Bayesian variable selection logistic regression: multivariate metaanalysis in GWAS

S. Banerjee & J. Söding

Max Planck Institute for Biophysical Chemistry

FEBRUARY 3, 2017

Simulation details

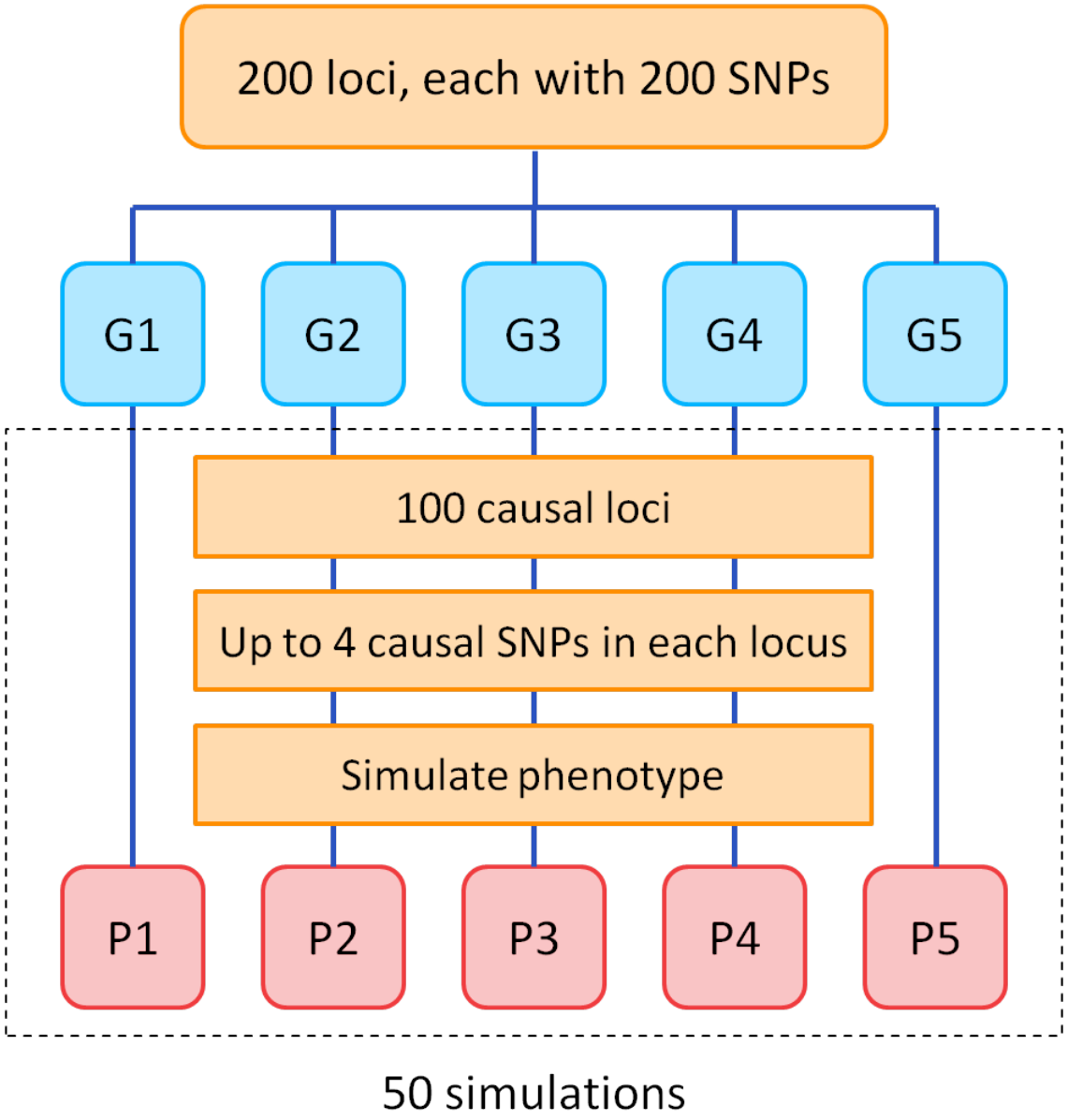
- Genotype: German Myocardial Infarction Family Study (GERMIFS)
- Five cohorts : G1, G2, G3, G4, G5
- Phenotype simulation:

