DOI: 10.1534/genetics.110.114280

Genomewide Multiple-Loci Mapping in Experimental Crosses by Iterative Adaptive Penalized Regression

Wei Sun,*,†,[1] Joseph G. Ibrahim* and Fei Zou*

*Department of Biostatistics and †Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599

Manuscript received January 17, 2010
Accepted for publication February 3, 2010

ABSTRACT

Genomewide multiple-loci mapping can be viewed as a challenging variable selection problem where the major objective is to select genetic markers related to a trait of interest. It is challenging because the number of genetic markers is large (often much larger than the sample size) and there is often strong linkage or linkage disequilibrium between markers. In this article, we developed two methods for genomewide multiple loci mapping: the Bayesian adaptive Lasso and the iterative adaptive Lasso. Compared with eight existing methods, the proposed methods have improved variable selection performance in both simulation and real data studies. The advantages of our methods come from the assignment of adaptive weights to different genetic makers and the iterative updating of these adaptive weights. The iterative adaptive Lasso is also computationally much more efficient than the commonly used marginal regression and stepwise regression methods. Although our methods are motivated by multiple-loci mapping, they are general enough to be applied to other variable selection problems.

IT is well known that complex traits, including many common diseases, are controlled by multiple loci (Hoh and Ott 2003). However, multiple-loci mapping remains one of the most attracting and most difficult problems in genetic studies, mainly due to the high dimensionality of the genetic markers as well as the complicated correlation structure among genotype profiles (throughout this article, we use the term "genotype profile" to denote the genotype profile of one marker, instead of the genotype profile of one individual). Suppose a quantitative trait and the genotype profiles of p_0 markers (*e.g.*, single-nucleotide polymorphisms, SNPs) are measured in n individuals. We treat this multiple-loci mapping problem as a linear regression problem,

$$y_i = b_0 + \sum_{j=1}^{p} x_{ij} b_j + e_i, \qquad (1)$$

or $\mathbf{y} = \mathbf{b}_0 + \mathbf{Xb} + \mathbf{e}$ in a matrix form, where $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{X} = (x_{ij})_{n\times p}$, $\mathbf{b}_0 = b_0\mathbf{1}_{1\times p}$, $\mathbf{b} = (b_1, \ldots, b_p)^T$, $\mathbf{e} = (e_1, \ldots, e_n)^T$, and $\mathbf{e} \sim N(\mathbf{0}_{n\times 1}, \sigma^2 I_{n\times n})$. y_i is the trait value of the ith individual, and b_0 is the intercept. Here p is the total number of covariates. If we consider only the main effect of each SNP, $p = p_0$; and if we consider the main effects and all the pairwise interactions, $p = p_0 + p_0(p_0 - 1)/2$. x_{ij} is the value of the jth covariate of individual i. The specific coding of x_{ij} depends on the study design and the inheritance model. For example, if additive inheritance is assumed, the main effect of a SNP can be coded as 0, 1, and 2 on the basis of the number of minor alleles. The major objective of multiple-loci mapping is to identify the correct subset model, *i.e.,* to identify those j's, such that $b_j \neq 0$, and estimate the b_j's.

Marginal regression and stepwise regression are commonly used for multiple-loci mapping. Permutation-based thresholds for model selection have been used for these two methods (Churchill and Doerge 1994; Doerge and Churchill 1996). Broman and Speed (2002) proposed a modified Bayesian information criterion (BIC) for model selection, which was further written as a penalized LOD score criterion and implemented within a forward–backward model selection framework (Manichaikul *et al.* 2009). The threshold of the penalized LOD score is also estimated by permutations.

Several simultaneous multiple-loci mapping methods have been developed, among which two commonly used approaches are Bayesian shrinkage estimation and Bayesian model selection. Most existing Bayesian shrinkage methods are hierarchical models based on the additive linear model specified in Equation (1), with covariate-specific priors: $p(b_j|\sigma_j^2) \sim N(0, \sigma_j^2)$. The coefficients are shrunk because their prior mean values are 0. The degree of shrinkage is controlled by the prior of the covariate-specific variance σ_j^2. An inverse-Gamma prior,

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.110.114280/DC1.

[1]*Corresponding author:* University of North Carolina, 4105E McGavran Greenberg Hall, Chapel Hill, NC 27599. E-mail: wsun@bios.unc.edu

$$p(\sigma_j^2 \mid \delta, \tau) = \text{inv-Gamma}(\delta, \tau) = \frac{\tau^\delta}{\Gamma(\delta)}(\sigma_j^2)^{-1-\delta}\exp\left(\frac{-\tau}{\sigma_j^2}\right), \quad (2)$$

leads to an unconditional prior of b_j as a Student's t distribution (Yi and Xu 2008). We refer to this method as the Bayesian t. Another choice is an exponential prior,

$$p\left(\sigma_j^2 \mid \frac{a^2}{2}\right) = \text{Exp}\left(\frac{a^2}{2}\right) = \frac{a^2}{2}\exp\left(-\frac{a^2}{2}\sigma_j^2\right), \quad (3)$$

where a is a hyperparameter. In this case, the unconditional prior of b_j is a Laplace distribution, $p(b_j) = (a/2)e^{-a|b_j|}$ that is closely related to the Bayesian interpretation of the Lasso (Tibshirani 1996); therefore, it has been referred to as the *Bayesian Lasso* (Yi and Xu 2008). Park and Casella (2008) constructed the Bayesian Lasso using a similar but distinct prior: $p(b_j|\sigma_j^2) \sim N(0, \sigma_e^2\sigma_j^2)$ and $p(\sigma_j^2 \mid a^2/2) = \text{Exp}(a^2/2)$. Hans (2009) proposed another Bayesian Lasso method, with more emphasis on prediction than on variable selection. Several general Bayesian model selection methods have been applied for multiple-loci mapping, for example, the stochastic search variable selection (George and McCulloch 1993) and the reversible-jump Markov chain Monte Carlo (MCMC) methods (Richardson and Green 1997). One example is the composite model space approach (CMSA) (Yi 2004).

We propose two variable selection methods: the Bayesian adaptive Lasso (BAL) and the iterative adaptive Lasso (IAL). The BAL is a fully Bayesian approach while the IAL is an expectation conditional maximization (ECM) algorithm (Meng and Rubin 1993). Both the BAL and the IAL are related to the adaptive Lasso (Zou 2006), which extends the Lasso (Tibshirani 1996) by allowing covariate-specific penalties. The adaptive Lasso enjoys the oracle property (Fan and Li 2001); *i.e.*, the covariates with nonzero coefficients will be selected with probability tending to 1, and the estimates of nonzero coefficients have the same asymptotic distribution as the correct model. However, the adaptive Lasso requires consistent initial estimates of the regression coefficients, which are generally not available in the high dimension, low sample size (HDLSS) setting. Huang *et al.* (2008) showed that with initial estimates obtained from the marginal regression, the adaptive Lasso still has the oracle property in the HDLSS setting under a partial orthogonality condition: the covariates with zero coefficients are weakly correlated with the covariates with nonzero coefficients. However, in many real-world problems, including the multiple-loci mapping problem, the covariates with zero coefficients are often strongly correlated with some covariates with nonzero coefficients. The BAL and the IAL extend the adaptive Lasso in the sense that they do not require any informative initial estimates of the regression coefficients so that they can be applied in the HDLSS setting, even if there is high correlation among the covariates.

After we completed an earlier version of this article, we noticed an independent work on extending the adaptive Lasso from a Bayesian point of view (Griffin and Brown 2007). There are several differences between Griffin and Brown's approach and our work. First, Griffin and Brown (2007) did not study the fully Bayesian approach, while we have implemented and carefully studied the BAL. Second, Hoggart *et al.* (2008) implemented Griffin and Brown's approach in HyperLasso, a coordinate descent algorithm, which is different from the IAL at both model setup and implementation. We showed in our simulation and real data analysis that the IAL has significantly better variable selection performance than the HyperLasso. The differences between the IAL and the HyperLasso are further elaborated in the discussion.

In this article, we focus on the genomewide multiple-loci mapping in experimental crosses of inbred strains (*e.g.*, yeast segregants, F_2 mice) where typically thousands of genetic markers are genotyped in hundreds of samples. Another situation is genomewide association studies (GWAS) in outbred populations where millions of markers are genotyped in thousands of individuals. Multiple-loci mapping in experimental crosses and GWAS presents different challenges. In experimental crosses, genotype profiles have higher correlations in longer genomic regions, which give simultaneous multiple-loci mapping methods more advantages than the marginal regression or stepwise regression. In contrast, GWAS data have higher dimensionality but the linkage disequilibrium (LD) blocks often have limited sizes. We focus on the experimental cross in this study, but similar methods can also be applied to GWAS data, although preselection or mapping chromosome-by-chromosome may be necessary to reduce the dimension of the covariates.

The remainder of this article is organized as follows. We first introduce the BAL and the IAL in the following two sections, respectively, and then evaluate them and several existing methods by extensive simulations in simulation studies. A real data study of multiple-loci mapping of gene expression traits is presented in the gene expression QTL study section. We summarize and discuss the implications of our methodology in the discussion.

THE BAL

The BAL is a Bayesian hierarchical model. The priors are specified as

$$p(b_0) \propto 1, \quad p(\sigma_e^2) \propto \frac{1}{\sigma_e^2}, \tag{4}$$

$$p(b_j \mid \kappa_j) = \frac{1}{2\kappa_j}\exp\left(-\frac{|b_j|}{\kappa_j}\right), \tag{5}$$

$$p(\kappa_j \mid \delta, \tau) = \text{inv-Gamma}(\kappa_j;\, \delta,\, \tau) = \frac{\tau^\delta}{\Gamma(\delta)}\kappa_j^{-1-\delta}\exp\left(-\frac{\tau}{\kappa_j}\right), \tag{6}$$

where $\delta > 0$ and $\tau > 0$ are two hyperparameters. The unconditional prior of b_j is

$$\begin{aligned} p(b_j) &= \int_0^\infty \frac{\tau^\delta}{2\Gamma(\delta)}\,\kappa_j^{-2-\delta}\exp\frac{-(|b_j| + \tau)}{\kappa_j}\,d\kappa_j \\ &= \frac{\tau^\delta\delta}{2}(|b_j| + \tau)^{-1-\delta}, \end{aligned} \tag{7}$$

which we refer to as a *power* distribution with parameters δ and τ. From this unconditional prior, we can see that smaller τ and larger δ lead to bigger penalization. In practice, it could be difficult to choose specific values for the hyperparameters δ and τ. Following a similar rationale of Yi and Xu (2008), we suggest a joint improper prior $p(\delta, \tau) \propto \tau^{-1}$ and let the data estimate δ and τ. We suggest the prior of τ^{-1} because in our application in the HDLSS setting, we encourage larger penalty and hence smaller τ and larger δ.

The posterior distribution of all the parameters is given by

$$\begin{aligned} &p(\mathbf{b},\, b_0,\, \sigma_e^2,\, \kappa_1,\, \ldots,\, \kappa_p \mid \mathbf{y}, \mathbf{X}) \\ &\quad\propto p(\mathbf{y} \mid \mathbf{b},\, \mathbf{X},\, b_0,\, \sigma_e^2)p(\sigma_e^2)p(b_0)\prod_{j=1}^{p} p(b_j \mid \kappa_j) \\ &\qquad\times\, p(\kappa_j \mid \delta, \tau)p(\delta, \tau) \\ &\quad\propto \frac{1}{\sigma_e^{2+n}}\exp\left[\frac{-\text{rss}}{2\sigma_e^2}\right]\frac{\tau^{\delta p - 1}}{(\Gamma(\delta))^p} \\ &\qquad\times \prod_{j=1}^{p}\kappa_j^{-2-\delta}\exp\left(-\frac{|b_j| + \tau}{\kappa_j}\right), \end{aligned} \tag{8}$$

where rss indicates residual sum of squares, *i.e.*, $\sum_{i=1}^n (y_i - b_0 - \sum_{j=1}^p x_{ij}b_j)^2$. We sample from this posterior distribution using a Gibbs sampler, which is presented in supporting information, File S1, Section A.

Our model can also be explained from the perspective of penalized least squares. Using the unconditional prior to replace the conditional prior, and taking the negative logarithm of the posterior probability, we obtain a penalized least-squares objective function: $\text{rss}/(2\,\sigma_e^2) + (1 + \delta)\log(|b_j| + \tau)$, with a log penalty for $|b_j|$. Another form of log penalty is $\lambda \log(|b_j|)$, which has been introduced by Zou and Li (2008) as the limit of bridge penalty $|b_j|^\gamma$, as $\gamma \to 0$. Comparing these two forms of log penalties, λ and $1 + \delta$ are both tuning parameters that control the size of the penalty, and our log penalty has an extra parameter τ to give a minimum penalty when $b_j = 0$. In contrast, if $b_j = 0$, the log penalty by Zou and Li (2008) is infinite, which is not a problem for their one-step estimate, given the initial estimate of b_j is not zero, but will cause problems in our iterative algorithms. Friedman (2008) proposed a log penalty, $\lambda \log((1-\beta)|b_j| + \beta)$, with $0 < \beta < 1$, which is equivalent to the log penalty that we use. Friedman's motivation is that this log penalty bridges the L_1 penalty (Lasso) and the L_0 penalty (all-subset selection) as β changes from 1 to 0. In contrast, we motivate this log penalty and solve the penalized least-squares problem from a Bayesian point of view.

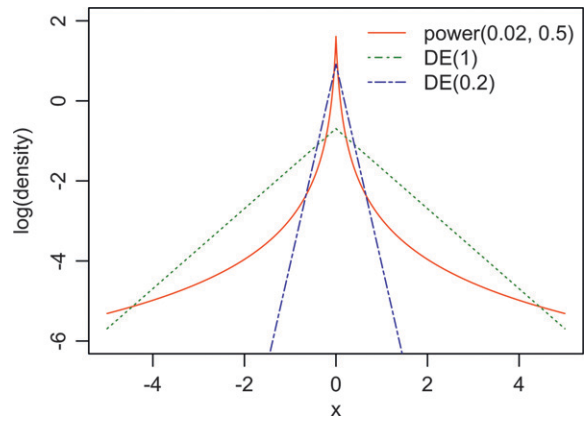


Figure 1.—Comparison of the power distribution $f(x; \tau, \delta) = (\tau^\delta\delta/2)(|x| + \tau)^{-1-\delta}$, given $\tau = 0.02$ and $\delta = 0.1$, and the Laplace (*i.e.*, double exponential, DE) distribution $f(x; \kappa) = (1/2\kappa)\exp(-|x|/\kappa)$, given $\kappa = 1$ or 0.2. We plot the density in log scale for better illustration. The power distribution tends to have a higher peak at zero and heavier tails for larger values.

