

Incorporating Functional Annotations for Fine-Mapping Causal Variants in a Bayesian Framework Using Summary Statistics

Wenan Chen,* Shannon K. McDonnell,* Stephen N. Thibodeau,[†] Lori S. Tillmans,[†] and Daniel J. Schaid*¹

*Department of Health Sciences Research, Division of Biostatistics, and [†]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota 55905

ORCID ID: 0000-0003-0379-2368 (W.C.)

ABSTRACT Functional annotations have been shown to improve both the discovery power and fine-mapping accuracy in genome-wide association studies. However, the optimal strategy to incorporate the large number of existing annotations is still not clear. In this study, we propose a Bayesian framework to incorporate functional annotations in a systematic manner. We compute the maximum *a posteriori* solution and use cross validation to find the optimal penalty parameters. By extending our previous fine-mapping method CAVIARBF into this framework, we require only summary statistics as input. We also derived an exact calculation of Bayes factors using summary statistics for quantitative traits, which is necessary when a large proportion of trait variance is explained by the variants of interest, such as in fine mapping expression quantitative trait loci (eQTL). We compared the proposed method with PAINTOR using different strategies to combine annotations. Simulation results show that the proposed method achieves the best accuracy in identifying causal variants among the different strategies and methods compared. We also find that for annotations with moderate effects from a large annotation pool, screening annotations individually and then combining the top annotations can produce overly optimistic results. We applied these methods on two real data sets: a meta-analysis result of lipid traits and a *cis*-eQTL study of normal prostate tissues. For the eQTL data, incorporating annotations significantly increased the number of potential causal variants with high probabilities.

KEYWORDS Bayesian fine mapping; annotations; summary statistics; causal variants

A large amount of annotation information of genomic elements has been generated, such as the Encyclopedia of DNA Elements (The ENCODE Project Consortium 2012). Regulatory elements, such as those marked by DNase I hypersensitive sites (DHSs), have been shown to be enriched with associations and explain a large proportion of heritability (Maurano *et al.* 2012; Gusev *et al.* 2014). Incorporation of these annotations into statistical association analyses can improve both the power in genome-wide association study (GWAS) discovery (Pickrell 2014) and the accuracy in fine mapping underlying causal variants (Quintana and Conti 2013; Kichaev *et al.* 2014; Wen *et al.* 2015). However, the number of annotations is often very large and many of them

may not be informative for the underlying causal genetic variants. Currently, there is no systematic and effective way to incorporate a large number of annotations in association analyses, which is crucial to the full use of the available information. In practice, usually only a small number of annotations are considered (Quintana and Conti 2013; Wen *et al.* 2015). A recent proposed algorithm, PAINTOR, suggests a one-by-one test of each annotation selecting the top 4–5 for inclusion. This has potential to waste the information from the remaining annotations, which may together provide significant predictive power of the causal status. Another approach (Pickrell 2014) considered a large number of annotations, however, it has two limitations. First, it assumes the maximal number of causal variants in each genomic region locus is one, which may lead to suboptimal results (Hormozdiari *et al.* 2014; Kichaev *et al.* 2014; Chen *et al.* 2015). Second, the process of annotation selection may not fully explore the available annotation information. In this study, we propose a Bayesian framework to incorporate a large number of annotations simultaneously. Moreover, we extend our previous

Copyright © 2016 by the Genetics Society of America

doi: 10.1534/genetics.116.188953

Manuscript received March 5, 2016; accepted for publication September 7, 2016; published Early Online September 21, 2016.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.188953/-/DC1.

¹Corresponding author: Division of Biostatistics, Harwick 7, Mayo Clinic, 200 First St. SW, Rochester, MN 55905. E-mail: schaid@mayo.edu

fine-mapping method, CAVIARBF, under this framework, which enables the use of summary statistics. Simulation results show that our proposed method achieves better or similar performance compared to PAINTOR and *fgwas* under a variety of different annotation selection strategies.

Methods

A general Bayesian hierarchical model to incorporate functional annotations

We first consider one region or locus for fine mapping and later extend it to multiple loci. For the locus of interest, suppose there are m variants and n individuals. We assume that only a subset of the m variants is causal. Denote the phenotypes by $Y_{n \times 1}$ and genotypes by $G_{n \times m}$. The phenotypes can be quantitative traits or binary traits. The effect size of each variant is denoted by $\beta_{m \times 1}$, which depends on the causal states, denoted by $c_{m \times 1}$, where the elements of vector c are either 1, indicating the corresponding variant is causal, or 0 for non-causal variants. For the i th variant, if $c_i = 1$, β_i is assumed to be normally distributed with a mean of zero and variance specified by σ_a^2 (Servin and Stephens 2007; Guan and Stephens 2008). For quantitative traits, Y and β_i also depend on a random variable τ that specifies their variances, for which we assign a noninformative prior (Servin and Stephens 2007; Guan and Stephens 2008). If $c_i = 0$, then $\beta_i = 0$. We assume the causal states represented by c depend on the annotations of the variants. For example, we can model the relationship between the annotation and the probability of being causal using a logistic model. Suppose there are d annotations for each variant and the annotation matrix is denoted by $A_{m \times d}$ and $A_{i \cdot}$ denotes the i th row. Denote the annotation effects on the causal states by $\gamma = (\gamma_0, \gamma_1^T)^T$, where T means transpose, γ_0 is the intercept, and γ_1 is a vector of annotation effects. We assume a normal prior distribution on the annotation effects with alternate distributions considered later. The full model can be written as follows, assuming quantitative traits and a logistic model linking annotations to causal states:

$$Y|\mu, G, \beta, \tau \sim N\left(\mu + G\beta, \frac{1}{\tau}I_n\right),$$

$$\beta_i|\tau, c_i \sim \begin{cases} N(0, \sigma_a^2 \frac{1}{\tau}) & \text{if } c_i = 1 \\ 0 & \text{if } c_i = 0 \end{cases}, i = 1 \dots, m,$$

$$c_i|\gamma_0, \gamma_1, A_{i \cdot} \sim \text{Bernoulli}\left(\frac{1}{1 + e^{-(\gamma_0 + A_{i \cdot} \gamma_1)}}\right), i = 1 \dots, m,$$

$$\gamma_1 \sim N(0, \lambda^{-1}I_{d \times d}).$$

We assign noninformative priors on μ , τ , and γ_0 . Specifically, $\mu|\tau \sim N\left(0, \frac{\sigma_a^2}{\tau}\right)$, $\tau \sim \Gamma(\kappa_1, \kappa_2)$, and $\gamma_0 \sim N(0, \lambda_0^{-1})$, where $\Gamma(\kappa_1, \kappa_2)$ is the Gamma distribution with the shape parameter κ_1 and the inverse scale parameter κ_2 , and we let $\sigma_a^2 \rightarrow \infty, \kappa_1, \kappa_2, \lambda_0 \rightarrow 0$.

The above probabilistic model forms a hierarchical model, or directed graphical model (Bishop 2006), as shown in Figure 1. In Figure 1, the nodes are random variables, where open circles represent unobserved/hidden random variables, shaded circles represent observed random variables, small solid circles represent fixed parameters, and the arrows show the dependencies between two random variables. The arrow leaves from the node on which the receiving node depends, for example, the causal state c depends on the annotation A and the annotation effects γ . Variables with noninformative priors are not shown in Figure 1. When using summary statistics, we need only to replace Y with the Z -test statistic, and replace G with the correlation matrix Σ , which will be described later. Compared to the previous model in CAVIARBF (Chen *et al.* 2015), this model has additional levels to incorporate annotations, enclosed by the box in Figure 1. Compared to PAINTOR (Kichaev *et al.* 2014) and the practical inference procedure of FM-QTL (Wen *et al.* 2015), this model has an extra level specifying the prior of the annotation effects γ , which turns out to be critical and enables the incorporation of a large number of annotations simultaneously into the model. For now we assume λ is a known fixed value, and later we will discuss ways to select the optimal parameter.

Exact Bayes factors using summary statistics for quantitative traits

We previously showed (Chen *et al.* 2015) that the fine-mapping method CAVIAR (Hormozdiari *et al.* 2014) is approximately equivalent to BIMBAM (Servin and Stephens 2007; Guan and Stephens 2008). Based on this equivalence, we developed an approximate Bayesian framework for fine mapping multiple causal variants in a locus using only summary statistics. When we assume at most one causal variant in a locus, it reduces to the approximate Bayes factor (ABF) derived by Wakefield (2009), and it produces exactly the same Bayes factor as in *fgwas* (Pickrell 2014) when the variance of the effect size is estimated (see *Appendix*). The approximation depends on a key assumption that the explained variance by each model is very small compared to the total variance of the phenotype. This assumption is reasonable for most GWAS hit regions due to the small effect sizes. However, for fine mapping expression quantitative trait loci (eQTL), this assumption may not hold very well because it is not uncommon that the gene expression trait is moderately or highly correlated with certain variants close to the gene. Fortunately, we found that for quantitative traits the exact Bayes factors using a D_2 prior (Servin and Stephens 2007) can still be calculated using only the marginal test statistics and correlation matrix. We assume a linear model as follows:

$$y = \mu + X_c \beta + \varepsilon,$$

where $\varepsilon \sim N(0, \frac{1}{\tau}I_n)$, I_n denotes the $n \times n$ identity matrix, and n is the number of individuals. X_c is the coded genotype matrix from putative causal variants. The D_2 prior assumes

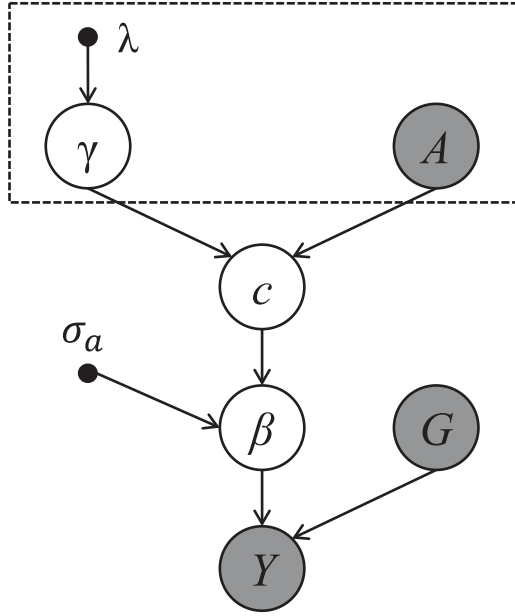


Figure 1 Hierarchical models incorporating functional annotations. The box indicates the added modeling block related to annotations. Shaded \circ indicates observed data. A , the annotation matrix of variants; c , the causal configuration vector of variants; G , full genotype matrix; Y , phenotype vector; β , effect size vector of genotypes on the phenotype; σ_a , the parameter specifying the prior distribution of β ; γ , the effect size vector of annotations on the causality state; λ , the parameter specifying the prior distribution of γ .

that β has a prior normal distribution of $N[0, v_c(1/\tau)]$, where v_c is a diagonal matrix and β and ε are independent. The j th marginal test statistic z_j is calculated from the model

$$y = \mu + X_{cj}\beta_j + \varepsilon_j,$$

where X_{cj} is the j th column of X_c . Denote the correlation matrix by Σ_x . Let z^* be the column vector where the j th component is $(z_j/\sqrt{z_j^2 + n - 2})$, and let S be a diagonal matrix where the j th entry is the unbiased sample variance of X_{cj} . The exact Bayes factor can be calculated as follows:

$$BF = \left| [(n-1)v_c S]^{-1} + \Sigma_x \right|^{-1/2} \left| (n-1)v_c S \right|^{-1/2} \times \left[1 - z^{*T} \left\{ [(n-1)v_c S]^{-1} + \Sigma_x \right\}^{-1} z^* \right]^{-n/2}.$$

The detailed proof is shown in the *Appendix*. Besides the exact calculation, to make it robust to different magnitudes of effect sizes, we also added the option in CAVIARBF to set multiple prior distributions for β using different variance values and to calculate the averaged Bayes factor as in BIMBAM (Servin and Stephens 2007; Guan and Stephens 2008).

Maximum a posteriori estimate using expectation maximization

In general, techniques such as Markov chain Monte Carlo (MCMC) can be used for inference based on the model in

Figure 1. However, this can be time consuming. Instead, we calculate a point estimate of the annotation effects γ , the *maximum a posteriori* (MAP) estimate, and use it for all other inferences. We may sacrifice some accuracy by using the MAP estimate to represent the full posterior distribution of γ , but we gain computational speed. Given Y, G, A , and specified λ, λ_0 , the posterior probability of γ is

$$p(\gamma|Y, G, A, \lambda, \lambda_0) \propto p(\gamma, Y, G, A, \lambda, \lambda_0) \propto p(Y|G, A, \gamma)p(\gamma|\lambda)p(\gamma_0|\lambda_0).$$

Therefore,

$$\begin{aligned} \ln p(\gamma|Y, G, A, \lambda, \lambda_0) &= \ln p(Y|G, A, \gamma) + \ln p(\gamma|\lambda) \\ &\quad + \ln p(\gamma_0|\lambda_0) + \text{constant} \\ &= \ln p(Y|G, A, \gamma) - \frac{1}{2}\lambda\|\gamma_1\|_2^2 + \frac{m}{2}\ln\lambda \\ &\quad - \frac{1}{2}\lambda_0\gamma_0^2 + \frac{1}{2}\ln\lambda_0 + \text{constant}, \end{aligned}$$

where $\|x\|_2^2 = \sum_{i=1}^q x_i^2$, x is a $q \times 1$ vector with elements x_i , and $\|\cdot\|_2$ is the l_2 norm of a vector. The term constant represents all other terms which are not a function of γ .

To maximize $\ln p(\gamma|Y, G, A, \lambda, \lambda_0)$ with respect to γ , we only need to maximize

$$L = \ln p(Y|G, A, \gamma) - \frac{1}{2}\lambda\|\gamma_1\|_2^2 - \frac{1}{2}\lambda_0\gamma_0^2.$$

Because $\lambda_0 \rightarrow 0$, L is simplified to

$$L = \ln p(Y|G, A, \gamma) - \frac{1}{2}\lambda\|\gamma_1\|_2^2.$$

This also shows the well-known correspondence between a Bayesian model using a Gaussian prior and a penalized likelihood model with l_2 /ridge penalty.

To maximize L , notice that the first part of the equation is a likelihood function with missing data, or hidden random variables. Therefore we can use the expectation-maximization (EM) algorithm to obtain the maximum likelihood estimate (MLE) of γ . For example, the causal status configuration c can be considered missing. Even though the effect size parameter β is also unobserved, it can be integrated out analytically for quantitative traits or using approximations for binary traits (Servin and Stephens 2007; Guan and Stephens 2008; Chen *et al.* 2015). Compared to the EM method for MLE, the MAP estimate requires only a small change at the maximization (M) step. We first show how to use the EM method for MLE when λ is set to 0 (Kichaev *et al.* 2014; Wen *et al.* 2015). If c were observed, the full-data likelihood would be $p(Y, c|G, A, \gamma) = p(Y|G, c)p(c|A, \gamma)$, where we have used the dependence relationship in Figure 1. Therefore,

$$\ln p(Y, c|G, A, \gamma) = \ln p(Y|G, c) + \ln p(c|A, \gamma).$$

Denote by $p(c|Y, G, A, \gamma^t)$ the conditional distribution of c given the observed data Y, G, A and the estimate γ^t at iteration t . In

the expectation (E) step, we calculate the expectation of the log likelihood of full data on this conditional distribution as

$$E_c[\ln p(Y, c|G, A, \gamma)] = E_c[\ln p(Y|G, c)] + E_c[\ln p(c|A, \gamma)].$$

The first term is not a function of γ , so it is irrelevant to the M step where we maximize this expectation with respect to γ . Therefore, we only need to focus on the second term. Assume given A, γ , the probability of each variant being causal is independent from each other. In implementation, we use a logistic regression model which satisfies this assumption. Then we have:

$$E_c[\ln p(c|A, \gamma)] = \sum_{i=1}^m E_c[\ln p(c_i|A, \gamma)].$$

And because

$$E_c[\ln p(c_i|A, \gamma)] = p(c_i = 1|Y, G, A, \gamma^t) \ln p(c_i = 1|A, \gamma) + p(c_i = 0|Y, G, A, \gamma^t) \ln p(c_i = 0|A, \gamma),$$

we have

$$E_c[\ln p(c|A, \gamma)] = \sum_{i=1}^m [p(c_i = 1|Y, G, A, \gamma^t) \ln p(c_i = 1|A, \gamma) + p(c_i = 0|Y, G, A, \gamma^t) \ln p(c_i = 0|A, \gamma)].$$

In the M step, we update γ by $\gamma^{t+1} = \arg \max E_c[\ln p(c|A, \gamma)]$. If we use a logistic model that has $p(c_i = 1|A, \gamma) = (1/1 + e^{-(\gamma_0 + A\gamma_1)_i})$, $(\gamma_0 + A\gamma_1)_i$ is the i th component of $\gamma_0 + A\gamma_1$, and then the objective function is exactly the form of the log-likelihood function in a logistic regression model. The only difference is that the observed data, $p(c_i = 1|Y, G, A, \gamma^t)$, $i = 1, \dots, m$, are not ones or zeros, but in the range of $[0, 1]$. Therefore, for each iteration, the updated γ can be solved by fitting a logistic model treating $p(c_i = 1|Y, G, A, \gamma^t)$ as the response and A as the data matrix. This was also used in Wen *et al.* (2015). Because our aim is to find the MAP estimate instead of the MLE, the M step changes to:

$$\gamma^{t+1} = \arg \max_{\gamma} E_c[\ln p(c|A, \gamma)] - \frac{1}{2} \lambda \|\gamma_1\|_2^2. \quad (1)$$

The objective function to maximize has the same form of a penalized log likelihood for logistic regression, with the response being $p(c_i = 1|Y, G, A, \gamma^t)$. In our implementation, we use the R package *glmnet* (Friedman *et al.* 2010) to solve the penalized logistic regression model.

For each iteration, $p(c_i = 1|Y, G, A, \gamma^t)$ needs to be recalculated. It is actually the posterior inclusion probability (PIP) for variant i , and can be calculated based on existing methods. Specifically, we first calculate the prior probability of each SNP being causal given the annotations and the estimated annotation effect size using $\text{prior}_i = p(c_i = 1|A, \gamma) = 1/1 + e^{-(\gamma_0 + A\gamma_1)_i}$. Then, any method that can calculate $p(c_i = 1|Y, G, \text{prior})$ can be used. For example, given the prior probability of each

SNP being causal, the original CAVIARBF can output the PIPs for each variant using summary statistics (Chen *et al.* 2015). It needs to be calculated for each iteration because the prior probability of each variant being causal changes depending on the current estimate γ^t .

Extension to LASSO and elastic net penalties

In Equation 1, we can change the l_2 penalty to other penalties, for example, l_1 (e.g., LASSO) penalty or elastic net (ENET), which uses both l_1 and l_2 penalties. These penalties correspond to different prior distributions of γ_1 . Specifically, denote each component of γ_1 by γ_{1i} . If we assume a Laplace prior, $p(\gamma_{1i}|\lambda) = \frac{\lambda}{2} e^{-\lambda|\gamma_{1i}|}$, $i = 1, \dots, m$, and γ_{1i} are independent of each other, then for the MAP estimate of γ , the objective function to maximize becomes

$$L = \ln p(Y|G, A, \gamma) - \lambda \|\gamma_1\|_1,$$

where the l_1 penalty is $\|\gamma_1\|_1 = \sum_{i=1}^m |\gamma_{1i}|$. For ENET penalty, the corresponding objective function is

$$L = \ln p(Y|G, A, \gamma) - \lambda \left[(1 - \alpha) \frac{1}{2} \|\gamma_1\|_2^2 + \alpha \|\gamma_1\|_1 \right],$$

where α is a parameter that determines the mixture of l_1 and l_2 penalties. The corresponding prior is $(\gamma_{1i}|\lambda) \propto e^{-\lambda((1-\alpha)\frac{1}{2}\gamma_{1i}^2 + \alpha|\gamma_{1i}|)}$, assuming the γ_{1i} is independent. Since only the penalty term changes, the same EM method can be applied except that in the M step, a different penalized model is solved. In our implementation, the R package *glmnet* is used to fit all three different penalized models.

Using summary statistics

The same framework can be applied if we have only the summary statistics. Denote the marginal test statistics for individual variants as $Z_{m \times 1}$, of which each component approximately follows a normal distribution. Denote the correlation matrix among variants as $\Sigma_{m \times m}$. As shown in Chen *et al.* (2015), in the fine-mapping setting, using (Y, G) is approximately equivalent to using (Z, Σ) . Therefore when summary statistics are used, simply replace G with Σ , and Y with Z . The posterior probability of γ is

$$p(\gamma|Z, \Sigma, A, \lambda, \lambda_0) \propto p(Z|\Sigma, A, \gamma) p(\gamma_1|\lambda) p(\gamma_0|\lambda_0).$$

To calculate the MAP estimate of γ , similarly as using (Y, G) , we take the limit $\lambda_0 \rightarrow 0$, and maximize the following penalized likelihood

$$L = \ln p(Z|\Sigma, A, \gamma) - \frac{1}{2} \lambda \|\gamma_1\|_2^2,$$

which assumes an l_2 penalty. A similar EM method can be used to obtain the MAP estimate. Specifically, for the M step, we only need to replace $p(c_i = 1|Y, G, A, \gamma^t)$ with $p(c_i = 1|Z, \Sigma, A, \gamma^t)$, which can be calculated using CAVIARBF. Using the l_2 penalty or ENET penalty is also straightforward.

Multi-loci extension

When there are multiple loci of interest, a joint model aggregating annotation information across all loci may be more powerful, as proposed in Kichaev *et al.* (2014) and Wen *et al.* (2015). We consider two scenarios here. For the first scenario, there is a single trait for all loci. This corresponds to fine mapping multiple GWAS hit regions for a single trait. For the second scenario, each locus has its own independent trait. For example, in fine mapping of eQTL, each locus has its own targeted gene expression trait. Figure 2 shows the hierarchical model of K loci incorporating annotation effects with subscripts denoting loci indices. All loci share the same annotation effect γ , which further depends on the parameter λ . Figure 2A corresponds to the model for fine mapping of multiple GWAS hit regions for a single trait and Figure 2B corresponds to the model for fine mapping of eQTLs, where Y_1, \dots, Y_K are gene expression traits and G_1, \dots, G_K are variants usually in a *cis* region of the corresponding genes. We assume independence among Y_1, \dots, Y_K given γ and the corresponding $(A_1, G_1), \dots, (A_K, G_K)$. The two models in Figure 2, A and B, require both the phenotype and individual genotype data. Figure 2C shows the model using only summary statistics, which can be applied for both scenarios. It assumes that the marginal test statistics Z_1, \dots, Z_K are independent from each other given γ and the corresponding $(A_1, \Sigma_1), \dots, (A_K, \Sigma_K)$. For fine mapping GWAS hit regions for a single trait, this assumption is usually reasonable if the genotypes of GWAS hit regions are independent, *e.g.*, they are not too close to each other, or on different chromosomes. For the eQTL model, the independence assumption of Y_1, \dots, Y_K implies the independence assumption of Z_1, \dots, Z_K . If we define $G = \{G_1, \dots, G_K\}$, $Z = \{Z_1, \dots, Z_K\}$, $\Sigma = \{\Sigma_1, \dots, \Sigma_K\}$, and $A = \{A_1, \dots, A_K\}$, $c = \{c_1, \dots, c_K\}$, $\beta = \{\beta_1, \dots, \beta_K\}$, and $Y = \{Y_1, \dots, Y_K\}$ for the eQTL model, then Figure 2, A and B, can be collapsed to Figure 1, and Figure 2C can be collapsed to Figure 1 with Y and G replaced by Z and Σ , respectively.

