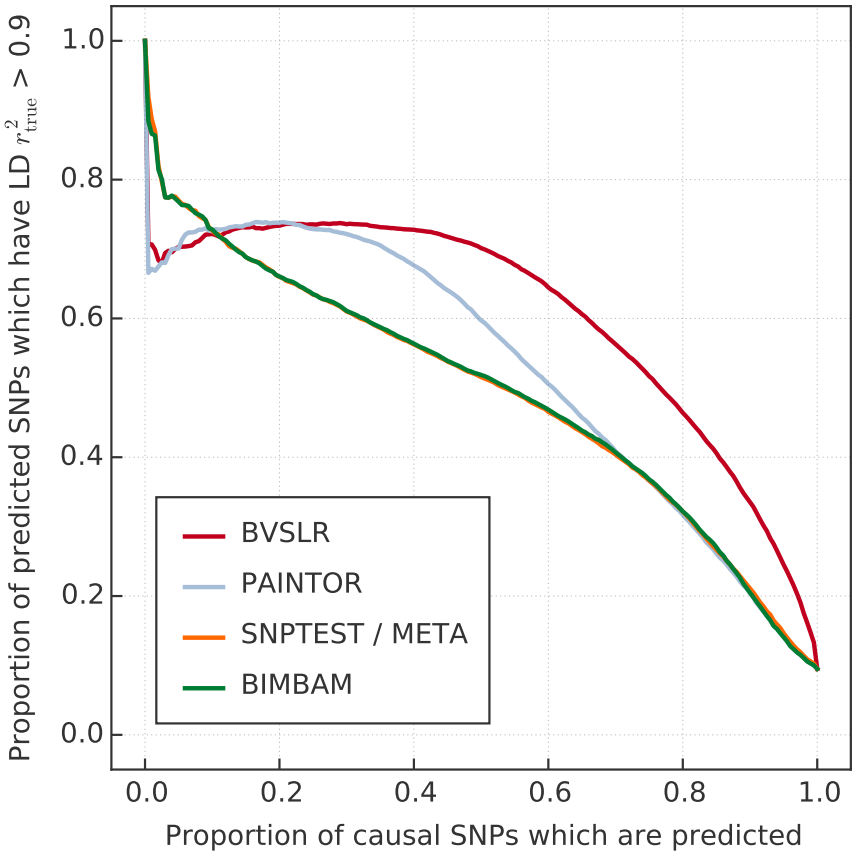
Finemapping causal variants



- Comparable to PAINTOR up to 20% recall

BVSLR predicts SNPs in strong LD with actual ones





Challenges of BVSLR

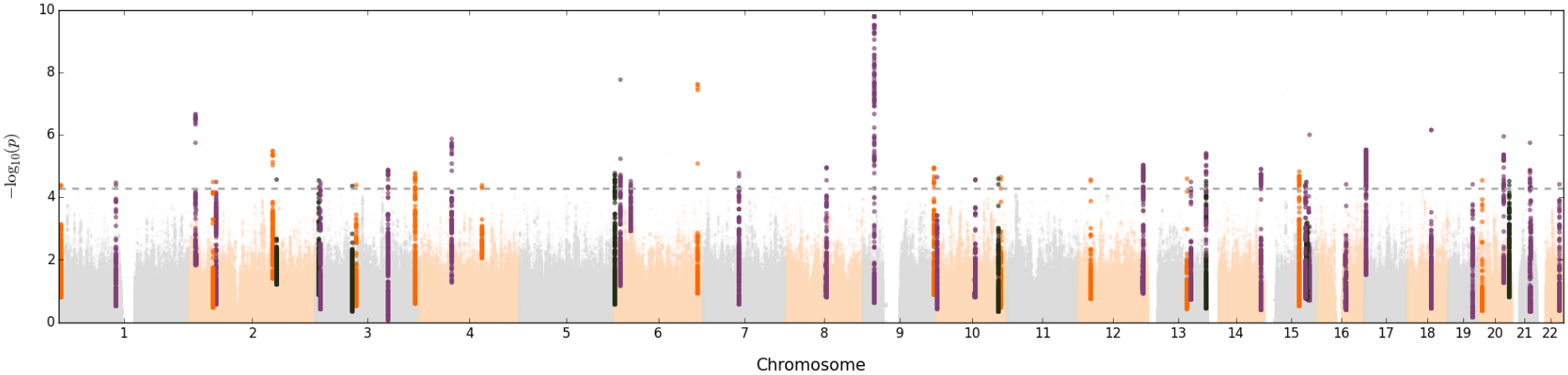
- Computational time (summing over $\mathbf{z}$-states)
- New summary statistics
- Prior assumptions