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

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

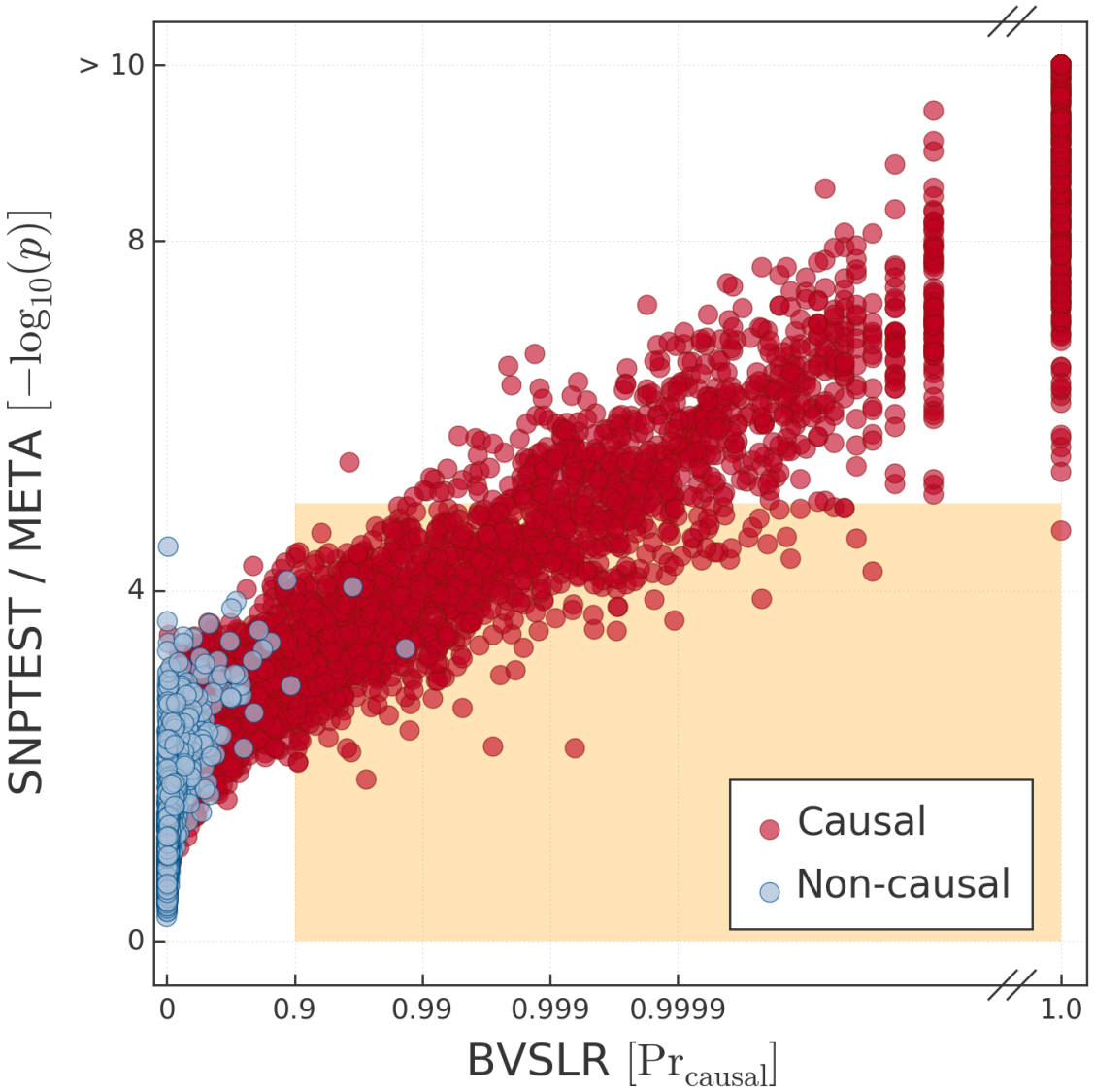
Cases are sampled from Y_n exceeding the threshold of normal distribution truncating the proportion of k (disease prevalence)