The same EM method can be used for the MAP estimate of γ . For example, in case of using summary statistics, the likelihood term in the penalized likelihood to maximize is

$$\ln p(Z|\Sigma, A, \gamma) = \sum_{i=1}^K \ln p(Z_i | \Sigma_i, A_i, \gamma).$$

The only change in the EM method is that $p(c_i = 1|Z, \Sigma, A, \gamma^t)$ now depends on observed data from all loci. However, we can show that $p(c_i = 1|Z, \Sigma, A, \gamma^t) = p(c_i = 1|Z_j, \Sigma_j, A_j, \gamma^t)$, where j is the locus where variant i is in, as follows:

$$\begin{aligned} p(c_i|Z, \Sigma, A, \gamma^t) &= \frac{p(c_i, Z, \Sigma, A, \gamma^t)}{p(Z, \Sigma, A, \gamma^t)} \\ &= \frac{p(\gamma^t) p(c_i, Z_j, \Sigma_j, A_j | \gamma^t) \prod_{l=1, l \neq j}^K p(Z_l, \Sigma_l, A_l | \gamma^t)}{p(\gamma^t) \prod_{l=1}^K p(Z_l, \Sigma_l, A_l | \gamma^t)} \\ &= \frac{p(c_i, Z_j, \Sigma_j, A_j | \gamma^t)}{p(Z_j, \Sigma_j, A_j | \gamma^t)} = p(c_i | Z_j, \Sigma_j, A_j, \gamma^t). \end{aligned}$$

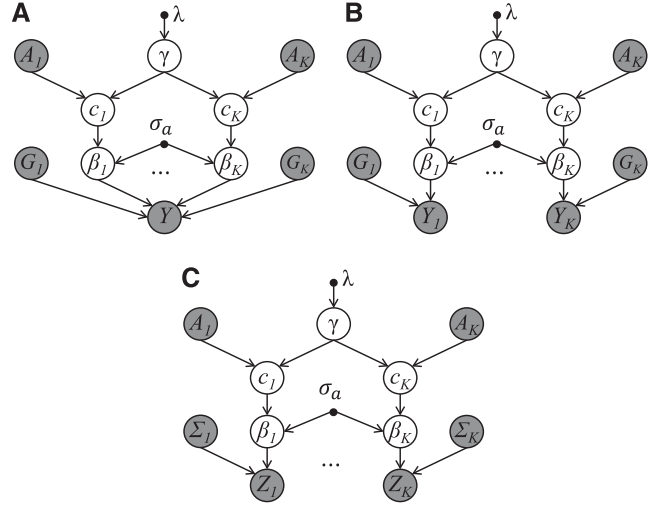


Figure 2 Hierarchical models of K loci incorporating functional annotations. All K loci share the same annotation effects. Shaded \circ indicates observed data. The definitions of symbols are the same as in Figure 1 except the extra subscript indicating the locus index. There are three different models depending on the input data: (A) full genotype and a single phenotype, *e.g.*, a GWAS data set; (B) full genotype and multiple phenotypes for each locus, *e.g.*, a eQTL data set; (C) marginal test statistics and correlations among variants for each locus, which can be derived from either a GWAS data set or an eQTL data set.

In the above, we have assumed that $Z_l, \Sigma_l, A_l, l = 1, \dots, K$ are independent from each other given γ . This means the PIPs can be calculated based on data from each locus, which reduces the computational cost. Similarly, for the eQTL model, we can use $p(c_i = 1|Y, G, A, \gamma^t) = p(c_i = 1|Y_j, G_j, A_j, \gamma^t)$ to compute the PIPs.

Penalty parameters selection using Akaike information criterion, Bayesian information criterion, and cross validation

In the above, we assumed parameter λ is a known fixed value. For real data analysis we need to make it data adaptive for better model fitting. To select the best parameter, we fit the model using different penalty parameters λ and choose the best model based on several criteria, including the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and cross validation (CV). Here the use of AIC and BIC are more heuristic than mathematically strict, however, they may provide easy-to-calculate ways to compare different models. For AIC, we use the formula

$$\text{AIC} = \text{loglik} - df(\lambda),$$

where loglik is the log likelihood of the observed data using the MAP estimate of γ , which is the first term of the penalized likelihood L . $df(\lambda)$ is the effective number of parameters, which will be explained soon. For BIC, we use the formula

$$\text{BIC} = \text{loglik} - \frac{\ln m}{2} df(\lambda),$$

where m is the number of variants. We follow the calculation of effective number of parameters/d.f. in Park and Hastie (2008) and Tibshirani and Taylor (2012). Specifically, for the l_2 penalty,

$$df(\lambda) = \text{tr} \left[(A^T W A + \Lambda)^{-1} A^T W A \right],$$

where $\text{tr}()$ is the trace function on a matrix, Λ is a $(d+1) \times (d+1)$ diagonal matrix with the diagonal entries being λ except the first one being 0. W is a diagonal matrix with diagonal entry $W_{ii} = \hat{\mu}_i (1 - \hat{\mu}_i)$, where $\hat{\mu}_i$ is the final fitted PIPs. For the l_1 penalty,

$$df(\lambda) = \text{rank}(A_c),$$

where A_c is a submatrix of A , with columns selected corresponding to nonzero values in γ . For ENET penalty, based on results derived by Park and Hastie (2008) and Tibshirani and Taylor (2012), we use

$$df(\lambda) = \text{tr} \left\{ [A_c^T W A_c + (1-\alpha)\Lambda]^{-1} A_c^T W A_c \right\}.$$

Compared to AIC and BIC, CV is more time consuming but usually achieves better performance. When we have the genotype- and phenotype-level data, *i.e.*, Y , G are available, we can partition the individuals into subsets, and do CV over different values of λ (and α for ENET penalty). When we only have summary statistics, we can partition the loci into subsets and do CV, as used in Pickrell (2014). The partition based on loci assumes independence among loci. For CV using partitioned loci, we calculate the MAP estimate of γ using the training data, and then use it to calculate the log likelihood of the testing data. The best parameter is chosen based on the maximum of likelihood on the testing data, *i.e.*, the maximum of the sum of the log likelihood of all the testing data.

Significance of individual annotations and estimating marginal effect sizes

Because the above framework incorporates all annotations into the model, whether an annotation is included in the final model and the estimated effect sizes depend on all the available annotations. For screening a large number of annotations, it is common practice to first assess their marginal association significance and effect size. The top ranked annotations may provide functional insight into the underlying mechanism, and can be further combined in the final model (Pickrell 2014). This may produce overly optimistic results and needs validation on external independent data sets. To perform this type of screening, we use the likelihood-based method by setting λ to 0; therefore, no parameter selection is needed and it can be computed quite quickly. To test the significance of individual annotations, we choose the likelihood ratio test of nested models, similarly as in PAINTOR (Kichaev *et al.* 2014). Denote the log likelihood as L_{null} for the null model without any annotation. Denote the log likelihood of adding a single annotation in the model as L_a . Then,

under the null hypothesis that the annotation is not associated with the causal states, $2(L_a - L_{\text{null}})$ approximately follows a $\chi^2(1)$ distribution. To verify the null distribution, we took the eQTL data (see later eQTL data analysis) and simulated 10^5 independent annotations from the standard normal distribution. Supplemental Material, Figure S1 shows the Q-Q plot under the null hypothesis, where the observed and theoretical quantiles match well.

Methods settings

We mainly compared the performance of CAVIARBF with PAINTOR, which can also incorporate functional annotations and showed very competitive performance over other methods (Kichaev *et al.* 2014). Because PAINTOR does not provide a direct way to simultaneously incorporate all annotations, several different strategies of selecting informative annotations were tested. We also compared our method with *fgwas*, assuming one causal variant and focusing on the different annotation selection scheme, where the Bayes factors are almost the same (they are exactly the same under certain settings). We implemented the forward-backward annotation selection proposed in Pickrell (2014) for both *fgwas* and CAVIARBF for the comparison. For the forward-backward annotation selection, we defined the significance for individual annotations using the P -value threshold 0.05. This may result in false inclusion of irrelevant annotations at the first step simply due to randomness given the large number of annotations. The later backward exclusion might reduce this effect. However, we still see obvious inflation of PIP and decreased performance under the null annotations using this search scheme. Because FM-QTL only considers as few annotations as PAINTOR, it is not compared here. All tested analysis methods are described in Table 1.

Without any specific change of settings, here is the default setting for CAVIARBF: We used fivefold CV, when CV is used for penalty parameter selection. We set the maximal number of causal variants in each locus to 3, and the parameter $\sigma_a = 0.1$. For CAVIARBF_ENET_CV, the grid search set for λ and α was set to $\{2^{-15}, 2^{-13}, \dots, 2^5, 10^2, 10^3, 10^4, 10^5, 10^6\} \times \{0, 0.2, 0.3, 0.5, 0.7, 0.8, 1\}$ for simulated data sets. For real data analysis, λ was searched from $\{2^{-2}, 2^{-1}, \dots, 2^{10}\}$ for a finer grid search. For CAVIARBF_L2_CV and CAVIARBF_L1_CV, the search set for λ was the same as that in CAVIARBF_ENET_CV. PAINTOR version 2.1 was used in all simulations and we set the maximal number of causal variants to 3. For lipid data analysis, we assumed the genotype effect sizes follow a normal distribution on the standardized genotypes (mean, 0; variance, 1), to be consistent with the assumption in PAINTOR (Chen *et al.* 2015). But for the eQTL analysis, we assumed the genotype effect size follows a normal distribution on the original scale of genotypes, *i.e.*, the counts of the specified allele.

Data simulation and evaluation

We first simulated a large number of annotations associated with each variant and some of them were correlated. Specifically,

Table 1 Different methods compared in this study

Name format	Examples	Annotation usage	Explanation
<method>_non	CAVIARBF_non, PAINTOR_non, fgwas_non	No	Fit the intercept using the traditional logistic regression in the EM method. For fgwas_non, no annotation is used
CAVIARBF_glm	CAVIARBF_glm	Yes	Fit all annotations using the traditional logistic regression in the EM method
CAVIARBF_<penalty>_<criterion>	CAVIARBF_L2_CV, CAVIARBF_L1_CV, CAVIARBF_ENET_CV, CAVIARBF_ENET_AIC, CAVIARBF_ENET_BIC	Yes	Use penalized models. L2, l_2 penalty; L1, l_1 penalty; ENET, elastic net penalty; <criterion> can be AIC, BIC, and CV
CAVIARBF_fb_CV	CAVIARBF_fb_CV	Yes	Assume one causal variant and use the forward-backward annotation selection proposed in fgwas
<method>_top<k>	PAINTOR_top40	Yes	First, annotations are ranked by their individual increase in likelihood compared to the model without any annotation. Then the top k annotations are used
<method>_top<k>t	PAINTOR_top10t, CAVIARBF_top10t	Yes	All annotations are ranked as in top<k>. Then we sequentially select the annotations with additional two thresholds: (1) the absolute correlations between already selected annotations and the annotation to select are less than a threshold, 0.2 in our simulations; (2) the individual P -value of the annotation to select against the null model without any annotation is less than a threshold, 0.05 in our simulations. This method may not be able to select exactly k annotations due to the additional constraints
<method>_step<k>t	PAINTOR_step5t, PAINTOR_step10t	Yes	The difference between step<k>t and top<k>t is that before selecting a new annotation to include, all remaining candidate annotations are ranked based on their individual contributions in likelihood to the model with all current selected annotations. Specifically, let L_0 be the log likelihood from the model using the current selected annotations and L_i be the log likelihood from the model using the current selected annotations plus candidate annotation i . Then the difference $L_i - L_0$ is the contribution of adding annotation i . It starts with the candidate annotation with the largest contribution and checks the threshold in the same way as in top<k>t. Once an annotation is selected, all the remaining candidate annotations will be evaluated based on the increase of likelihood for the next selection. This method may not be able to select exactly k annotations due to the additional constraints
fgwas	fgwas	Yes	fgwas with annotation and the forward-backward annotation selection

10 groups of annotations were simulated. Each group had 10 correlated annotations, and annotations between groups were independent. Annotations in a group were simulated from a multivariate normal distribution with the pairwise correlation coefficient of 0.4. Then, the probability of being causal for each variant was simulated based on the annotations using a logistic model. Specifically, $p(c_i = 1|A, \gamma) = (1/1 + e^{-(A\gamma)_i})$, where A is the $m \times 101$ annotation matrix including the intercept and all m variants, γ is the annotation effect size vector, $(A\gamma)_i$ is the i th component of $A\gamma$. Due to different annotation effects reported (Carbonetto and Stephens 2013; Kichaev *et al.* 2014; Wen *et al.* 2015), we considered two different settings for the annotation coefficient γ : a small number of annotations with large effects, and a large number of annotations with moderate effects. For the first setting we chose 5 out of 100 annotations with an annotation coefficient/effect size of $\log 5$. For the second setting we chose 40 out of 100 annotations with an annotation coefficient of $\log 1.5$. The rest of the annotations had a coefficient of zero except

the intercept. To ensure the number of causal variants in each locus was not too large, we set the intercept in γ to $\logit(0.01/n_l)$, where n_l is the number of variants in each locus, 35 for multiple loci data simulation. We only kept samples where the maximal number of causal variants in each locus is no more than 5, because having >5 causal variants in a locus may not be very realistic, and we used three maximal causal variants in all simulations and would not like the true underlying model to deviate from our assumption too much. For fine-mapping analysis, similarly to the simulation in Kichaev *et al.* (2014), we only focused on the loci with at least one causal variant. Therefore, the number of loci used for analysis was often smaller than the initial simulated loci. We simulated two different numbers of initial loci: 20 and 100. The average numbers of loci with 1, 2, 3, 4, and 5 causal variants across 100 data sets were summarized in Table S1.

Next we simulated the genotypes for multiple loci. We used HAPGEN (Su *et al.* 2011) to simulate genotypes, using the CEU population from the 1000 Genomes Project (Abecasis

et al. 2012) as the reference panel. We first randomly sampled multiple loci on chromosome 8 and made sure the distance between loci was no less than 1 cM. For each locus, we randomly selected a region of n_l common SNPs (minor allele frequency >0.01) and simulated $n = 10,000$ individuals. The genotypes were additively coded as 0, 1, and 2. Quantitative traits were simulated based on the genotypes and specified causal variants using a linear model. Let $G_{n \times l}$ be the sampled genotype matrix of causal variants in a locus, and $\beta_{l \times 1}$ the corresponding coefficients. Suppose $G_{n \times l}$ is centered with zero mean in each column, then the sample variance of the genotype component of this locus is $[(1/n)\beta^T G^T G \beta]$. We uniformly sampled a value between 1 and 2 and set it to the sample variance and obtained a solution of β , which made sure the variance explained by each locus was similar. The quantitative trait vector for n individuals was sampled from $N(\sum_{i=1}^r G_i \beta_i, \sigma_e^2)$, where r is the total number of loci and G_i and β_i are the causal genotype matrices and coefficients, respectively. We set σ_e^2 according to the total proportion of variance explained (PVE) by all loci. Specifically, we used the following definition of PVE (Guan and Stephens 2011; Zhou *et al.* 2013):

$$PVE = \frac{v}{v + \sigma_e^2},$$

where $v = sv(\sum_{i=1}^r G_i \beta_i)$. Function $sv()$ is the sample variance, *i.e.*, $sv(x) = (1/n)\sum_{i=1}^n (x_i - \bar{x})^2$, where x is a vector with each component denoted by x_i , and \bar{x} is the mean of all values in x . Therefore $\sigma_e^2 = ((1 - PVE)/PVE)v$. We set $PVE = 0.25$ for 100 initial loci and $PVE = 0.05$ for 20 initial loci. Note that the way we simulated genotype effects and the traits were different from that in Chen *et al.* (2015). Here we did not enforce the noncentrality parameters of causal variants to be in a certain range, which may not be easy to control when simulating a large number of loci. Similar as in Kichaev *et al.* (2014), we also did not require a significance threshold P -value of 5×10^{-8} for each locus.

One interesting question is whether we gain much by incorporating annotation information if we only have one GWAS hit, which is not uncommon in real data. To address this question, we also simulated 100 data sets with only one locus. We set $PVE = 0.01$ and $n_l = 250$. The simulation was the same as above except the following: When simulating 40 informative annotations with the effect size $\log 1.5$, the intercept was set to $\text{logit}(0.005/n_l)$, which ensured the number of causal variants was not too large. The total numbers of data sets with 1, 2, 3, 4, and 5 causal variants across 100 data sets were 31, 28, 21, 16, and 4, respectively, for five informative annotations. The distribution of causal variants was similar for 40 informative annotations.

We simulated 100 data sets for each scenario: two settings for annotation effect sizes and two settings for initial number of loci. To compare different methods, we calculated the proportion of causal variants included among 100 data sets when selecting the top SNPs ranked by PIPs.

We also simulated binary annotation with 100 initial loci when comparing CAVIARBF with *fgwas*. We first simulated the continuous annotations as described above and then dichotomized them into binary annotations using a random threshold. Everything else was the same as simulating continuous annotations. We also simulated scenarios under the null where no annotation was informative. The data were simulated in the same way as that with binary annotations except that the annotation effect size was zero.

Lipid data

The lipid data provides the meta-analysis results on four phenotypes: total cholesterol (TC), low-density lipoprotein cholesterol (LDL), high-density lipoprotein cholesterol (HDL), and triglycerides (TG) (Teslovich *et al.* 2010). To include as many SNPs as possible in the analysis, we downloaded the imputed results using ImpG (Pasaniuc *et al.* 2014) from <http://bogdan.bioinformatics.ucla.edu/software/impG/>. SNPs with potential ambiguous allele coding, *i.e.*, A/T or C/G, and SNPs with a meta-analysis sample size of $<80,000$ were removed before applying ImpG. We extracted the hit regions using a 100-kb window, the same sized region as used in the previous analysis (Kichaev *et al.* 2014), with the reported lead SNPs (Teslovich *et al.* 2010) in the center. We excluded several hit regions where SNPs had high correlations (>0.9 calculated using EUR populations from the 1000 Genomes Project) but showed large differences in Z statistics. In the end, we had 30 hit regions for LDL, 41 for HDL, 45 for TC, and 26 for TG.

Annotations used for lipid data

We downloaded the same annotation used in a previous study (Pickrell 2014) from <https://github.com/joepickrell/1000-genomes>. In brief, there are 451 annotations, where 450 are binary and only 1 is quantitative [distance to the nearest transcription start site (TSS)]. The annotations are mainly maps of DHSs in a wide range of primary cell types and cell lines. It also includes the predicted genome segmentations from six ENCODE cell lines (Hoffman *et al.* 2013). We normalized the quantitative annotation to have a mean of 0 and a variance of 1.

eQTL data

In the first stage, we identified all potential genes related to prostate cancer (PrCa)-risk SNPs for each risk interval. Specifically, from 146 reported PrCa-risk SNPs, we included a total of 6324 risk and in-linkage-disequilibrium (LD) tag SNPs ($r^2 > 0.5$) in 100 unique risk intervals. After quality control and normalization, both RNA sequencing (RNA-seq) data of normal prostate tissue and genotype data were obtained from 471 individuals with European ancestry. Samples were genotyped using Illumina Infinium 2.5 M bead arrays and then imputed using SHAPIT and IMPUTE2 with reference files from the 1000 Genomes Phase I integrated variant set. We analyzed all SNPs in the risk intervals for association with surrounding gene transcripts. In total there were 127,276

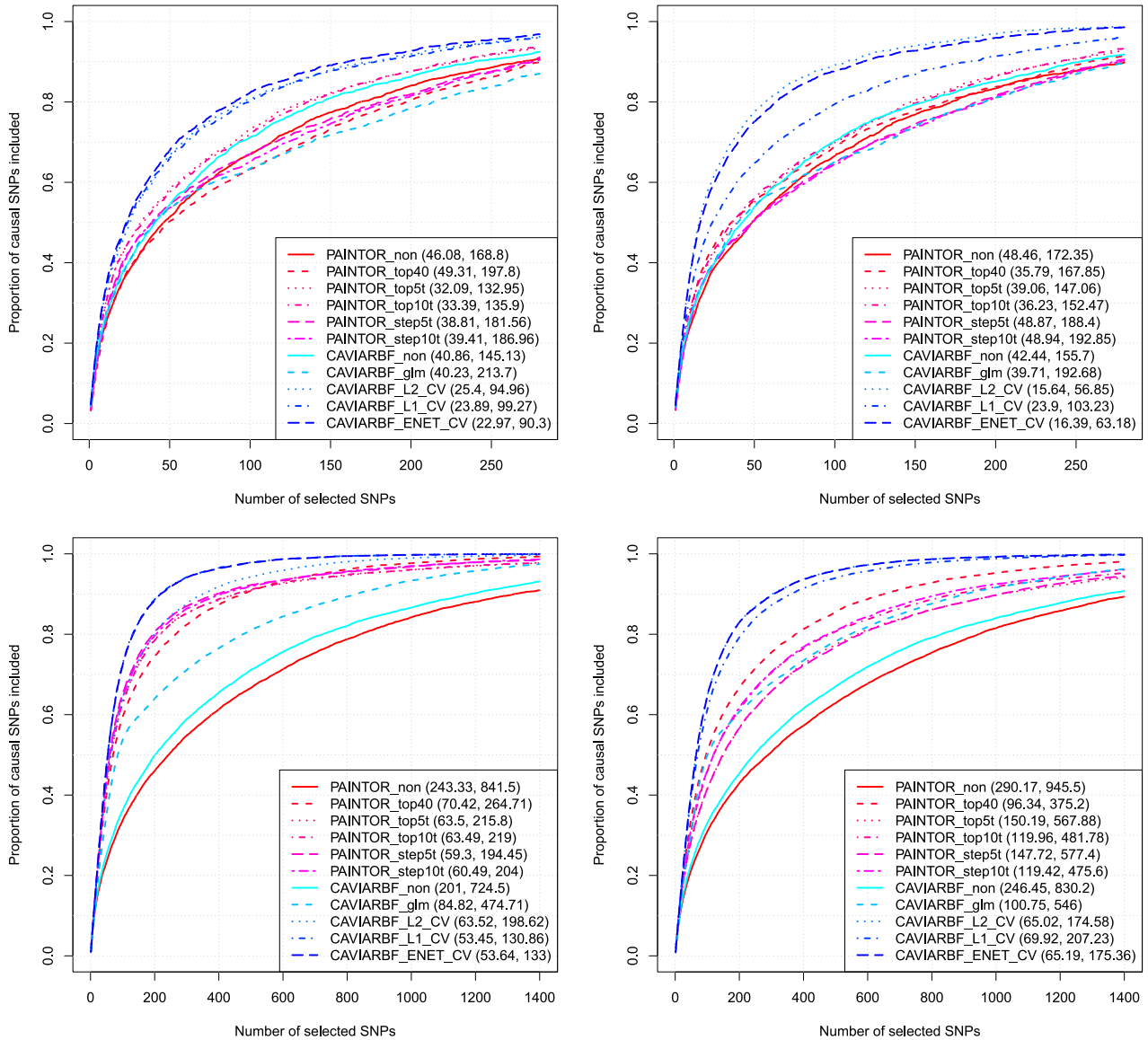


Figure 3 Proportion of causal variants identified by different methods. The top panels correspond to 20 initial loci (~12 causal loci) and the bottom 5panels correspond to 100 initial loci (~65 causal loci). The left column corresponds to 5 informative annotations with large effects, and the right column corresponds to 40 informative annotations with moderate effects. The numbers shown in the parentheses correspond to the average number of SNPs required to identify 50 and 80% of the causal SNPs, respectively.

SNP-gene pairs with 51 out of 100 risk intervals demonstrating a Bonferroni-significant eQTL signal (P -value $< 1.96 \times 10^{-7}$) and these were associated with 88 genes. More details on the data collection and analysis can be found in Thibodeau *et al.* (2015).

