Genome-wide Multiple Loci Mapping in Experimental Crosses by the Iterative Adaptive Penalized Regression

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

January 5, 2010

[*] Department of Biostatistics, University of North Carolina, Chapel Hill, NC, 27599

[**] Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599

[§] Corresponding author: wsun@bios.unc.edu

Keywords: Adaptive Lasso, Bayesian Adaptive Lasso, Iterative Adaptive Lasso, multiple loci mapping, quantitative trait loci (QTL), gene expression QTL (eQTL)

Abstract

Genome-wide multiple loci mapping can be viewed as a variable selection problem where the major objective is to select genetic markers related with a trait of interest. This is a challenging variable selection problem because the number of genetic markers is large (often much larger than the sample size) and there are often strong linkage or linkage disequilibrium between markers. In this paper, we developed two methods for genome-wide multiple loci mapping: the Bayesian adaptive Lasso and the iterative adaptive Lasso. Compared to the existing methods, the advantages of our methods come from the assignment of adaptive weights to different genetic makers, the iterative updating of these adaptive weights, and the ability to penalize most regression coefficients to be exactly zero. We evaluate these two methods as well as several existing methods in the application of genome-wide multiple loci mapping in experimental cross. Both large-scale simulation and real data analysis show that the proposed methods have improved variable selection performance. The iterative adaptive Lasso is also computationally much more efficient than the commonly used marginal regression and step-wise regression methods.

1 Introduction

It is well known that complex traits, including many common diseases, are controlled by multiple loci (Hoh and Ott, 2003). With the rapid advance of genotyping techniques, genome-wide high density genotype data can be measured accurately. However, multiple loci mapping remains one of the most attracting and most dif-

ficult problems in genetic studies, mainly due to the high dimensionality of the genetic markers as well as the complicate correlation structure among genotype profiles (throughout this paper, 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 Polymorphism's, SNPs) are measured in n individuals. We treat this multiple loci mapping problem as a linear regression problem:

$$y_i = b_0 + \sum_{j=1}^{p} x_{ij} b_j + e_i, \tag{1.1}$$

where y_i $(i = 1, ..., n)$ is the trait value of individual i, b_0 is the intercept, $e_i \sim N(0, \sigma_e^2)$. Here p is the total number of covariates. If we only consider the main effect of each SNP, $p = p_0$; and if we consider the main effects and all the pair-wise interactions of binary markers $p = p_0 + p_0(p_0 - 1)/2$. x_{ij} is the value of the j-th 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 based on the number of minor allele (i.e., the less frequent allele). Let $\mathbf{y} = (y_1, ..., y_n)^T$, $\mathbf{X} = (x_{ij})_{n \times p}$, $\mathbf{b}_0 = b_0 \mathbf{1}_{1 \times p}$, $\mathbf{b} = (b_1, ..., b_p)^T$, and $\mathbf{e} = (e_1, ..., e_n)^T$, equation (1.1) can be written as $\mathbf{y} = \mathbf{b}_0 + \mathbf{X}\mathbf{b} + \mathbf{e}$. We use this matrix form in some following derivations to simplify the notation. The major objective of multiple loci mapping is to identify the correct subset model, i.e., to identify those js, such that $b_j \neq 0$, and estimate the b_j's.

Marginal regression and step-wise regression are commonly used for multiple loci mapping. Permutation-based thresholds for model selection have been

proposed for marginal regression (Churchill and Doerge, 1994) and forward regression (Doerge and Churchill, 1996). Broman and Speed (2002) proposed a model selection criterion named BIC_δ, which was further rewritten into 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 estimates and Bayesian model selection. The existing Bayesian shrinkage methods are hierarchical models based on the additive linear model specified by equation (1.1), 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 related to the prior specification of the covariate-specific variance σ_j^2. An inverse-Gamma prior:

$$p(\sigma_j^2|\delta, \tau) = \text{inv-Gamma}(\delta, \tau) = \frac{\tau^\delta}{\Gamma(\delta)}(\sigma_j^2)^{-1-\delta} \exp(-\tau/\sigma_j^2), \tag{1.2}$$

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 for σ_j^2:

$$p(\sigma_j^2|a^2/2) = \text{Exp}(a^2/2) = \frac{a^2}{2} \exp\left(-\frac{a^2}{2}\sigma_j^2\right), \tag{1.3}$$

where a is a hyper-parameter. In this case, the unconditional prior of b_j is a Laplace distribution: $p(b_j) = \frac{a}{2}e^{-a|b_j|}$ 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 $p(\sigma_j^2|a^2/2) = \text{Exp}(a^2/2)$. 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) method (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 full Bayesian approach while the IAL is an (Expectation Conditional Maximization) ECM algorithm (Meng and Rubin, 1993). Both the BAL and the IAL are related with the adaptive Lasso (Zou, 2006), which extends the Lasso (Tibshirani, 1996) by allowing different penalization parameters for different regression coefficients. The adaptive Lasso enjoys the oracle property (Fan and Li, 2001), i.e., the covariates with nonzero coefficients will be selected with probability tend 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 where the number of covariates (p) is larger than the sample size (n). 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. 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 paper, 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 full Bayesian approach, while we have implemented and carefully studied the BAL. Secondly, 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 will be further elaborated in the Discussion section after we present our methods and results.

In this paper, we focus on the genome-wide multiple loci mapping in experimental cross of inbred strains (e.g., yeast sergeants, F2 mice) where typically thousands of genetic markers are genotyped in hundreds of samples. Another situation for multiple loci mapping is genome-wide association studies (GWAS) in outbred populations where millions of markers are genotyped in thousands of individuals. Multiple loci mapping in experimental cross and GWAS present different challenges. In experimental cross, genotype profiles have higher correlations in larger scale. Typically most markers from the same chromosome have correlated genotype profiles. Therefore the genotype profiles of QTL in experimental cross

are more often correlated. In contrast, GWAS data has higher dimensionality but the LD (linkage disequilibrium) blocks often have limited sizes. Therefore QTL genotype in GWAS are often independent or weakly correlated. We focus on the experimental cross because it is a situation where simultaneous multiple loci mapping methods have more advantages than the commonly used marginal regression or step-wise regression methods. However, we note this does not imply multiple loci mapping in experimental cross is in general easier than in GWAS.