Simulation details



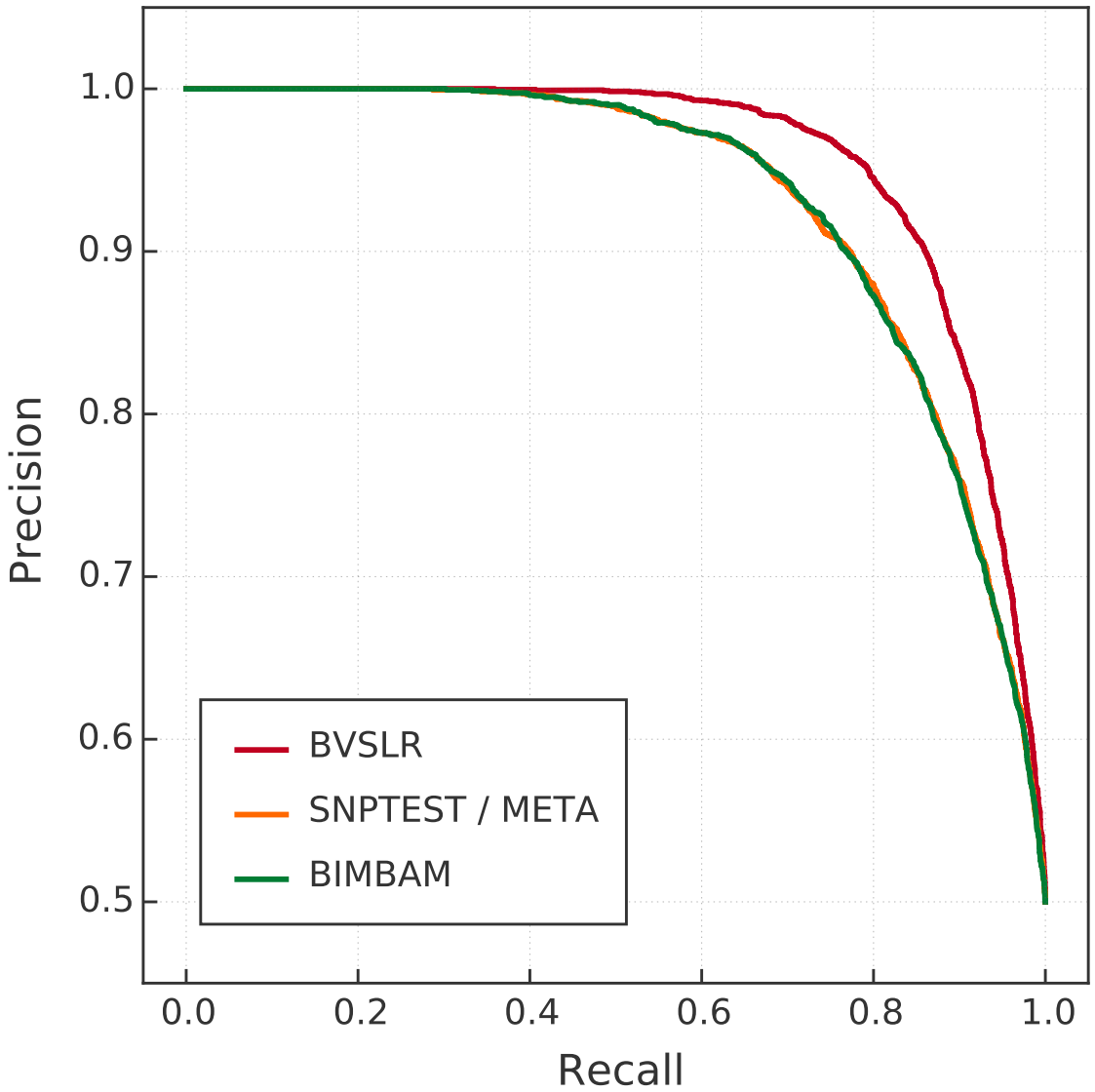
- BVSLR
- BIMBAM
- SNPTEST / META
- PAINTOR

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)