In the second stage, to fully identify all *cis*-eQTLs for these 88 target genes, we chose all candidate SNPs within 1.1 Mb of each gene's TSS or transcription end site positions. Because some candidate SNPs overlapped in multiple target genes, these violated our assumption of independence among loci. As a proof of principle, we selected 41 target genes with non-overlapping candidate SNPs in this analysis. The phenotype of each target gene is the residual from regressing normalized gene expression level on covariates: the histologic character-

istics, percent lymphocytes, percent epithelium, and 14 expression principal components. To reduce computational burden, we first filtered candidate SNPs using a P -value threshold of 1×10^{-4} and kept a maximum of 250 SNPs for each target gene/locus.

Annotations used for eQTL data

We collected 76 annotation tracks that may be related to *cis*-eQTLs in normal prostate tissue. These include predicted chromatin state using ChromHMM (Ernst and Kellis 2012) in breast and prostate normal and cancer cells (Taberlay *et al.* 2014), DHSs from the normal and cancer prostate cells from the ENCODE project (The ENCODE Project Consortium 2012), histone modifications such as histone acetylation,

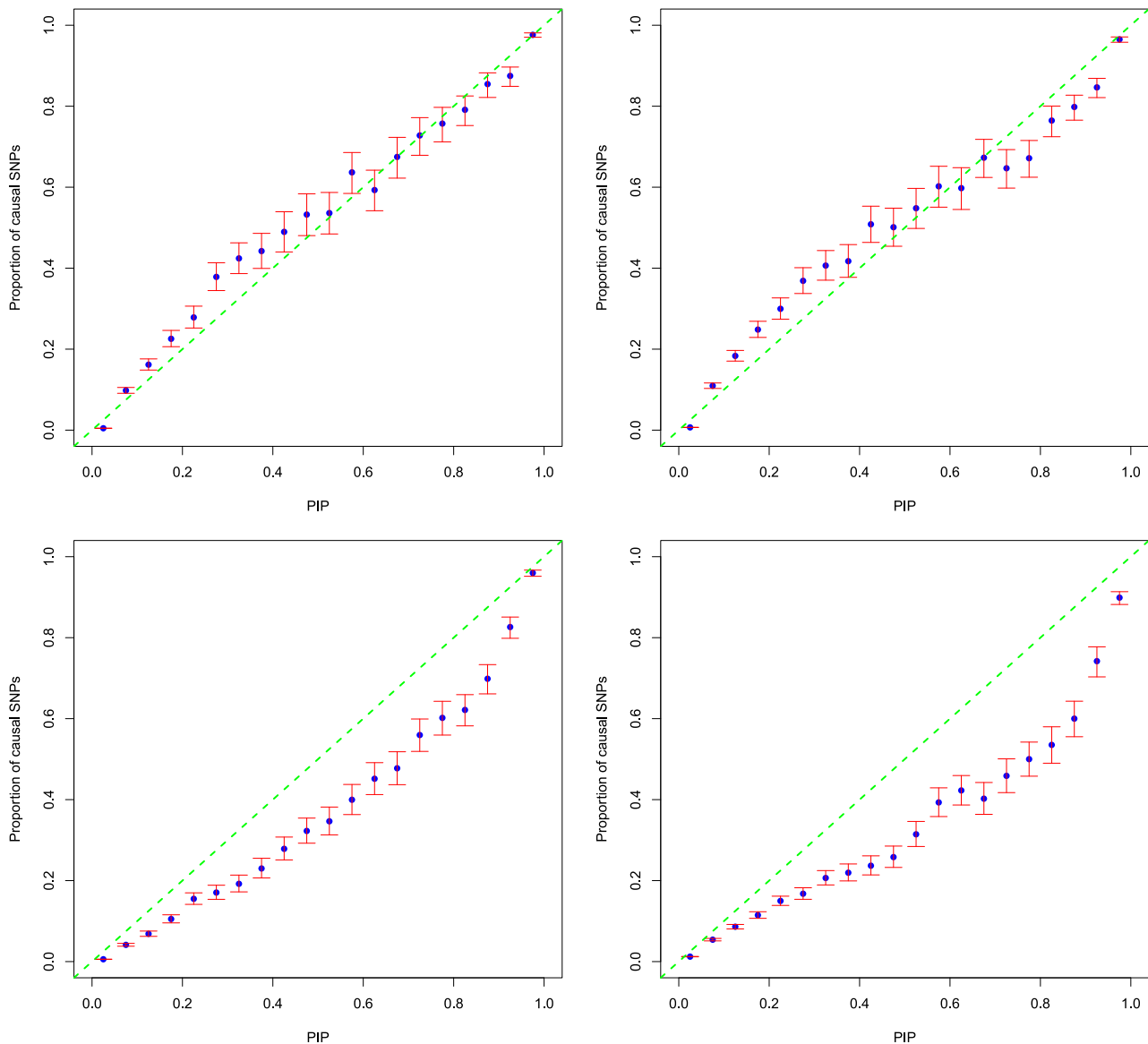


Figure 4 PIP calibration with 100 initial loci (~65 causal loci). The top two panels show results from CAVIARBF_ENET_CV, bottom two panels are from PAINTOR_top5t. Left two panels, 5 informative annotations with effect size $\log 5$; right two panels, 40 informative annotations with effect size $\log 1.5$. The x-axis shows the center of 20 bins of width 0.05. The y-axis is the proportion of causal SNPs. The blue points show the proportion of causal SNPs in each bin. The red bars show the 95% C.I. (Wilson inversion of score statistic) of the proportion assuming a binomial distribution in each bin. 100 data sets were used in each panel.

histone methylation (He *et al.* 2010; Wang *et al.* 2011; Bert *et al.* 2013; Hazelett *et al.* 2014), and transcription factor bindings (Tan *et al.* 2012; The ENCODE Project Consortium 2012; Sharma *et al.* 2013; Hazelett *et al.* 2014; Takayama *et al.* 2014). All annotations were aligned to human genome version 19 coordinates. For each category of the predicted chromatin state from ChromHMM, we created a binary annotation where 1 indicates the SNP is in that category and 0 otherwise. Other annotations are quantitative. For simplicity and proof of principle, we set the missing annotation values to 0. More advanced annotation imputation (Ernst and Kellis 2015) may achieve higher imputation accuracy. We normalized all annotations to have mean 0 and variance 1.

Data availability

The detailed annotation list and Gene Expression Omnibus track numbers are shown in File S2. The summary statistics and annotation matrix used of the eQTL data are available upon request. Software is available at <https://bitbucket.org/Wenan/caviarbf>.

Results

Proportion of causal variants identified by different methods

Results for all analyses are shown in Figure 3. The top panels correspond to 20 initial loci and the bottom panels correspond

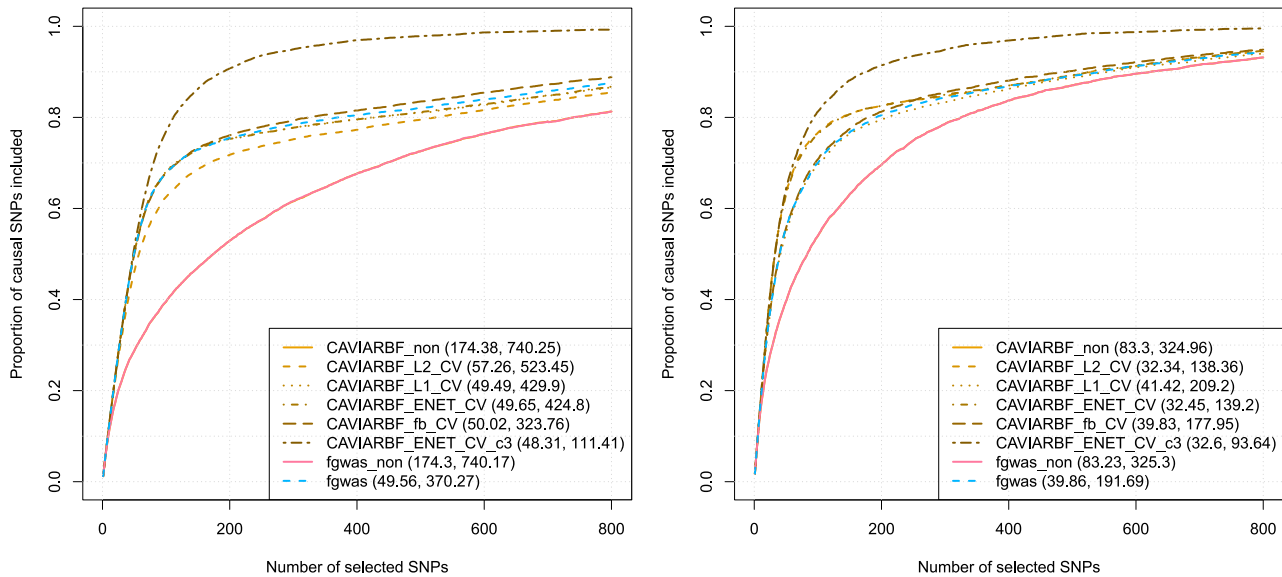


Figure 5 Proportion of causal variants identified by CAVIARBF and *fgwas*. The left panel corresponds to 5 informative annotations with large effects, and the right panel corresponds to 40 informative annotations with moderate effects. The numbers shown in the parentheses correspond to the average number of SNPs required to identify 50 and 80% of the causal SNPs, respectively. CAVIARBF_ENET_CV_c3 assumes three causal variants.

to 100 initial loci. The left column corresponds to five informative annotations with large effects, and the right column corresponds to 40 informative annotations with moderate effects. CAVIARBF applied using the ENET penalty with CV (CAVIARBF_ENET_CV) is the most robust and among the best methods across all scenarios. When the number of informative annotations is small and their effects are large, CAVIARBF using the ENET penalty usually has similar performance as that using the l_1 penalty. On the other hand, when the number of informative annotations is large and their effects are moderate, CAVIARBF using the ENET penalty usually has similar performance as that using l_2 penalty. This is consistent with the characteristics of these penalties (Tibshirani 1996; Zou and Hastie 2005). As expected, for both CAVIARBF and PAINTOR, including annotations increases the probability of identifying the causal variants. CAVIARBF_glm, which uses only the MLE without any penalty term when incorporating the annotations, has much lower accuracy compared to all penalized models, sometimes even lower than CAVIARBF_non (Figure 3, top two panels). This demonstrates the overfitting problem when fitting a large number of annotations without regularization. The same problem can be observed for PAINTOR. For example, PAINTOR including the top 40 annotations (PAINTOR_top40) has worse performance than PAINTOR applied with no annotation (PAINTOR_non) in the top left panel. This shows the importance of using penalties in the model when incorporating a large number of annotations. There is no single best annotation selection method for PAINTOR; the performance depends on the number of loci and the number of informative annotations. PAINTOR_top5t and PAINTOR_top10t appear to be good options in general, even though when the number of informative annotations and loci are large (bottom right panel),

simply choosing the top 40 annotations (PAINTOR_top40) achieves the best result for PAINTOR. The advantage of CAVIARBF_ENET_CV over PAINTOR is larger when the number of informative annotations is large. This illustrates the benefit of incorporating all annotations with regularization instead of only using the top annotations. When no annotation is used, CAVIARBF has slightly better performance than PAINTOR, although the magnitude of differences varies among different scenarios.

PIPs calibration

To assess the accuracy of estimated PIPs, we put SNPs into bins according to their PIPs and then compared the proportion of causal SNPs in each bin with the center PIP of that bin (Guan and Stephens 2011). Figure 4 shows the results using the data sets with 100 initial loci. Note that the initial loci include all causal and noncausal loci. The final data sets used in testing include only causal loci (Table S1). This follows the pattern of simulation used in PAINTOR (Kichaev *et al.* 2014). The left two panels represent the scenario of five informative annotations with an effect size of $\log 5$, the right two panels represent the scenario of 40 informative annotations with an effect size of $\log 1.5$. PIPs from CAVIARBF_ENET_CV, the top two panels with different annotation patterns, are calibrated well in general; while PIPs from PAINTOR_top5t, the bottom two panels, are overestimated. PIPs from CAVIARBF_ENET_CV show some downward bias in the middle, *e.g.*, from 0.2 to 0.5, resulting in conservative PIPs. However, this is preferable to upward bias which results in overly optimistic PIPs. When there are a large number of annotations with moderate effects (top right), CAVIARBF_ENET_CV shows a small upward bias, *e.g.*, from 0.7 to 0.9. PAINTOR_top5t shows consistent upward bias, which is larger in the middle and

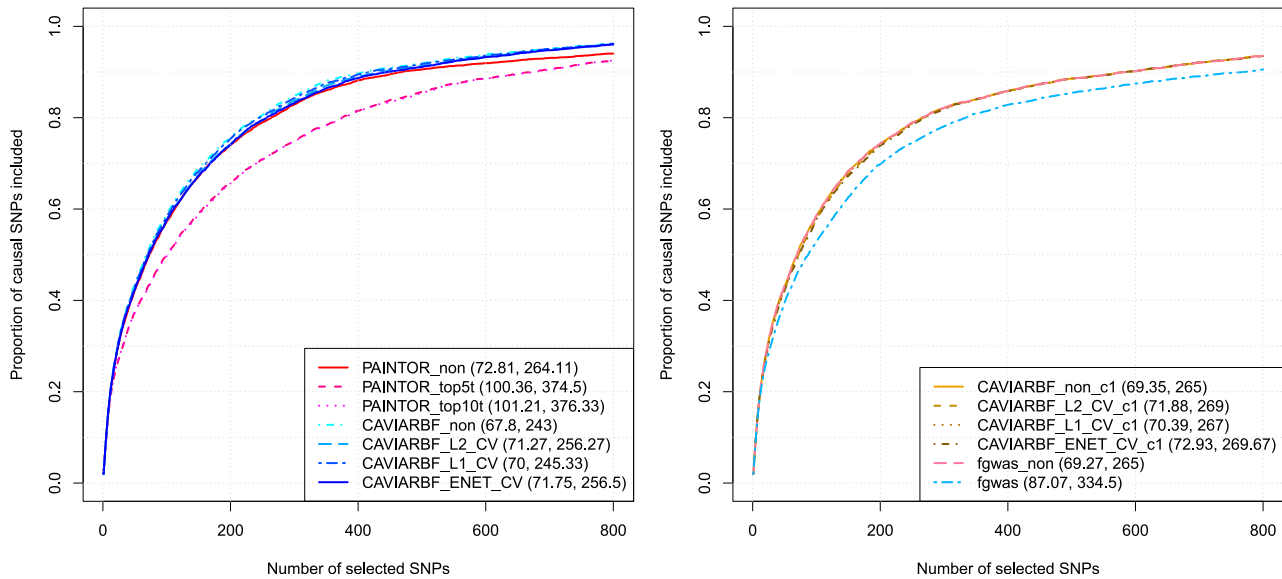


Figure 6 Proportion of causal variants identified when the annotation effect size is zero. The left panel shows comparison between PAINTOR and CAVIARBF assuming three maximal causal variants. The right panel shows comparison between *fgwas* and CAVIARBF assuming one causal variant.

becomes smaller when PIPs are close to one. We note that the number of loci harboring at least one causal variant has an influence on the calibration and accuracy of PIPs. Figure S2 and Figure S3 show the results for 40 and 20 initial loci, corresponding to ~ 25 and 12 loci on average harboring at least one causal variant (Table S1), respectively. For these small numbers of loci, even though CAVIARBF_ENET_CV always shows better calibration than PAINTOR_top5t, there exists upward bias, e.g., in the range of 0.6–0.9. Therefore, when incorporating annotations and the number of loci harboring at least one causal variant is < 20 , we need to interpret the PIPs with caution and consider the potential upward bias. We discuss potential sources of these biases in the Discussion section.

Comparison with *fgwas* assuming one causal variant

To make sure the Bayes factors were equivalent, we set W in *fgwas* to 0.01, σ_a to 0.1, and SNP variance as the weight input in CAVIARBF. A 10-fold CV was used for CAVIARBF as in *fgwas*. Figure 5 shows the results. When no annotation was used, CAVIARBF_non and *fgwas*_non had the same performance (CAVIARBF_non is overlapped and not visible). This also reflects their equivalence in Bayes factors. When there were five informative annotations with large effects (left panel), *fgwas* and CAVIARBF_ENET_CV showed similar results when the number of selected SNPs was small, and *fgwas* was slightly better when the number of selected SNPs was large. When there were 40 informative annotations with moderate effects (right panel), CAVIARBF_ENET_CV showed better performance than *fgwas*. CAVIARBF_fb_CV uses the same forward-backward annotation selection scheme as in *fgwas*. The performance was similar to *fgwas* in general, even though CAVIARBF_fb_CV was slightly better when the number of selected SNPs was large. The reason is not clear to us. It

might be due to a different optimization, where CAVIARBF uses the EM method and *fgwas* uses a gradient search. In both annotation scenarios, CAVIARBF_ENET_CV_c3, assuming three maximal causal variants, showed much better performance than those assuming one causal variant. The forward-backward annotation selection was faster than CAVIARBF_ENET_CV. It might be a good option when CAVIARBF_ENET_CV takes too much time or we know that only a small portion of annotations are informative among candidate annotations.

Performance under null annotations

Figure 6 (left panel) shows the performance of PAINTOR and CAVIARBF. CAVIARBF with different penalties showed similar or slightly decreased performance compared with CAVIARBF_non. CAVIARBF_ENET_CV performed similarly to CAVIARBF_non in terms of ranking causal variants when there were no informative annotations. On the other hand, the performance of PAINTOR_top5t or PAINTOR_top10t was far poorer than PAINTOR_non. Figure 6 (right panel) shows the comparison between *fgwas* and CAVIARBF assuming one causal variant. CAVIARBF with different penalties had similar performance, but *fgwas* had decreased performance compared to *fgwas*_non. Figure S4 shows the result of PIP calibration. Assuming three maximal causal variants, CAVIARBF_non (top left) shows good calibration, and CAVIARBF_ENET_CV (top right) has some inflated PIP estimation. The pattern is similar for CAVIARBF assuming one causal variant (middle panels). For *fgwas* without annotations, it shows good calibration (bottom left), however, when null annotations are included (bottom right), the PIP is more inflated than CAVIARBF. We also compared the number of SNPs with PIP > 0.9 . Figure S5 (top panel) shows the histogram of the difference between CAVIARBF_ENET_CV and CAVIARBF_non. Out of 100 data sets, 68 had exactly the same number between

CAVIARBF_ENET_CV and CAVIARBF_non. There are only 8 out of 100 data sets where CAVIARBF_ENET_CV has more than three SNPs than CAVIARBF_non. Given that the average number of causal variants was 50, the increase of the number of SNPs >0.9 for CAVIARBF_ENET_CV under the null was either zero or very small. We also compared the number of SNPs that reached 90% of the expected causal variants (called “90% confidence set” as in Table 2 in Kichaev *et al.* 2014, although it is not much related to a confidence level or a credible set if the number of causal variants in each locus is more than one). Figure S5 (bottom panel) shows the histogram of the difference. Out of 100, 40 had exactly no difference. However, there were also ~ 36 data sets with a reduction of SNPs >30 by CAVIARBF_ENET_CV. Therefore, the reduction of SNPs by the current implementation of CAVIARBF_ENET_CV does not provide convincing evidence of the informativeness of annotations. After examination of the difference between CAVIARBF_non and CAVIARBF_ENET_CV, we conclude this is due to falsely choosing a small penalty parameter instead of a large penalty parameter, even though the CV likelihood increased very little. Better ways to avoid this may be useful, such as choosing the largest penalty that achieves similar CV likelihood as the maximal CV likelihood.

Comparison of AIC, BIC, and CV for penalty parameters selection

We also compared performance using different model criteria including AIC, BIC, and CV (Figure S6). The performances of CV are better than that of using AIC or BIC. In terms of computing time, CV is more computationally intensive than AIC or BIC. Even though AIC and BIC are not the best options, incorporating annotations using AIC or BIC can still achieve improved performance than that without annotations. For example, the performance of methods using AIC is consistently better than CAVIARBF_non. In most situations, the performance of methods using AIC is better than or similar to those using BIC, which are usually close to CAVIARBF_non. This may be due to too much penalization in BIC. The only exception is when the number of informative loci is 5 and the total initial loci is 100 (bottom left panel), for methods using l_1 or ENET penalties, the performance using BIC is better than that using AIC.

Performance when there is only one GWAS hit

Figure 7 shows the results incorporating annotations when there is only one locus of interest. Because CV partitions loci, it requires at least two loci, so only AIC or BIC can be used for parameter selection. There is not much difference between CAVIARBF_non and CAVIARBF using annotations. This is reasonable because the number of causal variants in a single locus is very small, no more than 5, even though the number of noncausal variants is large. The information from the causal/noncausal contrast is limited, which makes it hard to determine informative annotations and make use of them. It is similar to the situation where there are only a few cases but

Table 2 Computing time of different methods in seconds

PAINTOR		CAVIARBF		fgwas	
<i>c</i> = 1					
PAINTOR_non	0.46	CAVIARBF_non	1.8	fgwas_non	0.2
PAINTOR_top10t	50	CAVIARBF_top10t	259	fgwas	454
		CAVIARBF_L2_CV	1,341		
		CAVIARBF_L1_CV	1,789		
		CAVIARBF_ENET_CV	10,727		
<i>c</i> = 3					
PAINTOR_non	9	CAVIARBF_non	10		
PAINTOR_top10t	1151	CAVIARBF_top10t	621		
		CAVIARBF_L2_CV	3,126		
		CAVIARBF_L1_CV	6,541		
		CAVIARBF_ENET_CV	46,184		

a large number of controls in a logistic regression, in which case it is difficult to estimate the effects of covariates or determine their significance. And this is exactly what happens in the computation: for each iteration of the EM method, a logistic regression is fitted with the PIPs as the outcome and annotations as the covariates. Because PAINTOR uses the MLE without regularization, simply choosing a large number of annotations, *e.g.*, PAINTOR_top40, severely deteriorates the performance. Even with careful screening of annotations considering correlation and significance levels, *e.g.*, PAINTOR_top10t or PAINTOR_step10t, the performance can still be much worse than that using no annotations. This also illustrates the advantage of the proposed penalized model or Bayesian framework: it is adaptive to the information available in the annotation. When there is very little information in the annotation, the penalized framework has similar results as that using no annotation. When there is enough information in the annotation, the penalized framework can use it and achieve better performance.