The BAL can be better understood by comparing it with the Bayesian Lasso. Recall that in the Bayesian Lasso, $b_j \mid \sigma_j^2 \sim N(0,\, \sigma_j^2)$, $\sigma_j^2 \sim \text{Exp}(a^2/2)$, and the unconditional prior for each b_j is a double-exponential distribution. The conditional normal prior resembles a ridge penalty and the unconditional Laplace prior resembles a Lasso penalty. Therefore an intuitive (albeit not accurate) explanation of the Gibbs sampler for the Bayesian Lasso is that it approaches a Lasso penalty by iteratively applying covariate-specific ridge penalties. Figure 1 in Park and Casella (2008) justifies this intuitive explanation: the coefficient paths of the Bayesian Lasso are a compromise between the coefficient paths of the Lasso and ridge regression. In contrast, an intuitive explanation of the BAL is that it approaches the log penalty by iteratively applying the adaptive Lasso penalty. Figure 1 illustrates the difference between the unconditional prior distribution of BAL and the Bayesian Lasso. The former has a higher peak at zero and heavier tails, which leads to more penalization for smaller coefficients and less penalization for larger coefficients. This shape can potentially be an advantage in the HDLSS setting

where strong penalty is needed, although as pointed by a reviewer, what shape works best depends on the data.

THE IAL

Because no point mass at zero is specified in the Bayesian shrinkage methods (including the BAL), the samples of the regression coefficients would not be exactly zero, and thus the Bayesian shrinkage methods do not automatically select variables. However, if we look for the mode of the posterior distribution, it could be exactly zero. This leads to the following ECM algorithm: the iterative adaptive Lasso. Specifically, under the setup of the BAL (Equations 4–6), we treat $\theta = (b_0, b_1, \ldots, b_p)$ as parameter and let $\phi = (\sigma_e^2, \kappa_1, \ldots, \kappa_p)$ be the missing data. The observed data are y_i and x_{ij}. The complete data log-posterior of θ is

$$l(\theta \mid \mathbf{y}, \mathbf{X}, \phi) = C - \frac{\text{rss}}{2\sigma_e^2} - \sum_{j=1}^{p} \frac{|b_j|}{\kappa_j}, \tag{9}$$

where C is a constant with respect to θ. Suppose in the tth iteration the parameter estimates are $\theta^{(t)} = (b_0^{(t)}, b_1^{(t)}, \ldots, b_p^{(t)})$. Then after some derivations (File S1, Section B), the conditional expectation of $l(\theta \mid \mathbf{y}, \mathbf{X}, \phi)$ with respect to the conditional density of $f(\phi \mid \mathbf{y}, \mathbf{X}, \theta^{(t)})$ is

$$Q(\theta \mid \theta^{(t)}) = C - \frac{\text{rss}/2}{\text{rss}^{(t)}/n} - \sum_{j=1}^{p} \frac{|b_j|}{(|b_j^{(t)}| + \tau)/(1+\delta)}, \tag{10}$$

where $\text{rss}^{(t)}$ is the residual sum of squares calculated on the basis of $\theta^{(t)} = (b_0^{(t)}, b_1^{(t)}, \ldots, b_p^{(t)})$. Comparing Equations 9 and 10, it is obvious that to obtain $Q(\theta \mid \theta^{(t)})$, we can simply let $\sigma_e^2 = \text{rss}^{(t)}/n$ and $\kappa_j = (|b_j^{(t)}| + \tau)/(1+\delta)$.

On the basis of the above discussions, the IAL is implemented as follows:

1. Initialization: We initialize $b_j (0 \le j \le p)$ with zero, initialize σ_e^2 by variance of y, and initialize $\kappa_j (1 \le j \le p)$ with $\tau/(1+\delta)$.
2. Conditional maximization (CM) step:
   a. Update b_0 by its posterior mode (see File S1, Section B for more details), $b_0 = (1/n)\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij} b_j\right)$.
   b. For $j = 1, \ldots, p$, update b_j by its posterior mode (see File S1, Section B),

$$\begin{cases} b_j = 0 & \text{if } -\frac{\sigma_j^2}{\kappa_j} \le \bar{b}_j \le \frac{\sigma_j^2}{\kappa_j} \\ b_j = \bar{b}_j - \frac{\sigma_j^2}{\kappa_j} & \text{if } \bar{b}_j > \frac{\sigma_j^2}{\kappa_j} \\ b_j = \bar{b}_j + \frac{\sigma_j^2}{\kappa_j} & \text{if } \bar{b}_j < \frac{-\sigma_j^2}{\kappa_j}, \end{cases}$$

where

$$\sigma_j^2 = \frac{\sigma_e^2}{\sum_{i=1}^{n} x_{ij}^2},$$

and

$$\bar{b}_j = \left(\sum_{i=1}^{n} x_{ij}^2\right)^{-1} \sum_{i=1}^{n} x_{ij}\left(y_i - b_0 - \sum_{k \ne j} x_{ik} b_k\right).$$

3. Expectation (E) step: With the updated b_j's, recalculate the residual sum of squares, rss, and do the following:
   a. Update σ_e^2: $\sigma_e^2 = \text{rss}/n$.
   b. Update κ_j: $\kappa_j = (|b_j| + \tau)/(1+\delta)$.

We say the algorithm is converged if the coefficient estimates $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_p$ have little change.

The above discussions proved that the IAL is an ECM algorithm, which guarantees its convergence. In each step of the ECM algorithm, the conditional log posterior is concave, so a local maximum is a global maximum, and thus it is computationally easy to maximize the conditional log posterior. However, the unconditional log prior is not concave; thus for some tuning parameters, the unconditional log posterior (that is, the concave log likelihood plus the nonconcave log prior) could be nonconcave, so that a local maximum many not be the global maximum. Therefore, similar to other EM algorithms, it is possible that the IAL identifies a local mode of the posterior. This is a common problem for EM algorithms and we address this issue by choosing appropriate initial values of the coefficients and appropriate tuning of the hyperparameters τ and δ.

In the HDLSS setting, especially where the covariates are highly correlated, initial estimates from ordinary least squares (OLS) or ridge regression are unavailable, unstable, or noninformative. Therefore we initialize all the coefficients by zero.

To decide δ and τ, we first consider this problem in an asymptotic point of view to show that theoretically, we can identify optimal δ and τ (see Theorem 1 in File S1, Section C). However, the theoretical results provide only a rough scale for δ and τ. In practice, we select the specific values of δ and τ by the BIC, followed by a variable filtering step. The BIC is written as

$$\text{BIC}_{\tau,\delta} = \log\frac{\text{rss}}{n} + \frac{\log(n)}{n}\text{d.f.}_{\tau,\delta}, \tag{11}$$

where $\text{d.f.}_{\tau,\delta}$ is the number of nonzero coefficients, an estimate of the degrees of freedom (Zou *et al.* 2007). Given the subset model selected by the BIC, the variable filtering step can be implemented by a single multiple-regression or stepwise backward selection. Multiple regression is computationally more efficient. However, occasionally, if two highly correlated covariates are both included in the model, it is possible that both of them are insignificant by multiple regression. In backward regression, however, after dropping one of them, the other one may become significant. We use backward regression for all the numerical results in this article. The P-value cutoff for variable filtering can be set as $0.05/p_E$, where p_E is the effective number of independent tests.

A conservative choice is $p_E = p$, the total number of covariates. In this article, we estimate p_E by the number of independent tests; see Sun and Wright (2010) and File S1, Section D for details. Our approach is closely related to the screening and cleaning method proposed by Wasserman and Roeder (2009). Wasserman and Roeder (2009) use Lasso, marginal regression, or forward stepwise regression accompanied with cross-validation for screening and use a multiple regression for cleaning. We use the IAL for screening and use multiple regression or backward regression for cleaning. As pointed out by an anonymous reviewer, a ridge regression using all the selected covariates may also be an appropriate choice for cleaning.

In contrast to the BIC plus variable filtering approach, an alternative strategy is to apply an extended BIC, which provides larger penalty for bigger models (Chen and Chen 2008). The simulation results in the next section show that the extended BIC leads to slightly worse variable selection performance. Our explanation is that the extended BIC is valid asymptotically and it is conservative when the sample size is relatively small ($n =$ 360 in our simulation). Compared with the extended BIC, the ordinary BIC tends to select larger models with all or most of the true discoveries plus some false discoveries (Chen and Chen 2008), which can be filtered out by the variable filtering step.

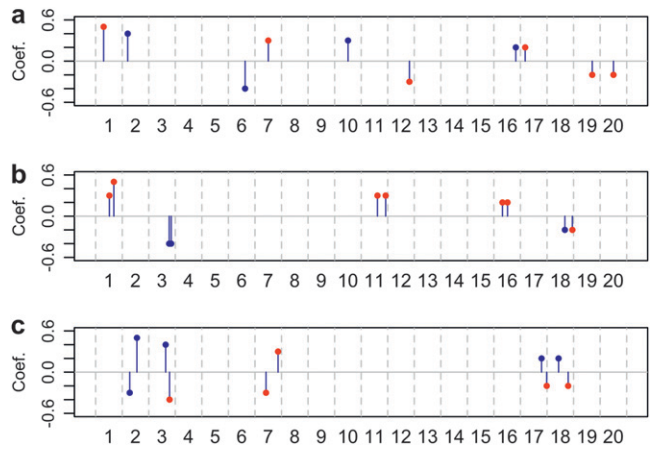


Figure 2.—The locations and effect sizes of the QTL in the simulation study. The markers labeled with red are among the 800 markers with "missing data". (a) Situations 1 and 2. (b) Situations 3 and 4. The genetic distances/R^2 between two QTL from chromosomes 1, 3, 11, 16, and 18 are 15 cM/0.63, 7 cM/0.68, 28 cM/0.25, 17 cM/0.56, and 26 cM/0.40, respectively, where R^2 denotes the correlation square. (c) Situations 5 and 6. The genetic distances/R^2 between two QTL from chromosomes 2, 3, 7, 17, and 18 are 25 cM/0.31, 13 cM/0.59, 41 cM/0.22, 17 cM/0.55, and 33 cM/0.30, respectively. The means (standard deviations) of the proportion of trait variance explained by the 10 QTL in the six simulation situations are 0.31 (0.02), 0.48 (0.03), 0.44 (0.03), 0.62 (0.04), 0.17 (0.01), and 0.29 (0.02), respectively.

SIMULATION STUDIES

We first use simulations to evaluate the variable selection performance of 10 methods: marginal regression, forward regression, forward–backward regression (with penalized LOD as the model selection criterion), the CMSA, the adaptive Lasso (with initial regression coefficients from marginal regression), the IAL, the HyperLasso, and three Bayesian shrinkage methods: the Bayesian *t*, the Bayesian Lasso, and the BAL. See File S1, Section E for the implementation details.

Simulation setup: We first simulate a marker map of 2000 markers from 20 chromosomes of length 90 cM, with 100 markers per chromosome [using function sim.map in R/qtl (Broman *et al.* 2003)]. The chromosome length is chosen to be close to the average chromosome length in the mouse genome. Next we simulate genotype data of the 360 F_2 mice on the basis of the simulated marker map (using function sim.cross in R/qtl). As expected, the markers from different chromosomes have little correlation, while the majority of the markers within the same chromosome are positively correlated (Figure S1). In fact, given the genetic distance of two SNPs, the expected R^2 between two SNPs in this F_2 cross can be explicitly calculated (Figure S2). For example, the R^2's of two SNPs 1, 5, and 10 cM apart are 0.96, 0.82, and 0.67, respectively. Finally, we randomly choose 10 markers as QTL and simulate quantitative traits in six situations with 1000 simulations per situation. Given the 10 QTL, the trait is simulated on the basis of the linear model in Equation 1, where genotype (x_{ij}) is coded by the number of minor alleles. The QTL effect sizes across the six situations are listed below:

1. Unlinked QTL: One QTL per chromosome, with effect sizes 0.5, 0.4, −0.4, 0.3, 0.3, −0.3, 0.2, 0.2, −0.2, and −0.2; $\sigma_e^2 = 1$. Recall that σ_e^2 is the variance of the residual error.
2. QTL linked in coupling: Two QTL per chromosome, with effect sizes of the QTL for each chromosome as (0.5, 0.3), (−0.4, −0.4), (0.3, 0.3), (0.2, 0.2), and (−0.2, −0.2); $\sigma_e^2 = 1$.
3. QTL linked in repulsion: Two QTL per chromosome, with effect sizes of the QTL for each chromosome as (0.5, −0.3), (0.4, −0.4), (0.3, −0.3), (0.2, −0.2), and (0.2, −0.2); $\sigma_e^2 = 1$.

Situations 2, 4, and 6 are the same as situations 1, 3, and 5, respectively, except that $\sigma_e^2 = 0.5$. The locations and effect sizes of the QTL in each situation are illustrated in Figure 2. To mimic the reality that the genotype of a QTL may not be observed, we randomly select 1200 markers with "observed genotype profiles" and use only these 1200 markers in the multiple-loci mapping. The information loss is limited due to the high density of the markers. In fact, the vast majority of the 800 markers with missing genotype can be tagged with $R^2 >$ 0.8 by at least one marker with observed genotype (Figure S3).

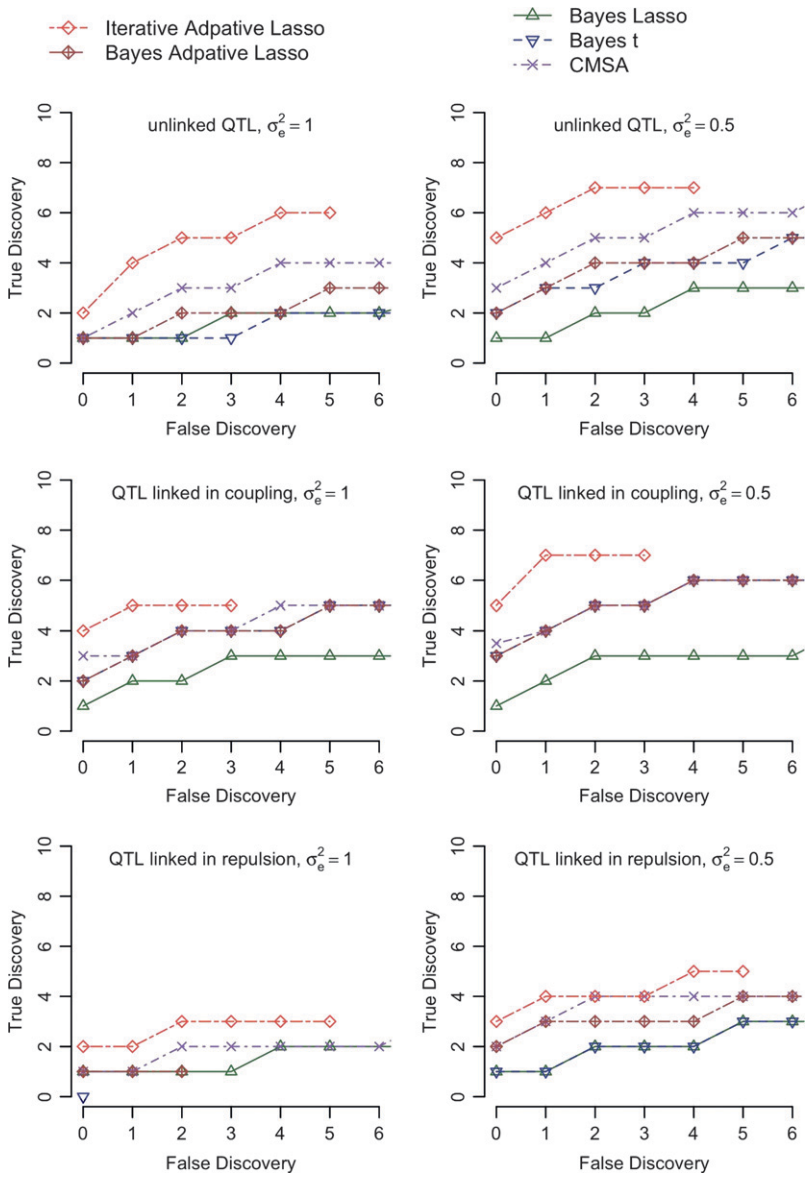


FIGURE 3.—Comparison of the number of true discoveries *vs.* the total number of false discoveries in the simulation study.