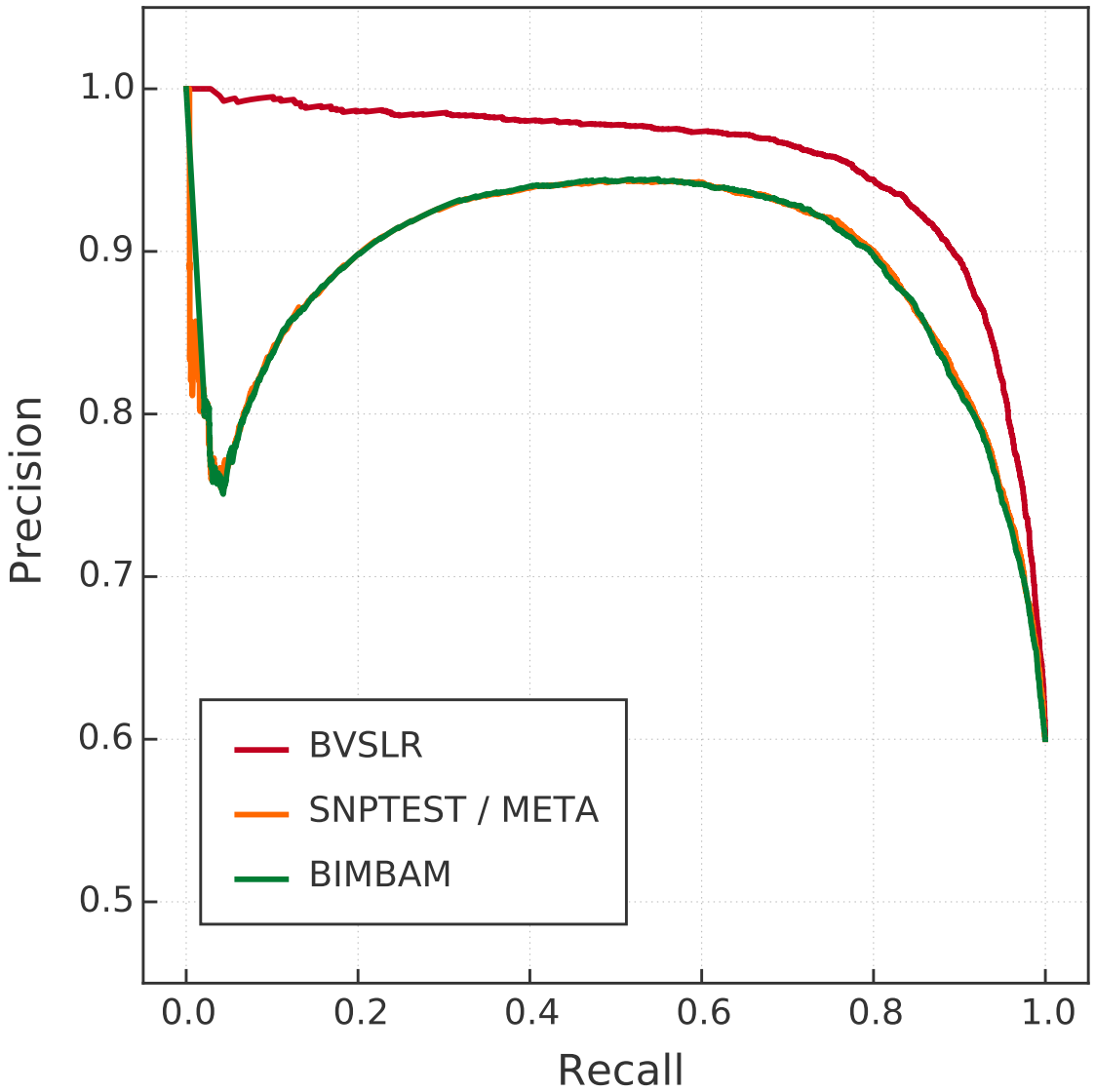
Prediction of causal loci



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

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

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