Fine mapping when the true correlation matrix is not available

Sometimes only the marginal Z statistics resulting from association analysis are available, for example, meta-analysis results based on marginal Z statistics of multiple studies without direct access to the underlying genotype data. If we assume there is at most one causal variant for each locus, CAVIARBF reduces to the ABF (Wakefield 2009). Under this assumption, both CAVIARBF and PAINTOR do not need the correlation matrix. Alternatively we can choose a reference population similar to the population being studied, *e.g.*, from the 1000 Genome Project, and use the correlation calculated from the reference population as an approximation. Applying penalization/shrinkage on the correlation matrix may also improve the performance (Wen and Stephens 2010; Chen and Schaid 2014; Pasaniuc *et al.* 2014). Specifically in our simulation, we added a small positive value to the main diagonal of the correlation matrix calculated from the haplotypes of 379 EUR subjects in the 1000 Genome Project. We

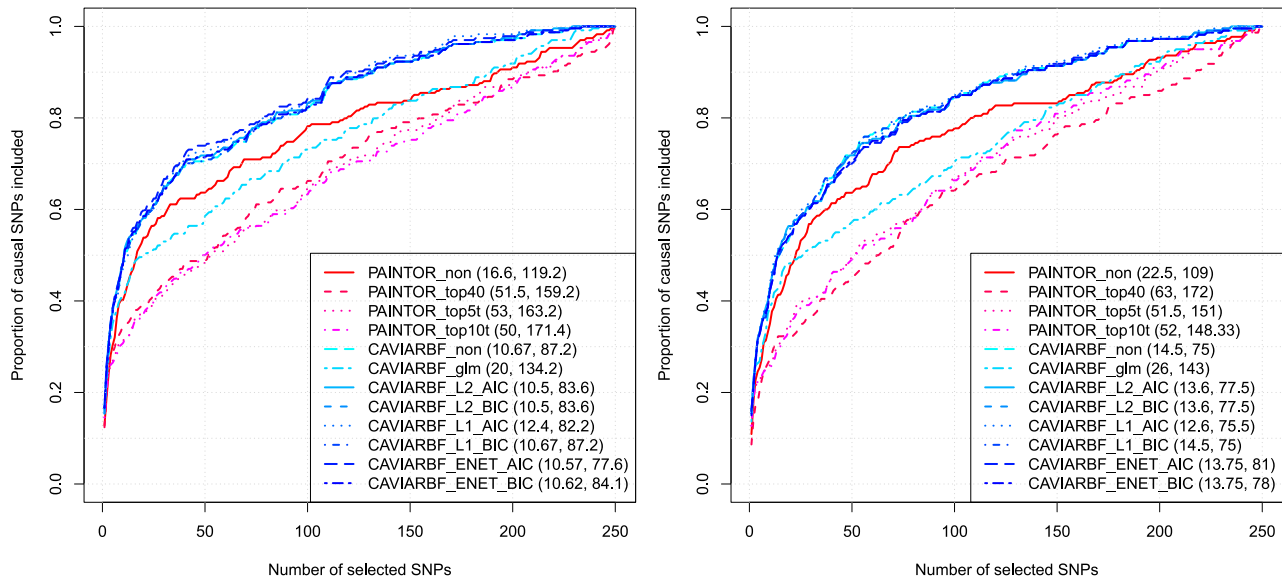


Figure 7 Proportion of causal variants identified when there is only one locus. The left column corresponds to 5 informative annotations with large effects, and the right column corresponds to 40 informative annotations with moderate effects. The numbers shown in the parentheses correspond to the average number of SNPs required to identify 50 and 80% of the causal SNPs, respectively.

empirically set this positive value to 0.2; using other values showed similar conclusion.

To compare different options for the unknown correlation matrix, we simulated data sets as described above using 40 initial loci and set the *PVE* to 0.1. In total, 100 data sets were simulated for each type of annotation setting and ~25 loci on average harboring causal variants were generated (Table S1). This number of loci matches better with that of the lipid data we will analyze later than using 20 or 100 initial loci.

We used either one or three maximal number of causal variants, denoted by *c1* or *c3*, respectively. Results are shown in Figure 8. The left two panels correspond to five informative annotations with an effect size of $\log 5$, and the right two panels correspond to 40 informative annotations with an effect size of $\log 1.5$. We first compare different options of PAINTOR and CAVIARBF using the top10t strategy (the top two panels). Unsurprisingly, for both CAVIARBF and PAINTOR, using the true correlation matrix (CAVIARBF_top10t_c3 and PAINTOR_top10t_c3) perform the best compared to other options. When the true correlation matrix is not available, CAVIARBF using a well-matched reference panel with shrinkage (CAVIARBF_top10t_c3_EUR0.2) achieves the best performance. Using the shrinkage on the reference correlation matrix significantly improves the performance compared to no shrinkage (CAVIARBF_top10t_c3_EUR). Even though assuming at most one causal variant is not optimal, it is better than using the reference panel without shrinkage. This can be useful if a well-matched reference panel is not available. For the same option, CAVIARBF has similar or slightly better performance than PAINTOR. Next we compare the top10t strategy with ENET using CV (the bottom two panels). CAVIARBF using the ENET to incorporate all

annotations (CAVIARBF_ENET_CV_c3_EUR0.2) further improves the performance of CAVIARBF_top10t_c3_EUR0.2, and is very close to the performance assuming known correlation matrix (CAVIARBF_ENET_CV_c3). When assuming at most one causal variant, ENET using CV (CAVIARBF_ENET_CV_c1) is in general still better than the top10t strategy (CAVIARBF_top10t_c1). We noted that PAINTOR 2.1 uses a new default strategy of setting the noncentrality parameters (NCP). Switching to the old setting of NCP showed a slight improvement but did not change the general pattern of the results.

Figure 9 shows the PIP calibration of CAVIARBF_ENET_CV_c3_EUR0.2 (two top panels), CAVIARBF_ENET_CV_c1 (two middle panels), and PAINTOR_top10t_c1 (two bottom panels). The left column corresponds to five informative annotations with effect size $\log 5$, and the right column corresponds to 40 informative annotations with an effect size of $\log 1.5$. In general, CAVIARBF_ENET_CV_c3_EUR0.2 and CAVIARBF_ENET_CV_c1 show slightly better calibration than PAINTOR_top10t_c1, especially when the annotation's effect size is moderate (the right column). The PIP of CAVIARBF_ENET_CV_c3_EUR0.2 and CAVIARBF_ENET_CV_c1 is a little inflated when its value is >0.5 . For high PIPs, e.g., $\text{PIP} > 0.95$, PIPs of all three methods are closer to the true proportion of causal variants in the data sets.

In summary, it is best to use the true correlation matrix when it is available. When the true correlation matrix is not available but there is a well matched reference panel, the correlation matrix from the reference panel with shrinkage achieves the best performance. To find a well-matched reference panel, we can use the continental populations from the 1000 Genomes Project, e.g., matching samples with European ancestry with the EUR population. When the (imputed) genotypes of a subsample are available, we can use correlations

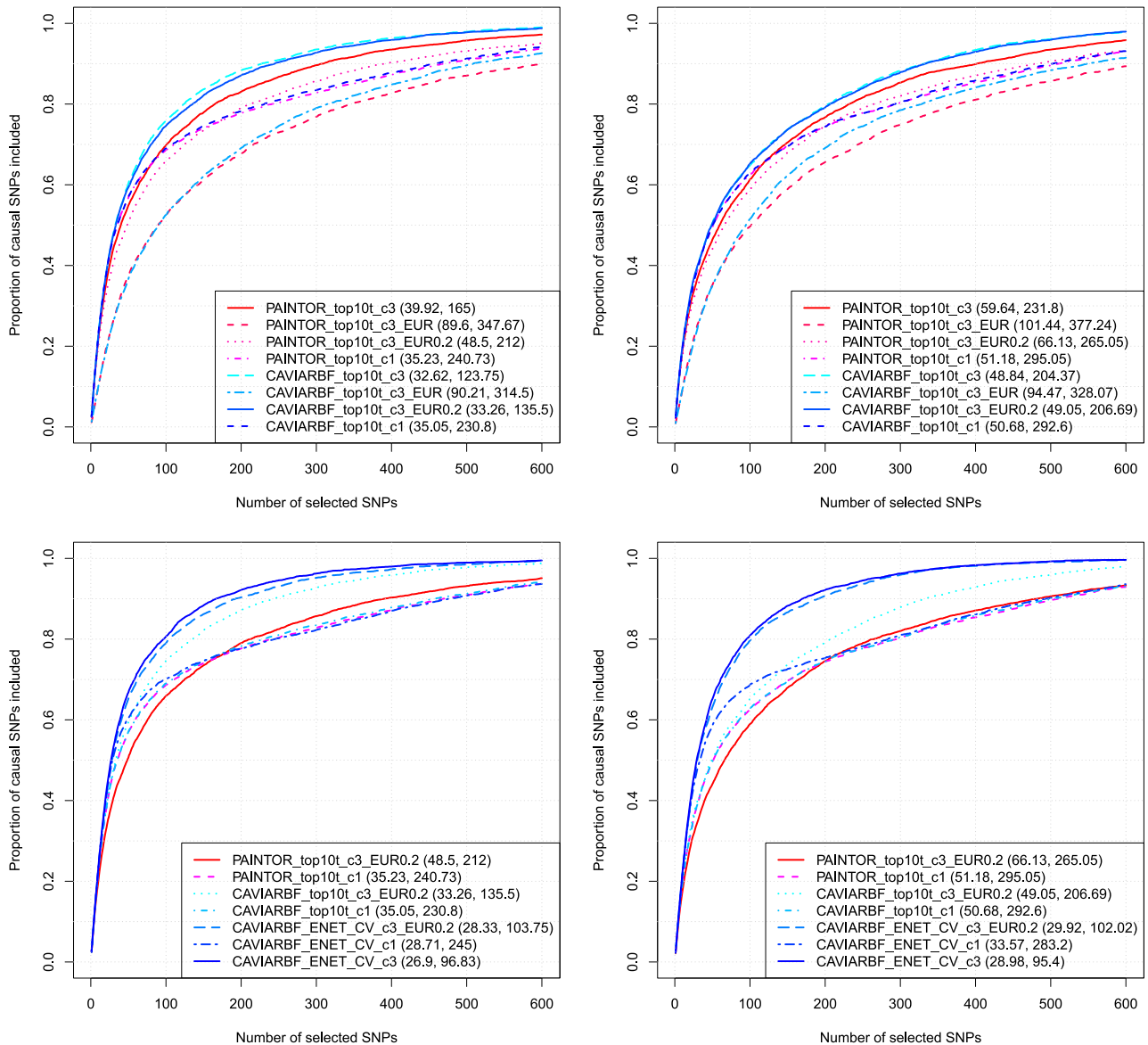


Figure 8 Proportion of causal variants identified with and without the true correlation matrix or with reference correlation matrix. The two panels on the left correspond to 5 informative annotations with effect size $\log 5$, and the two on the right correspond to 40 informative annotations with effect size $\log 1.5$. The numbers shown in the parentheses correspond to the average number of SNPs required to identify 50 and 80% of the causal SNPs, respectively. c3 and c1 indicates that the maximal number of causal variants is 3 or 1, respectively. EUR means using the correlation matrix calculated from the EUR population in the 1000 Genomes Project and EURO.2 means adding 0.2 to the main diagonal of the correlation matrix calculated from the EUR population.

computed from the subsample as an approximation of that from the full data. Otherwise, assuming at most one causal variant and using ENET to incorporate all annotations seems to be the current best practice.

Estimated annotation effect sizes and bootstrap-based confidence intervals

Figure S7 shows the distribution of estimated annotation effect sizes averaged over 100 data sets for CAVIARBF_ENET_CV. For null annotations, the median of the estimate is close to zero with both positive and negative variations; for

informative annotations, the estimate is more likely have the correct sign but with shrinkage of the effect size. It is clear that the estimated coefficients reflect the underlying annotation status in the averaged plot. However, in real data analysis, we need to distinguish random variation from true nonzero annotation effect size. We tried to use a bootstrap method to estimate the confidence interval of estimated annotations effect sizes. Specifically, we resampled the loci with replacement and calculated the confidence interval using the percentile method. Even though bootstrap-based confidence intervals can be problematic for penalized models

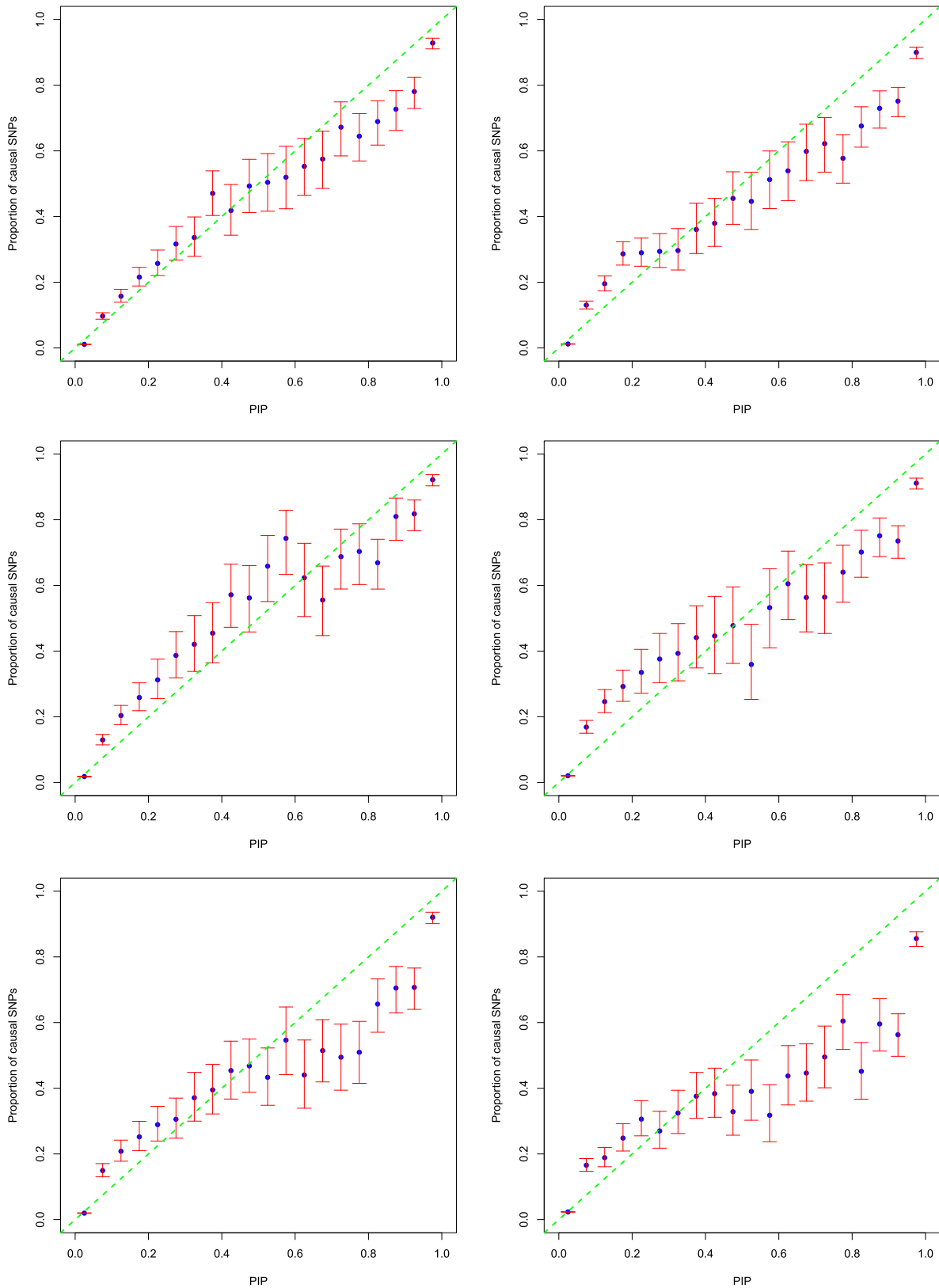


Figure 9 PIP calibration of CAVIARBF_ENET_CV_c3_EUR0.2 (two top panels), CAVIARBF_ENET_CV_c1 (two middle panels), and PAINTOR_top10t_c1 (two bottom panels). c3 and c1 indicates that the maximal number of causal variants is 3 or 1, respectively. EUR0.2 means adding 0.2 to the main diagonal of the correlation matrix calculated from the EUR population. The left column of panels, 5 annotations with effect size $\log 5$; the right column of panels, 40 annotations with effect size $\log 1.5$. The bins and calculation of C.I.s are the same as in Figure 4.

(Kyung *et al.* 2010), we still tried to assess how helpful it could be for our model. Table S2 shows the coverage of 90% bootstrap confidence intervals based on 20 data sets and averaged over different categories of SNPs. For five informative annotations with large effects, the coverage for nonnull SNPs does not reach the specified level. One challenge is the biased (shrunk) estimation resulting from penalized models. For 40 informative annotations, the coverage for nonnull SNPs is close to the nominal. However, it can be hard to distinguish the nonnull from null annotations based on the confidence intervals. One potential reason may be the small effect sizes. Figure S8 shows the individual bootstrap confidence intervals for each SNP. In our applications in this article, we simply rank annotations based on the size of the effects (corresponding to the standardized annotations). Better ways to rank the annotations and distinguish between null and nonnull annotations will need further investigation. For example, a fully Bayesian approach might provide a valid interval (credible interval).

Computing time

Table 2 shows the computing time of each method on a simulated data set with 34 loci, each with 35 SNPs, and 100 annotations. The central processing unit is Intel(R) Xeon(R) E5-4640 2.40 GHz. For CAVIARBF with penalized models, 10-fold CV was used. The parameter search space was the same as in the simulation. Because CAVIARBF is now provided in an R package using Rcpp to interface with the C++ code, it may appear slower when the R code takes a large portion of the time. This might explain that CAVIARBF_top10t is slower than PAINTOR_top10t for one maximal causal variant but faster for three maximal causal variants. When a penalized model is applied, CAVIARBF takes much longer than *fgwas*, but is still practical for fine mapping. For example, assuming three maximal causal variants in each locus, CAVIARBF_ENET_CV takes ~13 hr to finish. Application of the sampling-based methods instead of exhaustive enumeration of causal configurations can further reduce the computing time (Benner *et al.* 2016).

Lipid data fine mapping

The lipid meta-analysis summary statistics have been used in previous studies for either improving the power of GWAS discovery (Pickrell 2014) or fine mapping (Kichaev *et al.* 2014). Therefore it is natural to test our proposed method on this data set and compare it with other methods (see *Methods* for more details). After some initial investigation of the marginal Z-test statistics from meta-analysis and the correlation matrix from EUR population in the 1000 Genomes Project, we found that there are inconsistencies, see more in *Discussion*. To alleviate spurious findings due to mismatched Z statistics and the LD matrix, we set the maximal number of causal variants to 1, therefore not influenced by the potential mismatches. The sample sizes for each variant are relatively stable, ranging from 80,000 to ~100,000. Therefore we simply set the sample size parameter to 80,000 for CAVIARBF; setting to 100,000 showed similar results. Table 3 shows the

Table 3 Top-ranked annotations for HDL trait

Cell line	Type	CAVIARBF_ENET_CV		CAVIARBF_Individual_Annotation		CAVIARBF_top10t		PAINTOR_Individual_Annotation		PAINTOR_top10t	
		γ	standardized γ	P-value	γ	P-value	γ	P-value	γ	P-value	γ
—	distance to TSS	-0.053	-0.053	0.290	-0.210	0.000	0.186	0.313	0.000	0.000	0.000
K562	repressed	-0.060	-0.030	0.041	-0.755	-0.369	0.096	0.650	0.000	0.000	
hepg2	repressed	-0.045	-0.023	0.062	-0.715	0.000	0.121	0.641	0.000	0.000	
LCL	repressed	-0.038	-0.019	0.138	-0.512	0.000	0.243	0.455	0.000	0.000	
K562	transcribed	0.037	0.016	0.116	0.634	0.000	0.140	-0.672	0.000	0.000	
hela	repressed	-0.029	-0.014	0.220	-0.477	0.000	0.316	0.444	0.000	0.000	
H1hesc	repressed	-0.029	-0.014	0.185	-0.507	0.000	0.315	0.424	0.000	0.000	
fintestine (Lg)	DHS	0.051	0.013	0.005	1.721	0.000	0.012	-1.760	0.000	0.000	
fMuscle (trunk)	DHS	0.038	0.010	0.014	1.641	0.000	0.032	-1.743	0.000	0.000	
fMuscle (arm)	DHS	0.039	0.010	0.010	1.738	1.401	0.019	-1.877	-1.064	0.000	
Conjunctival fibroblasts	DHS	0.043	0.008	0.005	1.945	0.000	0.019	-1.940	0.000	0.000	
fThymus	DHS	0.034	0.007	0.016	1.822	0.000	0.019	-1.886	0.000	0.000	
Villous mesenchymal fibroblast	DHS	0.031	0.007	0.031	1.468	0.000	0.035	-1.601	0.000	0.000	
hepg2	TSS	0.033	0.005	0.004	2.145	1.041	0.005	-2.249	-1.341	0.000	
—	coding exons	0.028	0.004	0.008	2.190	1.552	0.017	-2.238	0.000	0.000	
—	3' UTR exons	0.025	0.003	0.022	2.211	2.007	0.010	-2.482	-2.316	0.000	
—	nonsynonymous	0.024	0.002	0.009	1.923	0.000	0.003	-3.038	-2.394	0.000	

Rows in boldface font indicate the top 10 annotations by different methods. To be consistent with the annotation parameterization as in CAVIARBF, the sign of effect sizes from PAINTOR need to be flipped.

top 10 ranked annotations by different methods for HDL trait and ranked by the absolute value of the standardized coefficients from CAVIARBF_ENET_CV. The top-ranked annotations are generally different among different methods. For example, Repressed (k562), a repressed region in the k562 cell line, is ranked as the second top annotation by CAVIARBF_ENET_CV and included by CAVIARBF_top10t, but not by other methods. DHSs in the fMuscle (arm) cell line is top ranked by all methods considered. Among the top 10 ranked annotations by CAVIARBF_ENET_CV, Repressed (k562), Repressed (hepg2), Transcribed (k562), Repressed (hela), DHSs (fetal large intestine), and DHSs (fetal muscle), are also identified as top annotations in Pickrell's results (Pickrell 2014). The effect sizes from CAVIARBF_ENET_CV are much smaller than that from fitting each annotation individually. This is because the effect size from CAVIARBF_ENET_CV is the additional effect after accounting for other correlated annotations. The penalization used in the model also shrinks the effect size. Due to the correlation among annotations, top annotations from the top10t strategy are in general not the same as the top 10 annotations by individual *P*-values. For CAVIARBF and PAINTOR, top annotations and annotations from the top10t strategy are similar, showing three common annotations with similar effect size. We note that the top annotations in Table 3 are not all the same as those from Figure 1 in Pickrell (2014) or as those in Table 6 from Kichaev *et al.* (2014). This may be due to several differences among the analyses: First, the methods used are different, even though they share some similarities. Second, the final data sets used for fine mapping are different. In Pickrell (2014), all regions were used to fit the model no matter whether there were significant signals. Even though both our and the analyses of Kichaev *et al.* (2014) were restricted to regions showing significant signals, the regions selected were different; 41 regions in our analysis *vs.* 37 in Kichaev *et al.* (2014), and the maximal number of causal variants assumed may also differ. The individual annotation results for LDL, TC, and TG can be found in File S1.

Next we compared the top ranked variants by different methods (Table 4). CAVIARBF with and without annotations (CAVIARBF_ENET_CV *vs.* CAVIARBF_non) do not show much difference, even though for rs2923084, the PIP is ~ 0.95 from CAVIARBF_ENET_CV and 0.86 from CAVIARBF_non. CAVIARBF_ENET_CV has similar PIPs as CAVIARBF_top10t for most SNPs; but for rs1800961, the PIP is ~ 0.06 from CAVIARBF_ENET_CV but ~ 0.58 from CAVIARBF_top10t. This is likely due to overoptimism of the top10t strategy. To illustrate this, we analyzed the simulated data sets with moderate annotation effects and 40 initial loci, setting the maximal number of causal variants to 1. We calculated the percentage of causal variants where both PIPs from CAVIARBF_non and CAVIARBF_ENET_CV fell within 0–0.1 and PIPs from CAVIARBF_top10t fell within 0.5–0.7. The percentage is 0.23 (17 out of 74). When adding the condition that PAINTOR_top10t fell within 0.8–1 and PAINTOR_non

Table 4 Top-ranked variants with PIPs of different methods for HDL trait

Chr	rsid	Position	Dosage allele ^a	Imputation quality ^b	z	CAVIARBF_ENET_CV	CAVIARBF_non	CAVIARBF_top10t	PAINTOR_top10t	PAINTOR_non
11	rs964184	116648917	C	0.986	-13.858	1.0000	1.0000	1.0000	1.0000	1.0000
16	rs2925979	81534790	C	1	-6.700	0.9979	0.9838	0.9959	0.9018	0.7769
19	rs4420638	45422946	G	1	9.423	0.9955	0.9952	0.9895	0.9237	0.9710
11	rs2923084	10388782	G	1	5.465	0.9458	0.8566	0.9673	0.3826	0.2019
4	rs13107325	103188709	T	1	6.516	0.9196	0.9077	0.9853	0.9377	0.5851
8	rs9987289	9183358	G	1	-10.309	0.9022	0.9013	0.9013	0.5744	0.5744
20	rs1800961	43042364	T	0.670	3.246	0.0661	0.0700	0.5837	0.9080	0.1207

Variables have at least one PIP > 0.9 from different methods. Rows in boldface font indicate at least one method outputs PIP < 0.9 . The variants are sorted by PIPs from CAVIARBF_ENET_CV in descending order. Chr, chromosome.