The remainder of this paper is organized as follows. We first introduce the BAL and the IAL in Sections 2 and 3, respectively. We then evaluate them and several representative existing methods by extensive simulations in Section 4. A real data study of multiple loci mapping of gene expression traits is presented in section 5. Finally, we summarize and discuss the implications of our methodology in Section 6.

2 The Bayesian adaptive Lasso (BAL)

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

$$\begin{aligned} p(b_0) &\propto 1, \quad p(\sigma_e^2) \propto 1/\sigma_e^2, && (2.1)\\ p(b_j|\kappa_j) &= \frac{1}{2\kappa_j}\exp\left(-\frac{|b_j|}{\kappa_j}\right), && (2.2)\\ p(\kappa_j|\delta,\tau) &= \text{inv-Gamma}(\kappa_j;\delta,\tau) = \frac{\tau^\delta}{\Gamma(\delta)}\kappa_j^{-1-\delta}\exp\left(-\frac{\tau}{\kappa_j}\right), && (2.3)\end{aligned}$$

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

$$p(b_j) = \int_0^\infty \frac{\tau^\delta}{2\Gamma(\delta)}\kappa_j^{-2-\delta}\exp\left(-(|b_j|+\tau)/\kappa_j\right)d\kappa_j = \frac{\tau^\delta\delta}{2}(|b_j|+\tau)^{-1-\delta}, \quad (2.4)$$

which we refer to as a *power* distribution with parameter δ and τ. From this unconditional prior, we can see that larger δ and smaller τ lead to bigger penalization.

In practice, it could be difficult to choose specific values for the hyper-parameters δ and τ. We suggest a joint improper prior $p(\delta, \tau) \propto \tau^{-1}$, and let the data estimate δ and τ. Thus the posterior distribution of all the parameters is given by

$$\begin{aligned} & p(\mathbf{b}, b_0, \sigma_e^2, \kappa_1, ..., \kappa_p | \mathbf{y}, \mathbf{X}) \\ \propto \; & p(\mathbf{y}|\mathbf{b}, \mathbf{X}, b_0, \sigma_e^2) p(\sigma_e^2) p(b_0) \prod_{j=1}^{p} p(b_j|\kappa_j) p(\kappa_j|\delta, \tau) p(\delta, \tau) \\ \propto \; & \frac{1}{\sigma_e^{2+n}} \exp\left[-\mathtt{rss}/(2\sigma_e^2)\right] \frac{\tau^{\delta p - 1}}{(\Gamma(\delta))^p} \prod_{j=1}^{p} \kappa_j^{-2-\delta} \exp\left(-\frac{|b_j| + \tau}{\kappa_j}\right). \quad (2.5) \end{aligned}$$

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

The BAL can be better understood by comparing it with the Bayesian Lasso. Recall that in the Bayesian Lasso, $b_j|\sigma_j^2 \sim N(0, \sigma_j^2)$, $\sigma_j^2 \sim \text{Exp}(a^2/2)$, 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: it approaches a Lasso penalty (which is common to all the covariates) 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 is 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 power penalty in equation (2.4) by iteratively applying covariate-specific Lasso penalty, i.e., the adaptive Lasso penalty. Figure 1 illustrates that the power distribution provides better penalty than the Laplace distribution (i.e., the Lasso penalty) since it has higher peak at zero and heavier tails, which leads to more penalization for smaller coefficients and less penalization for larger coefficients.

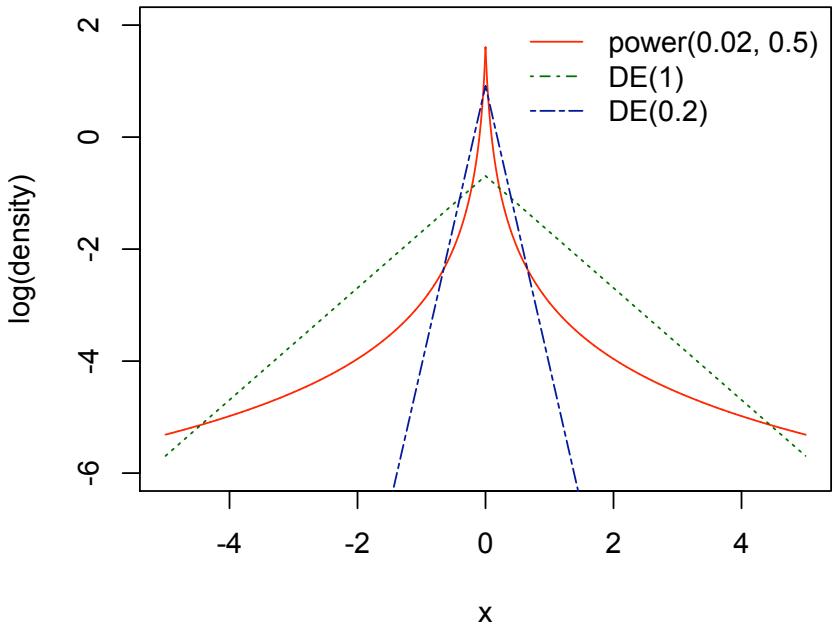

Figure 1: Comparison of the power distribution: $f(x;\tau,\delta) = \frac{\tau^{\delta}\delta}{2}(|x|+\tau)^{-1-\delta}$, given $\tau = 0.02$ and $\delta = 0.1$, and the Laplace (i.e., double exponential, DE) distribution $f(x;\kappa) = \frac{1}{2\kappa}\exp(-|x|/\kappa)$, given κ=1 or 0.2. We plot the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