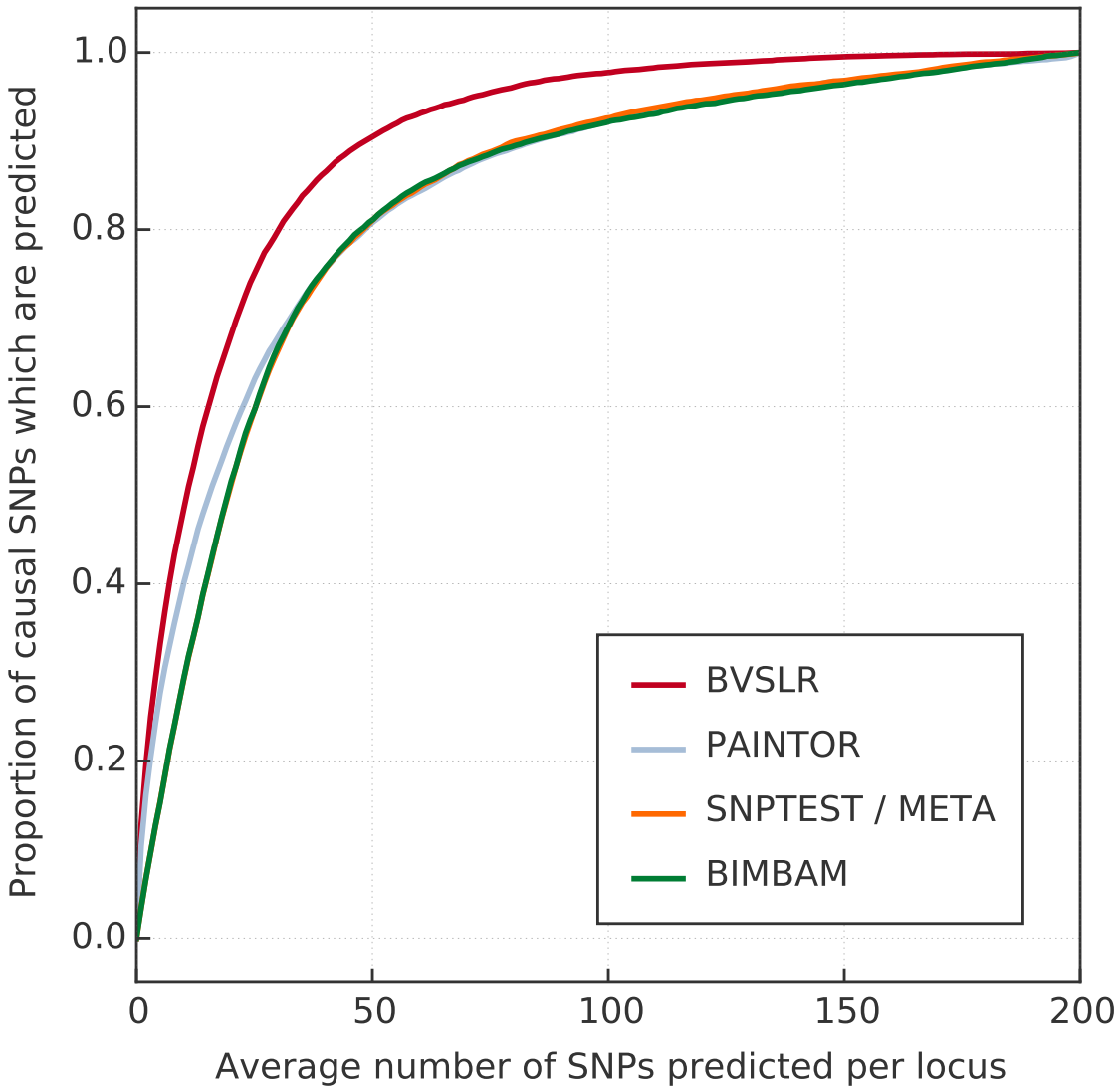


$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

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

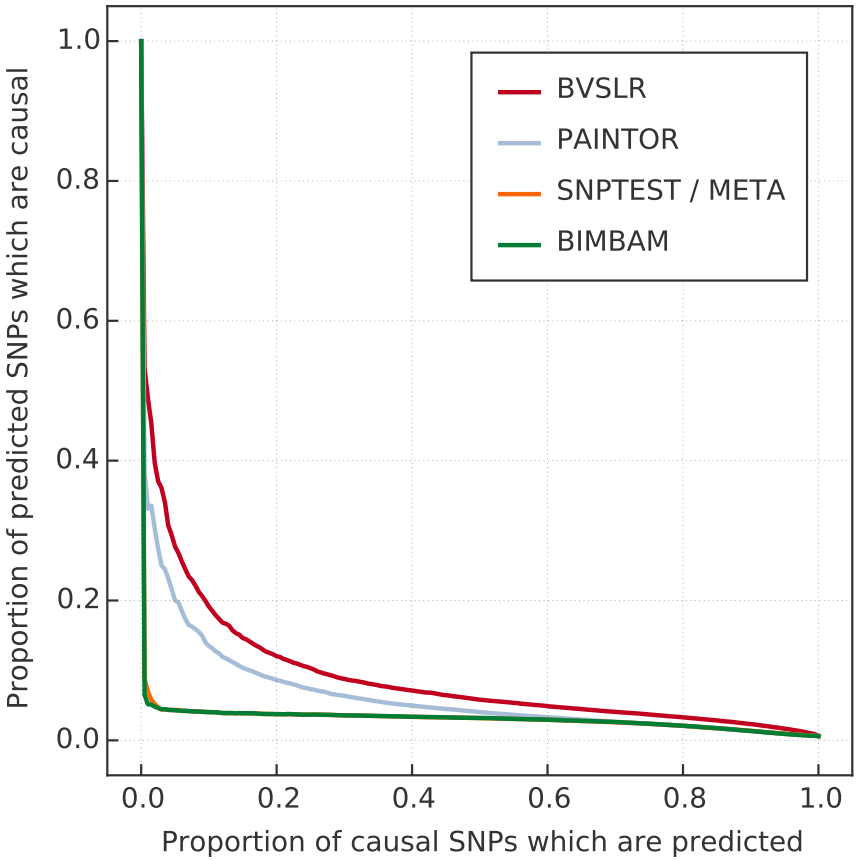
- 8 loci (out of 200) in LD with each other were introduced in the simulation