One important aspect in the implementation of Bayesian methods is the diagnosis of the convergence of the MCMC. The results in Figure S4, Figure S5, and Figure S6 suggest the convergence of the Bayesian Lasso, Bayesian *t*, and BAL, respectively.

Results: We divide the methods to be tested into two groups: the Bayesian methods that do not explicitly carry out variable selection (since most coefficients remain nonzero) and the stepwise regression, adaptive Lasso, and HyperLasso that explicitly select a subgroup of variables. The IAL is classified into the first group if we use the ordinary BIC to select the hyperparameters; and it is classified into the second group if we use the extended BIC or ordinary BIC plus variable filtering.

For either group, we compare the performance of different methods by comparing the number of true discoveries and false discoveries across different cutoffs of coefficient size or posterior probability. Given a cutoff, we can obtain a final model. We count the number of true discoveries in the final model as follows. For each of the true QTL, we check whether any marker in the final model satisfies the following three criteria: (1) it is located on the same chromosome as the QTL, (2) it has the same effect direction (sign of the coefficient) as the QTL, and (3) the R^2 between this marker and the QTL is >0.8. Different cutoffs such as 0.7 and 0.9 lead to similar conclusions (results not shown). If there is no such marker, there is no true discovery for this QTL. If there is at least one such marker, the one with the highest R^2 with the QTL is recorded as a true discovery and is excluded from the true discovery searching of other QTL. After the true discoveries of all the QTL are identified, the remaining markers in the final model are defined as false discoveries. These false discoveries are further divided into two classes: false discoveries linked to at least one QTL (linked false discoveries) and false discoveries unlinked to any QTL (unlinked false discoveries). A false discovery is linked to a

QTL if it satisfies the above three criteria. We summarize the results of each method by an receiver operating characteristic (ROC)-like curve that plots the median number of true discoveries *vs.* the median number of false discoveries across different cutoff values. The methods with ROC-like curves closer to the upper-left corner of the plot have better variable selection performance because they have less false discoveries and more true discoveries. It is possible that a few cutoff values correspond to the same median of false discoveries but different medians of true discoveries. In this case, the largest median of true discoveries is plotted to simplify the figure. In other words, these ROC-like curves illustrate the best possible performance of these methods. Forward regression outperforms marginal regression in all situations. Therefore we omit the results for marginal regression for readability of the figures.

We first compare the Bayesian methods and the IAL (with ordinary BIC). If the linked false discoveries are counted as false discoveries, the IAL has apparent advantages in all situations. Approximately, the performance of these methods can be ranked as IAL $\geq$ CMSA $\geq$ BAL $\geq$ Bayesian t $\geq$ Bayesian Lasso (Figure 3). If the linked false discoveries are counted as true discoveries, the performances of different methods are not well separated (Figure S7). Overall the IAL, the BAL, and the CMSA have similar and superior performance, and the Bayesian Lasso has inferior performance. We point out that the comparison of the IAL and the Bayesian methods should be interpreted with caution. The posterior mode estimates of parameters by the IAL provide less information than the fully Bayesian methods. For example, the Bayesian methods provide the distribution of the regression coefficients or the distribution of the probability of being included in the model, while the IAL cannot provide such information.

Next we compare the stepwise regression method, the HyperLasso, the adaptive Lasso (with initial estimates from marginal regression), and the IAL with extended BIC or ordinary BIC plus variable filtering. These methods tend not to select the unlinked false discoveries. In fact, the ROC-like curve for each of these methods is exactly the same whether we count unlinked false discoveries as true discoveries or not. If the linked false discoveries are treated as true discoveries, an additional fine-mapping step is needed to pinpoint the location of the QTL in a cluster of linked markers. Therefore, in general, methods that avoid linked false discoveries should be preferred. As shown in Figure 4, the IAL with ordinary BIC plus variable filtering has the best performance while the HyperLasso has the worst performance in all the situations. When the QTL are linked in repulsion, the HyperLasso has no power at all. The adaptive Lasso has similar performance to the IAL when the signal is strong (*i.e.*, QTL linked in coupling); otherwise it has significantly worse performance than the IAL. The stepwise regression and the IAL using the extended BIC have slightly worse performance than the IAL using ordinary BIC plus variable filtering.

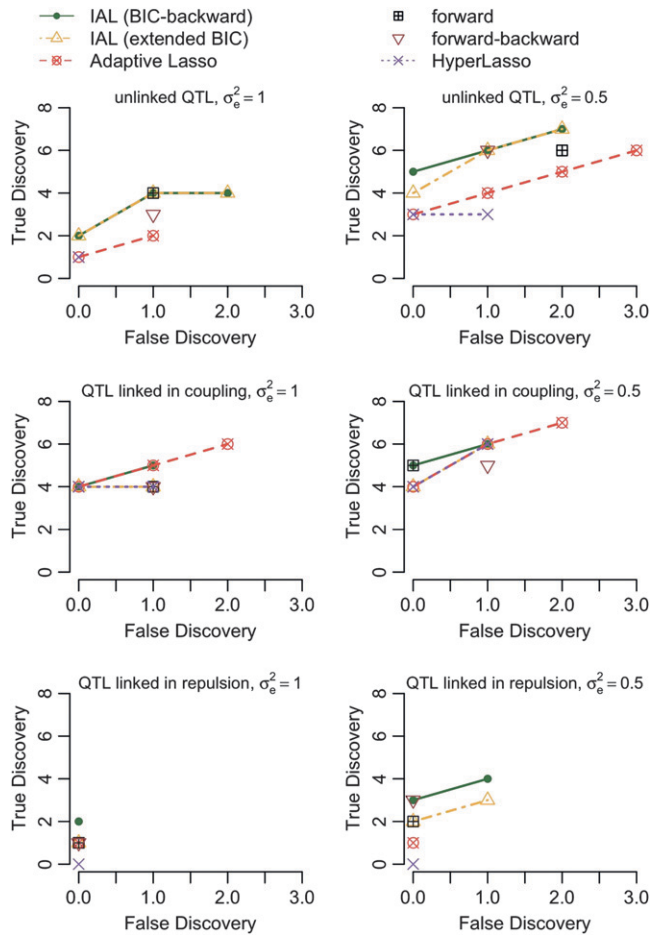


FIGURE 4.—Comparison of the number of true discoveries *vs.* the total number of false discoveries in the simulation study.

We have compared different methods by ROC-like curves and use somewhat *ad hoc* rules to define true/false discoveries. In practice, we may need different criteria to define the true/false discoveries. For example, if there is no missing covariate, we may define the true discovery as the identification of the exact covariate, instead of a highly correlated one. In fact, for prediction, a small number of false discoveries with small coefficients may not affect the prediction accuracy.

GENE EXPRESSION QTL STUDY

QTL study of one particular trait may favor one method by chance. To evaluate our method in real data in a comprehensive manner, we study the gene expression QTL (eQTL) of thousands of genes. The expression of each gene, like other complex traits, is often controlled by multiple QTL (BREM and KRUGLYAK 2005). Therefore multiple-loci mapping has important applications for eQTL studies. In this section, we study eQTL data with >6000 genes and 2956 SNPs in 112 yeast segregants (BREM and KRUGLYAK 2005; BREM *et al.* 2005). The gene

TABLE 1

The number of genes with a certain number of eQTL

Method	Total no. of genes with at least 1 eQTL	No. of genes with			
		1 eQTL	2 eQTL	3 eQTL	>3 eQTL
IAL	3199	1934	771	301	193
Marginal	3289	2536	667	82	4
Forward	3298	2365	734	171	28
Forward–backward	3294	2089	724	294	183
HyperLasso	128	95	20	4	8

expression data are downloaded from Gene Expression Omnibus (GEO) (GSE1990). The expression of 6229 genes is measured in the original data. We drop 129 genes that have >10% missing values and impute the missing values in the remaining 6100 genes by R function impute.knn (Troyanskaya *et al.* 2001). The genotype data were obtained from Rachel Brem. Fifteen SNPs with >10% missing values are excluded from this study, and the missing values in the remaining SNPs are imputed using the function fill.geno in R/qtl (Broman *et al.* 2003). The neighboring SNPs with the same genotype profiles are combined, resulting in 1027 genotype profiles. With >6000 genes, it is extremely difficult, if not impossible, to examine the QTL mapping results gene by gene to filter out possible linked false discoveries. Therefore, the Bayesian methods that generate lots of linked false discoveries were not applied to these eQTL data.

We apply the IAL, marginal regression, forward regression, forward–backward regression, and HyperLasso to these yeast eQTL data to identify multiple eQTL of each gene separately. In other words, we examine the performances of these methods across 6100 traits. The permutation *P*-value cutoff for marginal regression and stepwise regression is set at 0.05. The parameters δ and τ of the IAL are selected by ordinary BIC followed by backward regression, with a *P*-value cutoff 0.05/412, which is based on a conservative estimate of 412 independent tests (see File S1, Section D). For the HyperLasso, the "lambda" parameter is set at $\sqrt{n}\Phi^{-1}(1 - 0.05/412/2)$, and the "shape" parameter is set to be 1. The IAL and stepwise regressions have similar power to identify the genes with at least one eQTL (Table 1). Apparently, the IAL is the most powerful method in terms of identifying multiple eQTL per gene, and the HyperLasso has least power to identify either single eQTL or multiple eQTL per gene (Table 1).

Next we focus on the results of the IAL. Many previous studies have identified one-dimensional (1-D) eQTL hotspots. A 1-D eQTL hotspot is a genomic locus that harbors the eQTL of several genes. Similarly, if the expressions of several genes are associated with the same k loci, these k loci are referred to as a k-D eQTL hotspot. The results of the IAL reveal several 1-D eQTL hotspots (Figure 5), as well as many eQTL hotspots of higher dimensionality. We illustrate the 2-D eQTL hotspots in a two-dimensional plot where one point corresponds to one 2-D eQTL and the x, y coordinates of the point are the locations of the two eQTL (Figure 6). Comparing Figures 5 and 6, it is interesting that a 1-D eQTL can be further divided into several groups on the basis of the results of 2-D eQTL, which is consistent with the finding of Zhang *et al.* (2010).

We divide the whole yeast genome into 600 bins of 20-kb regions, which lead to $600 \times 599/2 = 179{,}700$ bin pairs as potential "2-D eQTL hotspots". Eleven bin pairs are linked to >15 genes (Table S1). The cutoff is chosen arbitrarily so that we can focus on a relatively small group of 2-D hotpots with definite significant enrichment. Due to space limits, we discuss in detail only the largest 2-D hotspot located at chromosome (Chr)15, 160–180 kb and Chr15, 560–580 kb. There are 46 genes linked to these two loci simultaneously, and among them 16 are involved in "generation of precursor metabolites and energy" (*P*-value 3.60×10^{-13}). A closer look reveals that 41 of the 46 genes are linked to one marker block at Chr15, 170,945–180,961 bp, and one marker at Chr15, 563,943 bp. One potential causal gene nearby Chr15, 171–181 kb is PHM7 (Zhu *et al.* 2008), and one potential causal gene nearby Chr15, 564 kb is CAT5 (Yvert *et al.* 2003). Interestingly, both PHM7 and CAT5 are among the 46 genes linked to both loci.

There are also several cases that one group of genes linked to three loci (~35.8 million possible three-loci combinations, Table S2) or even four loci (~5.3 billion possible four-loci combinations, Table S3). For example, three genes, KGD2, SDH1, and SDH3 are all linked to four loci: Chr2, 240–260 kb; Chr13, 20–40kb; Chr15, 160–180 kb; and Chr15, 560–580 kb. Interestingly, all three genes are involved in an "acetyl-CoA catabolic process" (*P*-value 1.93×10^{-7}).

DISCUSSION

In this article, we proposed two variable selection methods, namely the BAL and the IAL. These two methods extend the adaptive Lasso in the sense that they do not require any informative initial estimates of the regression coefficients. The BAL is implemented by MCMC. Through extensive simulations, we observe the BAL has apparently better variable selection perfor-

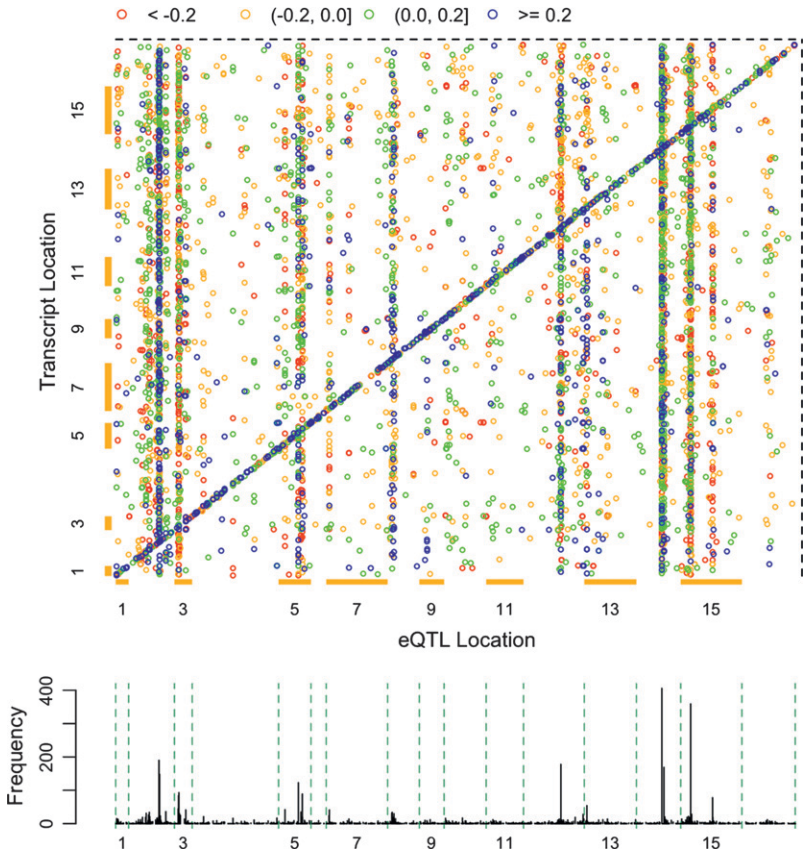


FIGURE 5.—Illustration of the eQTL mapping results in 112 yeast segregants. In the top panel, each point corresponds to a 1-D eQTL result, where the *x*-coordinate is the location of the eQTL and the *y*-coordinate is the location of the gene. Different colors indicate different sizes of the regression coefficients. The diagonal band indicates the *cis*-eQTL, where expression of one gene is associated with the genotype of a nearby marker. The vertical bands indicate 1-D eQTL hotspots. In the bottom panel, the number of genes linked to each marker is plotted. Several 1-D eQTL hotspots are apparent for those markers that harbor the eQTL of hundreds of genes.

mance than the Bayesian Lasso, slightly better performance than the Bayesian *t*, and slightly worse performance than the CMSA. The IAL, which is an ECM algorithm, aims at finding the mode of the posterior distribution. The IAL has uniformly the best variable performance among all the 10 methods we tested. Coupled with a variable filtering step, type I error of the IAL can be explicitly controlled.

