Penalized Multi-Marker versus Single-Marker Regression Methods for Genome-Wide Association Studies of Quantitative Traits

Hui Yi*,§, Patrick Breheny†, Netsanet Imam*, Yongmei Liu** and Ina Hoeschele*‡

* Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA
§ PhD Program in Genetics, Bioinformatics and Computational Biology, Virginia Tech, Blacksburg, VA 24061, USA
† Department of Biostatistics, University of Iowa, Iowa City, IA 52240, USA
** Departments of Epidemiology & Prevention and Internal Medicine, Division of Public Health Sciences, Translational Research Institute, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA
‡ Department of Statistics, Virginia Tech, Blacksburg, VA 24061, USA
Running Title:
Penalized vs Single Marker Regression GWAS

Keywords:
Genome-wide Association Study, penalized regression, false discovery rate, linkage disequilibrium

Corresponding Author:
Ina Hoeschele
Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, VA 24061-0477
(540) 231-3135
inha@vbi.vt.edu
ABSTRACT

The data from genome-wide association studies (GWAS) in humans are still predominantly analyzed using single marker association methods. As an alternative to Single Marker Analysis (SMA), all or subsets of markers can be tested simultaneously. This approach requires a form of Penalized Regression (PR) as the number of SNPs is much larger than the sample size. Here we review PR methods in the context of GWAS, extend them to perform penalty parameter and SNP selection by False Discovery Rate (FDR) control, and assess their performance in comparison with SMA. PR methods were compared with SMA using realistically simulated GWAS data with a continuous phenotype and real data. Based on these comparisons our analytic FDR criterion may currently be the best approach to SNP selection using PR for GWAS. We found that PR with FDR control provides substantially more power than SMA with genome-wide type-I error control but somewhat less power than SMA with Benjamini-Hochberg FDR control (SMA-BH). PR with FDR based penalty parameter selection controlled the FDR somewhat conservatively while SMA-BH may not achieve FDR control in all situations. Differences among PR methods seem quite small when the focus is on SNP selection with FDR control. Incorporating linkage disequilibrium into the penalization by adapting penalties developed for covariates measured on graphs can improve power but also generate more false positives or wider regions for follow-up. We recommend the Elastic Net with a mixing weight for the Lasso penalty near 0.5 as the best method.
INTRODUCTION

The goal of genome-wide association studies (GWAS) in humans and model organisms is to select a small subset of DNA markers, typically Single Nucleotide Polymorphisms (SNPs), which are in strong Linkage Disequilibrium (LD) with functional polymorphisms affecting a biomedical/clinical trait of interest. The selected markers are then replicated in other GWAS, fine-mapped and further validated. GWAS may be viewed as a large-scale variable selection problem, with several millions of common SNPs, measured directly or imputed, available in current studies in humans.

GWAS practitioners still strongly rely on single marker association analysis (SMA), including linear regression for continuous phenotypes, with control of the genome-wise error rate (GWER; a special case of the family-wise error rate FWER), which accounts for the multiplicity of the entire genome. Assuming that a biomedical trait of interest is affected by multiple polymorphisms with ‘detectable’ effects, SMA fits an incorrect model, and it is sensible to consider alternative methods which test all or subsets of markers simultaneously. This approach requires a form of Penalized Regression (PR) as the number of SNPs is much larger than the sample size.

A major practical issue with the use of penalized regression is the determination of “optimal” values for the tuning parameter(s) and the lack of an error rate associated with the selection of SNPs. Common approaches for tuning parameter value determination include cross-validation (CV) (e.g., Kohavi (1995)) and the use of a model selection criterion such as Akaike’s Information Criterion (AIC; Akaike (1974; 1977)), the Bayesian Information Criterion (BIC; Schwarz (1978)), or the Extended Bayesian
Information Criterion (EBIC) (Chen and Chen 2008). These approaches work well for prediction but not for the goal of variable selection as most of these criteria lead to an unacceptable number of false positives and do not provide a measure of error associated with the selected SNPs. This distinction is important as the best predictive models may contain covariates that are not important, while the model with all important covariates may not be the best predictive model.