^a Dosage value is the count of this allele.

^b This is the r2pred output from ImpG indicating the quality of imputed z statistics, 1 indicates the original z statistics from meta-analysis.

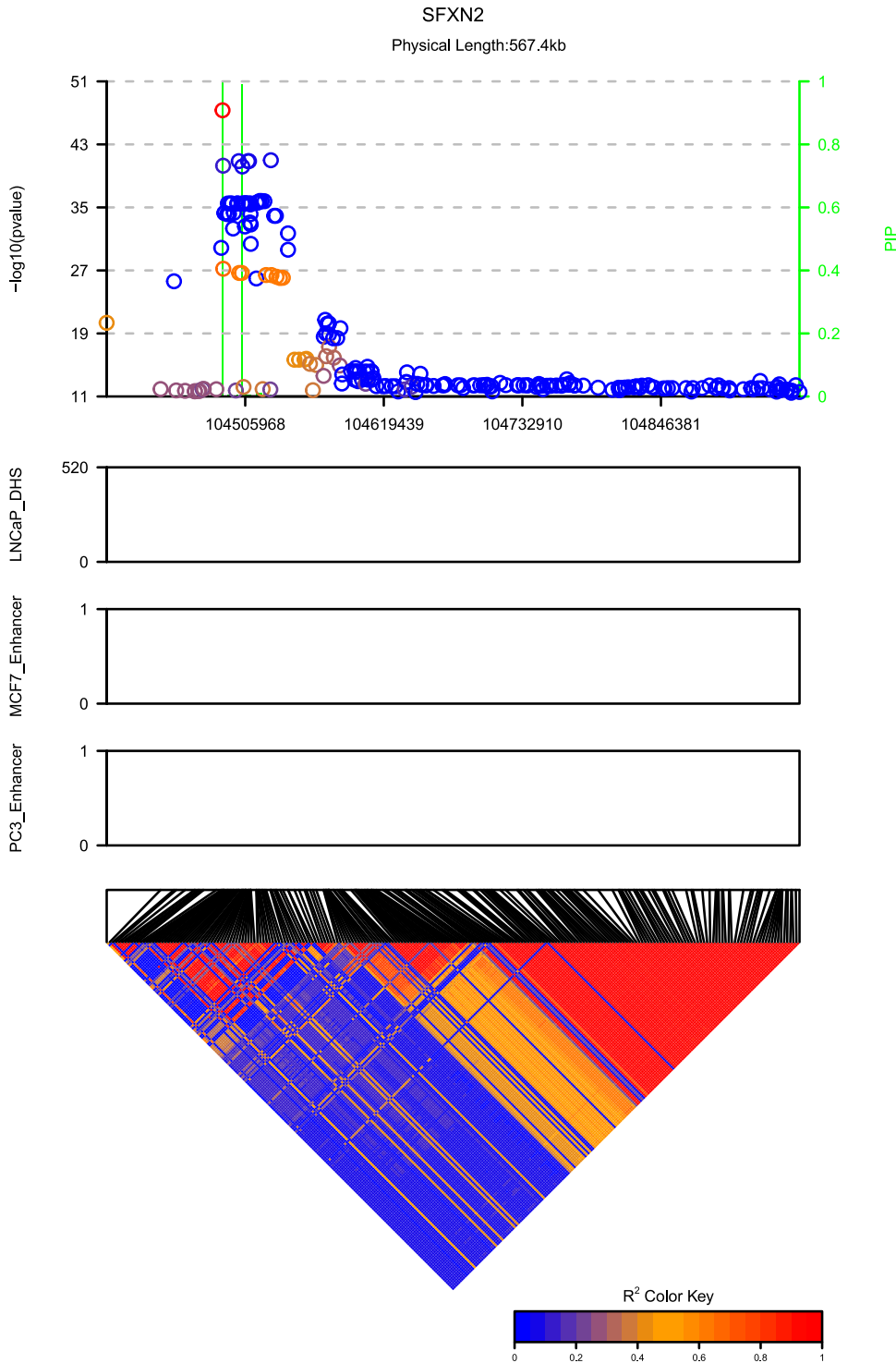


Figure 10 P-values, PIPs, annotations, and the LD for *cis*-eQTL analysis of SFXN2. The green lines are PIPs. The LD between the peak SNP and the remaining SNPs are color coded in the top panel. The three middle panels illustrated the annotations for individual SNPs. The bottom diagonal matrix shows the LD pattern among all SNPs in this locus.

fell within 0–0.2, the percentage is 0.15 (7 out of 47). This illustrates the overoptimism pertaining to the top10t strategy. On the other hand, for SNPs where both PIPs from CAVIARBF_non and CAVIARBF_top10t fell within 0–0.1 and PIPs from CAVIARBF_ENET_CV fell within 0.5–0.7, the percentage of causal variants is 0.57 (29 out of 51), calibrated well with the PIP range. Comparing CAVIARBF and PAINTOR, for some SNPs there are large PIP differences

even though in general PIPs are similar. For example, for rs2923084, the PIP from CAVIARBF_non is ~0.86, but it is ~0.20 from PAINTOR_non. Since the causal state needs confirmation from laboratory experiments, we cannot conclude which method is more accurate on this SNP, but being aware of the differences may be helpful for future application and comparison. Results for LDL, TC, and TG can be found in [Table S3](#).

Table 5 Top-ranked annotations for the eQTL study

Data ID	Cell line	Type	CAVIARBF_ENET_CV		CAVIARBF_Individual_Annotation		CAVIARBF_top10t
			γ	Standardized γ	P-value	γ	γ
GSM698597	LNCAp	AR	0.0760	0.5885	2.16×10^{-03}	0.4260	0.2595
GSM736565	LNCAp	DHS	0.1193	0.1529	2.30×10^{-09}	0.2565	0.2272
GSM686926	LNCAp	FoxA1	-0.0667	-0.1112	5.76×10^{-01}	-0.1945	0.0000
GSE57498	MCF7	ChromHMM.Enhancer	-0.0847	-0.1085	5.47×10^{-02}	-0.5081	0.0000
GSE57498	PC3	ChromHMM.Enhancer	0.0836	0.1005	4.33×10^{-02}	0.1996	0.0000
GSM989640	LNCAp	NKX3-1	0.1077	0.0963	2.97×10^{-05}	0.2423	0.0000
GSM1410771	VCaP	RUNX1	0.0808	0.0955	3.37×10^{-04}	0.2369	0.0000
GSE57498	PC3	ChromHMM.Heterochromatin	-0.0820	-0.0941	3.45×10^{-03}	-0.3140	-0.2469
GSM816637	LNCAp	DHS	0.0688	0.0930	1.26×10^{-03}	0.1303	0.0000
GSM736603	LNCAp	DHS	0.0691	0.0849	2.99×10^{-08}	0.2658	0.0000
GSE57498	HMEC	ChromHMM.Repressed	0.0591	0.0657	3.65×10^{-02}	0.1652	0.1612
GSM686935	LNCAp	H3K4me3	0.0620	0.0641	1.28×10^{-03}	0.2240	0.1453
GSE57498	MCF7	ChromHMM.Transcribed	0.0532	0.0635	2.51×10^{-02}	0.2184	0.1576
GSE57498	MCF7	ChromHMM.CTCF	0.0364	0.0363	2.74×10^{-02}	0.2229	0.1850
GSM816634	LNCAp	DHS	0.0221	0.0294	1.57×10^{-04}	0.1662	0.0000
GSM699633	LNCAp	NKX3-1	0.0204	0.0236	7.31×10^{-04}	0.1252	0.0000
GSM1024742	PrEC	DHS	-0.0073	-0.0069	1.12×10^{-03}	0.2143	0.0000
GSE57498	PrEC	ChromHMM.Promoter	-0.0008	-0.0009	1.76×10^{-03}	0.2307	0.0000

Rows in boldface font indicate the top 10 annotations ranked by different methods. CAVIARBF_ENET_CV ranks annotations based on the absolute value of coefficients, CAVIARBF_Individual_Annotation ranks annotations based on individual P-values. LNCAp, a prostate adenocarcinoma cell line; PC3, a prostate cancer cell line; HMEC, a normal mammary epithelial cell line; MCF7, a breast cancer cell line; PrEC, a normal prostate epithelial cell line.

cis-eQTL fine mapping

The objective of this study was to identify *cis*-eQTLs for genes related to PrCa risk SNPs (Thibodeau *et al.* 2015) using genotypes and RNA-seq data from normal prostate tissue samples from 471 individuals. Figure 10 shows the advantage of using PIPs instead of P-values in identifying *cis*-eQTLs on gene *SFXN2*. It is difficult to identify the potential causal variants from the P-values due to the large number of SNPs in LD. The green lines show the PIPs by applying our proposed method by taking both the LD information and the annotation information into account. It is clear that there are at least two highly likely causal variants influencing the gene expression. Actually there are three potential causal variants, two of which are very close in position.

We set the maximal number of causal variants to 3 because many loci show multiple eQTL signals. This can be illustrated using the sum of PIPs for each loci, an estimation of expected causal variants. Figure S9 shows the histogram of added final PIPs (using CAVIARBF_ENET_CV) for each target gene, which supports the multiple eQTL signal assumption for many loci.

We first compare estimated PIPs using approximate and exact calculations on data sets showing different proportions of variance explained in gene expression traits. We selected two target genes, one with a relatively small PVE by individual variants and the other with a large PVE. The maximal number of causal variants was set to 3. We then ran different fine-mapping options/methods with results displayed in Figure S10. When variants explain a small proportion of the variance (left panel), methods using the approximation, such as PAINTOR and CAVIARBF using the approximate calculation, have similar PIPs as CAVIARBF using the exact calculation. In

this scenario, PIPs are also not sensitive to the parameter of σ_a . However, when variants explain a large proportion of the variance (right panel), PAINTOR and CAVIARBF using the approximation have similar PIPs, but their results are different from CAVIARBF using the exact calculation. PIPs are also sensitive to the parameter of σ_a . Therefore for all eQTL fine mapping, we did not run PAINTOR and used CAVIARBF with the exact calculation and chose a set of values for σ_a (0.1, 0.2, 0.4, 0.8, 1.6) to calculate the average Bayes factors, similar to those used in Wen *et al.* (2015).

The distance to TSS has been shown to associate with the enrichment of *cis*-eQTLs (Gaffney *et al.* 2012; Wen *et al.* 2015). As a validation, we tested the significance of this annotation in our eQTL data. Surprisingly, the P-value for the distance to the TSS is 0.3215, which appears inconsistent with previous findings. Plotting the final fitted PIPs vs. the distance to the TSS (Figure S11) and further examination provide an explanation. There is an obvious decreasing pattern for PIPs when the distance to the TSS is <200 kb, however, PIPs are almost flat further away. In other words, the size of the region selected around the TSS can influence the significance test. We verified this by using the same fitted PIPs but picking SNPs within different distances from the TSS. The P-values from linear regression testing are 0.311 for the threshold of 1.1 Mb, 0.0624 for 200 kb, 0.000304 for 100 kb and 0.0091 for 50 kb. This is consistent with previous significant results using 100 kb to the TSS as the boundary of selected regions (Gaffney *et al.* 2012; Wen *et al.* 2015).

Table 5 shows the results of top-ranked annotations by different methods and sorted by the magnitude of the estimated effect size by CAVIARBF_ENET_CV. Androgen receptor (AR) binding sites and DHSs in LNCAp cells are the top

Table 6 Top-ranked variants with PIPs of different methods for the eQTL study

Chr	Target gene	rsid	Position	Dosage allele ^a	z	CAVIARBF_ENET_CV	CAVIARBF_non	CAVIARBF_top10t
5	<i>IRX4</i>	rs12655062	1890877	A	-26.758	1.0000	1.0000	1.0000
10	<i>SFXN2</i>	rs2902548	104487382	T	-16.385	1.0000	1.0000	1.0000
10	<i>NCOA4</i>	rs10993994	51549496	T	22.665	1.0000	1.0000	1.0000
17	<i>HNF1B</i>	rs3760511	36106313	G	6.623	1.0000	0.0343	1.0000
22	<i>TBX1</i>	rs1978060	19749525	A	21.090	1.0000	1.0000	1.0000
2	<i>SEPT2</i>	rs115271170	242347614	T	29.950	1.0000	1.0000	1.0000
2	<i>RAB17</i>	rs13404216	238491258	A	-11.072	1.0000	0.9999	1.0000
X	<i>GJB1</i>	rs747181	70432708	A	-12.501	0.9999	0.9999	0.9999
5	<i>IRX4</i>	rs6890484	1890577	G	6.966	0.9989	0.9960	0.9960
6	<i>HLA-L</i>	rs114620415	31239821	G	-20.153	0.9986	0.9794	0.9971
6	<i>HLA-L</i>	rs116373865	31239296	G	-23.825	0.9983	0.9782	0.9967
6	<i>HLA-L</i>	rs114172808	31303275	G	12.760	0.9980	0.9968	0.9972
7	<i>HOTTIP</i>	rs10486567	27976563	A	6.358	0.9967	0.1844	0.9975
2	<i>C2orf43</i>	rs72784331	20877436	A	7.622	0.9951	0.8203	0.9963
10	<i>SFXN2</i>	rs3818708	104503584	T	-14.792	0.9899	0.1126	0.9985
17	<i>HNF1B</i>	rs4430796	36098040	G	-7.334	0.9881	0.0835	1.0000
21	<i>TMPRSS2</i>	rs8134378	42893758	A	-14.832	0.9858	0.3305	0.9578
21	<i>TMPRSS2</i>	rs56095453	42893808	A	7.872	0.9841	0.3941	0.9756
17	<i>FAM57A</i>	rs2644715	632686	A	11.924	0.9800	0.9542	0.9552
17	<i>FAM57A</i>	rs2474694	618039	A	-19.853	0.9643	0.7060	0.7953
4	<i>BMPR1B</i>	rs6853490	95544718	G	-10.289	0.9602	0.0319	0.9862
5	<i>COMMD10</i>	rs804152	115670932	C	10.097	0.9577	0.9558	0.9586
1	<i>LOC284581</i>	rs9438393	205782718	G	52.062	0.9500	0.6991	0.9143
2	<i>NOL10</i>	rs7567304	10790029	T	12.020	0.9331	0.9001	0.9329
14	<i>SYNJ2BP</i>	rs1044527	70833819	G	13.428	0.9044	0.5466	0.6407
10	<i>SFXN2</i>	rs2297451	104503459	A	13.710	0.8727	0.0393	0.9808
No. of SNPs (PIP >0.9)						25	14	24

Variants have at least one PIP >0.9 from different methods. Rows in boldface font indicate at least one method outputs PIP <0.9. The variants are sorted by PIPs from CAVIARBF_ENET_CV in descending order. Chr, chromosome.

^a Dosage value is the count of this allele.

two ranked annotations by CAVIARBF_ENET_CV; where DHSs in LNCaP cells, the second top annotation, has a P -value of 3.43×10^{-09} , the smallest P -value among all annotations. These two annotations are also selected by CAVIARBF_top10t. Most annotations selected by CAVIARBF_top10t have a large standardized effect size γ in CAVIARBF_ENET_CV, even though in general the annotations selected by these two methods are not the same. Other top-ranked annotations from CAVIARBF_ENET_CV include the enhancer region, transcription factor binding sites, the heterochromatin state, the repressed and promoter region, and histone modification annotations. Similarly as in the lipid data, due to the correlation among annotations, annotations from the top10t strategy are in general not the same as the top 10 annotations by individual P -values. The effect sizes from CAVIARBF_ENET_CV are also smaller than that from fitting each annotation individually. In general, the sign of the effect sizes matched well with the function of annotations for both CAVIARBF_ENET_CV or CAVIARBF using individual annotations. For example, the heterochromatin state annotation has a negative value, therefore decreasing the probability of being eQTLs. This is intuitive because under the heterochromatin state, DNA sequences are tightly packed and hard to access. On the other hand, DHSs, except one from the PrEC cell line which is very small

in magnitude (-0.0073), and H3K4me3 all have positive values, consistent with their roles of DHSs and promoting gene regulation.

Finally, we compare the top ranked SNPs and PIPs using different methods in Table 6. Clearly, incorporating annotations increased the number of potential eQTLs with high PIPs. There are 25 SNPs with PIP >0.9 using CAVIARBF_ENET_CV, compared to 14 SNPs with PIP >0.9 using CAVIARBF without annotations (CAVIARBF_non). For most SNPs listed in Table 6, CAVIARBF_ENET_CV and CAVIARBF_top10t have similar PIPs. However, there is still a relatively large PIP difference for SNP rs1044527, with 0.9044 from CAVIARBF_ENET_CV and 0.6407 from CAVIARBF_top10t. Given the better performance of CAVIARBF_ENET_CV over the top10t strategy based on simulation studies, the results from CAVIARBF_ENET_CV may be more reliable. We further examined several SNPs with large increase of PIPs when incorporating functional annotations. The major annotation components are presented in Table 7. Several SNPs only rely on one major annotation component to increase the PIPs, e.g., rs3760511 within an androgen response element (Clinckemalie *et al.* 2013). Other SNPs combine multiple annotations together to improve the PIPs. Among these SNPs, rs10486567 is reported as a risk enhancer affecting both *NKX3-1* and *FOXA-AR* motifs (Hazelett

et al. 2014), SNP rs8134378 was reported within an androgen response element and reduces binding and transactivation by the androgen receptor (Clinckemalie *et al.* 2013).

Discussion

In this study, we proposed a framework to systematically incorporate a large number of annotations in association studies. By using CAVIARBF to calculate the PIPs, the framework only requires summary statistics. For quantitative traits, when the explained variance of the traits is high, *e.g.*, for some eQTLs, calculating Bayes factors from the summary statistics using approximation may produce misleading results. Instead we derived an exact calculation. Further investigation for binary traits with high explained variances may be useful.

Our method shares similarities with some existing methods, but also differs significantly from them. The most significant difference is that our method is a general framework to systematically incorporate functional annotations, and it uses the ENET in the penalized model with CV for parameter selection. In principle, any method that can output PIPs which take into account the prior probability of each variant being causal can be used within this framework. Compared to *fgwas* (Pickrell 2014), our method results in a more robust way to automatically select sparse annotations, it also allows multiple causal variants in each locus where *fgwas* assumes no more than one. The performance of *fgwas* is similar to that of CAVIARBF_L1_CV, except that CAVIARBF_L1_CV does not show decreased performance under the null annotations. There are also other differences: *fgwas* only supports binary annotations or a distance model using integers, while CAVIARBF can use any numeric annotations (binary or continuous). *fgwas* uses a penalized likelihood in CV for the testing fold while CAVIARBF uses the likelihood directly. Compared to the top5t or top10t strategy recommended by PAINTOR (Kichaev *et al.* 2014), the proposed method does not have the dilemma of choosing the significance threshold: if the *P*-value threshold is too stringent, we may lose informative annotations; if it is too liberal, we may include too many noninformative annotations. Fitting the model with top-ranked annotations without penalization may also result in overfitting (Pickrell 2014). FM-QTL (Wen *et al.* 2015) uses the full genotype data, assumes a small number of annotations and uses the same maximum likelihood-based method to incorporate annotations as in PAINTOR. iBRI (Quintana and Conti 2013) uses an L2 penalty with a fixed parameter, which may not be optimal. For PIP calculation, both FM-QTL and iBRI use an MCMC-based approach, which may be faster when there are a relatively larger number of causal variants in a large region. Therefore, an alternative is to use MCMC to perform genome-wide association analysis and fine mapping simultaneously, such as in Zhu and Stephens (2016). There are also other sampling-based methods showing highly reduced computing cost (Benner *et al.* 2016).

Table 7 Major annotation components that largely improve the PIPs of SNPs in the eQTL study

rsid	Annotations
rs3760511	AR binding in LNCaP
rs10486567	DHS, NKX3-1 binding in LNCaP
rs3818708	DHS in LNCaP
rs4430796	AR binding peak in LNCaP
rs8134378	DHS, NKX3-1, AR binding in LNCaP, RUNX1 binding in VCaP
rs56095453	DHS, NKX3-1 in LNCaP, RUNX1 binding in VCaP
rs6853490	DHS, NKX3-1 in LNCaP
rs1044527	RUNX1 binding in VCaP, NKX3-1, AR binding in LNCaP
rs2297451	DHS in LNCaP
rs2474694	DHS, H3K4me3 in LNCaP, RUNX1 binding in VCaP
rs9438393	H3K4me3, DHS in LNCaP, ChromHMM.Heterochromatin in PC3 and HMEC, ChromHMM.Promoter in PC3, RUNX1 binding in VCaP

The proposed model assumes independence among different loci. Some eQTLs may regulate more than one gene, thus are shared for more than one locus. This may create dependencies among loci but the extent and how much it will influence the final results are not clear. More investigation is needed to accommodate the overlapping scenario.

When the number of loci is small, *e.g.*, ~20 loci, the estimated PIP may be overly optimistic. One reason may be that we use a point estimation (MAP) from *glmnet* instead of a full Bayesian method. A full Bayesian method for the proposed graphical model may show better calibration. Another possible cause is that CV among loci may not be very effective if the number of loci is small. This may result in overfitting with the grid search to tune parameters. If full genotype data are available, CV by partitioning individuals instead of partitioning loci would be a better solution, and the assumption of independence between individuals is much easier to satisfy than the assumption of independence between loci. Otherwise, better methods may be needed to tune the parameters. In the extreme case, when the number of loci is very small, *e.g.*, <5, the benefit from annotations is not likely to be obvious. In this case we can perform fine mapping without using annotations.

When ranking annotations based on the MAP point estimate of γ , there is no direct quantification of the uncertainty of this estimate. To measure the estimation variance, we can use the bootstrap method by resampling the loci as previously proposed (Kichaev *et al.* 2014). Because we use CV to choose the best penalty parameters, different partitioning of the loci may select different penalty parameters resulting in different MAP estimates. In this case, we may need to run the CV several times and pick the penalty parameters with the highest frequency, *i.e.*, the mode of all selected penalty parameters. We ran CV five times for real data application in this paper, except the HDL data, where we ran CV 20 times to pick the mode because several parameters show similar frequencies and the mode is not very sharp.

The proposed framework can also be used to incorporate pathway information. One simple method is to code each

pathway as an annotation, where 1 indicates the variant is in the pathway and 0 otherwise. How much the pathway information can improve the performance needs further study.