Association with coronary artery diseases (CAD)

- 5 GERMIFS cohorts
- 6228 cases, 6854 controls
- Imputed with 1000G Phase 1

Association with coronary artery diseases (CAD)

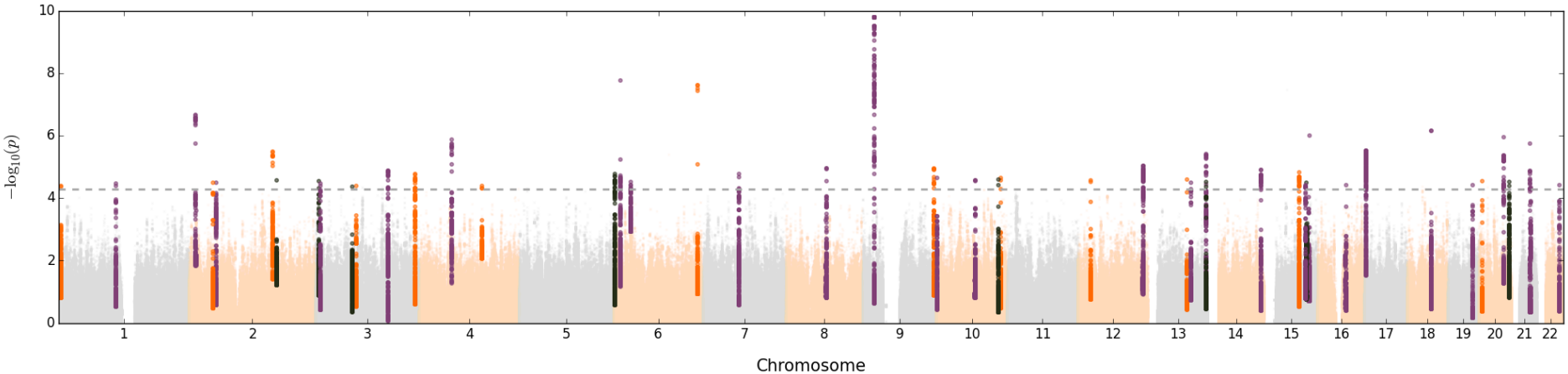
- 5 GERMIFS cohorts
- 6228 cases, 6854 controls
- Imputed with 1000G Phase 1



GWAS using SNPTEST / META

Association with coronary artery diseases (CAD)