The absence in PR of the control of an appropriate error rate such as the FWER or the False Discovery Rate FDR (Benjamini and Hochberg 1995; Sabatti and Freimer 2003) has been recognized by several authors who have employed different strategies including data permutation (Ayers and Cordell 2010), stability selection (Meinshausen and Bühlmann 2010) with FDR control (Ahmed et al. 2011), and multi-stage (data splitting) approaches (Meinshausen et al. 2009; Wasserman and Roeder 2009) to control FWER or FDR. However, these approaches require much more computation and may not be feasible on a genome-wide scale, and/or the provision of an error rate estimate comes at the expense of reduced power.

A second practical issue is that a number of PR methods differing in the penalty function have been proposed, and hence it is unclear whether any and which of these methods should be preferred in the context of variable selection in GWAS. Moreover, some authors (Kim and Xing 2009; Liu et al. 2011) have suggested that penalties should incorporate LD, but it is unclear whether such penalties produce any gain in power and/or a reduction in the false positive findings.
The purpose of this contribution is to review penalized regression methods in the context of GWAS, to extend selected PR methods by incorporating FDR control and to assess their performance in comparison with single marker (SNP) analysis, and to provide recommendations on the use of these methods in GWAS practice. PR methods are compared with SMA on realistically simulated GWAS, consisting of genotype data on single and multiple chromosomes and a continuous phenotype, and on real data.

MATERIALS AND METHODS

Data Simulation

To simulate GWAS data with realistic patterns of LD, we used the software HAPGEN2 (Su et al. 2011), which produces genotyped individuals by re-sampling from a set of reference haplotypes. We used the haplotypes of the 60 Caucasians of European origin (CEU) in HapMap2 (International HapMap Consortium, (2007)). For most simulation scenarios, SNP genotypes were simulated for a single chromosome (chromosome 21). For an additional simulation scenario, SNP genotypes were simulated for three chromosomes (19, 21 and 22). The number of SNPs genotyped across all 22 autosomes in the HapMap2 population is 3,849,034, and the numbers of SNPs on chromosomes 19, 21 and 22 are 56,607, 50,165 and 54,786, respectively. Following the simulation of each SNP dataset, SNPs were removed if they had a Minor Allele Frequency (MAF) below 0.01 or if their absolute correlation with another SNP exceeded 0.999, reducing the number of SNPs on average to 25,033, 21,519 and 22,199 for chromosomes 19, 21 and 22, respectively.
To determine suitable sample sizes for the data simulation, we performed power calculations using QUANTO (Gauderman and Morrison 2001). A functional or causal SNP affecting the phenotype will be referred to as a Quantitative Trait Locus (QTL). The sample sizes were determined as those needed to detect an isolated QTL by SMA, with a heritability (variance explained by the QTL over total variance) of 0.10, with a power of 50%, and with a p-value based significance threshold of \(5.5 \times 10^{-8} \times \frac{3849034}{50165}\) for chromosome 21 and \(5.5 \times 10^{-8} \times \frac{3849034}{(56607+50165+54786)}\) for the three-chromosome simulation, using the GWER p-value threshold of \(5.5 \times 10^{-8}\) (see below). The required sample sizes were \(N = 201\) and \(N = 222\), respectively.

For the chromosome 21 simulations, two scenarios were considered: (1) two isolated QTLs with heritability of 0.10 each and hence a total heritability of 0.20; (2) eight QTLs comprising a group of four QTLs with weak pairwise LD \((0.01 \leq r^2 \leq 0.1)\) and two groups of two QTLs each with within-group LD of \(r^2 \approx 0.5\). All QTLs had MAF \(> 0.05\). For scenario (2), the four QTLs in one group had a heritability of 0.05, and the remaining four QTLs had a heritability of 0.04. Taking into account LD among the 8 QTLs in scenario (2), the total heritability was 0.48 (computed as