In our lipid data analysis, we found inconsistencies between the available Z statistics and the LD from the EUR population in 1000 Genomes Project. For example, rs2144300 and rs4846914 are in perfect LD, correlation coefficient is 1, in EUR. However, the meta-analysis z -scores are 9.16 and 9.442. Without using annotation and assuming one maximal causal variant, this difference results in rs4846914 with PIP 0.61, much higher than rs2144300 with PIP 0.05, even though it makes more sense that these two should have the same or at least very similar PIPs. If the difference of z -scores is not due to computing error, there are two potential sources which may also be common for other meta-analysis results. First, there may be other populations that show different LD patterns from the EUR population. To address this problem, previous studies (Kichaev and Pasaniuc 2015; Wen *et al.* 2015) have developed different ways to handle multiple populations. Our framework can also be extended similarly, which is an interesting future direction. Second, and more likely to be the source, imputed genotypes or meta-analyzed z -scores were not computed for the same set of individuals, given that usually an r^2 threshold is used for imputed genotypes, *e.g.*, 0.3 for the meta-analysis results (Teslovich *et al.* 2010). This is discussed in detail in Zhu and Stephens (2016). Therefore, a better way is to report summary statistics using all imputed genotypes without applying a threshold. In our method, we assume the sample sizes are the same or at least similar. Strictly speaking, it is not fair to compare two SNPs if their sample size is different because the Bayes factors depend on the sample size. However, if the sample size difference is small, it might not influence the comparison much.

Acknowledgments

This research was supported by the United States Public Health Service; National Institutes of Health, contract grant number GM-065450 and CA-151254; and by the Department of Defense (W81XWH-11-1-0261). Informed consent was obtained from all subjects and the study was approved by the Mayo Clinic Institutional Review Board.

Literature Cited

- Abecasis, G. R., A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin *et al.*, 2012 An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56–65.
- Benner, C., C. C. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti *et al.*, 2016 FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 32: 1493–1501.
- Bert, S. A., M. D. Robinson, D. Strbenac, A. L. Statham, J. Z. Song *et al.*, 2013 Regional activation of the cancer genome by long-range epigenetic remodeling. *Cancer Cell* 23: 9–22.
- Bishop, C. M., 2006 *Pattern Recognition and Machine Learning*. Springer, New York.
- Carbonetto, P., and M. Stephens, 2013 Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease. *PLoS Genet.* 9: e1003770.
- Chen, W., and D. J. Schaid, 2014 PedBLIMP: extending linear predictors to impute genotypes in pedigrees. *Genet. Epidemiol.* 38: 531–541.
- Chen, W., B. R. Larrabee, I. G. Ovsyannikova, R. B. Kennedy, I. H. Haralambieva *et al.*, 2015 Fine mapping causal variants with an approximate bayesian method using marginal test statistics. *Genetics* 200: 719–736.
- Clinkemalie, L., L. Spans, V. Dubois, M. Laurent, C. Helsen *et al.*, 2013 Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. *Mol. Endocrinol.* 27: 2028–2040.
- Ernst, J., and M. Kellis, 2012 ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9: 215–216.
- Ernst, J., and M. Kellis, 2015 Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* 33: 364–376.
- Friedman, J., T. Hastie, and R. Tibshirani, 2010 Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33: 1–22.
- Gaffney, D. J., J. B. Veyrieras, J. F. Degner, R. Pique-Regi, A. A. Pai *et al.*, 2012 Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* 13: R7.
- Guan, Y., and M. Stephens, 2008 Practical issues in imputation-based association mapping. *PLoS Genet.* 4: e1000279.
- Guan, Y. T., and M. Stephens, 2011 Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Ann. Appl. Stat.* 5: 1780–1815.
- Gusev, A., S. H. Lee, G. Trynka, H. Finucane, B. J. Vilhjalmsson *et al.*, 2014 Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* 95: 535–552.
- Hazelett, D. J., S. K. Rhie, M. Gaddis, C. Yan, D. L. Lakeland *et al.*, 2014 Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS Genet.* 10: e1004102.
- He, H. H., C. A. Meyer, H. Shin, S. T. Bailey, G. Wei *et al.*, 2010 Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* 42: 343–347.
- Hoffman, M. M., J. Ernst, S. P. Wilder, A. Kundaje, R. S. Harris *et al.*, 2013 Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* 41: 827–841.
- Hormozdiari, F., E. Kostem, E. Y. Kang, B. Pasaniuc, and E. Skinn, 2014 Identifying causal variants at loci with multiple signals of association. *Genetics* 198: 497–508.
- Kichaev, G., and B. Pasaniuc, 2015 Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am. J. Hum. Genet.* 97: 260–271.
- Kichaev, G., W. Y. Yang, S. Lindstrom, F. Hormozdiari, E. Skinn *et al.*, 2014 Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 10: e1004722.
- Kyung, M., J. Gill, M. Ghosh, and G. Casella, 2010 Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* 5: 369–411.
- Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen *et al.*, 2012 Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337: 1190–1195.
- Park, M. Y., and T. Hastie, 2008 Penalized logistic regression for detecting gene interactions. *Biostatistics* 9: 30–50.

- Pasaniuc, B., N. Zaitlen, H. Shi, G. Bhatia, A. Gusev *et al.*, 2014 Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics* 30: 2906–2914.
- Pickrell, J. K., 2014 Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* 94: 559–573.
- Quintana, M. A., and D. V. Conti, 2013 Integrative variable selection via Bayesian model uncertainty. *Stat. Med.* 32: 4938–4953.
- Servin, B., and M. Stephens, 2007 Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* 3: e114.
- Sharma, N. L., C. E. Massie, A. Ramos-Montoya, V. Zecchini, H. E. Scott *et al.*, 2013 The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. *Cancer Cell* 23: 35–47.
- Su, Z., J. Marchini, and P. Donnelly, 2011 HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics* 27: 2304–2305.
- Taberlay, P. C., A. L. Statham, T. K. Kelly, S. J. Clark, and P. A. Jones, 2014 Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Res.* 24: 1421–1432.
- Takayama, K., T. Suzuki, T. Fujimura, T. Urano, S. Takahashi *et al.*, 2014 CtBP2 modulates the androgen receptor to promote prostate cancer progression. *Cancer Res.* 74: 6542–6553.
- Tan, P. Y., C. W. Chang, K. R. Chng, K. D. Wansa, W. K. Sung *et al.*, 2012 Integration of regulatory networks by NKX3–1 promotes androgen-dependent prostate cancer survival. *Mol. Cell. Biol.* 32: 399–414.
- Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou *et al.*, 2010 Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466: 707–713.
- The ENCODE Project Consortium, 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74.
- Thibodeau, S. N., A. J. French, S. K. McDonnell, J. Cheville, S. Middha *et al.*, 2015 Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. *Nat. Commun.* 6: 8653.
- Tibshirani, R., 1996 Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* 58: 267–288.
- Tibshirani, R. J., and J. Taylor, 2012 Degrees of freedom in lasso problems. *Ann. Stat.* 40: 1198–1232.
- Wakefield, J., 2009 Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* 33: 79–86.
- Wang, D., I. Garcia-Bassets, C. Benner, W. Li, X. Su *et al.*, 2011 Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474: 390–394.
- Wen, X., and M. Stephens, 2010 Using linear predictors to impute allele frequencies from summary or pooled genotype data. *Ann. Appl. Stat.* 4: 1158–1182.
- Wen, X., F. Luca, and R. Pique-Regi, 2015 Cross-population joint analysis of eQTLs: fine mapping and functional annotation. *PLoS Genet.* 11: e1005176.
- Zhou, X., P. Carbonetto, and M. Stephens, 2013 Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 9: e1003264.
- Zhu, X., and M. Stephens, 2016 Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *bioRxiv*, DOI: <http://dx.doi.org/10.1101/042457>.
- Zou, H., and T. Hastie, 2005 Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.* 67: 301–320.

Communicating editor: N. Yi

Appendix

CAVIARBF Reduces to the ABF When There Is at Most One Causal Variant

We derive the equivalence using the ABF Equation 3 in Chen *et al.* (2015). When $m = 1$, Equation 3 reduces to

$$\text{BF}_2 = |1 + nv|^{-(1/2)} \exp\left\{\frac{1}{2}z^2 \left[1 + (nv)^{-1}\right]^{-1}\right\}.$$

Therefore,

$$\text{BF}_2^{-1} = \sqrt{1 + nv} \exp\left(-\frac{1}{2}z^2 \frac{nv}{1 + nv}\right).$$

The corresponding ABF^{-1} (H1 vs. H0, from Equation 2 in Wakefield 2009) is

$$\text{ABF}^{-1} = \sqrt{\frac{V+W}{V}} \exp\left(-\frac{z^2}{2} \frac{W}{V+W}\right) = \sqrt{1 + \frac{W}{V}} \exp\left(-\frac{z^2}{2} \frac{W/V}{1 + W/V}\right).$$

Therefore, CAVIARBF in the case $m = 1$ and ABF have the same form. If $(W/V) = nv$, then CAVIARBF and ABF are exactly the same. This is the case in the application of ABF in *fgwas* (Pickrell 2014) when V is not available from data. For quantitative traits, it is estimated as $1/(2nf(1-f))$, where f is the allele frequency (this is half of the value in Equation 14 from Pickrell 2014, and is used in *fgwas* code). We have $(W/V) = nW2f(1-f)$. The Bayes factor in CAVIARBF and *fgwas* are exactly the same if $v = W2f(1-f)$. This can be done in CAVIARBF by setting $\sigma_a = \sqrt{W}$ and setting the weights to the variance of the SNP using $2f(1-f)$. For binary traits, $V = n/[2n_1n_2f(1-f)]$, where n is the total sample size, n_1 is the number of cases, and n_2 is the number of controls. Therefore $W/V = nn_1n_2/n^2W2f(1-f)$. The Bayes factor in CAVIARBF and *fgwas* are exactly the same if $v = [(n_1n_2/n^2)W2f(1-f)]$. This can be done in CAVIARBF by setting σ_a to $\sqrt{n_1n_2/n^2W}$ and setting the weights to the variance of the SNP using $2f(1-f)$. From the above derivation, we can also see that *fgwas* assumes the prior normal distribution of the variant effect size on the original scale as in BIMBAM instead of the standardized scale (Chen and Schaid 2014). The only difference between CAVIARBF and ABF is that in our previous derivation we assumed z was a score statistic but Wakefield assumed a Wald statistic, however, these two statistics are asymptotically equivalent under the null or small deviation from the null, thus very similar due to the large sample size and small effects in GWAS.

Exact Bayes Factors Using Marginal Statistics for Quantitative Traits

In this section we derive the exact Bayes factors using a D_2 prior (Servin and Stephens 2007), which only depends on the marginal statistics and the correlation matrix. Suppose X_c is an $n \times m$ matrix with coded genotypes, where n is the number of individuals and m is the number of coded columns from putative causal variants. Without loss of generality, we assume X_c is centered, *i.e.*, $1/n \sum_{i=1}^n X_{cij} = 0$, $j = 1, 2, \dots, m$, where X_{cij} is the element of row i and column j in X_c . Let s_j be the sample variance of the j th column, *i.e.*, $s_j = 1/(n-1) \sum_{i=1}^n X_{cij}^2$. The quantitative phenotype y is an $n \times 1$ vector. We assume a linear model as follows:

$$y = \mu + X_c \beta + \varepsilon, \quad (1)$$

where β is the effect size and $\varepsilon \sim N(0, \frac{1}{\tau} I_n)$. Here I_n denotes the $n \times n$ identity matrix. The D_2 prior assumes that β has a prior normal distribution $N(0, v_c(1/\tau))$, where v_c is a diagonal matrix and β and ε are independent. Let $v = \begin{bmatrix} 0 & \\ & v_c \end{bmatrix}$, where empty blocks are simply 0s. Then the Bayes factor using a D_2 prior (Servin and Stephens 2007) can be written as:

$$\text{BF} = n^{1/2} |v^{-1} + X^T X|^{-(1/2)} |v_c|^{-(1/2)} \left[\frac{y^T y - y^T X (v^{-1} + X^T X)^{-1} X^T y}{y^T y - n \bar{y}^2} \right]^{-(n/2)}, \quad (2)$$

where $X = [1, X_c]$, *i.e.*, the augmented matrix with all 1s in the first column, \bar{y} is the mean value of y . The superscript T means the transpose of the corresponding matrix or vector. Equation 2 can be further simplified to

$$\text{BF} = |v_c^{-1} + X_c^T X_c|^{-(1/2)} |v_c|^{-(1/2)} \left[1 - \frac{y^T X_c (v_c^{-1} + X_c^T X_c)^{-1} X_c^T y}{y^T y - n\bar{y}^2} \right]^{-(n/2)}.$$

Let S be the diagonal matrix with diagonal entries $s_j, j = 1, \dots, m$. Then the correlation matrix of X_c is $\Sigma_x = \frac{1}{n-1} S^{-(1/2)} X_c^T X_c S^{-(1/2)}$, therefore $X_c^T X_c = (n-1) S^{1/2} \Sigma_x S^{1/2}$. Then we have

$$v_c^{-1} + X_c^T X_c = (n-1) S^{1/2} \left\{ [(n-1)v_c S]^{-1} + \Sigma_x \right\} S^{1/2},$$

and

$$\frac{y^T X_c (v_c^{-1} + X_c^T X_c)^{-1} X_c^T y}{y^T y - n\bar{y}^2} = \frac{y^T X_c [(n-1)S]^{-(1/2)} \left\{ [(n-1)v_c S]^{-1} + \Sigma_x \right\}^{-1} [(n-1)S]^{-(1/2)} X_c^T y}{y^T y - n\bar{y}^2}.$$

Let z_j be the marginal test statistics corresponding to the j th column of X_c , denoted by X_{cj} , then $z_j = \hat{\beta}_j / \hat{\sigma}_j (X_{cj}^T X_{cj})^{-(1/2)}$, where $\hat{\beta}_j = (X_{cj}^T X_{cj})^{-1} X_{cj}^T y$, $\hat{\sigma}_j = \sqrt{1/(n-2)(y^T y - n\bar{y}^2 - \hat{\beta}_j^T X_{cj}^T y)}$. Because $s_j = \frac{1}{n-1} X_{cj}^T X_{cj}$, we have

$$z_j = \frac{[(n-1)s_j]^{-(1/2)} X_{cj}^T y}{\sqrt{\frac{1}{n-2} \left\{ y^T y - n\bar{y}^2 - [(n-1)s_j]^{-1} (X_{cj}^T y)^2 \right\}}}.$$

Then $y^T y - n\bar{y}^2$ can be written as

$$y^T y - n\bar{y}^2 = \frac{[(n-1)s_j]^{-1} (X_{cj}^T y)^2 (z_j^2 + n - 2)}{z_j^2}.$$

Therefore we have

$$\frac{[(n-1)s_j]^{-(1/2)} X_{cj}^T y}{\sqrt{y^T y - n\bar{y}^2}} = \frac{z_j}{\sqrt{z_j^2 + n - 2}},$$

where we have used the fact that $X_{cj}^T y$ has the same sign as z_j . Let z^* be the column vector where the j th component is $(z_j / \sqrt{z_j^2 + n - 2})$, the correlation coefficient between X_{cj} and y , then we have:

$$\frac{y^T X_c (v_c^{-1} + X_c^T X_c)^{-1} X_c^T y}{y^T y - n\bar{y}^2} = z^{*T} \left\{ [(n-1)v_c S]^{-1} + \Sigma_x \right\}^{-1} z^*.$$

Therefore the Bayes factor can be written using only the marginal test statistics and correlation matrix as follows:

$$\text{BF} = |[(n-1)v_c S]^{-1} + \Sigma_x|^{-(1/2)} |(n-1)v_c S|^{-(1/2)} \left[1 - z^{*T} \left\{ [(n-1)v_c S]^{-1} + \Sigma_x \right\}^{-1} z^* \right]^{-(n/2)}.$$

Figure S1
Q-Q plot for $\chi^2_{v=1}$

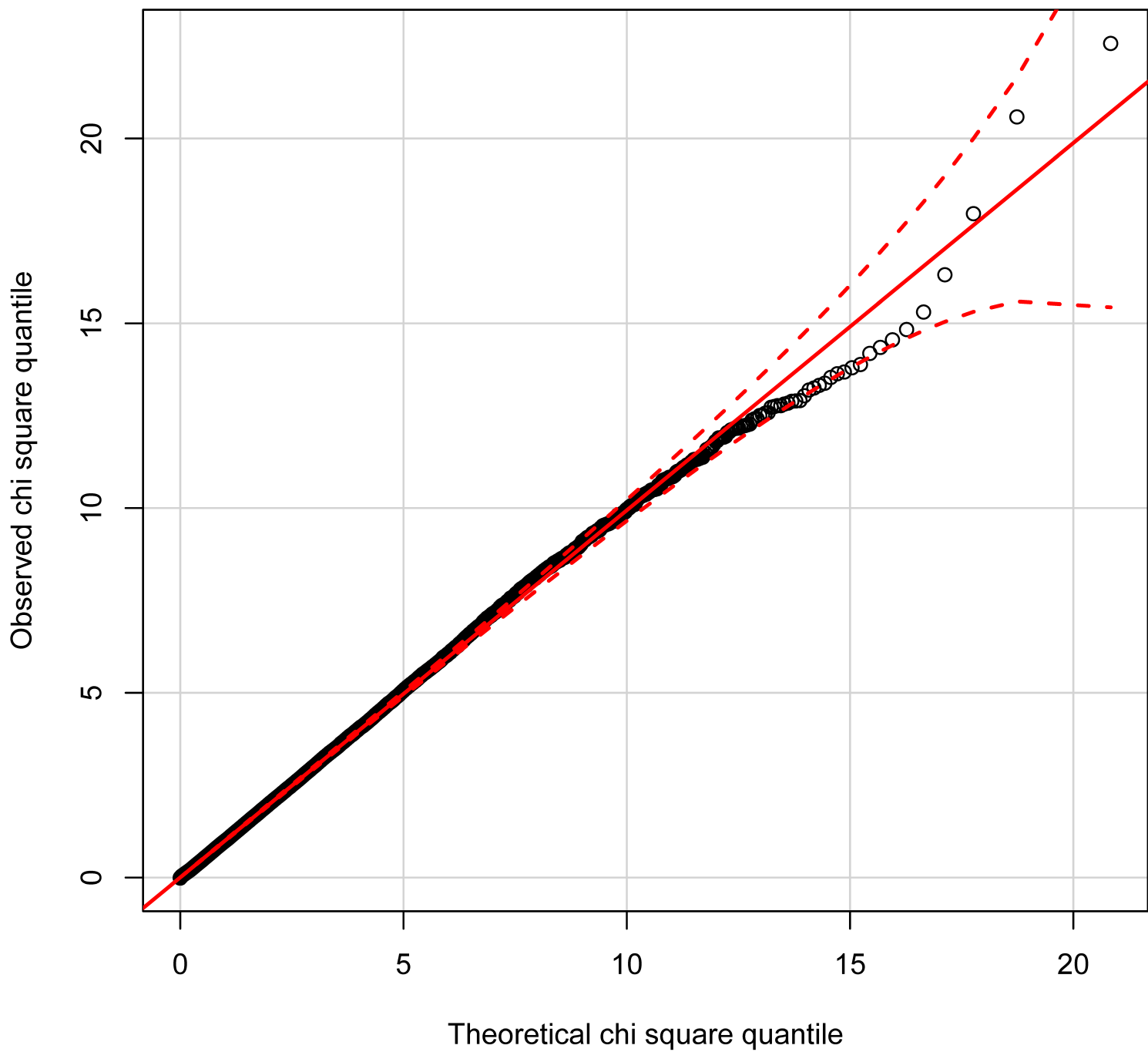


Figure S2

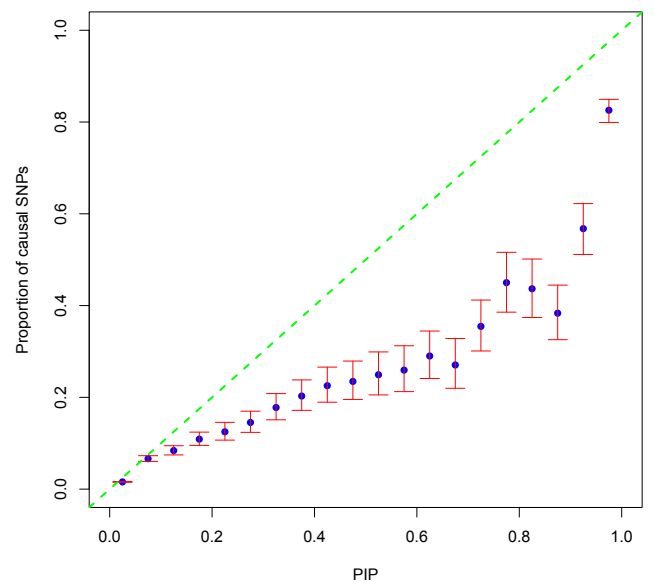
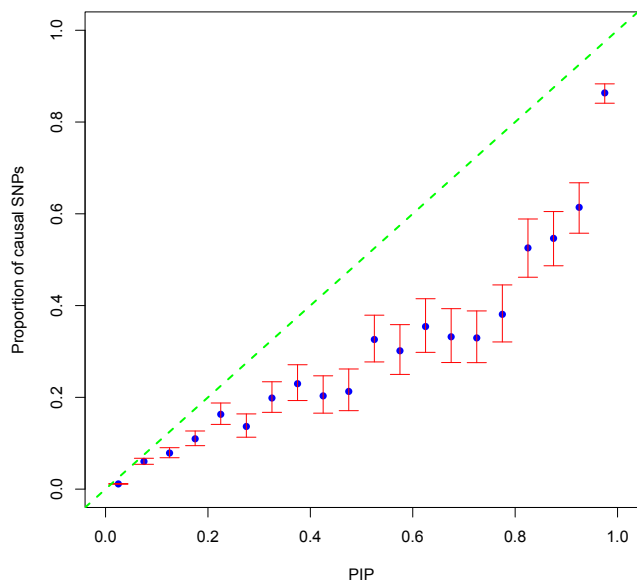
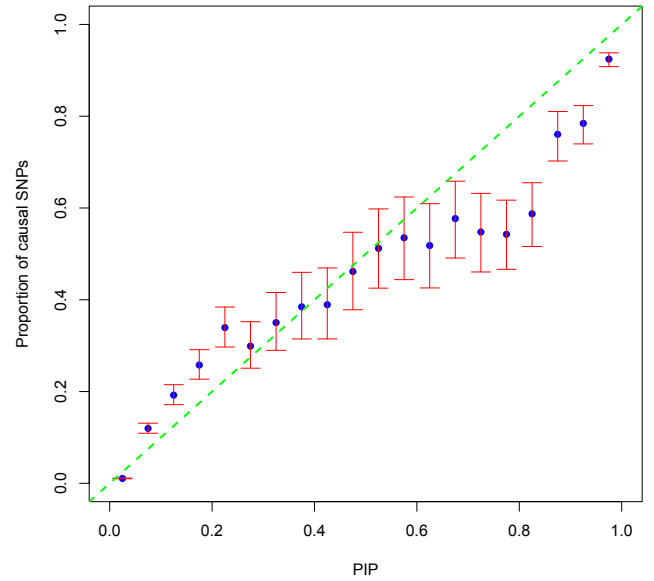
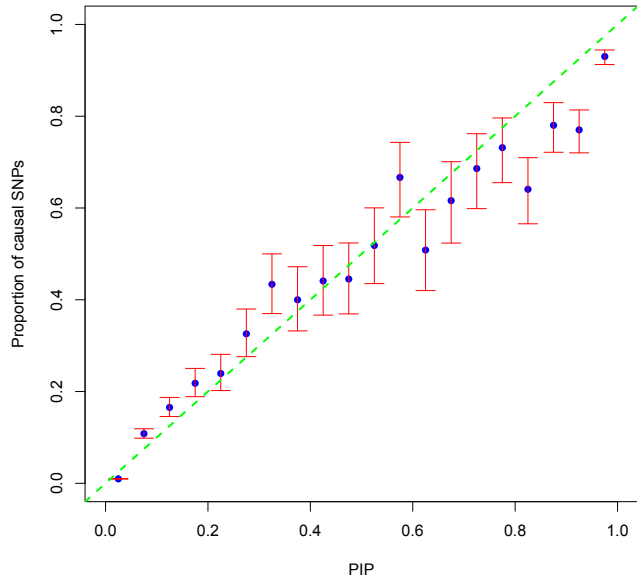