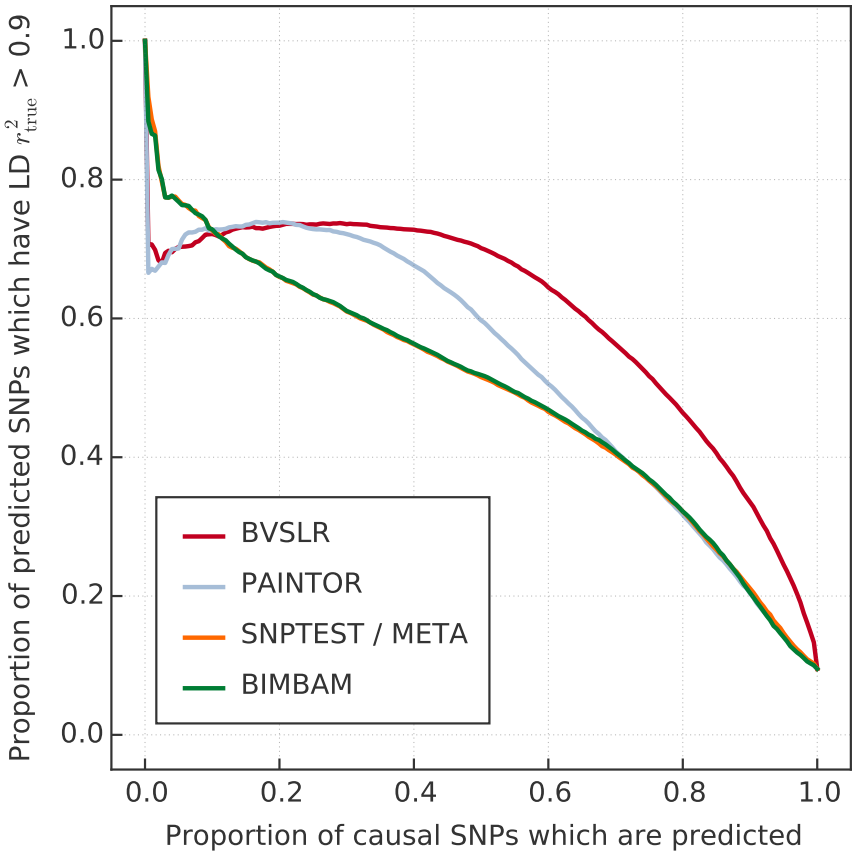
Finemapping causal variants



- Comparable to PAINTOR up to 20% recall

BVSLR predicts SNPs in strong LD with actual ones



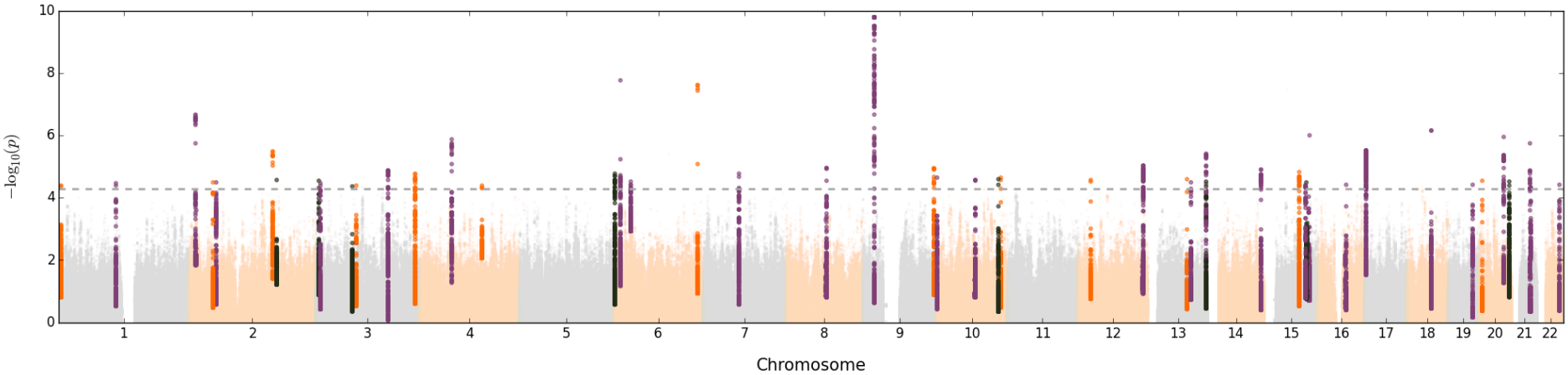


Association with coronary artery diseases (CAD)