\[
h^2 = \frac{\text{Var} \left( \sum_{k=1}^{8} X_k \beta_k \right)}{\text{Var} \left( \sum_{k=1}^{8} X_k \beta_k \right) + \sigma^2_e},
\]

where \(X_k\) is the allelic dose of SNP \(k\) in individual \(i\), \(\beta_k\) is the regression coefficient for SNP \(k\), and \(\sigma^2_e\) is the residual variance). Moreover, for scenario (2) the “effective” heritability of each individual QTL in the context of the single marker model was \(\approx 0.10\) due to LD among QTLs. Hence, SMA had the same expected power for QTL detection under both scenarios. A total of 200
replicate simulations were performed, in which the QTL positions and effects were kept constant. The phenotype was simulated based on an additive model with $Q$ QTL, $Y = \sum_{j=1}^{Q} X_j \beta_j + e$, where $Y$ denotes the phenotype. Heritability of a single QTL $j$ was computed as $h_j^2 = \beta_j^2 \text{var}(X_j) / \text{var}(Y)$. “Effective” heritability (increased due to LD with other QTL) was computed as above but by replacing $\beta_j$ with

$$\tilde{\beta}_j = \frac{\text{cov}(Y, X_j) / \text{var}(X_j) = \left[ \beta_j \text{var}(X_j) + \sum_{j' \neq j}^{Q} \beta_{j' \text{cov}}(X_j, X_{j'}) \right] / \text{var}(X_j)}{\text{var}(X_j)}.$$  

For the multi-chromosome simulation, chromosome 19 did not harbor any QTL, chromosome 21 harbored the 8 QTLs of scenario (2), and chromosome 22 harbored two isolated QTL with heritability of 0.10 and MAF > 0.05 as in scenario (1). The total heritability was 0.68.

**Single Marker Regression**

For single marker regression, variable selection is performed by choosing a cut-off value for the p-values, determined by some method of multiple testing control, most commonly the genome- (family-) wise error rate (GWER), which accounts for the multiplicity of the entire genome. The GWER-based p-value threshold is obtained by estimating an “effective number of independent tests” (e.g., The International HapMap (2005); Dudbridge and Gusnanto (2008)) by permutation or analytic approximation. The International HapMap Consortium permutation-based estimated significance threshold is $5.5 \times 10^{-8}$ for two-sided tests of SNPs, which we used here. As an alternative to the stringent GWER threshold, we investigate the False Discovery Rate (FDR; Benjamini...
and Hochberg (1995)). For FDR control, we use the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg 1995) shown to control the FDR under positive-regression dependence (Benjamini and Yekutieli 2001), and this condition is believed to hold for GWAS (Sabatti and Freimer 2003). We also use Benjamini and Yekutieli's (2001) BY approach which ensures FDR control under any form of dependence but can be very conservative. Additionally, we consider the local FDR (Efron 2005; Efron and Tibshirani 2002), where the p-values, assumed to represent a mixture of null and alternative hypotheses, are transformed to z-scores which are modeled by a mixture distribution, or

$$f(z) = \pi_0 f_0(z) + (1 - \pi_0) f_1(z) \quad [1]$$

where $f_0(z)$ is the density of $z$ under the null hypothesis, $f_1(z)$ is the density under the alternative hypothesis, $f(z)$ is the (overall) mixture density of $z$, and $\pi_0$ is the proportion of true null hypotheses. Based on [1], the local FDR is

$$locFDR(z) = \frac{\pi_0 f_0(z)}{f(z)} \quad [2]$$

which is also a posterior probability that the null hypothesis is true given $z$. The FDR is a lower bound on $locFDR$ because it is the expectation of $locFDR$ within a tail area (Efron 2005). Sun and Cai (2007) developed an oracle testing procedure that minimizes the false negative rate (the expected proportion of false negatives among all non-rejections) subject to a constraint on the FDR. This procedure is a thresholding rule based on the $locFDR$ and is implemented with an adaptive procedure that is asymptotically valid and optimal if the estimates of $\pi_0$, $f_0$ and $f$ in [2] are consistent. We estimated $f(z)$ and $\pi_0$
using the simple non-parametric approach described in Storey et al. (2005). Despite the correlation structure of the z-values, \( f_0 \) was fit well by the N(0,1) null distribution (see Figure S1 as a typical representative of the data simulation replicates). We also used the consistent estimators of \( \pi_0, f_0 \) and \( f \) in [2] of Jin and Cai (2007). We will refer to this procedure with the two sets of estimators for \( \pi_0, f_0 \) and \( f \) as LFDR1 and LFDR2. The proportion of true nulls \( \pi_0 \) is expected to be just below or at 1.0 in the context of GWAS.