3 The Iterative Adaptive Lasso (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, so that 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 (2.1)-(2.3)), we treat $\theta = (b_0, b_1, ..., b_p)$ as the parameter of interest and let $\phi = (\sigma_e^2, \kappa_1, ..., \kappa_p)$ be the missing data. The observed data are y_i and x_{ij}. We are interested in the following complete data log-posterior of θ:

$$l(\theta|\mathbf{y}, \mathbf{X}, \phi) = C - \texttt{rss}/(2\sigma_e^2) - \sum_{j=1}^{p} \frac{|b_j|}{\kappa_j}, \tag{3.1}$$

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 after some derivations (Supplementary Materials Section B), the conditional expectation of $l(\theta|\mathbf{y}, \mathbf{X}, \phi)$ with respect to the conditional density of $f(\phi|\mathbf{y}, \mathbf{X}, \theta^{(t)})$ is

$$Q(\theta|\theta^{(t)}) = C - (\texttt{rss}/2)/\left(\texttt{rss}^{(t)}/n\right) - \sum_{j=1}^{p} \frac{|b_j|}{\left(|b_j^{(t)}| + \tau\right)/(1+\delta)}, \tag{3.2}$$

where $\texttt{rss}^{(t)}$ is the residual sum of squares calculated based on $\theta^{(t)} = (b_0^{(t)}, b_1^{(t)}, ..., b_p^{(t)})$. Comparing equation (3.1) and (3.2), it is obvious that in order to obtain $Q(\theta|\theta^{(t)})$, we can simply update $\sigma_e^2 = \texttt{rss}^{(t)}/n$ and $\kappa_j = (|b_j^{(t)}| + \tau)/(1+\delta)$.

Based on the above discussions, the IAL is implemented as follows:

1. Initialization: We initialize $b_j (0 \leq j \leq p)$ with zero, initialize σ_e^2 by variance of y, and initialize $\kappa_j (1 \leq j \leq p)$ with $\tau/(1+\delta)$.

2. Conditional Maximization (CM) step:

   (a) Update b_0 by its posterior mode (see Supplementary Materials Section B),

   $$b_0 = (1/n) \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij} b_j \right). \tag{3.3}$$

   (b) For $j = 1, ..., p$, update b_j by its posterior mode (see Supplementary Materials Section B),

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

   where

   $$\sigma_j^2 = \frac{\sigma_e^2}{\sum_{i=1}^{n} x_{ij}^2}, \text{ and } \bar{b}_j = \left(\sum_{i=1}^{n} x_{ij}^2 \right)^{-1} \sum_{i=1}^{n} x_{ij} \left(y_i - b_0 - \sum_{k \neq j} x_{ik} b_k \right) \tag{3.4}$$

3. Expectation (E) step:

   With the updated b_j's, recalculate the residual sum of squares, `rss`, and

   (a) Update σ_e^2: $\sigma_e^2 = \texttt{rss}/n$.

   (b) Update κ_j: $\kappa_j = (|b_j| + \tau)/(1+\delta)$.

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

Being a valid ECM algorithm only guarantees that IAL identifies a local mode of the posterior probability. In order to obtain desirable variable selection performance, two issues need to be considered. (1) How to choose the initial values of the coefficients? (2) How to choose the hyper-parameter τ and δ. In the HDLSS setting, especially where the covariates are highly correlated, initial estimates from ordinary least squares or ridge regression are either unavailable, unstable, or none-informative. Therefore we simply initialize all the coefficients by zero. To decide δ and τ, we first examine the choice of δ and τ in an asymptotic point of view to estimate their magnitudes.

Theorem 1. Consider the multiple linear regression problem formulated in equation (1.1) 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. $|\text{corr}(\mathbf{x}_j, \mathbf{y} | \tilde{\mathbf{b}}_{-j}^{(t+1)})| > c$, then $p(\hat{b}_j^{(t+1)} \neq 0) \to 1$.

See the Supplementary Materials Section C for the proof.

Theorem 1 can be explained 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 $0 < d < 1/2$.

Theorem 1 only provides the magnitude of $(1+\delta)/\tau$. In practice, we select δ and τ by the BIC, followed by a backward filtering. The BIC is written as

$$\text{BIC}_{\tau,\delta} = \log(\texttt{rss}/n) + \frac{\log(n)}{n} df_{\tau,\delta}, \tag{3.5}$$

where $df_{\tau,\delta}$ is the number of nonzero coefficients, an estimate of the degrees of freedom (Zou *et al.*, 2007). Given the τ and δ selected by BIC, we can obtain a subset model, which usually includes a small number of covariates. Given the subset model, the backward filtering starts from the covariate with the smallest coefficient (in terms of absolute value) and iteratively test each covariate following the ascending order of their coefficients (absolute values). In each test, the model with/without the covariate is compared by an F-test. If the covariate can be dropped (i.e., the p-value is not small enough), the next covariate is tested; otherwise the backward filtering is terminated and the remaining covariates are the selected variables. The p-value cutoff can be set as $0.05/p_E$, where p_E is the effective number of independent tests. A conservative choice is to set $p_E = p$, the total number of tests. In this paper, we estimate p_E by quantifying the relation between nominal p-value and permutation p-value, see Sun and Wright (2009) and Supplementary Materials Section D for details. Note this backward filtering step is computationally efficient for the IAL since the IAL only keeps a small number of covariates with non-zero coefficients. Similar backward filtering approach can be applied to the Bayesian methods. However, we did not apply backward filter-

ing to the Bayesian methods for two practical concerns. First, few coefficients from the Bayesian methods will be exactly zero, thus backward filtering is computationally intensive. Second, most coefficients from the Bayesian methods are close to 0, thus their ranks are not quite informative. This will affect the stability of the backward filtering algorithm since it relies on these ranks.