The IAL differs from the HyperLasso (Griffin and Brown 2007; Hoggart *et al.* 2008) in at least two aspects. First, the HyperLasso specifies inverse gamma distribution for $\kappa_j^2/2$, and the resulting unconditional posterior relies on a numerical function. In contrast, we specify inverse gamma distribution for κ_j, and it has a much simpler unconditional posterior (Equation 7). The difference is not trivial since it leads not only to convenient theoretical studies in Theorem 1 (File S1, Section C), but also to better numerical stability. For example, the HyperLasso becomes unstable for small shape parameter while IAL is stable for all possible values of δ and τ. Second, we select δ and τ by BIC and further filter out covariates with insignificant effects. In contrast, the HyperLasso directly assigns a large penalization to control the type I error. As shown in the *Results* section, the strong penalization of HyperLasso leads to little power to detect relatively weaker signals.

The IAL is computationally very efficient. For example, it takes 4 hr to carry out the multiple-loci mapping for the yeast eQTL data with 6100 genes and 1017 markers. In contrast, marginal regression, forward regression, and forward–backward regression take ~60, 100, and 200 hr. All of the computation was done using a Dual Xenon 2.0-Ghz Quadcore server. One additional computational advantage of the IAL is that the type I error is controlled by the computationally efficient variable filtering step. The IAL results can be reused for different type I errors. In contrast, for the stepwise regression, all the computation needs to be redone for each type I error.

Our results seem to contradict the results of Yi and Xu (2008) that the Bayesian Lasso has adequate variable selection performance. This inconsistency can be explained by the fact that we are studying the variable selection problem with a much denser marker map. It is known that the Lasso does not have variable selection consistency if there are strong correlations between the covariates with zero and nonzero coefficients (Zou 2006). Since the Bayesian Lasso has similar penalization characteristics to the Lasso (Park and Casella 2008) and the denser marker map leads to higher correlations among genotype profiles, it is not surprising that the Bayesian Lasso has inferior performance in our simulations. In fact, in our simulations, the Bayesian Lasso overpenalizes the regression coefficients (Figure S8). This is consistent with the findings that "Lasso has had to choose between including too many variables or overshrinking the coefficients" (Radchenko and James 2008, p. 1310). In contrast, the Bayesian *t*, the BAL, and the IAL have increasingly smaller penalization on the

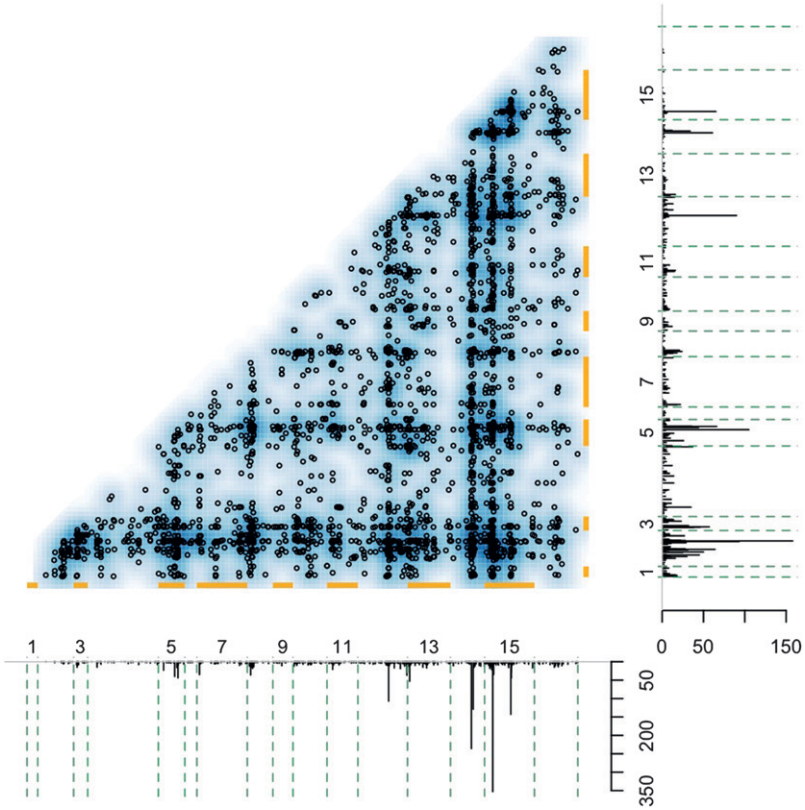


FIGURE 6.—The distribution of the locus pairs linked to the same gene. Here each point corresponds to a 2-D eQTL, and the background color reflects the density of the distribution. If the expression of one gene is linked to more than two loci, we plot each pair of linked loci. For example, if one gene is linked to three markers 1, 2, and 3, which are located at positions p_1, p_2, and p_3, respectively, this gene corresponds to three points in the figure, (p_1, p_2), (p_1, p_3), and (p_2, p_3).

coefficients estimates. The IAL seems to provide unbiased coefficient estimates. This leads to an assumption that the IAL has the oracle property, which warrants further theoretical study.

Due to the computational cost and the need to further filter out false discoveries, the Bayesian shrinkage methods are less attractive for large-scale computation such as eQTL studies. However, there is room to improve these Bayesian shrinkage methods, such as the equi-energy sampler approach (KOU *et al.* 2006). Alternative prior distribution for hyperparmeters may also lead to better variable selection performance of the BAL. These strategies are among our future studies. Specifically, for the gamma prior in the BAL, we have tried the strategy to set the hyperparameters δ and τ as fixed numbers. In general, the results are not very sensitive to the choices of δ and τ, and no combination of δ and τ leads to significantly better results than assigning the joint prior for δ and τ; *i.e.*, $p(\delta, \tau) \propto \tau^{-1}$.

In the current implementation, we handle missing genotype data by imputing them first (using the Viterbi algorithm implemented in R/qtl) and then taking the imputed values as known. A more sophisticated approach for the BAL is to take the genotype data as unknown and sample them within MCMC; and for the IAL, we can summarize its results across multiple imputations (SEN and CHURCHILL 2001). However, these sophisticated approaches are computationally more intensive and are mainly designed for relatively sparse marker maps. The current high-density SNP arrays often have high-confidence genotype call rates >98% (RABBEE and SPEED 2006). Imputation methods are also an active research topic. Haplotype information from related or unrelated individuals can be used to obtain accurate genotype imputation (MARCHINI *et al.* 2007). Therefore simply imputing the genotype data and then taking it as known may be sufficient for many studies using high-density SNP arrays, although careful examination of missing data patterns is always important.

We have mainly discussed our method in a linear regression framework. Extension to the generalized linear model (*e.g.*, logistic regression for binary responses) is possible. The generalized linear model can be solved by iterated reweighted least squares. Similar to the approach used in FRIEDMAN *et al.* (2009), our method can be plugged in to solve the least-squares problem within the loop of iterated reweighted least squares. However, this approach is computationally intensive. YI and BANERJEE (2009) proposed an intriguing and efficient EM algorithm for multiple-loci mapping by generalized linear regression. In their EM algorithm, many correlated covariates can be simultaneously updated, which has the advantage of accommodating the correlation among the covariates. However, our method uses a different model setup and cannot adopt the same approach of YI and BANERJEE (2009). Computationally efficient extension of our method to a generalized linear model warrants further studies.

We tested the robustness of our methods by additional sets of simulations where the traits are log or

exponentially transformed (results not shown). The conclusion is that our methods are robust to mild violations of linearity assumption. However, we still expect that the additive linear model assumption is severely violated in some situations, *e.g.*, when epistatic interactions are present. As mentioned in the Introduction, we can include pairwise interactions in our model, similar to the approaches used by Zhang and Xu (2005) and Yi *et al.* (2007). In practice, prioritizing interactions by the significance of main effects or biological knowledge may help to reduce the multiple-testing burden and to improve the power. How to penalize the interaction term also warrants further study. Grouping the interaction terms and the corresponding main effects together and applying group penalties (Yuan and Lin 2006) may be a better approach than penalizing the main effects and interactions separately.

In summary, we have developed iterative adaptive penalized regression methods for genomewide multiple-loci mapping problems. Both theoretical justifications and empirical evidence suggest that our methods have superior performance than the existing methods. Although our work is motivated by genetic data, our methods are general enough to be applied to other HDLSS problems.

The authors thank two anonymous reviewers and the associate editor for their constructive comments and suggestions. The authors also thank Jun Liu for his valuable suggestions for the implementation of the Bayesian adaptive Lasso. This research was supported by National Institute of Mental Health (NIMH) grants 1 P50 MH090338-01 and 1 RC2 MH089951-01. J. G. Ibrahim's research was partially supported by National Institutes of Health (NIH) grants GM70335 and CA74015. F. Zou's research was partially supported by NIH grant GM074175-03.

LITERATURE CITED

Brem, R. B., and L. Kruglyak, 2005 The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc. Natl. Acad. Sci. USA **102**(5): 1572–1577.

Brem, R. B., J. D. Storey, J. Whittle and L. Kruglyak, 2005 Genetic interactions between polymorphisms that affect gene expression in yeast. Nature **436**(7051): 701–703.

Broman, K. W., and T. P. Speed, 2002 A model selection approach for the identification of quantitative trait loci in experimental crosses. J. R. Stat. Soc. Ser. B **64:** 641–656.

Broman, K. W., H. Wu, S. Sen and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889–890.

Chen, J., and Z. Chen, 2008 Extended Bayesian information criteria for model selection with large model spaces. Biometrika **95**(3): 759.

Churchill, G., and R. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

Doerge, R., and G. Churchill, 1996 Permutation tests for multiple loci affecting a quantitative character. Genetics **142:** 285–294.

Fan, J., and R. Li, 2001 Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96:** 1348–1360.

Friedman, J., 2008 Fastsparse regression and classification. Technical report, Stanford University, Stanford, CA.

Friedman, J., T. Hastie and R. Tibshirani, 2009 Regularization paths for generalized linear models via coordinate descent. Technical report, Department of Statistics, Stanford University, Stanford, CA.

George, E., and R. McCulloch, 1993 Variable selection via Gibbs sampling. J. Am. Stat. Assoc. **88:** 881–889.

Griffin, J., and P. Brown, 2007 Bayesian adaptive lassos with non-convex penalization. Technical Report, University of Kent, Kent, UK.

Hans, C., 2009 Bayesian lasso regression. Biometrika **96:** 1–11.

Hoggart, C., J. Whittaker, M. De Iorio and D. Balding, 2008 Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. PLoS Genet. **4**(7):e1000130.

Hoh, J., and J. Ott, 2003 Mathematical multi-locus approaches to localizing complex human trait genes. Nat. Rev. Genet. **4:** 701–709.

Huang, J., S. Ma and C.-H. Zhang, 2008 Adaptive Lasso for sparse high-dimensional regression models. Stat. Sin. **18:** 1603–1618.

Kou, S., Q. Zhou and W. Wong, 2006 Equi-energy sampler with applications in statistical inference and statistical mechanics. Ann. Stat. **34**(4): 1581–1619.

Manichaikul, A., J. Moon, S. Sen, B. Yandell and K. Broman, 2009 A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics **181:** 1077–1086.

Marchini, J., B. Howie, S. Myers, G. McVean and P. Donnelly, 2007 A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. **39:** 906–913.

Meng, X.-L., and D. B. Rubin, 1993 Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika **80**(2): 267–278.

Park, T., and G. Casella, 2008 The Bayesian Lasso. J. Am. Stat. Assoc. **103:** 681–686.

Rabbee, N., and T. Speed, 2006 A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics **22:** 7–12.

Radchenko, P., and G. M. James, 2008 Variable inclusion and shrinkage algorithms. J. Am. Stat. Assoc. **103:** 1304–1315.

Richardson, S., and P. J. Green, 1997 On Bayesian analysis of mixtures with an unknown number of components (with discussion). J. R. Stat. Soc. Ser. B **59:** 731–792.

Sen, S., and G. Churchill, 2001 A statistical framework for quantitative trait mapping. Genetics **159:** 371–387.

Sun, W., and F. A. Wright, 2010 A geometric interpretation of the permutation p-value and its application in eQTL studies. Ann. Appl. Stat. (in press). (http://www.bios.unc.edu/∼ wsun/research.htm).

Tibshirani, R., 1996 Regression shrinkage and selection via the Lasso. J. R. Stat. Soc Ser. B. **58:** 267–288.

Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie *et al.*, 2001 Missing value estimation methods for DNA microarrays. Bioinformatics **17**(6): 520–525.

Wasserman, L., and K. Roeder, 2009 High dimensional variable selection. Ann. Stat. **37**(5A): 2178.

Yi, N., 2004 A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. Genetics **167:** 967–975.

Yi, N., and S. Banerjee, 2009 Hierarchical generalized linear models for multiple quantitative trait locus mapping. Genetics **181**(3): 1101.

Yi, N., and S. Xu, 2008 Bayesian LASSO for quantitative trait loci mapping. Genetics **179:** 1045–1055.

Yi, N., D. Shriner, S. Banerjee, T. Mehta, D. Pomp *et al.*, 2007 An efficient Bayesian model selection approach for interacting quantitative trait loci models with many effects. Genetics **176:** 1865–1877.

Yuan, M., and Y. Lin, 2006 Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B **68**(1): 49–67.

Yvert, G., R. B. Brem, J. Whittle, J. M. Akey, E. Foss *et al.*, 2003 Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat. Genet. **35**(1): 57–64.

Zhang, W., J. Zhu, E. E. Schadt and J. S. Liu, 2010 A Bayesian partition method for detecting pleiotropic and epistatic eQTL modules. PLoS Comput. Biol. **6**(1): e1000642.

Zhang, Y., and S. Xu, 2005 A penalized maximum likelihood method for estimating epistatic effects of QTL. Heredity **95:** 96–104.

Zhu, J., B. Zhang, E. Smith, B. Drees, R. Brem *et al.*, 2008 Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat. Genet. **40**(7): 854.

Zou, H., 2006 The adaptive Lasso and its oracle properties. J. Am. Stat. Assoc. **101:** 1418–1429.

Zou, H., and R. Li, 2008 One-step sparse estimates in nonconcave penalized likelihood models. Ann. Stat. **36:** 1509–1533.

Zou, H., T. Hastie and R. Tibshirani, 2007 On the "degrees of freedom" of the lasso. Ann. Stat. **35**(5): 2173–2192.

Communicating editor: H. Zhao

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.109.114280/DC1

Genomewide Multiple-Loci Mapping in Experimental Crosses by Iterative Adaptive Penalized Regression

Wei Sun, Joseph G. Ibrahim and Fei Zou

DOI: 10.1534/genetics.109.114280

File S1

Supporting Information

A: The Gibbs Sampler for the Bayesian adaptive Lasso

1. Initialization. We initialize $(b_0, b_1, ..., b_p)$ with zero and initialize $(\sigma_e^2, \kappa_1, ..., \kappa_p)$ with a positive number.