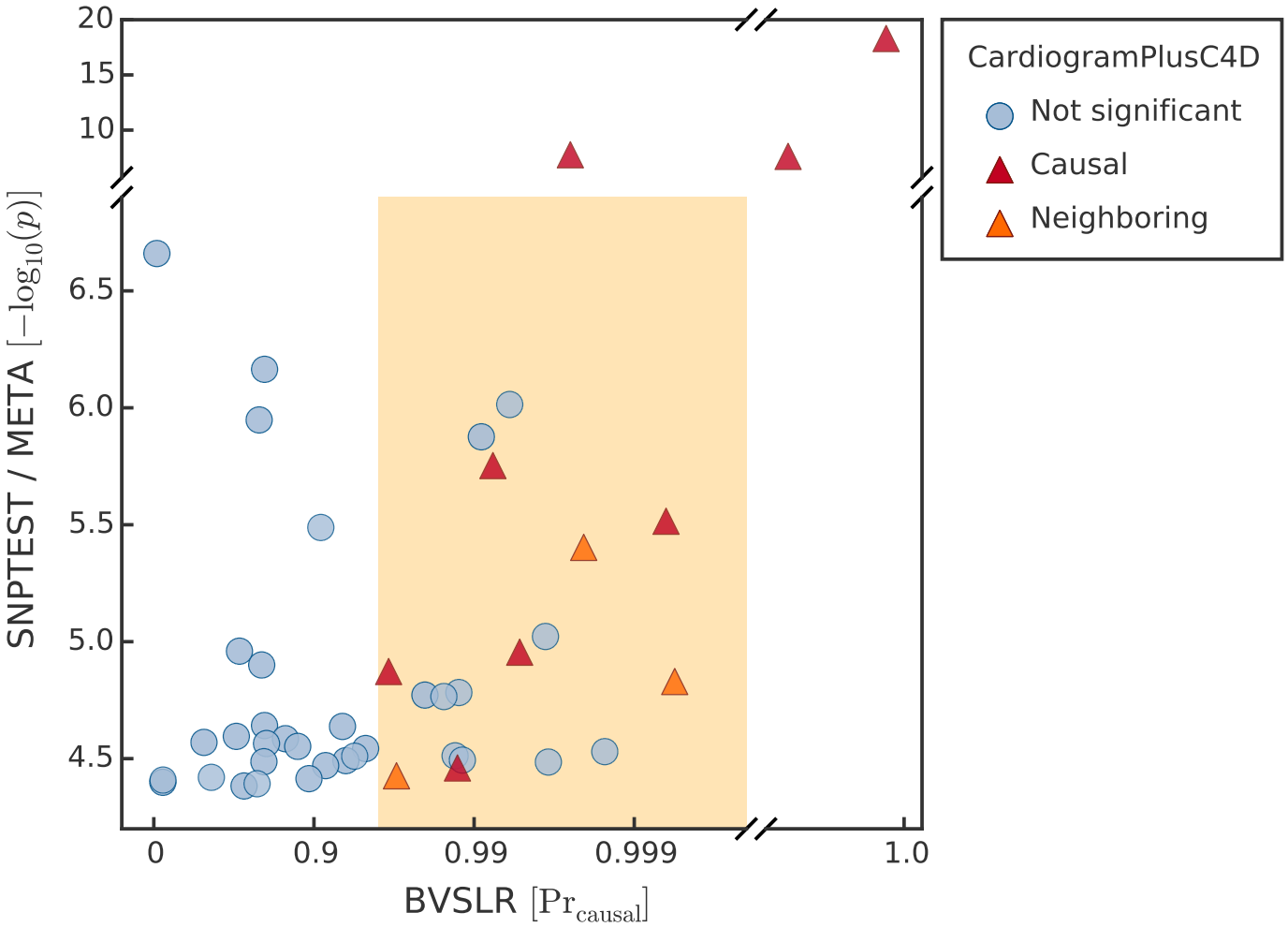
- 5 GERMIFS cohorts
- 6228 cases, 6854 controls
- Imputed with 1000G Phase 1