An alternative strategy is to apply an extended BIC, which provides larger penalty for bigger models (Chen and Chen, 2008). We compared the extended BIC with the ordinary BIC plus backward filtering in our simulations. The latter has better 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 can better distinguish the model missing some true discoveries with the model keeping all the true discoveries; however it is likely to include some false discoveries in addition to all the true discoveries (Chen and Chen, 2008). Nevertheless, these false discoveries can be filtered out by the backward filtering.

4 Simulation Studies

We first use simulations to evaluate the variable selection performance of ten methods: marginal regression, forward regression, forward-backward regression (with penalized LOD as the model selection criterion), the composite model space approach (CMSA), the adaptive Lasso (with initial regression coefficients from marginal regression), the iterative adaptive Lasso (IAL), the HyperLasso, and the three Bayesian shrinkage methods: the Bayesian t, the Bayesian Lasso, and the

Bayesian adaptive Lasso (BAL).

4.1 Simulation setup

We employ the R/qtl (Broman *et al.*, 2003) to simulate genotype of 360 F2 mice. 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). The chromosome length is chosen to be close to the average chromosome length in mouse genome. Then we simulate genotype data of the 360 F2 mice based on 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 (Supplementary Figure 1). In fact, given the genetic distance of two SNPs, the expected R^2 between two SNPs in this F2 cross can be explicitly calculated (Supplementary Figure 2). For example, the R^2s of two SNPs 1cM, 5cM, and 10cM apart are 0.96, 0.82, and 0.67, respectively. Finally, we choose 10 markers from the 2000 markers as QTL, and simulate quantitative traits in six situations with 1000 simulations per situation. Given the 10 QTL, the trait is simulated based on the linear model in equation (1.1), where genotype (x_{ij}) is coded by the number of minor alleles. The QTL effect sizes across the six situations are listed below:

1. Un-linked 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.

3. 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); $\sigma_e^2 = 1$.

5. 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 only use 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 (Supplementary Figure 3).

4.2 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 implementation 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 E and F. The R package BPrimm

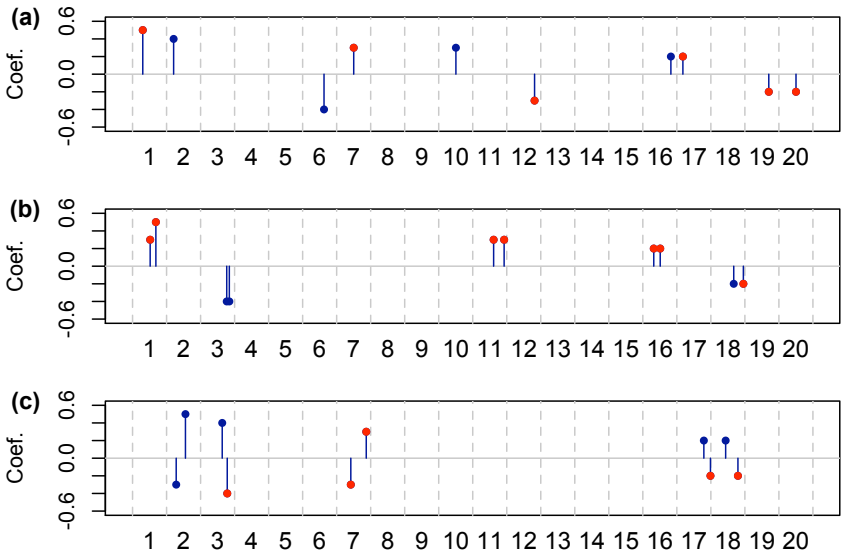

Figure 2: The locations and effect sizes of the QTL in simulation study. The markers labeled with red color are among the 800 markers with "missing data". (a) Situations 1 or 2. (b) Situations 3 or 4. The genetic distances/r^2 between two QTL from chromosome 6, 7, 10, 15, and 19 are 15cM/0.63, 7cM/0.68, 28cM/0.25, 17cM/0.56, and 26cM/0.40, respectively, where r^2 denotes the correlation square. (c) Situations 5 or 6. The genetic distances/r^2 between two QTL from chromosome 7, 8, 14, 18, and 19 are 25cM/0.31, 13cM/0.59, 41cM/0.22, 17cM/0.55, and 33cM/0.30, respectively. The mean (standard deviation) 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

can be downloaded from http://www.bios.unc.edu/∼wsun/software/.

We use 10,000 permutations to calculate permutation p-value 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). Some criteria can be used to dissect multiple QTL in one chromosome from the results of marginal regression. However the implementation of such criteria requires ad-hoc considerations and is beyond the scope of this paper. 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 plus backward filtering are implemented for 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 backward 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 software 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. As suggested by the author of the HyperLasso (personal communication), the "shape" parameter should be between 1 and 5. We tried three values 1, 3, and 5, and chose to set shape=1 since it gave slightly better performance than 3 and 5.

All the Bayesian methods use 10,000 burn-in iterations followed by 10,000 iterations to obtain 1000 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 all the 5 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.

4.3 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 non-zero), and the step-wise 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 hyper-parameters; and it is classified into the second group if we use the extended BIC or ordinary BIC plus backward 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, (3) the R^2 between this marker and the QTL is larger than 0.8. The cutoff 0.8 is chosen based on the R^2 between the markers with observed genotype profiles and those with unobserved genotype profiles so that the vast majority of the unobserved markers can be tagged. 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 un-linked 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 ROC-like curve that plots the median number of true discoveries versus 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 performance of different methods are not well-separated (Supplementary Figure 7). Overall the IAL, BAL and the CMSA have similar and superior performance, and the adaptive Lasso and the Bayesian Lasso have inferior performance.