**Penalized Regression Methods**

All penalized regression methods for the multiple regression model

\[
y_i = \sum_{k=1}^{p} x_{ik} \beta_k + e_i
\]

(assuming a centered response variable \( y \) and standardized SNP covariates \( x \)) can be represented by the following estimator:

\[
\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2N} \sum_{i=1}^{N} \left( y_i - \sum_{k=1}^{p} x_{ik} \beta_k \right)^2 + P(\lambda, \bar{\beta}) \right\}
\]

[3]

where \( \hat{\beta} \) is a vector of \( p \) regression coefficients, \( y_i \) is the continuous trait measurement on individual \( i \), \( x_{ik} \) is allelic dose of SNP \( k \) in individual \( i \), \( e_i \) is a residual, \( N \) is sample size, \( P(\cdot) \) is a penalty function, and \( \lambda \) is a vector of tuning or penalty parameters with typically one or two components. PR methods differ in the specification of \( P(\cdot) \). Variable selection, by setting unimportant coefficients to zero, occurs if the penalty has a singularity at zero. Below we only consider such penalties. Given the extensive literature, we do not review these methods in detail but rather present their penalty functions and discuss some properties relevant to their application to GWAS.
Desirable properties for penalized regression methods have been established by several authors and include sparsity (variable selection is enabled by automatically setting small coefficients to zero), continuity (an estimator which is continuous in data avoids instability in model prediction), asymptotic unbiasedness (bias should be low, in particular for large true coefficients), and the (strong) oracle property (Fan and Li 2001; Fan and Peng 2004; Kim et al. 2008; Zhang 2010). The Lasso (Tibshirani 1996) performs poorly with regard to the unbiasedness property, prompting the development of other penalties to overcome this problem. The Adaptive Lasso (Zou 2006), the Smoothly Clipped Absolute Deviation SCAD (Fan and Li 2001), and the Minimax Concave Penalty MCP (Zhang 2010) possess the (strong) oracle property.

**Lasso, Elastic Net and Adaptive Lasso**

The Lasso (Tibshirani 1996) employs the L1 penalty, or

\[
P_{\text{Lasso}} (\hat{\beta}, \hat{\lambda}) = \lambda \sum_{k=1}^{p} |\beta_k|, \quad \lambda \geq 0
\]  \[4\]

If there are groups of variables with strong pairwise correlations, it has been empirically observed that the Lasso tends to select only one variable from each group. This behavior is in contrast with single marker regression, which will select all markers in sufficiently strong LD with a QTL.

The Elastic Net EN (Zou and Hastie 2005) combines the Lasso (L1) penalty, enabling it to perform variable selection, with the Ridge (L2) penalty, enabling it to deal with multi-collinearity, or
\[ P_{EN}(\lambda, \tilde{\beta}) = \lambda_1 \lambda_2 \sum_{k=1}^{p} |\beta_k| + \lambda_1 (1 - \lambda_2) \sum_{k=1}^{p} \beta_k^2; \quad \lambda_1 \geq 0, \ 0 \leq \lambda_2 \leq 1 \]  

[5]

In contrast with the Lasso, the EN may be expected to co-select SNPs in strong LD.

Here, EN analysis was performed with three different values for \( \lambda_2 \): 0.9, 0.5 and 0.3.

Lasso and EN are implemented in the R package \textit{glmnet} (Friedman et al. 2010).

