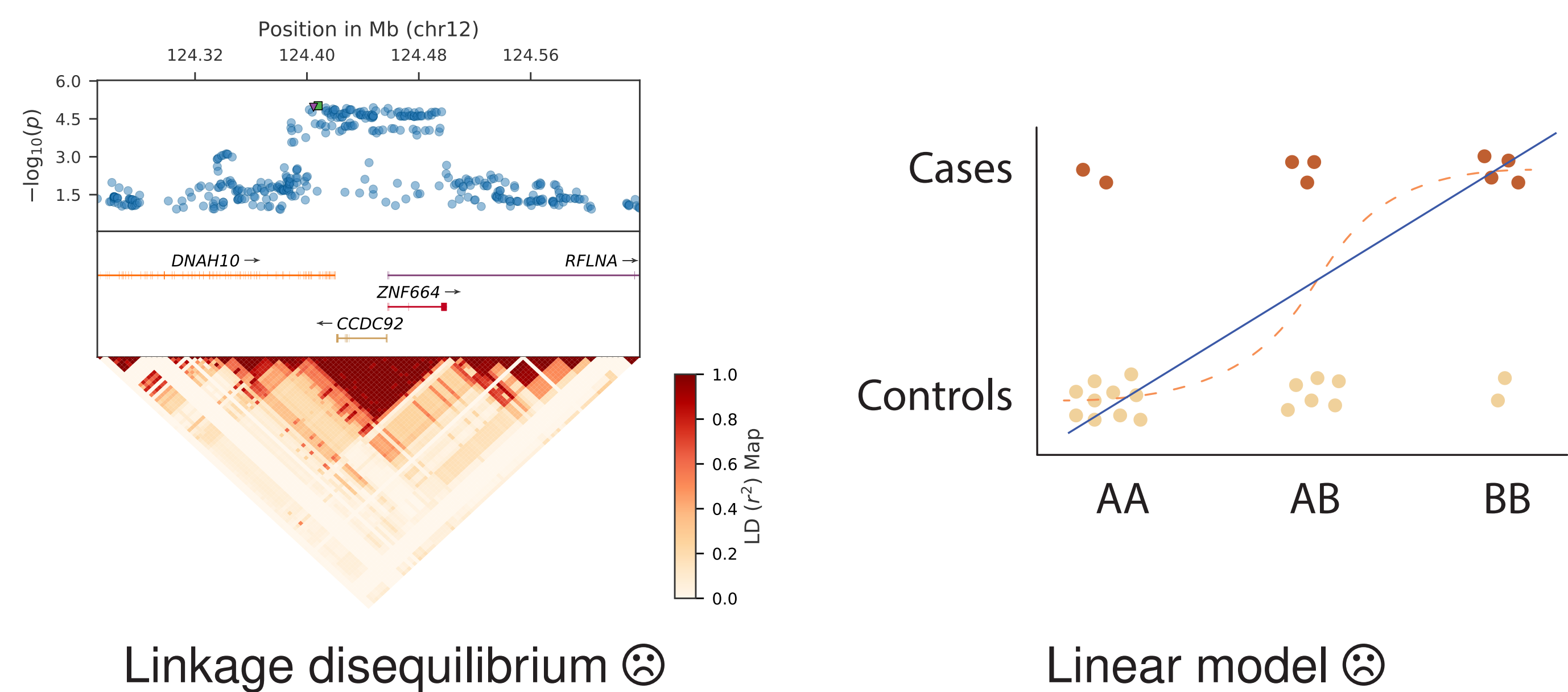


1. Post-GWAS analyses using multiple regression could not utilize the benefits of logistic model



Multiple logistic regression should work much better in the non-linear regimes.
Challenges for using multiple logistic regression:

- The integration for the maximum likelihood cannot be solved analytically.
- MCMC sampling is computationally intractable.
- Solutions using Laplace and linear approximations essentially makes it a linear model.
- Using multiple loci together for the analysis.