Next we compare the step-wise regression method, the HyperLasso, the adaptive Lasso (with initial estimates from marginal regression) and the IAL with extended BIC or ordinary BIC plus backward 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 avoid linked false dis-

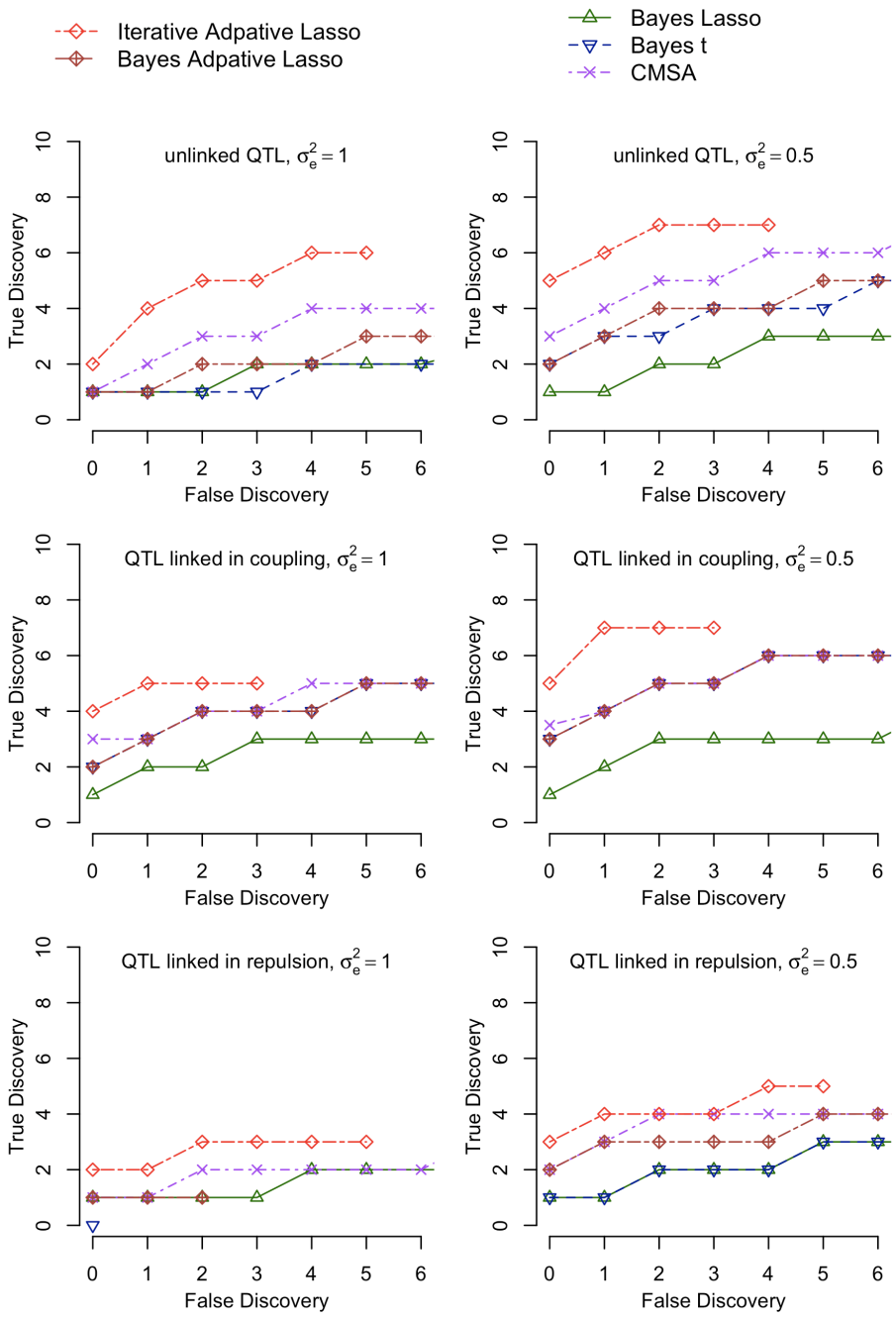

Figure 3: Comparison of the number of true discoveries vs. the total number of false discoveries in simulation study.

coveries should be preferred. As shown in Figure 4, the IAL with ordinary BIC plus backward filtering has the best performance while the HyperLasso has the worst performances in all the situations. When the QTL are linked in repulsion, the HyperLasso has no power at all. The adaptive Lasso has similar performance as the IAL when the signal is strong (i.e., QTL linked in coupling), otherwise it has significantly worse performance than the IAL. The step-wise regression and the IAL using the extended BIC have slightly worse performance than the IAL using ordinary BIC plus backward filtering.

5 Gene expression QTL study

QTL study of one particular trait may favor one method by chance. In order 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 an eQTL data with more than 6000 genes and 2956 SNPs in 112 yeast segregants (Brem and Kruglyak, 2005; Brem *et al.*, 2005). The gene expression data is downloaded from Gene Expression Omnibus (GEO, GSE1990). The expression of 6229 genes are measured in the original data. We drop 129 genes that have more than 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 is from Dr. Rachel Brem. Fifteen SNPs with more than 10% missing values are excluded from this study, and the