2. Update b_0: Let $\Theta_{-\text{parameter}_1}$ denote all of the parameters except parameter_1. The conditional posterior distribution of b_0 is

$$p(b_0|\Theta_{-b_0}) \sim N(\bar{b}_0, s_0^2), \tag{1}$$

where $\bar{b}_0 = (1/n)\sum_1^n (y_i - \sum_{j=1}^p x_{ij} b_j)$, and $s_0^2 = (1/n)\sigma_e^2$.

3. Update b_j: The conditional posterior distribution of b_j is

$$p(b_j|\Theta_{-b_j}) \quad \propto \quad \exp\left[-\frac{(b_j - \bar{b}_j)^2}{2\sigma_j^2} - \frac{|b_j|}{\kappa_j}\right], \tag{2}$$

where

$$\sigma_j^2 = \frac{\sigma_e^2}{\sum_{i=1}^n x_{ij}^2}, \text{ and } \bar{b}_j = \left(\sum_{i=1}^n x_{ij}^2\right)^{-1} \sum_{i=1}^n x_{ij}\left(y_i - b_0 - \sum_{k\neq j} x_{ik} b_k\right). \tag{3}$$

Although $p(b_j|\Theta_{-b_j})$ has no closed form, it is log-concave. Thus we sample b_j by Adaptive Rejection Sampling (ARS) [Gilks, 1992]. Note that whenever one of the b_j's is updated, it is used immediately for updating the other b_j's.

4. Update σ_e^2:

$$p(\sigma_e^2|\Theta_{-\sigma_e^2}) \propto (\sigma_e^2)^{-1-n/2} \exp\left[-\texttt{rss}/(2\sigma_e^2)\right], \tag{4}$$

which is inv-Gamma$(\sigma_e^2; n/2, \texttt{rss}/2)$.

5. Update κ_j:

$$p(\kappa_j|\Theta_{-\kappa_j}) \propto \kappa_j^{-2-\delta} \exp\left(-\frac{|b_j|+\tau}{\kappa_j}\right), \tag{5}$$

which is inv-Gamma$(\kappa_j; 1+\delta, |b_j|+\tau)$.

6. Update τ:

$$p(\tau|\Theta_{-\tau}) \propto \tau^{p\delta-1} \exp\left(-\tau \sum_{j=1}^{p} \kappa_j^{-1}\right). \tag{6}$$

Therefore $\tau|\Theta_{-\tau} \sim \text{Gamma}(p\delta, \sum_{j=1}^{p} \kappa_j^{-1})$.

7. Update δ:

$$p(\delta|\Theta_{-\delta}) \propto \left(\frac{\tau^p}{\prod_{j=1}^{p} \kappa_j}\right)^{\delta} \Gamma(\delta)^{-p}. \tag{7}$$

It is easy to show that $p(\delta|\Theta_{-\delta})$ is log-concave, so we sample δ using ARS.

B: Derivations of Iterative Adaptive Lasso

CM step of the iterative adaptive Lasso

As shown in Section A, the conditional posterior distribution of b_j is

$$p(b_j|\Theta_{-b_j}) \quad \propto \quad \exp\left[-\frac{(b_j-\bar{b}_j)^2}{2\sigma_j^2} - \frac{|b_j|}{\kappa_j}\right], \tag{8}$$

where

$$\sigma_j^2 = \frac{\sigma_e^2}{\sum_{i=1}^{n} x_{ij}^2}, \text{ and } \bar{b}_j = \left(\sum_{i=1}^{n} x_{ij}^2\right)^{-1} \sum_{i=1}^{n} x_{ij} \left(y_i - b_0 - \sum_{k\neq j}^{p} x_{ik} b_k\right). \tag{9}$$

Let ζ_j be the mode of the conditional posterior distribution of b_j, then $\zeta_j = \arg\min_{b_j} f(b_j)$, where $f(b_j) = (b_j - \bar{b}_j)^2/(2\sigma_j^2) + |b_j|/\kappa_j$. Therefore ζ_j can be solved by letting the

derivative of $f(b_j)$ to be 0. One extra complexity arises because the derivative of $|b_j|/\kappa_j$ does not exist at $b_j = 0$. This problem can be circumvented by considering three situations: $\bar{b}_j > 0$, $\bar{b}_j = 0$, and $\bar{b}_j < 0$. It is easy to show that

$$\begin{cases} \zeta_j = 0 & \text{if } \bar{b}_j = 0 \\ \zeta_j \in [0, \bar{b}_j] & \text{if } \bar{b}_j > 0 \\ \zeta_j \in [\bar{b}_j, 0] & \text{if } \bar{b}_j < 0 \end{cases} .$$

- If $\bar{b}_j > 0$, we can first consider the derivatives of $f(b_j)$ in $(0, \bar{b}_j]$: $f'(b_j) = b_j/\sigma_j^2 - \bar{b}_j/\sigma_j^2 + 1/\kappa_j$, and $f''(b_j) = 1/\sigma_j^2$. If $\bar{b}_j \leq \sigma_j^2/\kappa_j$, $f'(b_j) > 0$, therefore $f(b_j)$ may achieve its minimum at 0. This is can be proved as follows.

  - If $0 < \bar{b}_j \leq \sigma_j^2/\kappa_j$, $\forall\ b_j \in [0, \bar{b}_j]$

    $$\begin{aligned} f(b_j) - f(0) &= (b_j - \bar{b}_j)^2/(2\sigma_j^2) + |b_j|/\kappa_j - \bar{b}_j^2/(2\sigma_j^2) = b_j^2/(2\sigma_j^2) - b_j\bar{b}_j/\sigma_j^2 + b_j/\kappa_j \\ &= b_j^2/(2\sigma_j^2) + (b_j/\sigma_j^2)(\sigma_j^2/\kappa_j - \bar{b}_j) \geq 0, \end{aligned}$$

    therefore $\zeta_j = 0$.

  - If $\bar{b}_j > \sigma_j^2/\kappa_j$, when $b_j = \bar{b}_j - \sigma_j^2/\kappa_j$ $f'(b_j) = 0$, and $f''(b_j) > 0$. Therefore $\zeta_j = \bar{b}_j - \sigma_j^2/\kappa_j$

- If $\bar{b}_j < 0$, similar to the situation of $\bar{b}_j > 0$, we can first consider the derivatives of $f(b_j)$ in $[\bar{b}_j, 0)$: $f'(b_j) = b_j/\sigma_j^2 - \bar{b}_j/\sigma_j^2 - 1/\kappa_j$, and $f''(b_j) = 1/\sigma_j^2$. If $\bar{b}_j \geq -\sigma_j^2/\kappa_j$, $f'(b_j) < 0$, therefore $f(b_j)$ may achieve its minimum at 0. This is can be proved as follows:

  - If $-\sigma_j^2/\kappa_j \leq \bar{b}_j < 0$, $\forall\ b_j \in [\bar{b}_j, 0]$

    $$\begin{aligned} f(b_j) - f(0) &= (b_j - \bar{b}_j)^2/(2\sigma_j^2) + |b_j|/\kappa_j - \bar{b}_j^2/(2\sigma_j^2) = b_j^2/(2\sigma_j^2) - b_j\bar{b}_j/\sigma_j^2 - b_j/\kappa_j \\ &= b_j^2/(2\sigma_j^2) - (b_j/\sigma_j^2)(\bar{b}_j + \sigma_j^2/\kappa_j) \geq 0, \end{aligned}$$

    therefore $\zeta_j = 0$.

  - If $\bar{b}_j < -\sigma_j^2/\kappa_j$, when $b_j = \bar{b}_j + \sigma_j^2/\kappa_j$ $f'(b_j) = 0$, and $f''(b_j) > 0$. Therefore $\zeta_j = \bar{b}_j + \sigma_j^2/\kappa_j$

Summarizing the above results, we have

$$\begin{cases} \zeta_j = 0 & \text{if } -\sigma_j^2/\kappa_j \leq \bar{b}_j \leq \sigma_j^2/\kappa_j \\ \zeta_j = \bar{b}_j - \sigma_j^2/\kappa_j & \text{if } \bar{b}_j > \sigma_j^2/\kappa_j \\ \zeta_j = \bar{b}_j + \sigma_j^2/\kappa_j & \text{if } \bar{b}_j < -\sigma_j^2/\kappa_j \end{cases}.$$

Note that this CM-step update is actually quite similar to the shooting method for the Lasso calculation [Fu, 1998].

E-step of the iterative adaptive Lasso

The complete data log-posterior is

$$l(\theta|\mathbf{y}, \mathbf{X}, \phi) = C - \texttt{rss}/(2\sigma_e^2) - \sum_{j=1}^{p} \frac{|b_j|}{\kappa_j}, \tag{10}$$

where C is a constant with respect to θ. Suppose in the t-th iteration, the parameter estimates are $\theta^{(t)} = (b_0^{(t)}, b_1^{(t)}, ..., b_p^{(t)})$. Then the conditional expectation with respect to the conditional density of $f(\phi|\mathbf{y}, \mathbf{X}, \theta^{(t)})$ is

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \int l(\theta|\mathbf{y}, \mathbf{X}, \phi) f(\phi|\mathbf{y}, \mathbf{X}, \theta^{(t)}) d\phi \\ &= \int \left(C - \texttt{rss}/(2\sigma_e^2) - \sum_{j=1}^{p} \frac{|b_j|}{\kappa_j} \right) f(\phi|\mathbf{y}, \mathbf{X}, \theta^{(t)}) d\phi. \end{aligned} \tag{11}$$

From the derivation of the Bayesian adaptive Lasso, we have

$$f(\phi|\mathbf{y}, \mathbf{X}, \theta^{(t)}) = \frac{C_f}{\sigma_e^{2+n}} \exp\left(-\texttt{rss}^{(t)}/(2\sigma_e^2)\right) \prod_{j=1}^{p} \kappa_j^{-2-\delta} \exp\left(-\frac{|b_j^{(t)}| + \tau}{\kappa_j} \right), \tag{12}$$

where $\texttt{rss}^{(t)}$ is the residual sum of squares calculated using $\theta^{(t)}$, C_f is the normalizing constant. By letting $\int f(\phi|\mathbf{y}, \mathbf{X}, \theta^{(t)}) d\phi = 1$, it is easy to show that

$$C_f = \left[\frac{1}{\Gamma(n/2)} \left(\texttt{rss}^{(t)}/2 \right)^{n/2} \right] \left[\prod_{j=1}^{p} \frac{\left(|b_j^{(t)}| + \tau \right)^{1+\delta}}{\Gamma(1+\delta)} \right]. \tag{13}$$

Therefore

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= C - (\texttt{rss}/2)\left(\texttt{rss}^{(t)}/2\right)^{n/2}\frac{1}{\Gamma(n/2)}\int \sigma_e^{-4-n}\exp\left(-\texttt{rss}^{(t)}/(2\sigma_e^2)\right)d\sigma_e^2 \\ &\quad -\sum_{j=1}^{p}\frac{|b_j|\left(|b_j^{(t)}|+\tau\right)^{1+\delta}}{\Gamma(1+\delta)}\int \kappa_j^{-3-\delta}\exp\left(-\frac{|b_j^{(t)}|+\tau}{\kappa_j}\right)d\kappa_j \\ &= C - (\texttt{rss}/2)/\left(\texttt{rss}^{(t)}/n\right) - \sum_{j=1}^{p}\frac{|b_j|}{\left(|b_j^{(t)}|+\tau\right)/(1+\delta)}. \end{aligned} \tag{14}$$

C: Choosing tuning parameters, an asymptotic point of view.

Theorem 1. Consider the multiple linear regression problem formulated with n samples. Assume the penalization parameters of the IAL satisfy $(1+\delta)/\tau = O(n^{1/2+d})$, where $0 < d < 1/2$. Denote the coefficient estimates in the t-th iteration as $\hat{\mathbf{b}}^{(t)}$. Let $\mathbf{X}_{-j}$ be $\mathbf{X}$ without the j-th column and let $\tilde{\mathbf{b}}_{-j}^{(t+1)}$ be the coefficient estimates (except b_j) before estimating $\hat{b}_j^{(t+1)}$.

(i) If $\hat{b}_j^{(t)} = 0$ and $\mathbf{x}_j \perp \mathbf{y}|\mathbf{X}_{-j}\tilde{\mathbf{b}}_{-j}^{(t+1)}$, then $p(\hat{b}_j^{(t+1)} = 0) \to 1$.

(ii) If $\exists\ c > 0$, s.t. $|\mathrm{corr}(\mathbf{x}_j, \mathbf{y}|\tilde{\mathbf{b}}_{-j}^{(t+1)})| > c$, then $p(\hat{b}_j^{(t+1)} \neq 0) \to 1$.

We first prove conclusion (i) of Theorem 1. Whether $\hat{b}_j^{(t+1)}$ is penalized to zero amounts to whether

$$|\bar{b}_j^{(t+1)}| \leq \frac{\sigma_e^2(1+\delta)}{n s_j^2(\tau + |\hat{b}_j^{(t)}|)}, \tag{15}$$

where $s_j^2 = \mathbf{x}_j^T\mathbf{x}_j/n$. Based on the assumption that $\mathbf{x}_j \perp \mathbf{y}|\mathbf{X}_{-j}\tilde{\mathbf{b}}_{-j}^{(t+1)}$, we have $\sqrt{n}\bar{b}_j^{(t+1)} = O(1)$. Given that $\hat{b}_j^{(t)} = 0$, $(1+\delta)/\tau = O(n^{1/2+d})$, equation (15) can be written as

$$\sqrt{n}|\bar{b}_j^{(t+1)}| \leq \frac{\sigma_e^2(1+\delta)}{\sqrt{n}s_j^2\tau} \Leftrightarrow O(1) \leq O(n^d). \tag{16}$$

thus $p(\hat{b}_j^{(t+1)} = 0) \to 1$.

Since $\exists\ c > 0$, s.t. $|\text{corr}(\mathbf{x}_j, \mathbf{y}|\tilde{\mathbf{b}}_{-j}^{(t+1)})| > c$, then $\bar{b}_j = O(1)$. Based on the assumption that $(1+\delta)/\tau = O(n^{1/2+d})$, where $0 < d < 1/2$, we have $(1+\delta)/(n\tau) = O(n^{-1/2+d}) = o(1)$, therefore asymptotically

$$|\bar{b}_j| > \frac{\sigma_e^2(1+\delta)}{n s_j^2 \tau}, \tag{17}$$

which means b_j will be selected into the model even if it is 0 in the previous step.

An intuitive explanation of Theorem 1 is as follows. First, we need to penalize the coefficients big enough so that if $\hat{b}_j = 0$ in the previous iteration, it remains 0 if $\mathbf{x}_j$ is uncorrelated with $\mathbf{y}$ given all the other coefficients estimates. This requires $(1+\delta)/\tau = O(n^{1/2+d})$ and $d > 0$. On the other hand, the penalization should be small enough so that we can select those $\mathbf{x}_j$ that are not independent with $\mathbf{y}$, given all the other covariates. This requires $(1+\delta)/\tau = O(n^{1/2+d})$ and $d < 1/2$. Combining these two conditions, we need $(1+\delta)/\tau = O(n^{1/2+d})$, where $0 < d < 1/2$.