2. B-LORE uses logistic model and sparsity-inducing priors

Probability of n^{th} individual with genotype \mathbf{x}_n to be diseased:

$$p(y_n = 1 | \mathbf{x}_n, \boldsymbol{\beta}) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_n)}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_n)}$$

Prior on effect sizes given hyperparameters π and σ , $p(\boldsymbol{\beta}_i | \pi, \sigma)$

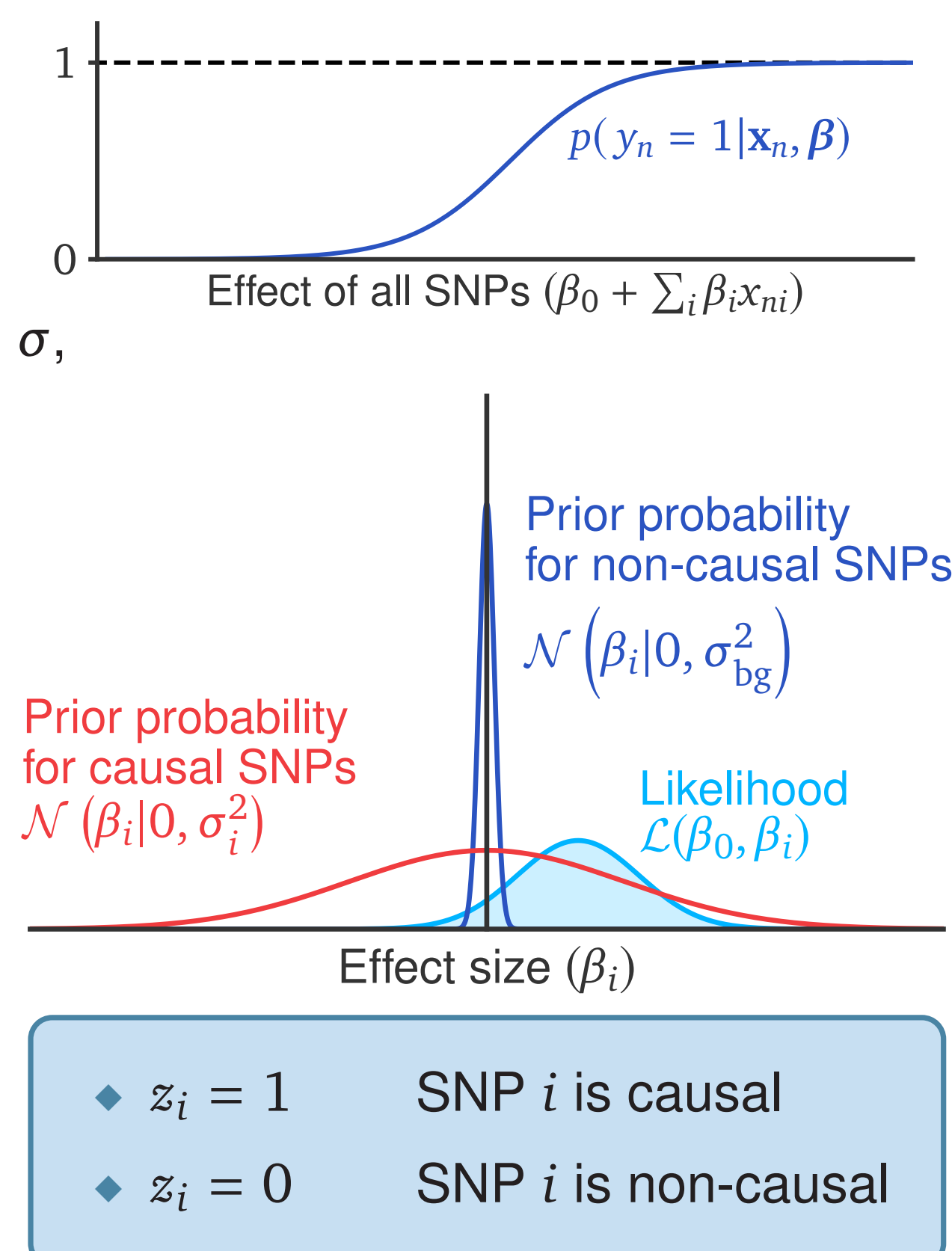
$$= \pi \mathcal{N}(\boldsymbol{\beta}_i | \mathbf{0}, \sigma^2) + (1 - \pi) \delta_0$$

$$= \sum_{z_i=0,1} \pi^{z_i} (1 - \pi)^{(1-z_i)} \mathcal{N}(\boldsymbol{\beta}_i | \mathbf{0}, \text{diag}(\sigma_{z,i}^2))$$

$$= \sum_{z_i=0,1} p(\mathbf{z} | \pi) \mathcal{N}(\boldsymbol{\beta}_i | \mathbf{0}, \text{diag}(\sigma_{z,i}^2))$$