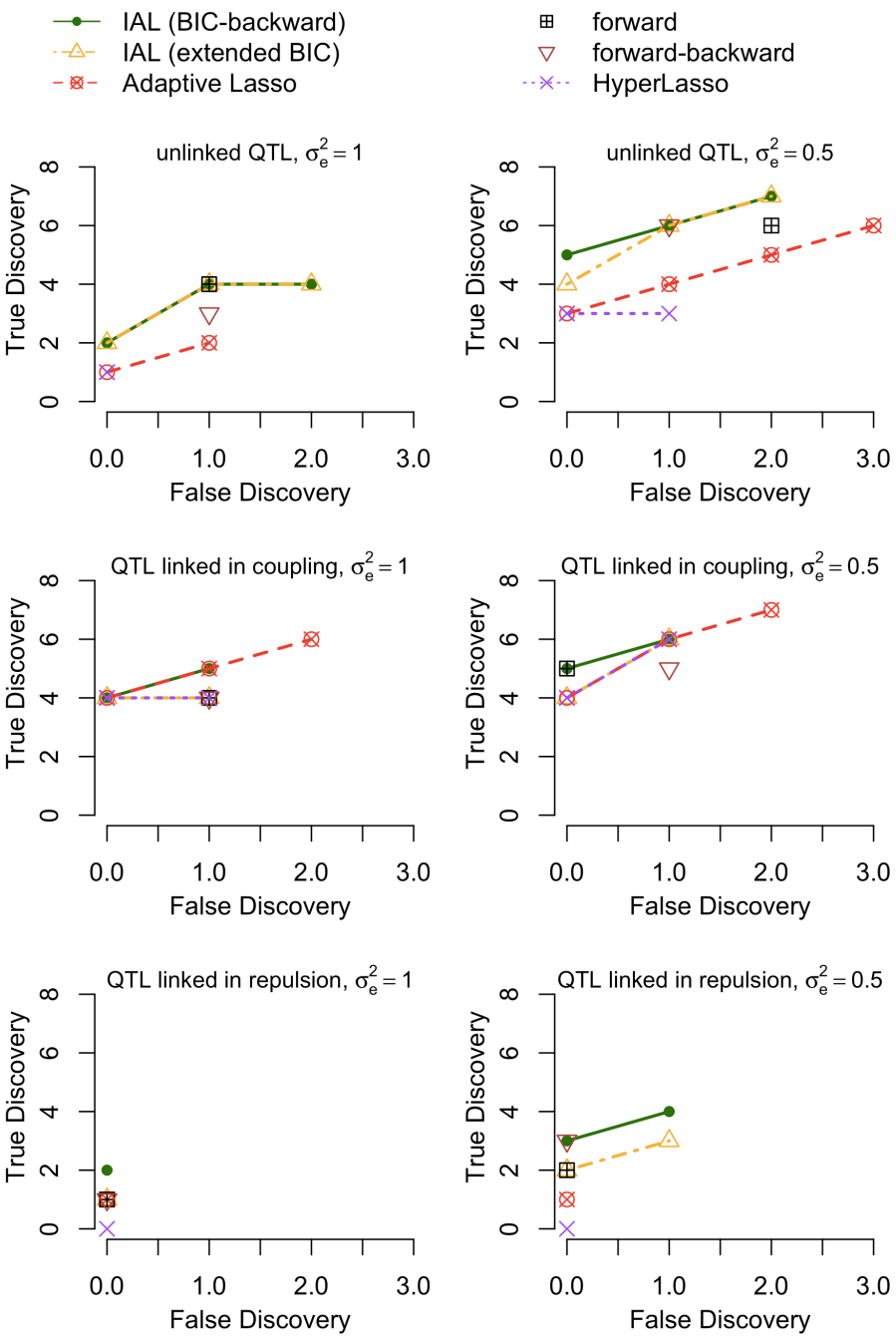

Figure 4: Comparison of the number of true discoveries vs. the total number of false discoveries in simulation study.

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 more than 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 this eQTL data.

We apply the IAL, marginal regression, forward regression, forward-backward regression, and HyperLasso to this 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 step-wise regression are set as 0.05. The parameters δ and τ of the IAL are selected by ordinary BIC followed by backward filtering. In the backward filtering step, we use a p-value cutoff 0.05/412, based on a conservative estimate of 412 independent tests (see the Supplementary Materials Section D). For the HyperLasso, the "lambda" parameter is set as $\sqrt{n}\Phi^{-1}(1 - 0.05/412/2)$, and the "shape" parameter is set to be 1. The IAL and step-wise 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 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 associ-

Table 1: The number of genes with certain number eQTL.

method	Total # of genes with at least one eQTL	The number 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

ated with the same k loci, these k loci is referred to as a k-D eQTL hotspot. The results of the IAL reveals 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 Figure 5 and Figure 6, it is interesting that a 1D-eQTL can be further divided into several groups based on the results of 2D-eQTL, which is consistent with the finding of Zhang *et al.* (2009).

We divide the whole yeast genome into 600 bins of 20kb regions, which lead to 600*599/2 = 179,700 bin pairs as potential "2D eQTL hotspots". Eleven bin pairs are linked to more than 15 genes (Supplementary Table 1). The cutoff is chosen arbitrarily so that we can focus on a relatively small group of 2D hotpots with definite significant enrichment. Due to space limit, we only discuss in deails the largest 2D hotspot located at chr15:160kb-180kb and chr15:560-580kb. There are 46 genes linked to these two loci simultaneously, and among them 16 are in-