D: The number of independent tests

We estimate the number of independent tests by examining the relation between nominal p-values and permutation p-values. Let p_p and p_n be permutation p-value and nominal p-value, respectively. We found that the relation between $\log_{10}(p_p)$ and $\log_{10}(p_n)$ can be fitted a linear regression [Sun and Wright, 2009]:

$$\log_{10}(p_p) = a + b\log_{10}(p_n) \tag{18}$$

The number of independent tests at nominal p-value p_n, can be estimated by p_p/p_n. Therefore, based on the above linear model:

$$p_p = 10^a p_n^b \Rightarrow p_p/p_n = 10^a p_n^{b-1} \tag{19}$$

We have found from both simulated data and real data that b is often smaller than 1, thus the number of independent tests increase as p_n decreases. We provided an geometric interpretation of this observation in [Sun and Wright, 2009]. Suppose association to K markers need to be tested. Any p-value smaller than $0.05/K$ passes

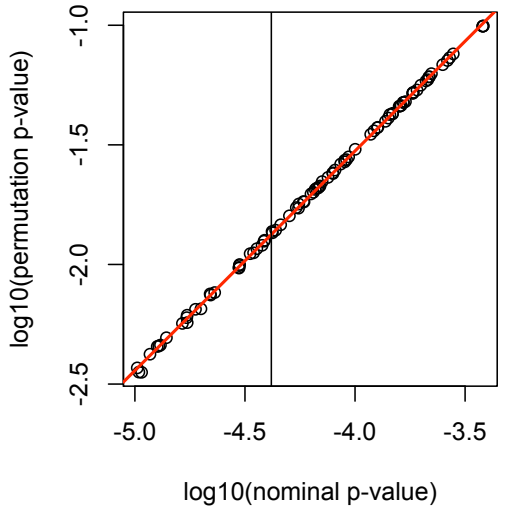
the Bonferroni correction, hence is significant. Therefore we only need to consider the p-values that are larger than $0.05/K$, and then the maximum number of independent tests we need to correct is $10^a(0.05/K)^{b-1}$.

In order to estimate a and b for our simulated data, we randomly choose 10 simulated traits, calculate their minimum nominal p-values and the corresponding permutation p-values in each chromosome and in each of the six simulation situations. Those pairs of nominal p-value and permutation p-value are then merged together. We use those pairs with nominal p-value larger than 10^{-5} and permutation p-value smaller than 0.1 to fit a linear model:

$$\log_{10}(p_p) = 2.1459 + 0.9179 \log_{10}(p_n) \tag{20}$$

As shown in the following figure, this linear relation (red line) capture the relation between $\log_{10}(p_p)$ and $\log_{10}(p_n)$ very well. At the nominal p-value cutoff 0.05/1200 (the vertical line in the following figure indicates $\log_{10}(0.05/1200)$), the number of independent tests is

$$10^{2.1459}(0.05/1200)^{0.9179-1} \approx 320.$$



For the yeast eQTL study, we have estimated in our previous work [Sun and Wright, 2009]

that

$$\log_{10}(p_p) = 2.52 + 0.978 \log_{10}(p_n) \tag{21}$$

Hence at the nominal p-value cutoff 0.05/1027, the number of independent tests is

$$10^{2.52}(0.05/1027)^{0.978-1} \approx 412.$$

E: The implementations of different methods

We have implemented marginal regression, forward regression, the adaptive Lasso (with marginal regression coefficients as initials), the Bayesian t, Bayesian Lasso, BAL, and IAL in an R package BPrimm (Bayesian and Penalized regression in multiple loci mapping). The computationally intensive parts are written by C. Our implementations of the Bayesian t and Bayesian Lasso are mainly based on [Yi and Xu, 2008], but with small modifications for the Bayesian t to further improve its computational efficiency. We leave the details of the Gibbs samplers for both methods in the Supplementary Materials Section F and G. The R package BPrimm can be downloaded from http://www.bios.unc.edu/∼wsun/software/.

We calculate permutation p-values (by 10,000 permutations) for both marginal and step-wise regression, and use permutation p-value 0.05 as cutoff. For marginal regression, we only keep the most significantly linked marker in each chromosome to eliminate redundant loci, a strategy that has been used elsewhere [Wang et al., 2006]. For forward regression, we use permutation-based residual empirical threshold (RET) to select variables [Doerge and Churchill, 1996]. We employ the function `stepqtl` in R/qtl [Broman et al., 2003] for the forward-backward regression with penalized LOD score as the model selection criterion. The function `stepqtl` allows user to add two loci into the model each time. We use this option for simulation situations 3-6 where two QTL are simulated from the same chromosome.

There are two options for the priors of the Bayesian Lasso: $p(b_j|\sigma_j^2) \sim N(0, \sigma_j^2)$ [Yi and Xu, 2008], and $p(b_j|\sigma_j^2) \sim N(0, \sigma_e^2\sigma_j^2)$ [Park and Casella, 2008], where σ_e^2 is the variance of the residual errors. The results we shall discuss are based on the former, while the latter yields similar results (data not shown). The Bayesian Lasso uses two hyperparameters r and s to specify the prior of $\kappa^2/2$ as Gamma(s, r).

Following [Yi and Xu, 2008], we set both r and s as small numbers such as $r = 0.01$ and $s = 0.01$. Values smaller than 0.1 yield similar results in terms of the number of true/false discoveries.

We use the implementation of CMSA in R/qtlbim [Yandell et al., 2007]. We choose not to carry out interval mapping because the genetic markers are already dense enough. The CMSA method requires an additional input, namely the expected number of QTL. We supply this parameter with the true number of simulated QTL. For each marker, we record its posterior probability belonging to the true model from the output of the CMSA.

Both extended BIC and ordinary BIC followed by variable filtering are implemented for tuning parameter selection of the IAL and the AL (with marginal regression coefficients as initials). The extended BIC is the ordinary BIC plus $2\gamma \log \tau(\mathcal{S}_j)$, where $\mathcal{S}_j$ indicates a model of size j and $\tau(\mathcal{S}_j) = \binom{p}{j}$ is the total number of models with j covariates. Following [Chen and Chen, 2008], we set $\gamma = 1 - 1/(2\kappa)$, where κ is solved from $p = n^\kappa$. In the BIC plus variable filtering approach, we use $0.05/p_E$ as p-value cutoff, where p_E is the effective number of independent tests. A conservative estimate of p_E is 320. See Supplementary Materials Section D and [Sun and Wright, 2009] for more details. In the implementation of the adaptive Lasso, given the weights estimated from marginal regression, the Lasso problem is solved by R function glmnet [Friedman et al., 2009]. A combinations of L_1 and L_2 penalty are allowed in R function glmnet, i.e., the elastic net penalty [Zou and Hastie, 2005]: $\sum_{j=1}^{p} \left[(1-\alpha)\beta_j^2/2 + \alpha|\beta_j|\right]$. The high correlations among the covariates may cause degeneracies for Lasso calculation. Following [Friedman et al., 2009], we choose to set $\alpha = 0.95$ to obtain a solution much like the Lasso, but removes the degeneracy problem.

The HyperLasso is downloaded from http://www.ebi.ac.uk/projects/BARGEN/. The "-linear" option is used to fit linear model. The "-iter" option is set as 50 to choose highest posterior mode among 50 runs of the HyperLasso. The "lambda" parameter is set as $\sqrt{n}\Phi^{-1}(1 - 0.05/p_E/2)$, where p_E is the effective number of independent tests, which is set as 320. The "shape" parameter is set as 1.

All the Bayesian methods use 10,000 burn-in iterations followed by 10,000 iterations to obtain 1,000 samples, one from every 10 iterations. To monitor the

convergence, we calculate the Gelman and Rubin scale reduction parameter (for 5 parallel chains) and the Geweke's statistic for each of the 1,200 coefficients (Supplementary Figure 4-6). For all the three Bayesian methods, the vast majority of the Gelman and Rubin statistics are smaller than 1.05, and the Geweke's statistic are approximately normally distributed. The auto correlation of the markers at the simulated QTL (or the marker that has the highest correlation with a QTL if the QTL genotype is not observed) is smaller than 0.15 for five chains. The default options in R/coda are used to calculate these diagnostic statistics. We note that "no diagnostic can 'prove' convergence of a MCMC" [Carlin and Louis, 2000]. However, these diagnostic statistics do suggest convergence of all the three Bayesian shrinkage methods.

F: Gibbs sampler for the Bayesian t

The following Gibbs sampler of the Bayesian t is mainly based on [Yi and Xu, 2008], with small modification to further improve its computational efficiency. The priors are specified as:

$$\begin{aligned} p(b_0) &\propto 1, && (22)\\ p(\sigma_e^2) &\propto 1/\sigma_e^2, && (23)\\ p(b_j|\sigma_j^2) &= N(0,\sigma_j^2), && (24)\\ p(\sigma_j^2) &= \text{inv-Gamma}(\delta,\tau) = \frac{\tau^\delta}{\Gamma(\delta)}(\sigma_j^2)^{-1-\delta}\exp(-\tau/\sigma_j^2), && (25) \end{aligned}$$

where $j = 1, ..., p$, indicating p covariates (markers). The posterior distribution of all the parameters is given by

$$\begin{aligned} & p(\mathbf{b}, b_0, \sigma_e^2, \sigma_1^2, ..., \sigma_p^2|\mathbf{y}, \mathbf{X})\\ \propto\ & p(\mathbf{y}|\mathbf{b}, \mathbf{X}, b_0, \sigma_e^2)P(\sigma_e^2)\prod_{j=1}^{p} p(b_j|\sigma_j^2)p(\sigma_j^2)\\ \propto\ & \left(\frac{\tau^\delta}{\Gamma(\delta)}\right)^p \frac{1}{\sigma_e^{2+n}}\exp\left[-\mathtt{rss}/(2\sigma_e^2)\right]\prod_{j=1}^{p}\left[(\sigma_j^2)^{-3/2-\delta}\exp\left(-\frac{b_j^2+2\tau}{2\sigma_j^2}\right)\right]. \quad (26) \end{aligned}$$

It is easy to show that the full conditional posterior distributions of (b_0, b_j, σ_e^2, σ_j^2) are

$$b_0|\Theta_{-b_0} \sim N\left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}b_j\right), \frac{1}{n}\sigma_e^2\right), \quad (27)$$

$$b_j|\Theta_{-b_j} \sim N\left(\frac{\sum_{i=1}^{n} x_{ij}\left(y_i - b_0 - \sum_{k\neq j}^{p} x_{ik}b_k\right)}{\sum_{i=1}^{n} x_{ij}^2 + \sigma_e^2/\sigma_j^2}, \frac{\sigma_e^2}{\sum_{i=1}^{n} x_{ij}^2 + \sigma_e^2/\sigma_j^2}\right), \quad (28)$$

$$\sigma_e^2|\Theta_{-\sigma_e^2} \sim \text{inv-Gamma}\left(n/2, \texttt{rss}/2\right), \quad (29)$$

$$\sigma_j^2|\Theta_{-\sigma_j^2} \sim \text{inv-Gamma}\left(1/2 + \delta, b_j^2/2 + \tau\right) \quad (30)$$

Assuming $\pi(\delta, \tau) \propto \tau^{-1}$, the posterior distribution for τ is thus given by

$$p(\tau|\Theta_{-\tau}) \propto \tau^{p\delta-1}\prod_{j=1}^{p}\exp\left(-\frac{\tau}{\sigma_j^2}\right) \propto \tau^{p\delta-1}\exp\left(-\tau\sum_{j=1}^{p}\sigma_j^{-2}\right). \quad (31)$$

Therefore $p(\tau|\Theta_{-\tau}) \sim \text{Gamma}\Big(p\delta, \sum_{j=1}^{p}\sigma_j^{-2}\Big)$. The posterior distribution for δ is

$$p(\delta|\Theta_{-\delta}) \equiv f(\delta) \propto \left(\frac{\tau^\delta}{\Gamma(\delta)}\right)^p \prod_{j=1}^{p}(\sigma_j^2)^{-\delta} \propto \left(\frac{\tau^p}{\prod_{j=1}^{p}\sigma_j^2}\right)^\delta \Gamma(\delta)^{-p}. \quad (32)$$

There is no closed form for this density, however, it is easy to show that this density is a log-concave function, and thus we sample δ using the Adaptive Rejection Sampling algorithm [Gilks, 1992] within the Gibbs sampler.

G: Gibbs sampler for the Bayesian Lasso

The following Gibbs sampler of the Bayesian Lasso is based on Park and Casella [Park and Casella, 2008] and Yi and Xu [Yi and Xu, 2008]. The priors are specified

as:

$$
\begin{aligned}
p(b_0) &\propto 1, && (33)\\
p(\sigma_e^2) &\propto 1/\sigma_e^2, && (34)\\
p(b_j|\sigma_j^2) &\sim N(0,\sigma_j^2), && (35)\\
p(\sigma_j^2|\kappa^2/2) &\sim \text{Exp}(\kappa^2/2) = \frac{\kappa^2}{2}\exp\left(-\frac{\kappa^2}{2}\sigma_j^2\right), && (36)\\
p(\kappa^2/2) &= \text{Gamma}(s,r) = \frac{r^s}{\Gamma(s)}\left(\kappa^2/2\right)^{s-1}\exp\left(-r\frac{\kappa^2}{2}\right), && (37)
\end{aligned}
$$

where $j = 1, ..., p$, indicating p covariates (markers). Note we model the distribution of $\kappa^2/2$ instead of κ, and in the prior of $\kappa^2/2$, s and r are the shape and rate parameters, respectively.