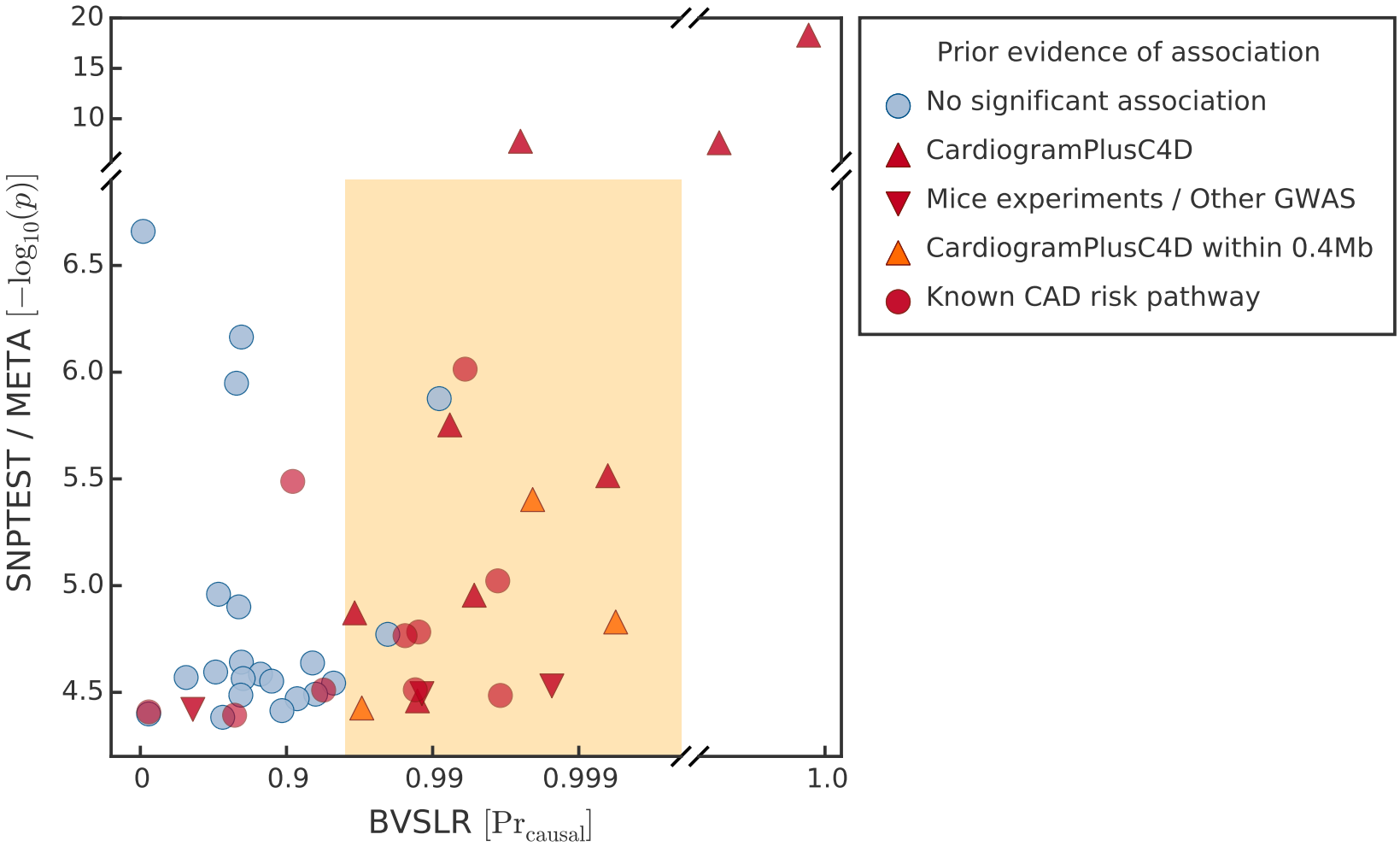
GWAS using SNPTEST / META

- Applied BVSLR on these 45 loci, selecting 400 SNPs at each locus.

BVSLR predictions



BVSLR predictions

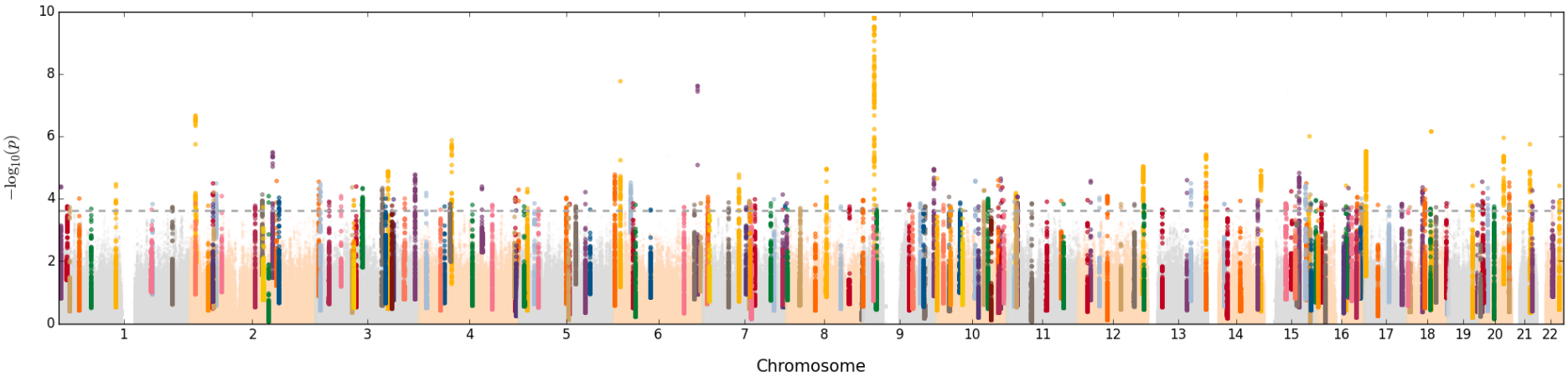


Top BVSLR predicted loci (not discovered in CardiogramPlusC4D)