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

Association with coronary artery diseases (CAD)

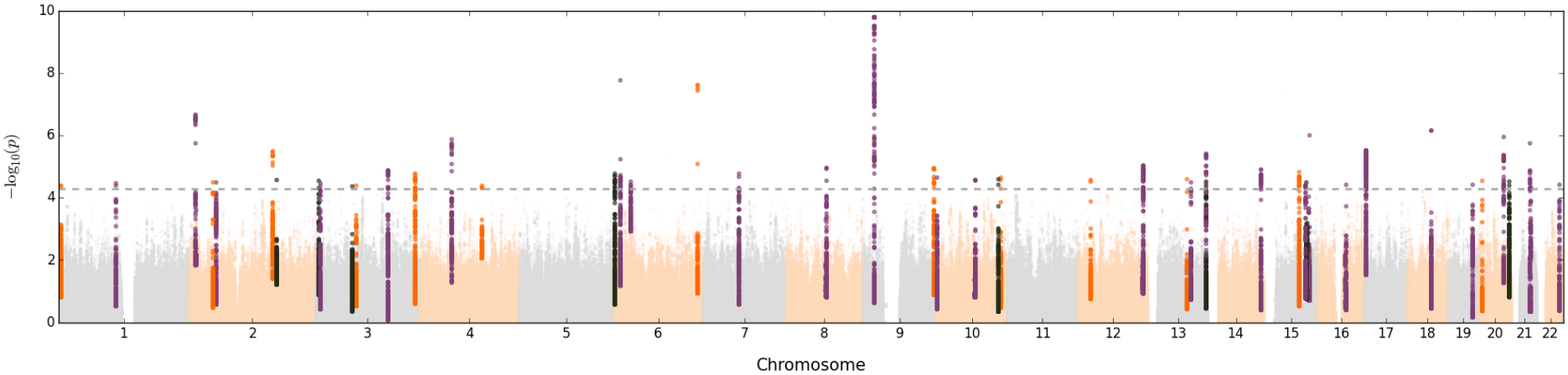
- 5 GERMIFS cohorts
- 6228 cases, 6854 controls
- Imputed with 1000G Phase 1