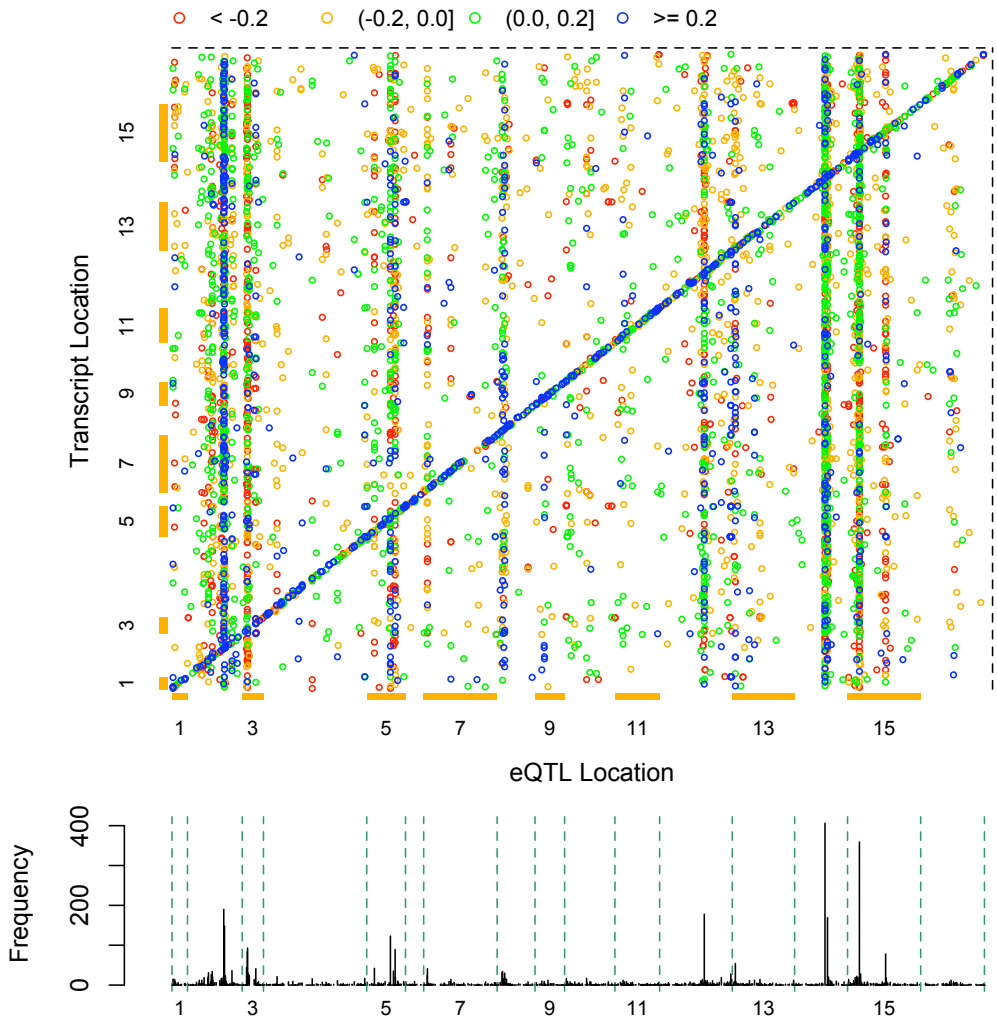

Figure 5: Illustration of the eQTL mapping results in 112 yeast sergeants. In the upper panel, each point corresponds to a 1-D eQTL result, where X-coordinate is the location of the eQTL and Y-coordinate is the location of the gene. Different colors indicate different sizes of the regression coefficients. The diagonal band indicates the cis-eQTL where expression of one gene is associated with the genotype of a nearby marker. The vertical band indicate 1D-eQTL hotspots. In the lower panel, the number of genes linked to each marker are plotted. Several 1-D eQTL hotspots are apparent for those markers that harbor the eQTL of hundreds of genes.

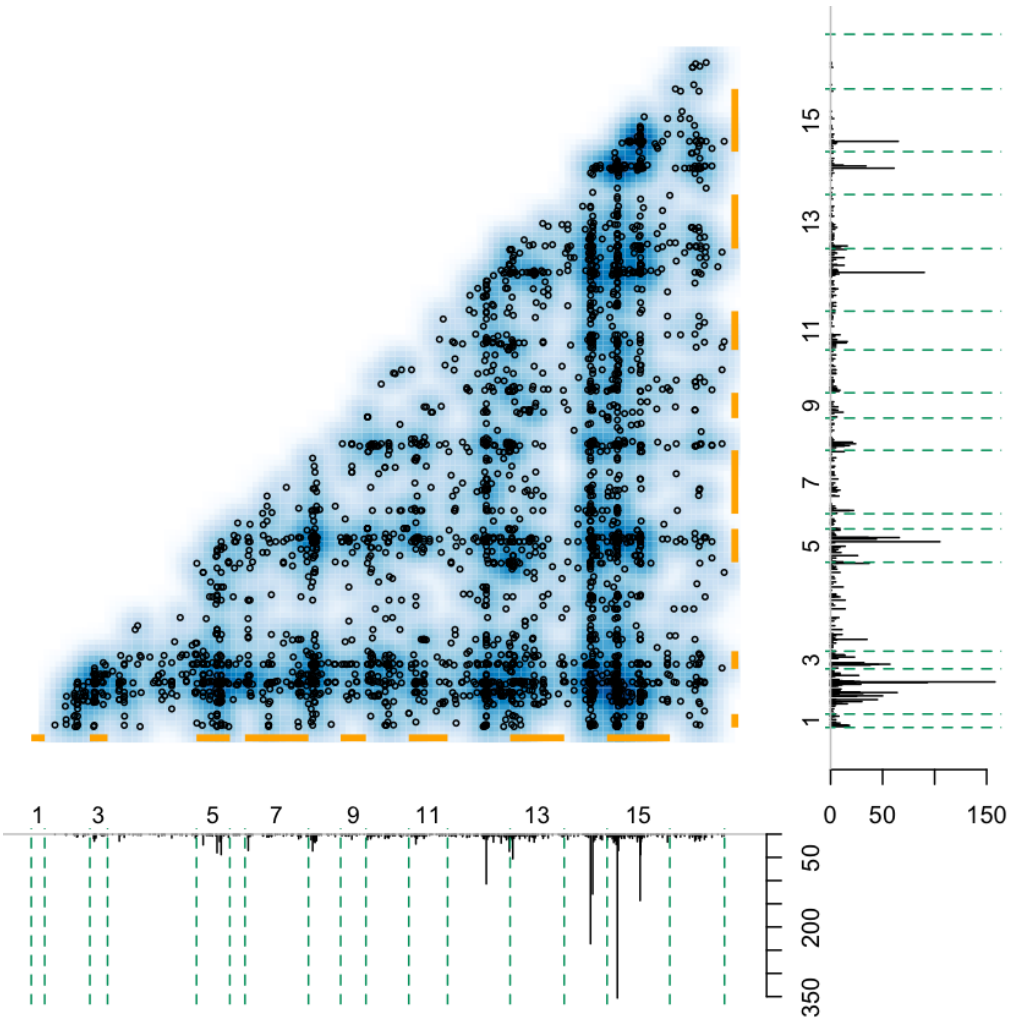

Figure 6: The distribution of the loci pairs linked to the same gene. Here each point corresponds to a 2-D eQTL, and the background color reflects the density of the distribution. If the expression of one gene is linked to more than two loci, we plot each pair of linked loci. For example, if one gene is linked to three markers 1, 2, and 3, which are located at position 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).