Region	Pr	Gene	Comments
6p21.3	0.998	C6orf10-BTNL2	GWAS for CAD in Han Chinese, 2012
15q25	0.997	IL-16	GWAS for CAD in Han Chinese, 2012
4q13.1	0.996	desert	–
15q25	0.994	AKAP13	C. hypertrophy (mice) / GWAS (BP in Koreans, 2011)
2p16	0.994	NRXN1	GWAS for CAD in OHGS1 + WTCCC2
3q28	0.992	IL1RAP	Involved in risk pathway
6p25	0.992	SERPINB	patented as biomarker for CVD
20q13.3	0.991	EDN3	GWAS hit for BP / CVD
7q11.22	0.990	AUTS2	–
12q24	0.982	ZNF664	GWAS hit for HDL-C, TG
13q21.1	0.981	ARHGEF1	Controls vascular tone and BP

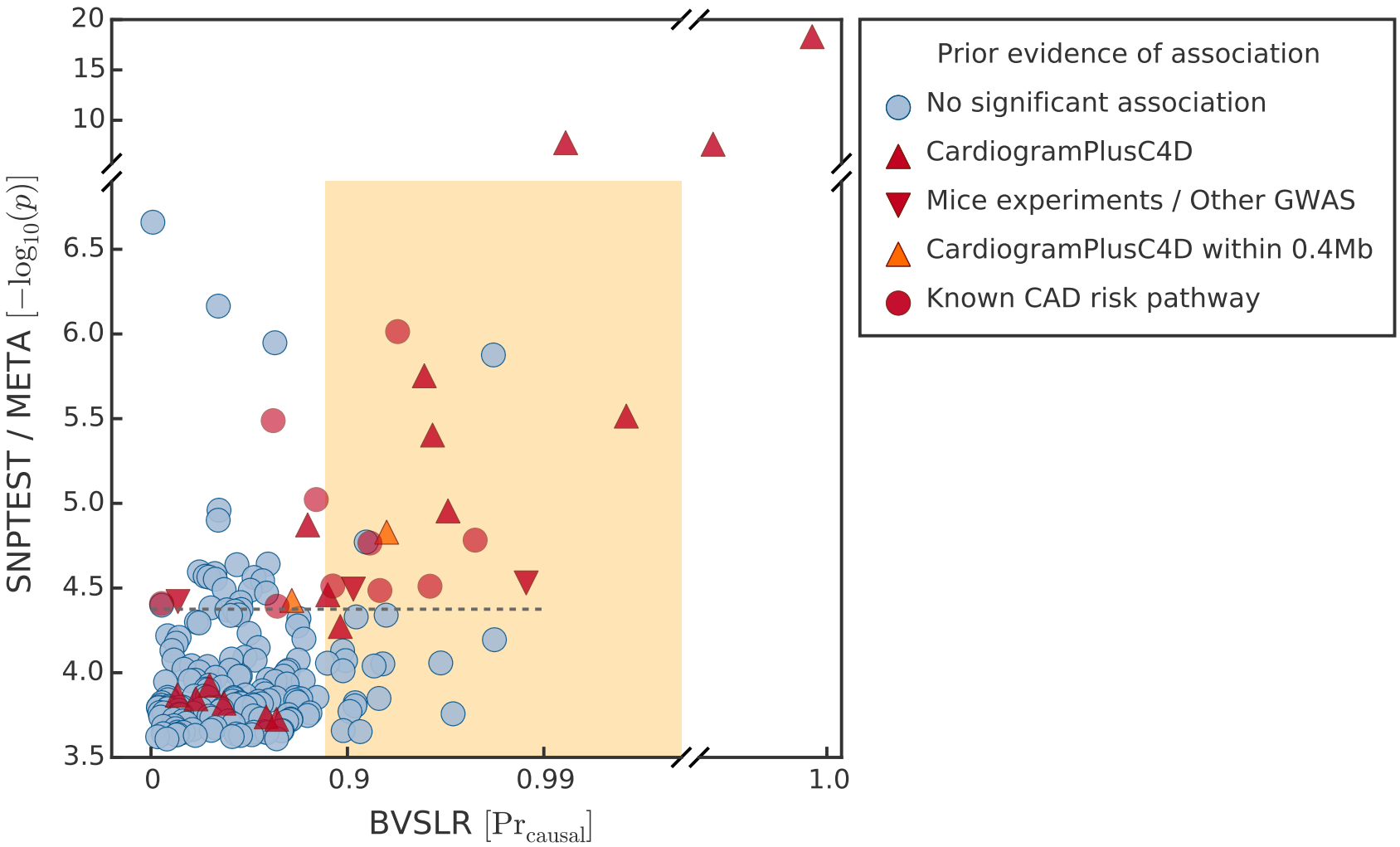
Extension to 200 loci

GWAS using SNPTEST / META



- Applied BVSLR on these 200 loci, selecting 400 SNPs at each locus.