Figure S3

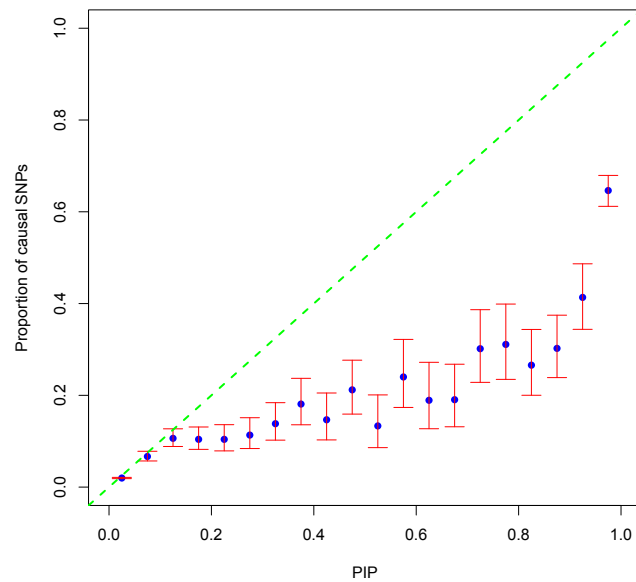
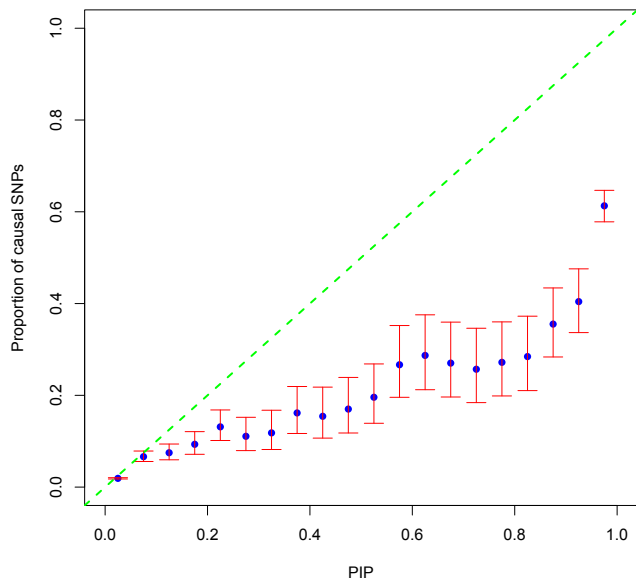
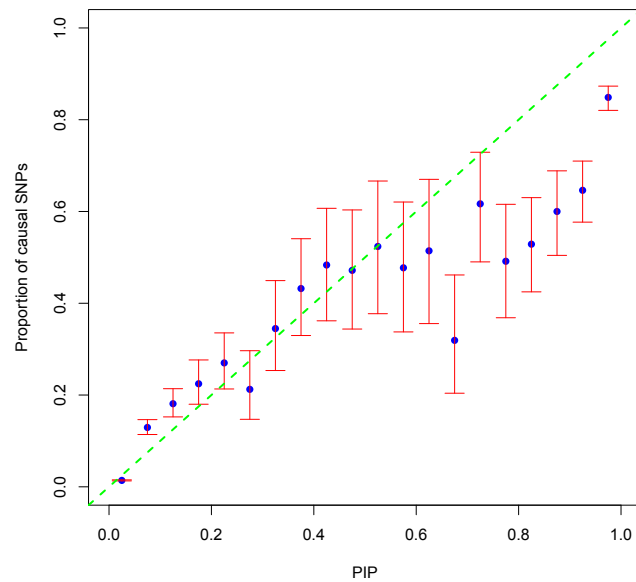
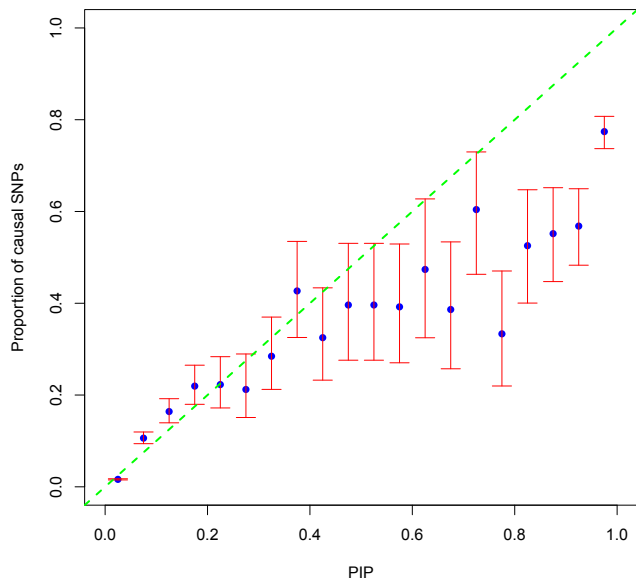


Figure S4

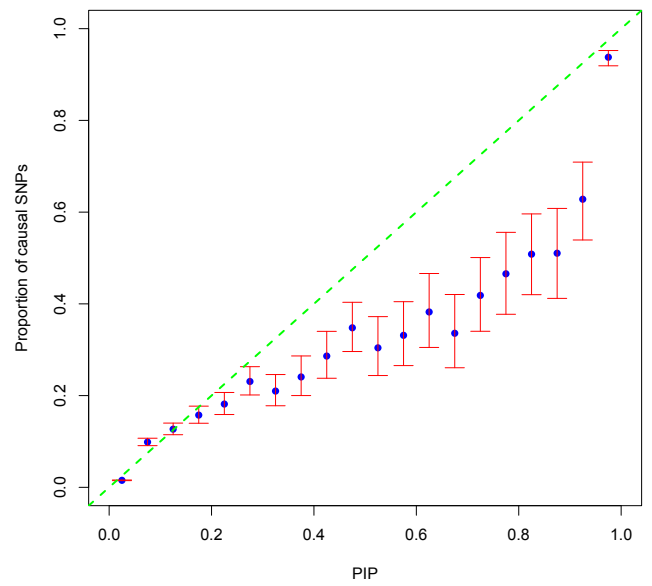
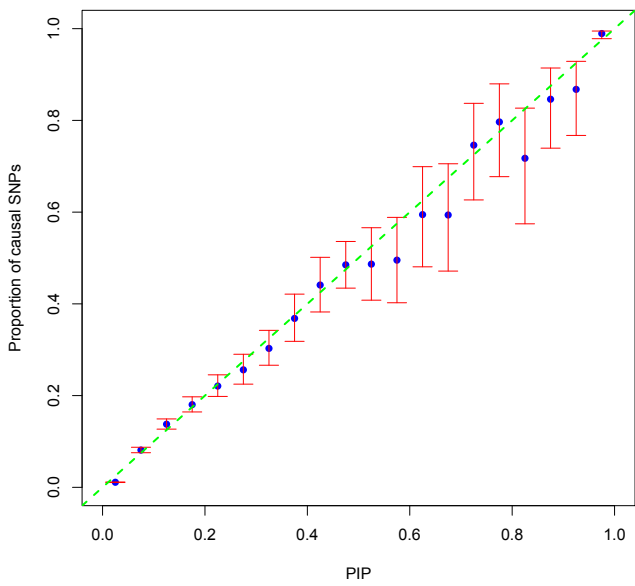
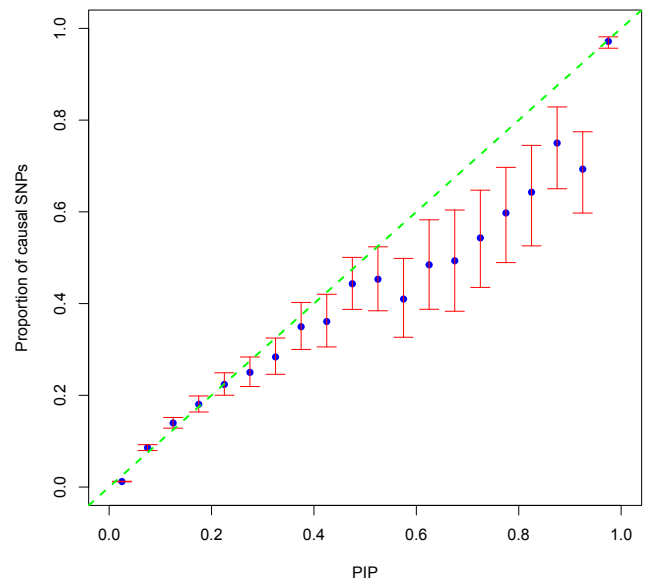
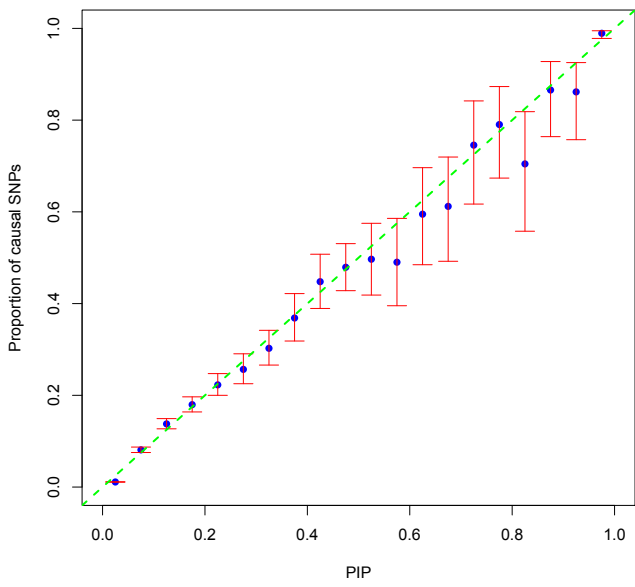
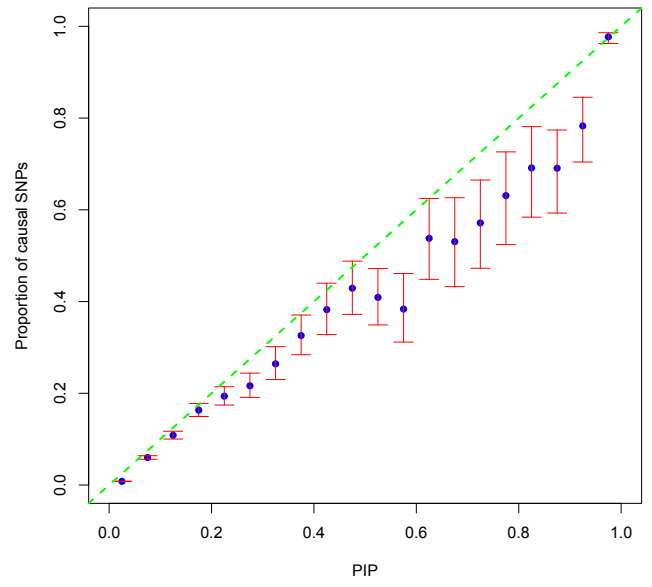
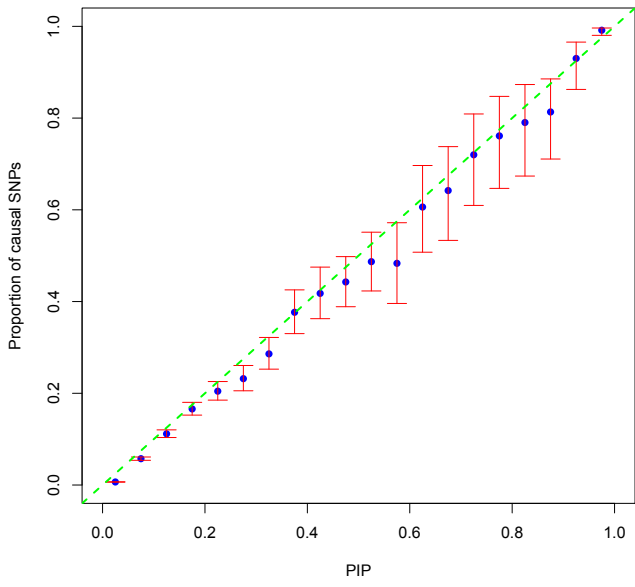
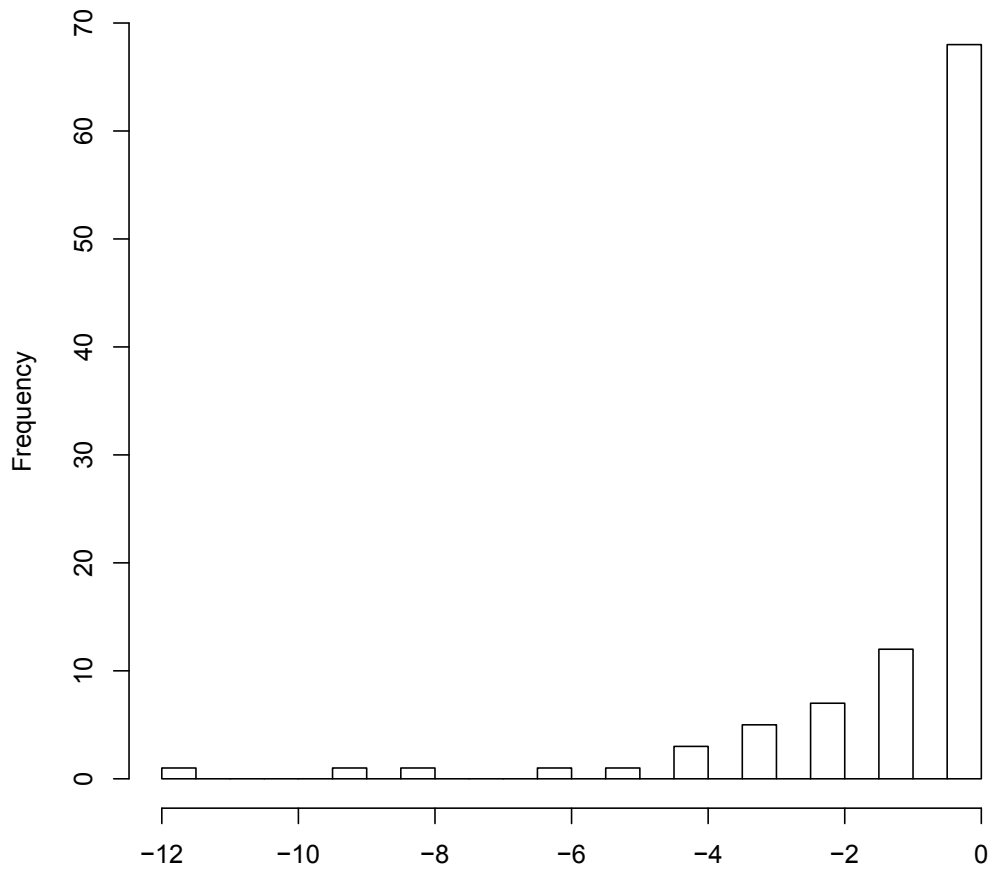


Figure S5

histogram of differences



histogram of differences

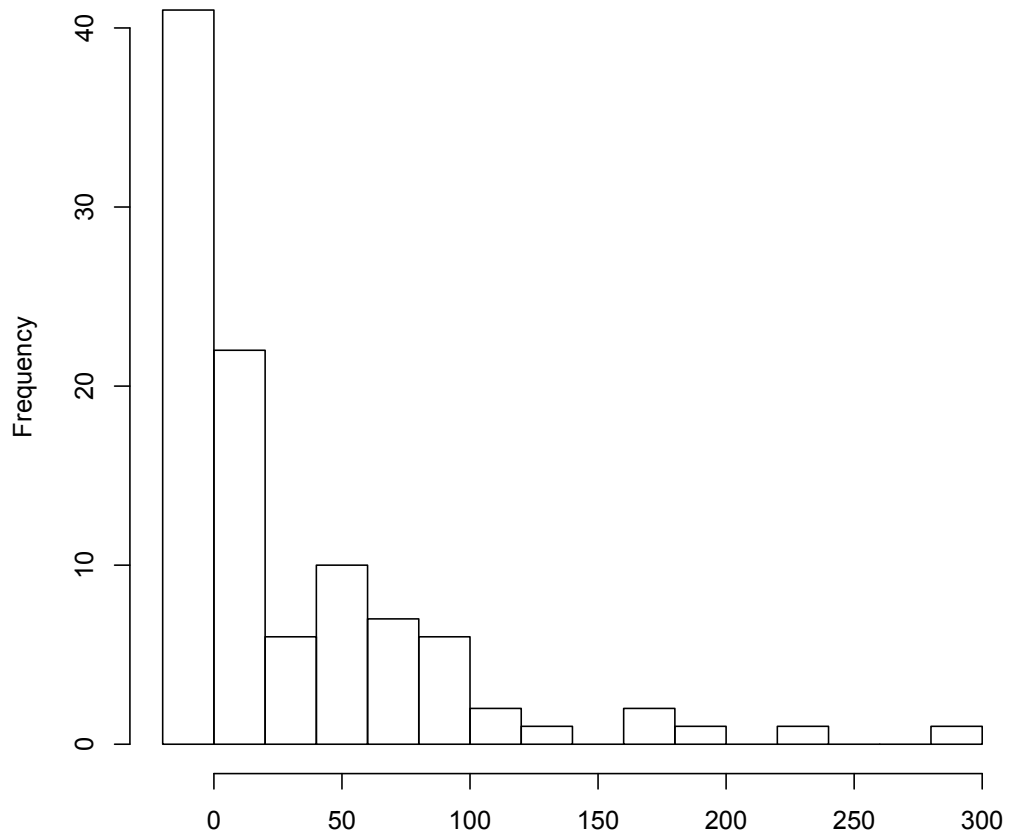


Figure S6

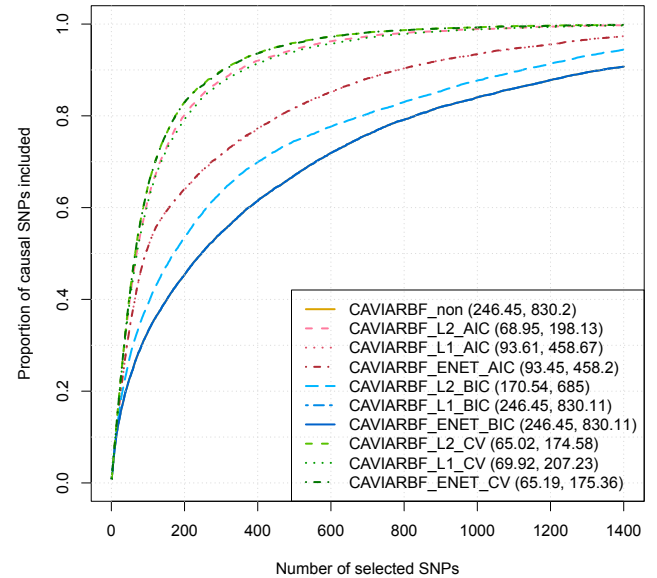
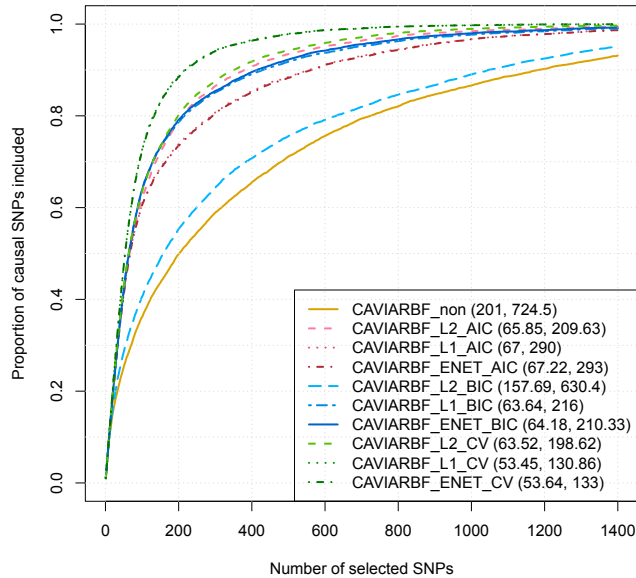
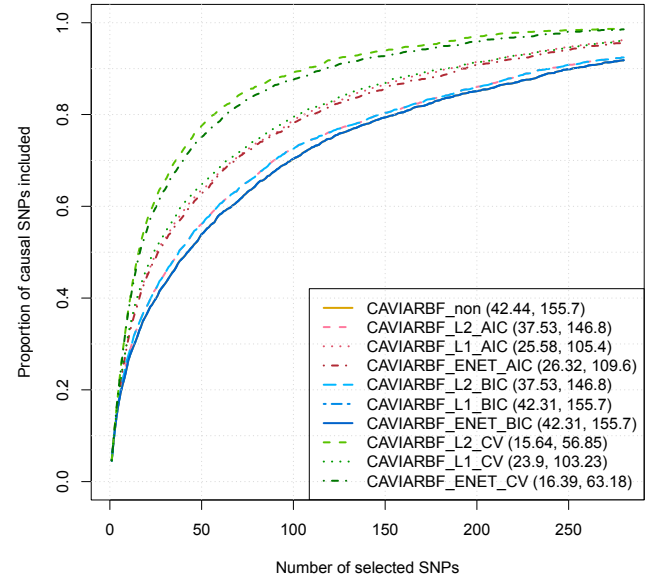
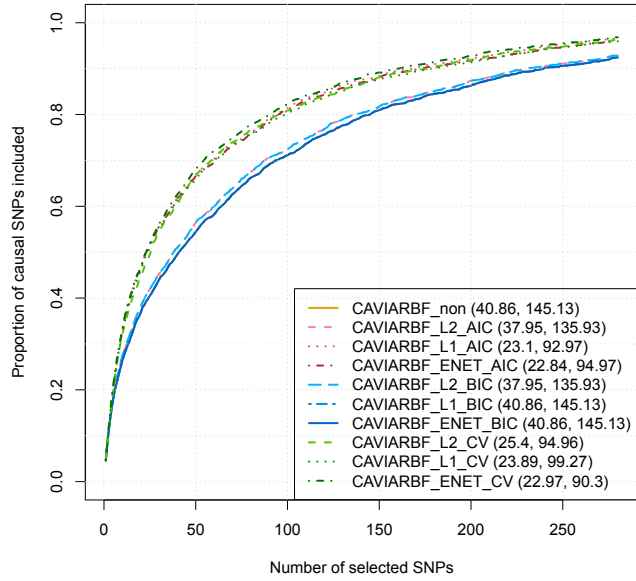