where, $\sigma_{z,i}^2 = z_i \sigma^2$

$z_i \in \{0, 1\} \Rightarrow$ Indicator variable of causality



3. We introduce the quasi-Laplace approximation

Evidence approximation: maximizing the marginal likelihood

$$\text{m}\mathcal{L}(\pi, \sigma) := p(\mathbf{y} | \mathbf{x}, \pi, \sigma) = \sum_{\mathbf{z}} p(\mathbf{z} | \pi) \int p(\mathbf{y} | \mathbf{x}, \boldsymbol{\beta}) \mathcal{N}(\boldsymbol{\beta} | \mathbf{0}, \text{diag}(\sigma_{\mathbf{z}}^2)) d\boldsymbol{\beta} \rightarrow \max$$

Quasi-Laplace approximation:

$$p(\mathbf{y} | \mathbf{x}, \boldsymbol{\beta}) \mathcal{N}(\boldsymbol{\beta} | \mathbf{0}, \text{diag}(\sigma_{\mathbf{z}}^2)) = p(\mathbf{y} | \mathbf{x}, \boldsymbol{\beta}) \mathcal{N}(\boldsymbol{\beta} | \mathbf{0}, \tilde{\sigma}^2 \mathbb{I}) \frac{\mathcal{N}(\boldsymbol{\beta} | \mathbf{0}, \text{diag}(\sigma_{\mathbf{z}}^2))}{\mathcal{N}(\boldsymbol{\beta} | \mathbf{0}, \tilde{\sigma}^2 \mathbb{I})}$$

$$\propto \mathcal{N}(\boldsymbol{\beta} | \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Lambda}}^{-1})$$

Benefits:

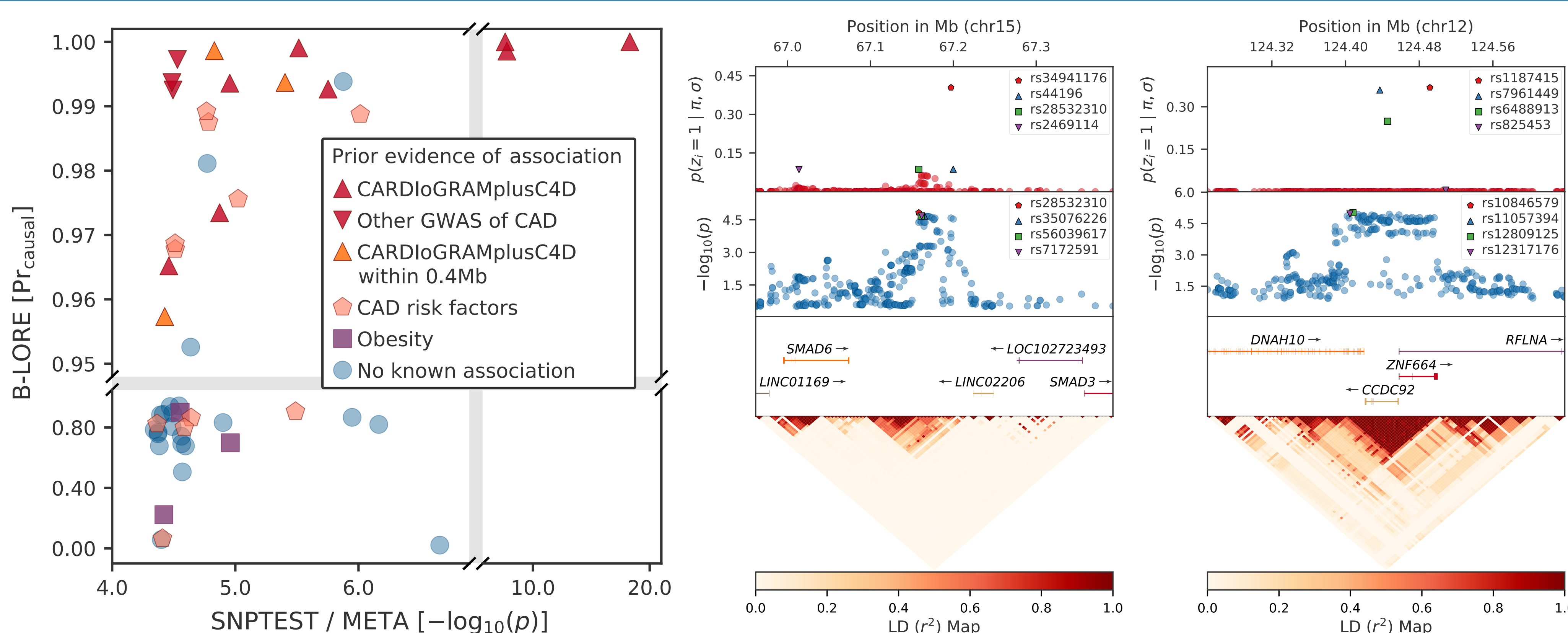
- The regularizer pulls the maximum of the regularized likelihood near to the mode of the integral, making it more accurate than Laplace approximation.
- Can be extended to multiple studies.
- Fast gradient-descent optimization.