Extension to 200 loci

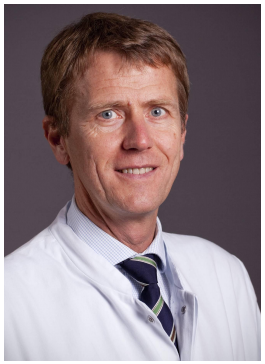


Essence

- Novel Bayesian method for GWAS
- Multivariate analysis in meta studies
- Precision
- Predicts new associations in CAD

Many thanks to . . .


Johannes Söding


Heribert Schunkert


Jeanette Erdmann

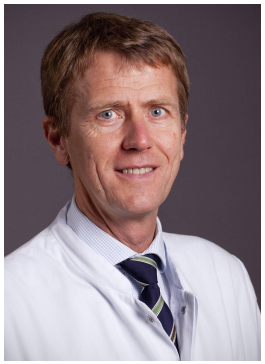
e:Med SYSTEMS MEDICINE

e:AtheroSysMed

Contact me:
saikat.banerjee@mpibpc.mpg.de

Many thanks to . . .


Johannes Söding


Heribert Schunkert


Jeanette Erdmann


AG Söding

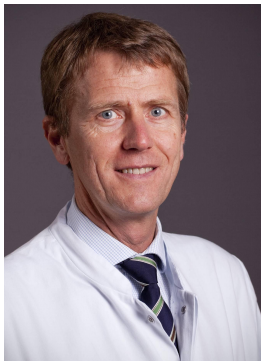
e:Med SYSTEMS MEDICINE

e:AtheroSysMed

Contact me:
saikat.banerjee@mpibpc.mpg.de

Many thanks to . . .


Johannes Söding


Heribert Schunkert


Jeanette Erdmann


AG Söding

Thank you!

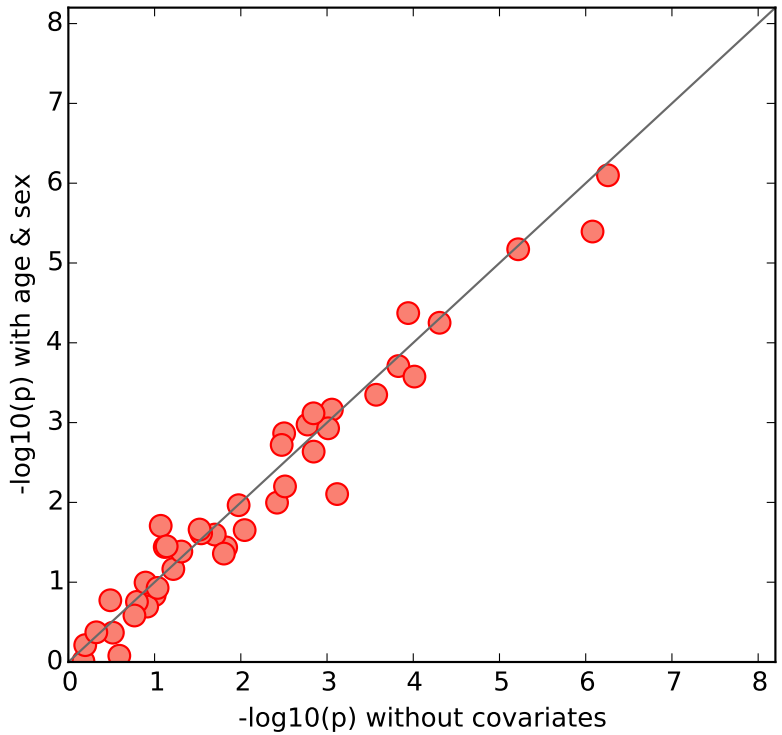
e:Med SYSTEMS MEDICINE

e:AtheroSysMed

Contact me:
saikat.banerjee@mpibpc.mpg.de

Risk correction can be improved

Pirinen *et al.* , Nat. Gen. 2012



Liability correction