The posterior distribution of all the parameters is

$$
\begin{aligned}
& p(\mathbf{b}, b_0, \sigma_e^2, \sigma_1^2, ..., \sigma_p^2, \kappa^2/2|\mathbf{y}, \mathbf{X})\\
\propto\ & p(\mathbf{y}|\mathbf{b}, \mathbf{X}, b_0, \sigma_e^2)p(\sigma_e^2)\prod_{j=1}^{p}\left[p(b_j|\sigma_j^2)p(\sigma_j^2|\kappa^2/2)\right]p(\kappa^2/2)\\
\propto\ & \frac{1}{\sigma_e^{2+n}}\exp\left[-\texttt{rss}/(2\sigma_e^2)\right]\prod_{j=1}^{p}\left[\sigma_j^{-1}\exp\left(-\frac{b_j^2}{2\sigma_j^2}-\frac{\kappa^2\sigma_j^2}{2}\right)\right]\\
& \frac{r^s}{\Gamma(s)}\left(\kappa^2/2\right)^{p+s-1}\exp\left(-r\frac{\kappa^2}{2}\right). && (38)
\end{aligned}
$$

Then the full conditional posterior distributions of (b_0, b_j, σ_e^2, σ_j^2, $\kappa^2/2$) are

$$
\begin{aligned}
b_0|\Theta_{-b_0} &\sim N\left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}x_{ij}b_j\right), \frac{1}{n}\sigma_e^2\right), && (39)\\
b_j|\Theta_{-b_j} &\sim N\left(\frac{\sum_{i=1}^{n}x_{ij}\left(y_i - b_0 - \sum_{k\neq j}^{p}x_{ik}b_k\right)}{\sum_{i=1}^{n}x_{ij}^2 + \sigma_e^2/\sigma_j^2}, \frac{\sigma_e^2}{\sum_{i=1}^{n}x_{ij}^2 + \sigma_e^2/\sigma_j^2}\right), && (40)\\
\sigma_e^2|\Theta_{-\sigma_e^2} &\sim \text{inv-Gamma}\left(n/2, \texttt{rss}/2\right), && (41)\\
\sigma_j^{-2}|\Theta_{-\sigma_j^2} &\sim \text{inv-Gauss}\left(\frac{\kappa}{|b_j|}, \kappa^2\right), && (42)\\
\kappa^2/2|\Theta_{-\kappa^2/2} &\sim \text{Gamma}\left(p+s, \sum_{j=1}^{p}\sigma_j^2 + r\right), && (43)
\end{aligned}
$$

where the density of an inverse Gaussian (inv-Gauss) distribution [Chhikara and Folks, 1989] is given by

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi}} x^{-3/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right). \tag{44}$$

An alternative setup is to assign the priors of b_j as

$$p(b_j|\sigma_j^2) \sim N(0, \sigma_e^2\sigma_j^2), \tag{45}$$

and leave all the other priors the same [Park and Casella, 2008]. This setup has the advantage that the joint posterior of $(\mathbf{b}, \sigma_e^2)$ has at most one mode. The conditional posterior distributions remain the same except that

$$b_j|\Theta_{-b_j} \sim N\left(\frac{\sum_{i=1}^n x_{ij}\left(y_i - b_0 - \sum_{k\neq j}^p x_{ik}b_k\right)}{\sum_{i=1}^n x_{ij}^2 + 1/\sigma_j^2}, \frac{\sigma_e^2}{\sum_{i=1}^n x_{ij}^2 + 1/\sigma_j^2}\right), \tag{46}$$

$$\sigma_e^2|\Theta_{-\sigma_e^2} \sim \text{inv-Gamma}\left((n+p)/2, \texttt{rss}/2 + \sum_{j=1}^p b_j^2/(2\sigma_j^2)\right), \tag{47}$$

$$\sigma_j^{-2}|\Theta_{-\sigma_j^2} \sim \text{inv-Gauss}\left(\frac{\kappa\sigma_e}{|b_j|}, \kappa^2\right). \tag{48}$$

References

[Broman et al., 2003] Broman, K., Wu, H., Sen, S., and Churchill, G., 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**:889–890.

[Carlin and Louis, 2000] Carlin, B. P. and Louis, T. A., 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC.

[Chen and Chen, 2008] Chen, J. and Chen, Z., 2008. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, **95**(3):759.

[Chhikara and Folks, 1989] Chhikara, R. S. and Folks, J. L., 1989. *The inverse gaussian distribution: theory, methodology, and applications*. Marcel Dekker, Inc., New York, NY, USA.

[Doerge and Churchill, 1996] Doerge, R. and Churchill, G., 1996. Permutation tests for multiple loci affecting a quantitative character. *Genetics*, **142**(285-294).

[Friedman et al., 2009] Friedman, J., Hastie, T., and Tibshirani, R., 2009. Regularization Paths for Generalized Linear Models via Coordinate Descent. Technical report, Department of Statistics, Stanford University.

[Fu, 1998] Fu, W. J., 1998. Penalized regressions: The bridge versus the Lasso. *Journal of Computational and Graphical Statistics*, **7**(3):397–416.

[Gilks, 1992] Gilks, W. R., 1992. Derivative-free adaptive rejection sampling for gibbs sampling. In Bernardo, J., Berger, J., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, pages 641–649. Oxford University Press.

[Liu, 1998] Liu, B., 1998. *Statistical genomics: linkage, mapping, and QTL analysis*. CRC Press.

[Park and Casella, 2008] Park, T. and Casella, G., 2008. The Bayesian Lasso. *Journal of the American Statistical Association*, **103**:681–686.

[Sun and Wright, 2009] Sun, W. and Wright, F. A., 2009. A geometric interpretation of the permutation p-value and its application in eQTL studies. *Annals of Applied Statistics*, **in press**:http://www.bios.unc.edu/∼wsun/research.htm.

[Wang et al., 2006] Wang, S., Yehya, N., Schadt, E. E., Wang, H., Drake, T. A., and Lusis, A. J., 2006. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet*, **2**(2):e15.

[Yandell et al., 2007] Yandell, B., Mehta, T., Banerjee, S., Shriner, D., Venkataraman, R., Moon, J., Neely, W., Wu, H., von Smith, R., and Yi, N., *et al.*, 2007. R/qtlbim: QTL with Bayesian Interval Mapping in experimental crosses. *Bioinformatics*, **23**:641–643.

[Yi and Xu, 2008] Yi, N. and Xu, S., 2008. Bayesian LASSO for Quantitative Trait Loci Mapping. *Genetics*, **179**:1045–1055.

[Zou and Hastie, 2005] Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, **67**(2):301–320.

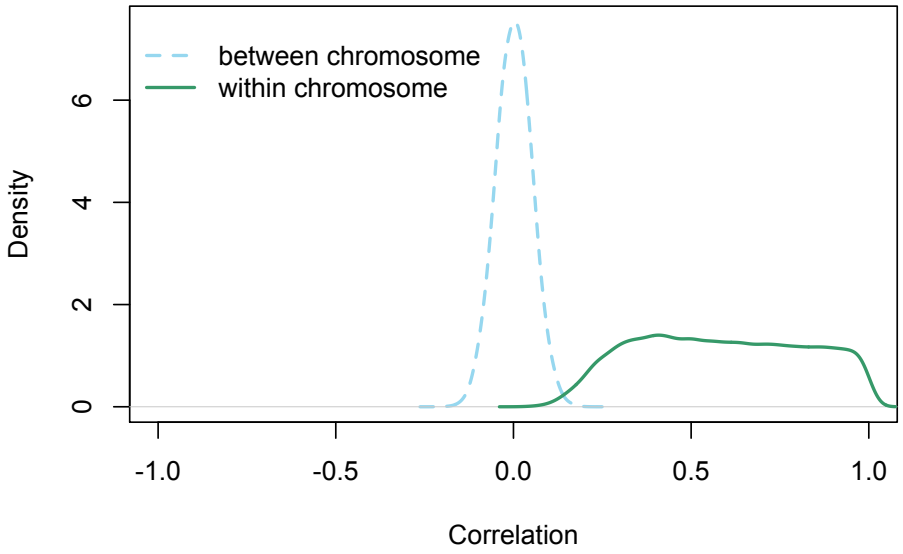


FIGURE S1.—The distribution of the correlations between genotype profiles from the same chromosome and the genotype profiles from different chromosomes.

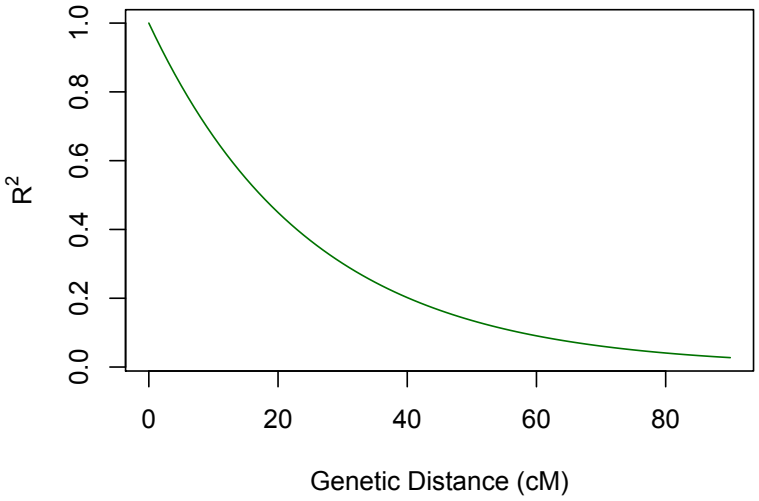


FIGURE S2.—The relation between genetic distance (centimorgan) and R^2 in F2 population. Con- sider two SNPs in diploid genome. Suppose the two alleles of the first SNP are A and a, and the two SNPs for the second SNP are B and b. From aabb to AABB, there are nine possible genotype combinations in F2 offsprings. Given genetic distance, we calculated recombination fraction by Haldane's mapping function, then the expected frequencies of different genotype combinations can be calculated (Table 6.8, page 172 of [Liu, 1998]). These expected frequencies provide the joint distribution of the two SNP's genotype, hence we can calculate their R^2.

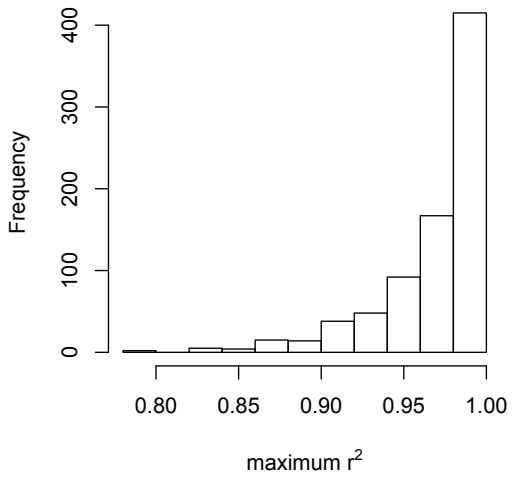


FIGURE S3.—For each of the 800 markers with "unobserved" genotype in our simulation, we calculate its R^2 with all the 1200 markers with observed genotype, and then take the maximum. This figure illustrates the distribution of 800 such maximum R^2.

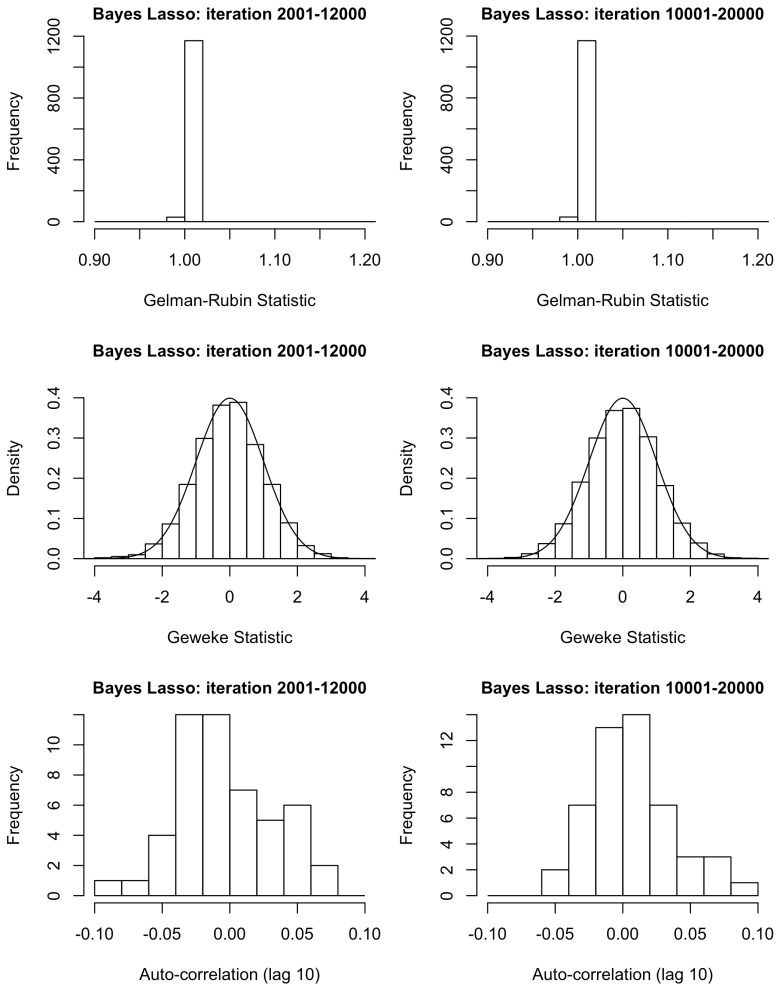


FIGURE S4.—The convergence diagnosis of the Bayesian Lasso. The Gelman-Rubin statistics for all the 1200 coefficients are calculated from 5 chains. The Geweke statistics distribution are based on 6000 coefficients from the 5 chains. The auto-correlation are calculated for the 50 covariates that are the QTL or have the highest correlation with the QTL (with unobserved genotype).

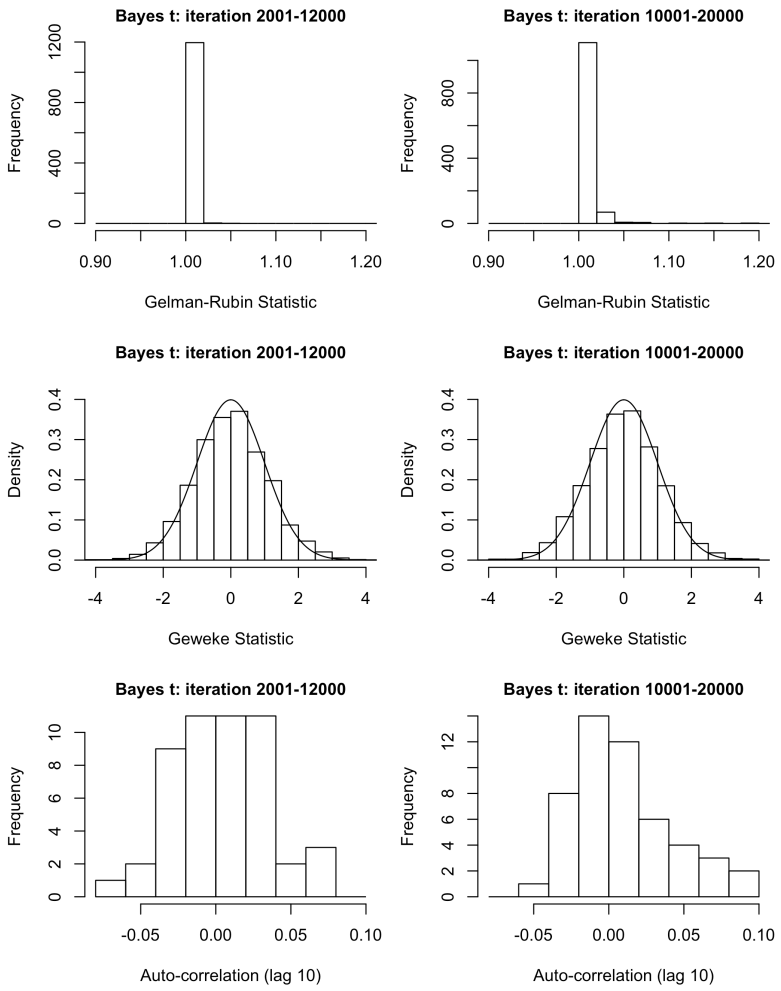


FIGURE S5.—The convergence diagnosis of the Bayesian t. The same statistics in the Web Figure 4 are plotted.