Figure S7

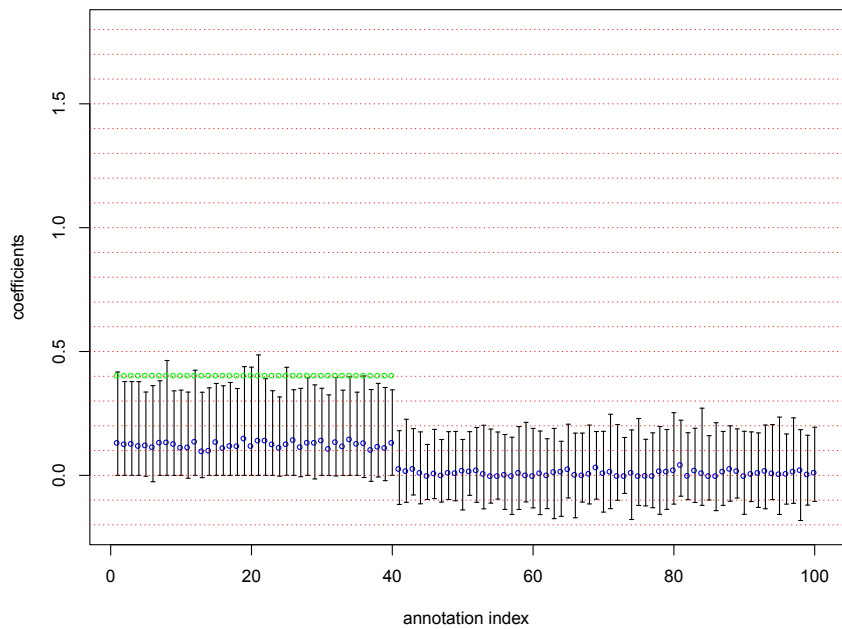
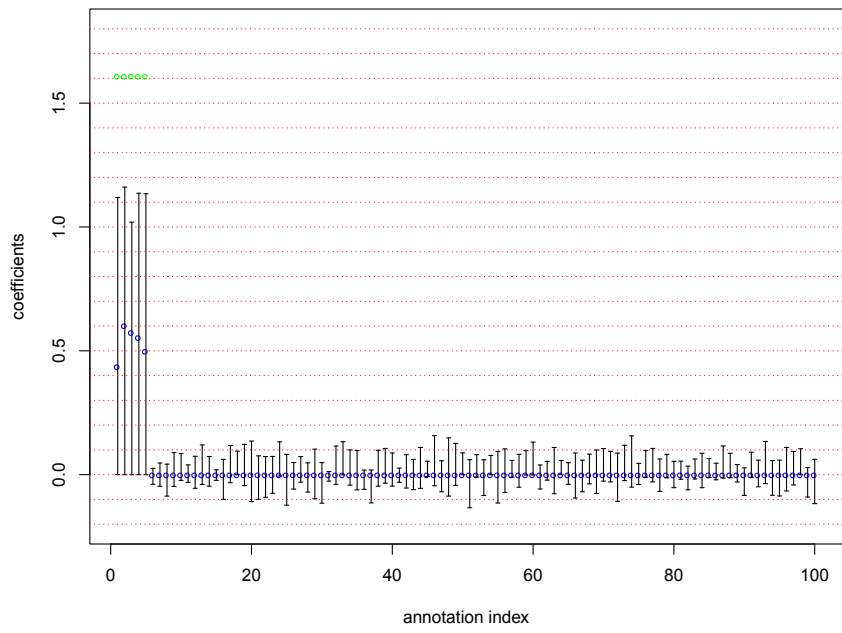
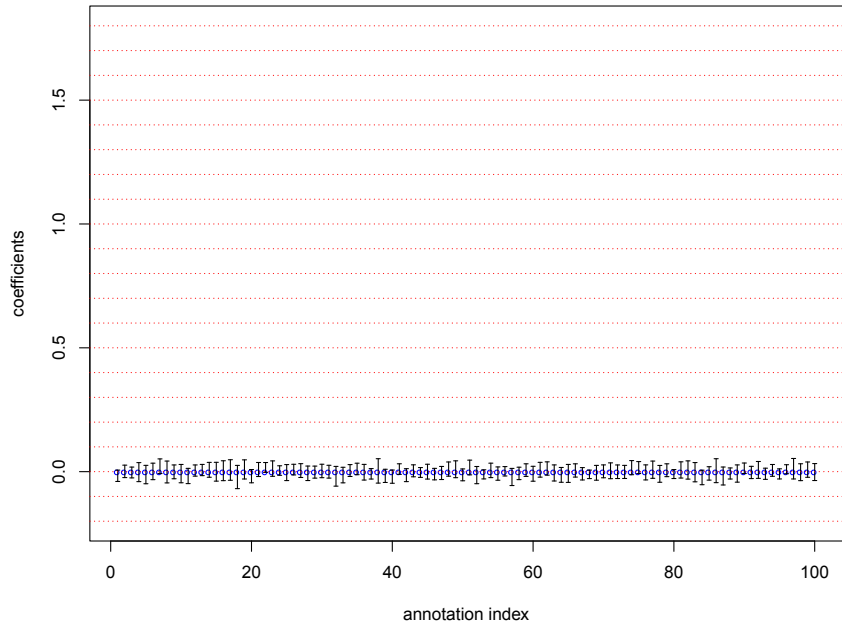


Figure S8

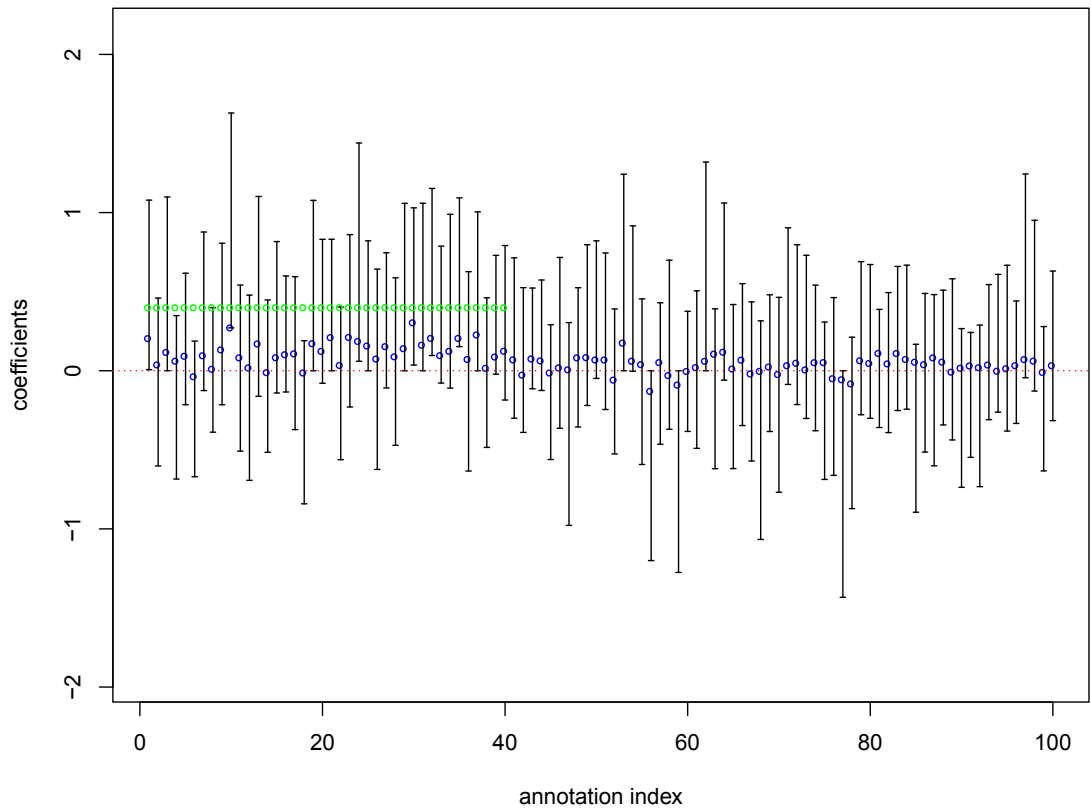
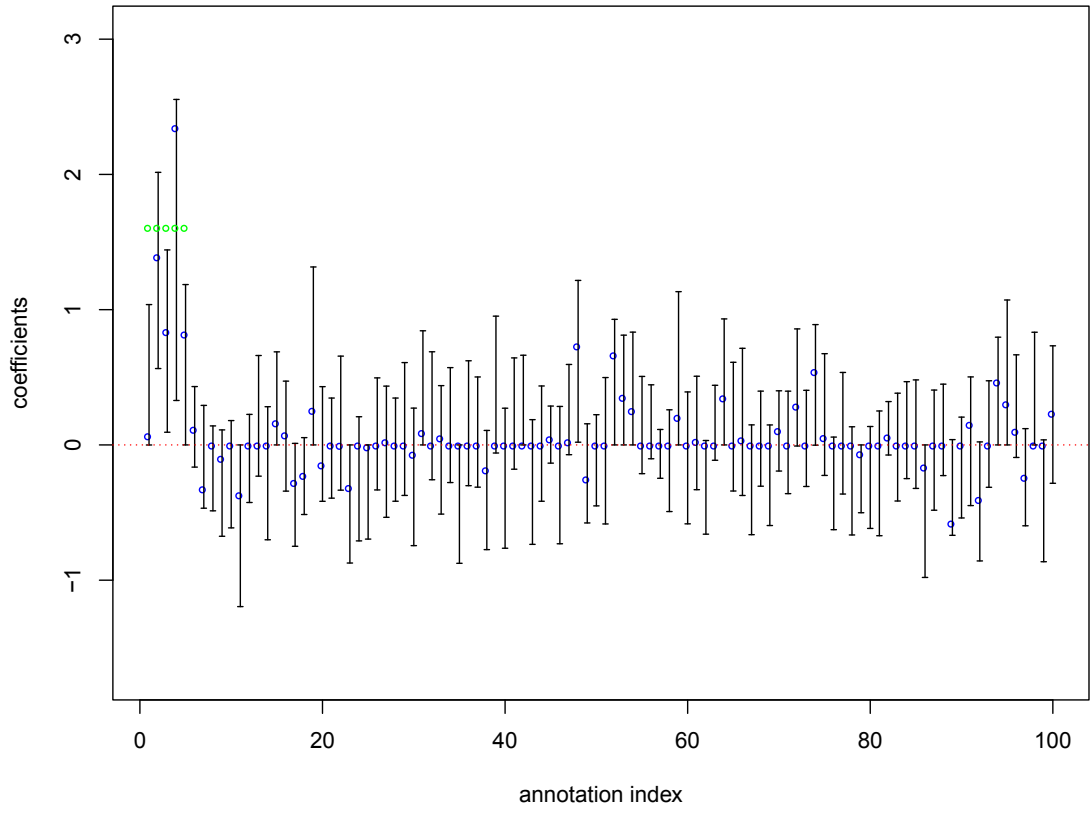


Figure S9

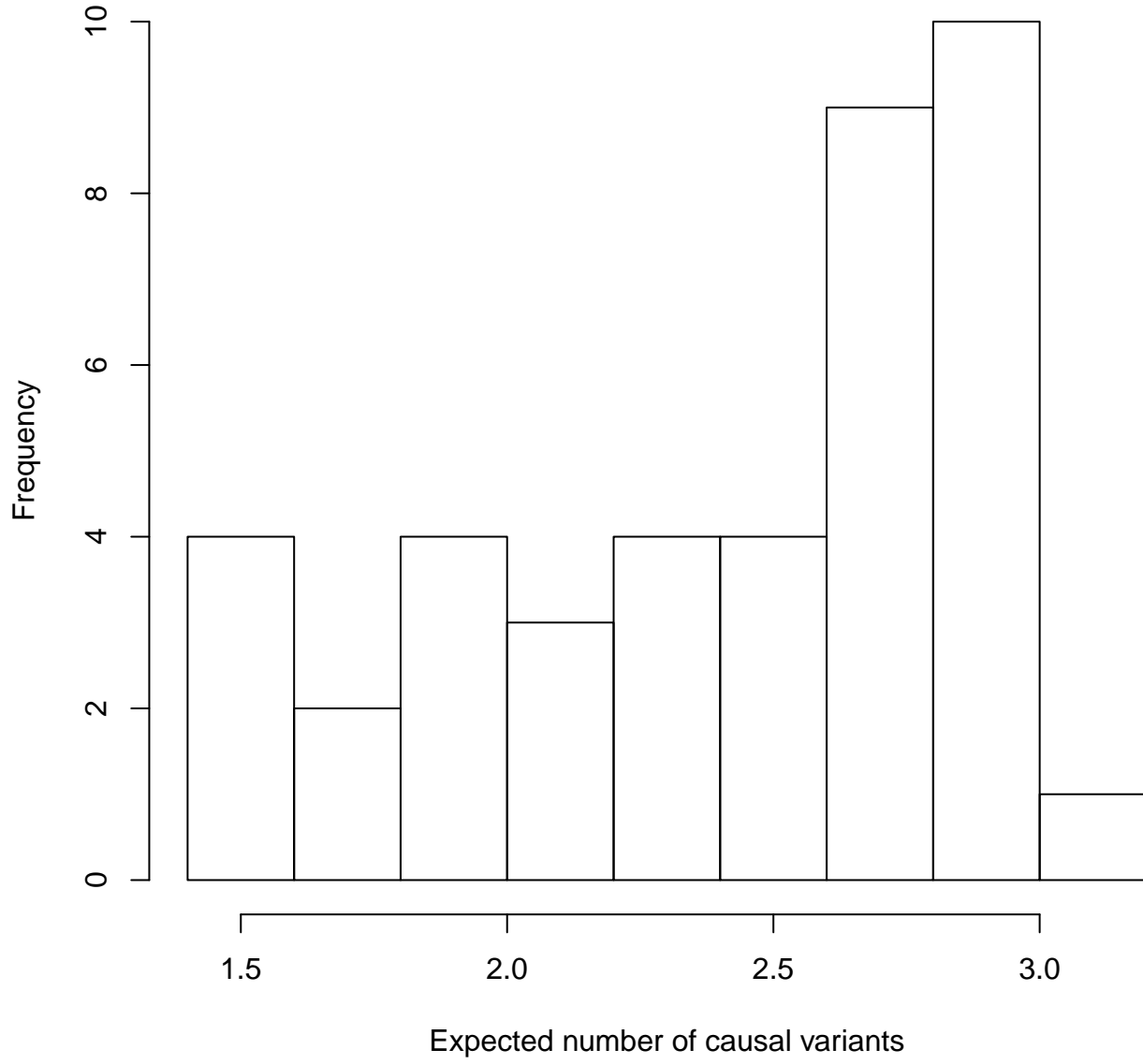


Figure S10

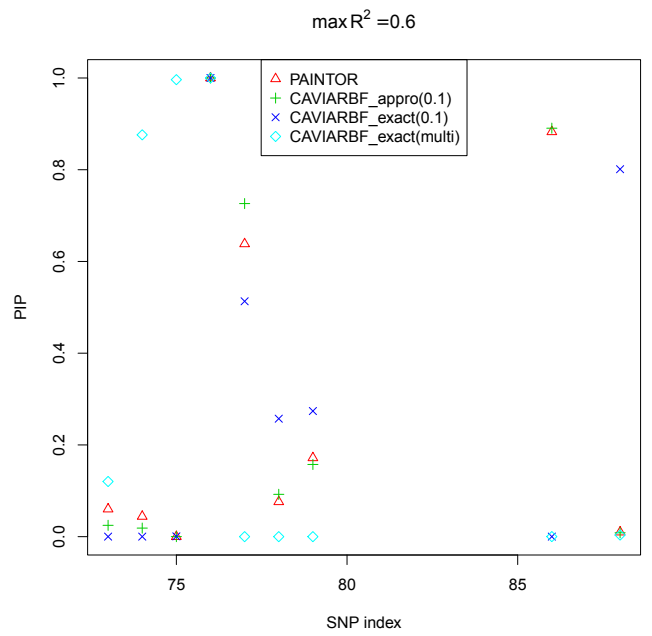
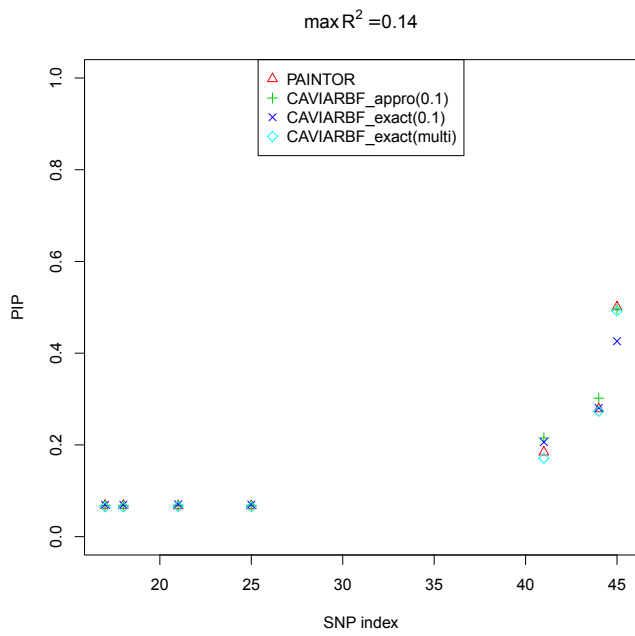


Figure S11

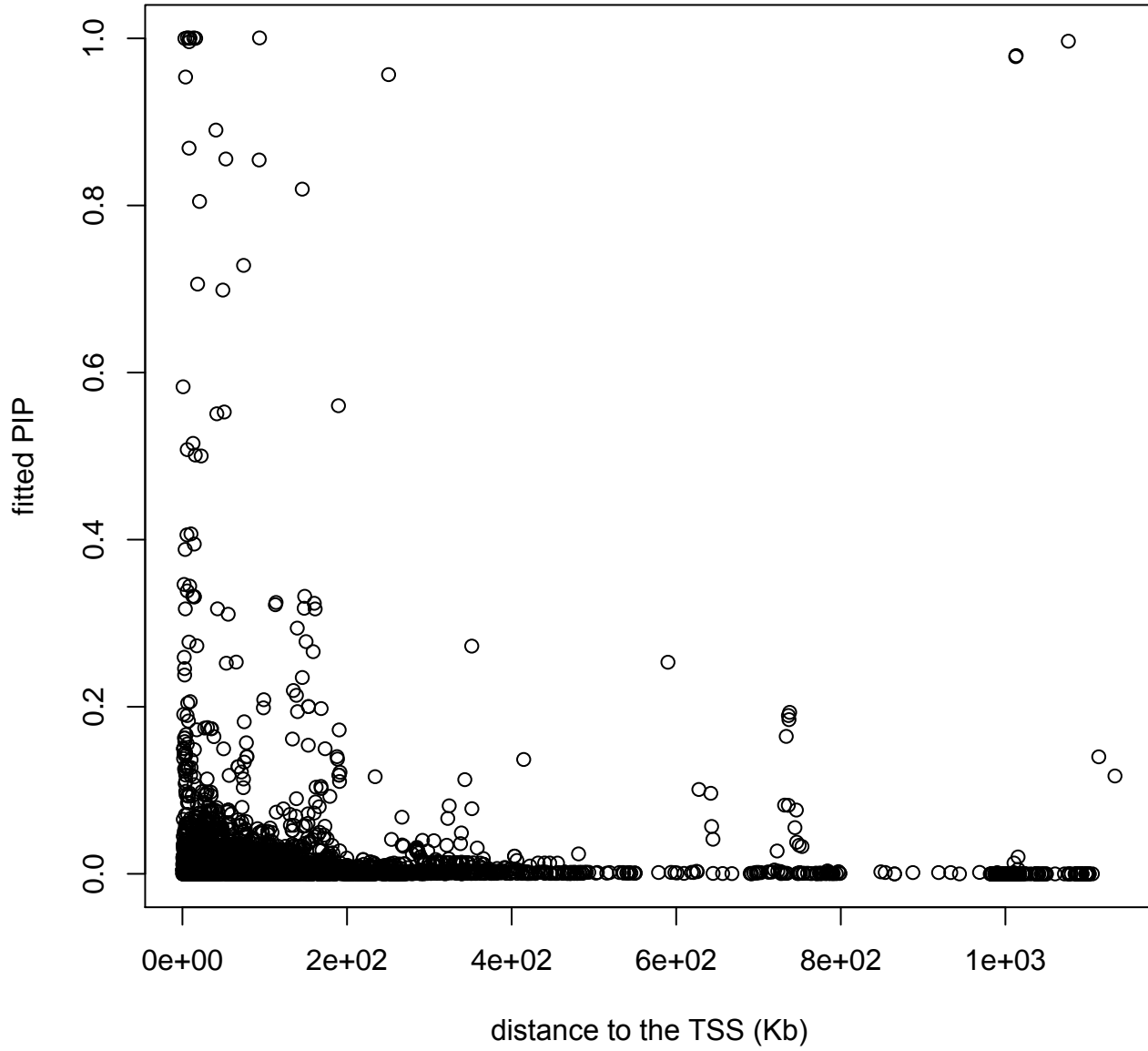


Table S1. Average number of loci with different number of causal variants across 100 simulated data sets.

Initial loci	Informative annotations	1	2	3	4	5	Total loci with causal variants
20	5	7.18	3.46	1.29	0.26	0.08	12.27
20	40	7.63	3.99	1.41	0.31	0.09	13.43
40	5	14.78	6.82	2.45	0.62	0.21	24.88
40	40	14.91	7.74	2.82	0.86	0.17	26.5
100	5	37.01	17.67	5.55	1.50	0.48	62.21
100	40	37.42	20.67	7.44	1.73	0.39	67.65
100*	0	30.83	7.45	1.13	0.17	0.02	39.6
100*	5	34.05	14.81	4.35	1.11	0.14	54.46
100*	40	32.43	8.56	1.76	0.27	0.00	43.02

*Binary annotations.

Table S2. Coverage of the 90% bootstrap confidence interval

	Non-null SNPs	Null SNPs	Overall
5 informative annotations	0.61	0.993	0.974
40 informative annotations	0.8725	0.988	0.942

Table S3. Top ranked annotations for LDL, TC and TG traits

chr	rsid	position	dosage allele*	imputation quality**	z	CAVIARBF _ENET_CV	CAVIARBF _non	CAVIARBF _top10t	PAINTOR _top10t	PAINTOR _non
LDL										
11	rs964184	116648917	C	0.986	10.492	1.000	1.000	1.000	1.000	0.999
16	rs2000999	72108093	A	1	-9.755	0.985	0.979	0.985	0.932	0.899
5	rs6882076	156390297	C	1	-9.748	0.967	0.959	0.970	0.820	0.764
1	rs629301	109818306	T	1	-27.854	0.966	0.968	0.987	0.968	0.928
1	rs2479409	55504650	A	1	11.061	0.935	0.922	0.985	0.942	0.736
20	rs6072249	39656342	G	0.943	-8.827	0.479	0.179	0.874	0.990	0.186
TC										
11	rs964184	116648917	C	0.986	15.739	1.000	1.000	1.000	1.000	1.000
2	rs1260326	27730940	C	1	10.731	0.991	0.979	0.999	0.999	0.962
16	rs2000999	72108093	A	1	-10.153	0.985	0.971	0.985	0.947	0.880
5	rs6882076	156390297	C	1	-10.940	0.984	0.982	0.985	0.876	0.854
2	rs7570971	135837906	A	1	-5.618	0.977	0.978	0.984	0.953	0.932
1	rs2479409	55504650	A	1	10.136	0.958	0.926	0.981	0.927	0.724
5	rs12916	74656539	C	1	-14.363	0.951	0.914	0.993	0.954	0.774
1	rs629301	109818306	T	1	-24.350	0.812	0.789	0.905	0.812	0.696
TG										
11	rs964184	116648917	C	0.986	32.244	1.000	1.000	1.000	1.000	1.000
15	rs138570705	44266730	A	0.924	-11.097	1.000	1.000	1.000	1.000	1.000
2	rs1260326	27730940	C	1	24.539	1.000	1.000	1.000	1.000	1.000
8	rs12678919	19844222	G	1	22.849	0.999	0.999	0.999	0.969	0.970
15	rs2412710	42683787	A	1	-5.624	0.920	0.914	0.702	0.359	0.774
10	rs2068888	94839642	A	1	5.582	0.569	0.567	0.920	0.909	0.445

*Dosage value is the count of this allele. **This is the r2pred output from ImpG indicating the quality of imputed z statistics, 1 indicates the original z statistics from meta-analysis. Variants have at least one PIP > 0.9 from different methods. Rows in bold indicate at least one method outputs PIP less than 0.9. The variants are sorted by PIPs from CAVIARBF_ENET_CV in descending order.

File S1: Annotation results for LDL, TC and TG. (.xlsx, 187 KB)

Available for download as a .xlsx file at:

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.188953/-/DC1/FileS1.xlsx>

File S2: Annotation list and GEO track numbers for eQTL study. (.xlsx, 12 KB)

Available for download as a .xlsx file at:

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.188953/-/DC1/FileS2.xlsx>