volved 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, 170,945-180,961bp, and one maker at Chr15, 563,943bp. One potential causal gene nearby chr15:171-181kb is PHM7 (Zhu *et al.*, 2008), and one potential causal gene nearby chr15:564kb 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 ($\sim$35.8 millions possible three loci combinations, Supplementary Table 2) or even four loci ($\sim$5,346 millions possible four loci combinations, Supplementary Table 3). For example, three genes KGD2, SDH1, SDH3 are all linked to four loci: chr2:240-260kb, chr13:20-40kb, chr15:160-180kb, and chr15:560-580kb. Interestingly, all these three genes are involved in "acetyl-CoA catabolic process" (p-value 1.93×10^{-7}).

6 Discussion

In this paper, we have proposed two variable selection methods, namely the Bayesian Adaptive Lasso (BAL) and the Iterative Adaptive Lasso (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 performance 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 ten methods we tested. Coupled with a backward filtering approach, 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 has a much simpler unconditional posterior (equation (2.4)). The difference is not trivial since it not only leads to convenient theoretical studies in the theorem 1, but also 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 by backward filtering. 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 lead to little power to detect relatively weaker signals.

The IAL is computationally very efficient. For example, it takes 4 hours to carry out the multiple loci mapping for the yeast eQTL data with 6100 genes and 1017 markers. In contrast, the marginal regression, forward regression, and forward backward regression take about 60, 100, and 200 hours. 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 backward filtering step. The IAL results can be reused for different type I errors. In contrast, for the step-wise regression, all the compu-

tation need to be redone for each type I error.

Our results seems contradict to 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 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 non-zero coefficients (Zou, 2006). Since the Bayesian Lasso has similar penalization characteristics as 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 over-penalizes the regression coefficients (Supplementary Figure 8). This is consistent with the findings that "Lasso has had to choose between including too many variables or over shrinking the coefficients" (Radchenko and James, 2008). In contrast, the Bayesian t, the BAL and the IAL have increasingly smaller penalization on the coefficients estimates. The IAL seems to provide unbiased coefficients estimates. This leads to an assumption that the IAL has the oracle property, which warrant 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 works. Although for the gamma prior in the BAL, we have tried the strategy to set the

hyper-parameters δ and τ as fixed numbers. In general, the results is not very sensitive to the choices of the 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 it first (using Viterbi algorithm implemented in R/qtl) and then take 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 rate larger than 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 discuss our method in a linear regression framework. Extension to the generalized linear model (e.g., logistic regression for binary responses) is straightforward. Generalized linear model can be solved by iterated re-weighted least squares. Similar to the approach used in Friedman *et al.* (2009), our method can be plugged-in to solve the least square problem with in the loop of iterated re-weighted least squares. Yi and Banerjee (2009) proposed an intriguing and efficient approach to apply generalized linear model for multiple loci mapping. We

cannot adopt similar approach since we use different penalties. Yi and Banerjee (2009) also proposed to study large number of markers chromosome by chromosome, which is plausible solution to apply our method to GWAS data where millions of genetic markers are available.

We have tested the robustness of our methods by additional sets of simulations where the traits are log or exponentially transformed (data not shown). The conclusion is that our methods are fairly robust to mild violations of the linear model assumption. Although it has been argued that genetic effects most often contribute to the complex traits additively (Hill *et al.*, 2008), one may still expect the additive linear model assumption is severely violated in some situations. For example, when epistatic interactions are present. As mentioned in the introduction section, it is straightforward to include the pair-wise 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 (e.g., genetic interaction or protein-protein interaction) 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 genome-wide 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 as well.

Acknowledgements

The authors want to thanks Dr. Jun Liu for valuable suggestions for the implementation of the Bayesian Adaptive Lasso. WS's research is supported in part by grant from National Institute Environmental Health Sciences (5 P30 ES10126-07). FZ's research is supported by NIH grant GM074175.

Supplementary Materials

References

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

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

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

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

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

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

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

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

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

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

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, pages 881–889.

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

Hill, W., Goddard, M., and Visscher, P. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet*, **4**(2), e1000008.

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

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

Huang, J., Ma, S., and Zhang, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statistica Sinica*, **18**, 1603–1618.

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

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

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

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

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

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

Radchenko, P. and James, G. M. (2008). Variable inclusion and shrinkage algorithms. *Journal of the American Statistical Association*, **103**, 1304–1315.

Richardson, S. and Green, P. J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion). *Journal of the Royal Statistical Society: Series B*, **59**, 731–792.

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

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.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. Royal. Statist. Soc B.*, **58**, 267–288.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, **17**(6), 520–525.

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

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

Yi, N. and Banerjee, S. (2009). Hierarchical Generalized Linear Models for Multiple Quantitative Trait Locus Mapping. *Genetics*, **181**(3), 1101.

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

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

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal Of The Royal Statistical Society Series B*, **68**(1), 49–67.

Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., Mackelprang, R., and Kruglyak, L. (2003). Trans-acting regulatory variation in Sac-

charomyces cerevisiae and the role of transcription factors. *Nat Genet*, **35**(1), 57–64.

Zhang, W., Zhu, J., Schadt, E. E., and Liu, J. S. (2009). A Bayesian Partition Method for Detecting eQTL Modules. *Manuscript*.

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

Zhu, J., Zhang, B., Smith, E., Drees, B., Brem, R., Kruglyak, L., Bumgarner, R., and Schadt, E. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature genetics*, **40**(7), 854.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418–1429.

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.

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