B-LORE
schema

1. Two-step optimization at each cohort to estimate $\tilde{\sigma}$ and $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Lambda}})$.

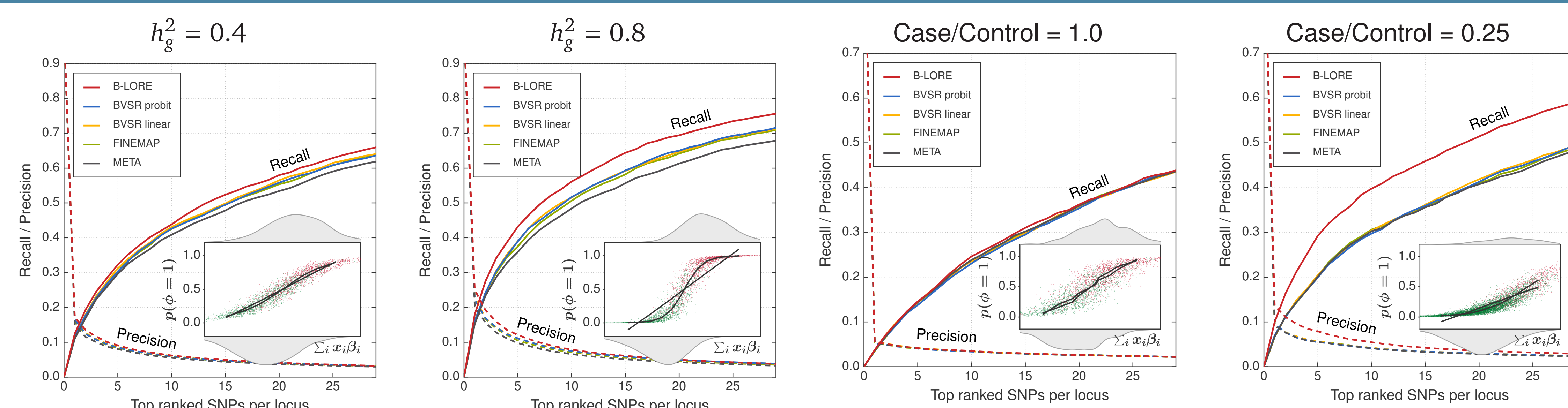
2. Estimation of hyperparameters (π, σ) .

5. Meta-analysis example: B-LORE discovers novel loci associated with coronary artery disease



Meta-analysis of 5 cohorts, Germal Myocardial Infarction Family Studies (GerMIFS I-V) – 6234 cases and 6848 controls.

6. Examples of non-linear regimes in case-control GWAS



4. Inference

Prediction of causality of each locus.

The probability for a locus to be causally associated with the disease is

$$\text{Pr}_{\text{causal}} = p(\text{locus is causal} | \boldsymbol{\phi}, \mathbf{X}, \hat{\pi}, \hat{\sigma}) = 1 - p(\mathbf{z} = \mathbf{0} | \boldsymbol{\phi}, \mathbf{X}, \hat{\pi}, \hat{\sigma})$$

Statistical finemapping of causal variants.

The posterior probability for SNP i to be causal is

$$p(z_i = 1 | \boldsymbol{\phi}, \mathbf{X}, \hat{\pi}, \hat{\sigma})$$

7. References

1. Banerjee *et al.* PLOS Genet 2018, doi:10.1371/journal.pgen.1007856
2. Servin *et al.* PLOS Genet 2007, doi:10.1371/journal.pgen.0030114
3. Guan *et al.* Ann Appl Stat 2011, doi:10.1214/11-AOAS455
4. CARDIoGRAMplusC4D Nat Genet 2015, doi:10.1038/ng.3396

8. Acknowledgement

We thank Prof. Dr. Jeanette Erdmann for helpful discussions. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1313A-2014).