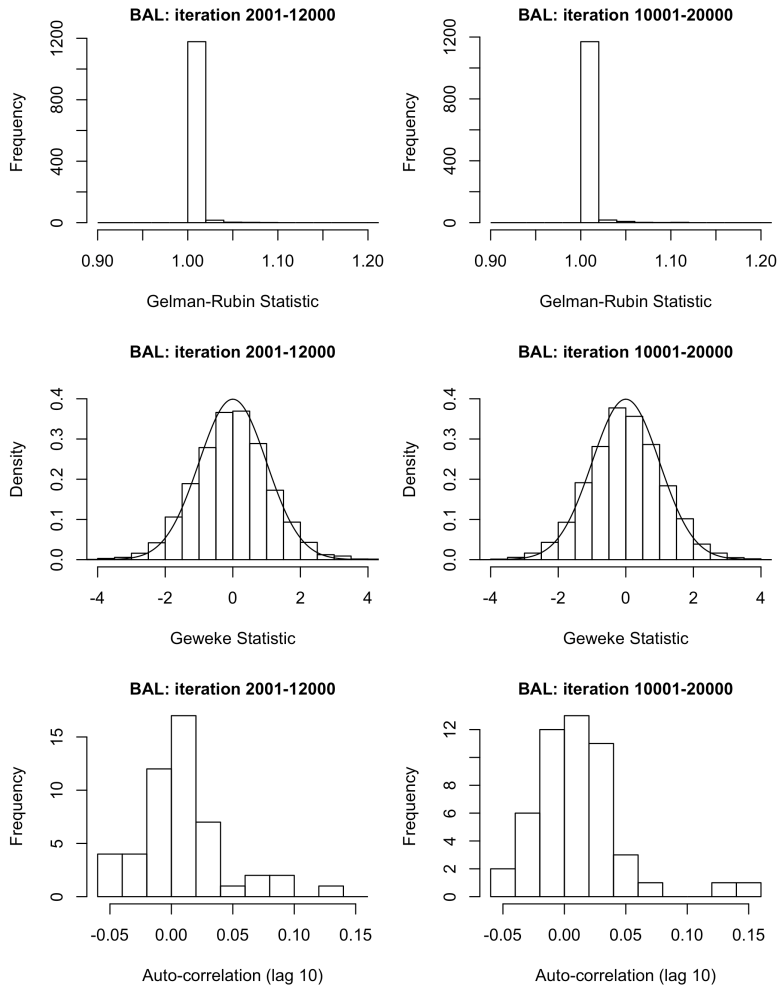


FIGURE S6.—The convergence diagnosis of the Bayesian adaptive Lasso. The same statistics in the Web Figure 4 are plotted.

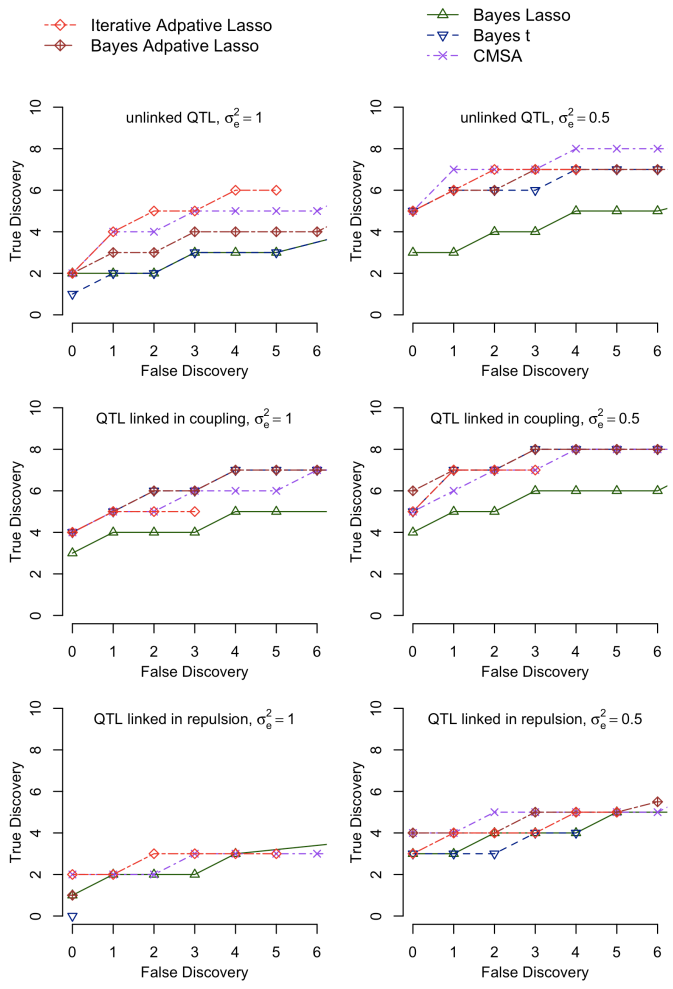


FIGURE S7.—Comparison of the number of true discoveries vs. the number of unlinked false discoveries from our simulations, for those method without explicit variable selection. For the IAL, the results before backward filtering.

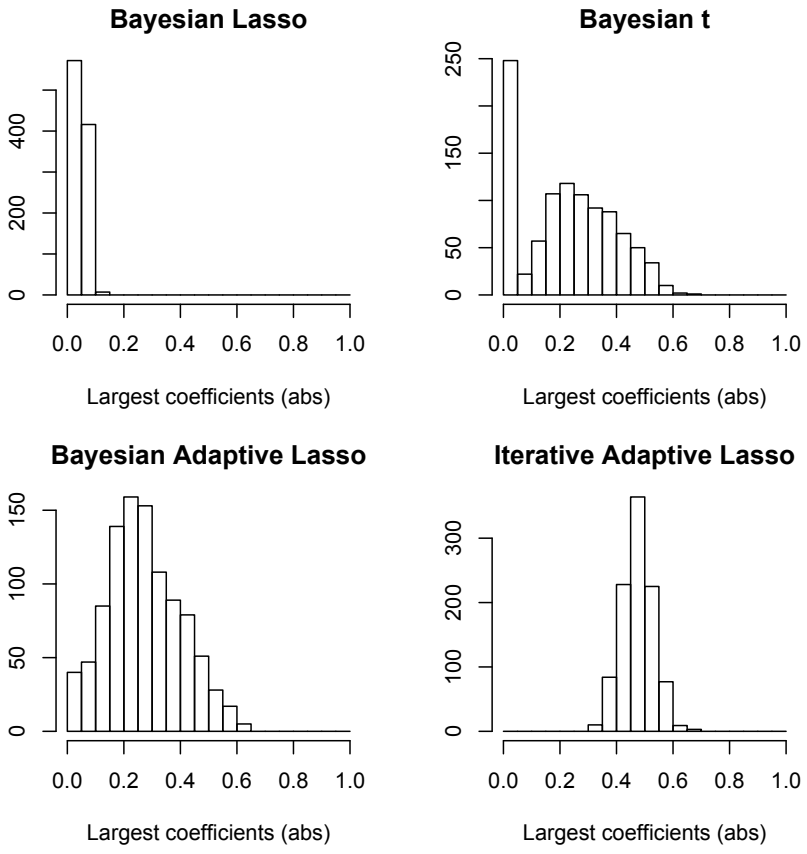


FIGURE S8.— Distributions of the largest coefficients (in absolute value) from each of the 1000 simulations in simulation situation 2 (unlinked QTL and σ^2_e =0.5).

TABLE S1

A list of 2D eQTL hotspots

QTL1	QTL2	Count	Gene Symbols	GO Term
15:160kb	15:560kb	46	ADE3 ALD4 ATP18 ATP20 ATP5 CAT5 COX5A COX7 DLD3 ECM16 ERS1 GPM3 INH1 KGD2 LSC1 LSP1 MDH1 MGM101 MIR1 NCE102 PEX2 PHM7 PIL1 POR1 RCK2 SDH1 SDH2 SDH3 SDH4 TCB2 TIM11 XKS1 YBR230C YCP4 YFR057W YGL188C YGR046W YJR120W YLL020C YLR224W YLR294C YML002W YMR148W YMR266W YNL274C YOL048C	generation of precursor metabolites and energy, 16 out of 46 genes, 1.42e-12
2:360kb	15:160kb	26	CRS5 CSI1 DIS3 DPH5 GAC1 HSP104 NMD5 ROM1 RTN2 SDP1 SDS23 SEN1 SOL4 SSE2 TPS2 UGP1 XBP1 YCR051W YGR052W YGR130C YHL021C YHR080C YKL036C YMR090W YNL194C YSC84	trehalose metabolic process, 3 out of 26 genes, 0.00065
5:420kb	15:160kb	25	AGX1 ATP5 CIN5 DAK1 DCS1 ECM16 GTT1 POR1 PRO1 SDH3 SYM1 TFS1 TSA2 YBR285W YDR070C YHR087W YJL161W YJR096W YJR120W YKL151C YLR252W YNL200C YNL274C YNR014W YPL014W	mitochondrion, 12 out of 25 genes, 0.00404
14:440kb	15:160kb	20	AAP1 CDC60 CLG1 DED81 DPS1 ERS1 FPR3 GUS1 MTM1 NAT1 NUP2 THS1 TYS1 UGP1 URE2 YDL203C YER182W YKR043C YMR244C-A YOL073C	tRNA aminoacylation, 5 out of 20 genes, 2.64e-06
12:660kb	14:440kb	18	ADI1 ARF2 CDC48 CPR5 FSH2 GET3 GSP2 HMX1 IZH2 LPP1 NDE1 PDR17 QRI5 TGL1 TRX2 UGP1 YDL237W YDR107C	NA
3:60kb	13:40kb	17	ALD5 BAT1 BAT2 CHA1 DIC1 ILV2 ILV3 ILV5 MCT1 OAC1 SAT4 XDJ1 YDR111C YDR112W YHR162W YJL213W YLR089C	branched chain family amino acid biosynthetic process, 5 out of 17 genes, 1.55e-09
4:1480kb	12:1040kb	17	YBL111C YBL113C YDR544C YEL074W YEL077C YFL066C YHL050C YHR218W YIL177C YJL225C YLL067C YRF1-1 YRF1-2 YRF1-3 YRF1-4 YRF1-6 YRF1-7	telomere maintenance via recombination, 6 out of 17 genes, 3.87e-11
12:660kb	14:480kb	16	CDC53 COX5A ERG26 ERV29 EUG1 KCC4 MUQ1 PMT3 VPS70 YDL086W YDL193W YEL047C YGR058W YJR114W YKL036C YNL058C	NA
14:480kb	15:160kb	16	ALR1 CBP2 COX5A DCS1 FUN14 MRPL38 MRPL51 MRPS18 SKO1 SSE2 TAH18 YCP4 YKL036C YML087C YPL105C YSC84	mitochondrion, 10 out of 16 genes, 0.00087
2:240kb	15:160kb	15	DNM1 GIP2 HXK1 HXT4 HXT6 HXT7 KGD2 SDH1 SDH3 SDH4 SER1 TYS1 YDL110C YHR097C YNL095C	acetyl-CoA catabolic process, 4 out of 15 genes, 8.72e-07
2:540kb	14:440kb	15	ARF2 CAP2 DSK2 ERG4 FSH2 IMD1 IMD2 ISR1 MBF1 MSP1 PFY1 UTH1 YGL199C YLR104W YOL073C	actin binding, 2 out of 15 genes, 0.00601

The eQTL location are the staring location of the 20kb bins. The gene ontology enrichment scores are calculated using SGD go term finder.

TABLE S2

A list of 3D eQTL hotspots.

Bin1		Bin2		Bin3			
Chr	Start	Chr	Start	Chr	Start	Count	Gene symbols
							YEL077C YFL066C YFL068W YHL050C YIL177C
4	1480k	12	1040k	15	560k	8	YJL225C YRF1-1 YRF1-2
2	240k	13	1b	15	160k	7	DNM1 HXT4 HXT6 HXT7 KGD2 SDH1 SDH3
2	360k	4	1480k	12	1040k	6	YEL077C YFL068W YIL177C YJL225C YRF1-1 YRF1-2
2	360k	4	1480k	15	560k	6	YEL077C YFL068W YIL177C YJL225C YRF1-1 YRF1-2
2	360k	12	1040k	15	560k	6	YEL077C YFL068W YIL177C YJL225C YRF1-1 YRF1-2
5	400k	15	160k	15	560k	6	ATP5 ECM16 POR1 SDH3 YJR120W YNL274C
2	360k	11	120k	15	160k	5	CSI1 HSP104 SOL4 UGP1 YKL036C
12	760k	12	1040k	15	560k	5	YEL077C YFL067W YFL068W YHL050C YJL225C
2	240k	15	160k	15	560k	4	KGD2 SDH1 SDH3 SDH4
2	520k	7	40k	14	480k	4	DBP7 PNO1 UTP4 UTP7
4	1480k	12	760k	12	1040k	4	YEL077C YFL068W YHL050C YJL225C
4	1480k	12	760k	15	560k	4	YEL077C YFL068W YHL050C YJL225C
13	1b	15	160k	15	560k	4	KGD2 MDH1 SDH1 SDH3
14	480k	15	160k	15	560k	4	COX5A ECM16 YCP4 YLR294C
2	200k	15	240k	15	560k	3	ATP1 ECM38 YBL100C
2	240k	13	1bp	15	560k	3	KGD2 SDH1 SDH3
2	280k	14	440k	15	160k	3	DED81 NAT1 THS1
2	360k	4	1480k	12	760k	3	YEL077C YFL068W YJL225C
2	360k	11	120k	12	640k	3	HSP104 UGP1 YKL036C
2	360k	12	640k	15	160k	3	HSP104 UGP1 YKL036C
2	360k	12	760k	12	1040k	3	YEL077C YFL068W YJL225C
2	360k	12	760k	15	560k	3	YEL077C YFL068W YJL225C
2	360k	14	480k	15	160k	3	SSE2 YKL036C YSC84
2	520k	3	40k	14	480k	3	KCC4 SSP120 YGR219W
2	520k	13	120k	14	440k	3	IMD1 IMD2 YAR075W
11	120k	12	640k	15	160k	3	HSP104 UGP1 YKL036C

The eQTL location are the staring location of the 20kb bins. We note that the group of genes YEL077C, YIL177C, YJL225C, YRF1-1, and YRF1-2 have very strong sequence similarity, hence they might co-hybridize on the microarray.

TABLE S3

A list of 4D eQTL hotspots

Bin1		Bin2		Bin3		Bin4			
Chr	Start	Chr	Start	Chr	Start	Chr	Start	Count	Gene symbols
2	360k	4	1480k	12	1040k	15	560k	5	YEL077C YIL177C YJL225C YRF1-1 YRF1-2
2	240k	13	20k	15	160k	15	560k	3	KGD2 SDH1 SDH3
2	360k	11	140k	12	660k	15	160k	3	HSP104 UGP1 YKL036C
4	1480k	12	780k	12	1040k	15	560k	3	YEL077C YHL050C YJL225C
2	360k	4	1480k	12	780k	12	1040k	2	YEL077C YJL225C
2	360k	4	1480k	12	780k	15	560k	2	YEL077C YJL225C
2	360k	12	780k	12	1040k	15	560k	2	YEL077C YJL225C
5	540k	13	40k	14	440k	15	500k	2	OST2 YOR102W

The eQTL location are the staring location of the 20kb bins.