GWAS using SNPTEST / META

Association with coronary artery diseases (CAD)

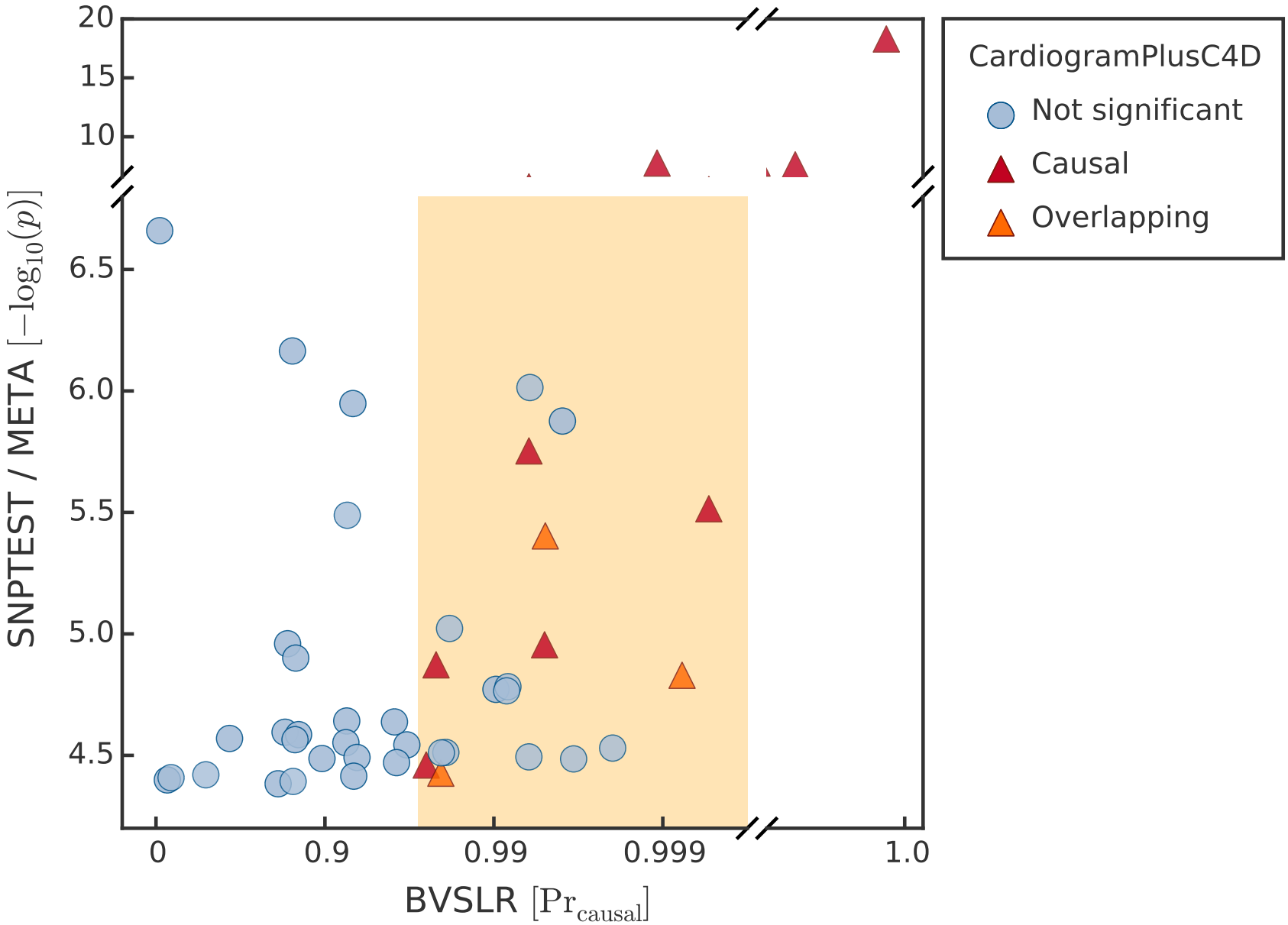
- 5 GERMIFS cohorts
- 6228 cases, 6854 controls
- Imputed with 1000G Phase 1



GWAS using SNPTEST / META

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

BVSLR predictions



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

Literature-based classification