Zou (2006) proposed the Adaptive Lasso to obtain an oracle procedure, which has the penalty

\[ P_{AdLasso}(\lambda, \tilde{\beta}) = \lambda \sum_{k=1}^{p} w_k |\beta_k|; \quad \lambda \geq 0, \ w_k = |\tilde{\beta}_k|^{-1} \]  

[6]

where the weights \( w_k \) are computed from an initial set of solutions for the coefficients.

The Adaptive Lasso makes sense intuitively by placing smaller weights on the important regressors to reduce the shrinkage applied to their coefficients. Here we used SMA, Lasso with CV and Ridge Regression with CV as the initial estimators.

\textbf{MCP and SCAD}

The MCP and SCAD penalties depend on two tuning parameters; both begin with the same penalization as the Lasso but increasingly reduce the penalization further away from zero (\textit{e.g.}, Breheny and Huang (2011)). Because in preliminary studies we found little difference between the MCP and SCAD results and SCAD is expected to perform somewhere between MCP and Lasso (Breheny and Huang 2011), here we only consider the MCP with penalty function
\[ P_{MCP}(\tilde{\lambda}, \beta_k) = \lambda_1 \int_0^{\lambda_1} \left[ 1 - \frac{x}{\lambda_1 \lambda_2} \right]_+ dx = \lambda_1 \left\{ \begin{array}{ll} \left( |\beta_k| - \frac{\beta_k^2}{2 \lambda_1 \lambda_2} \right) & \text{if } |\beta_k| \leq \lambda_1 \lambda_2 \\
\frac{\lambda_1 \lambda_2}{2} & \text{if } |\beta_k| > \lambda_1 \lambda_2 \end{array} \right. \] \tag{7}

For the second tuning parameter in MCP ($\lambda_2$), four values were evaluated (3, 10, 30, 100), with the largest value approaching the Lasso. MCP is implemented in the R package `ncvreg` (Breheny and Huang 2011).

**Selection of Tuning Parameter Values: Current State**

A major practical issue with the use of penalized regression for GWAS is the need to determine “optimal” values for the tuning parameter(s). The “usual” approaches include CV and the use of model selection criteria such as AIC, BIC and EBIC. While CV and AIC are generally useful for prediction rather than model/variable selection, BIC tends to select a true sparse model but does not achieve sufficient sparsity in high-dimensional very sparse settings such as linkage (QTL) mapping and GWAS for which modified BIC criteria have been proposed (e.g., Bogdan et al. (2004)). Moreover, the BIC criterion is a function of the degrees of freedom (df) of a model, and the df are not straightforward to obtain in PR (e.g., Ye (1998), Zou et al. (2007) and Zhang (2010)).

Given the shortcomings of model selection criteria, a superior approach should be obtained by combining PR with the control of an error rate such as FWER or FDR. Ayers and Cordell (2010) used data permutation to determine the value of the tuning parameter. Multi-stage strategies have also been suggested as an approach to performing error control with PR (Meinshausen et al. 2009; Wasserman and Roeder 2009). Meinshausen et al. randomly split the data in half multiple (B) times. Each time
the first half was used for variable screening using Lasso with CV and related methods, and the second half was used to compute adjusted p-values, based on which the FWER and the FDR can be controlled asymptotically. We used Lasso with CV for tuning parameter value selection in the screening step. We expected this method to have reduced power relative to our single step FDR controlling methods (see below) due to the data splitting and the use of multiple regression which reduces the significance of two (highly) correlated variables representing alternative hypotheses. Performing the two steps of screening and variable selection both on the entire dataset as in Sun et al. (2010) should improve power; these authors used the IAL in the screening step and backwards selection instead of multiple regression in the second (cleaning) step. However, there is no analytical proof of (asymptotic) FDR (or FWER) control, and we have therefore not further considered a two-step approach performed on the entire dataset.

Most recently, Sampson et al. (2013) have proposed a local FDR based method for selecting the value of the tuning parameter specifically for the Adaptive Lasso. The advantage of this approach is that it uses the local FDR based on which optimal oracle procedures were developed for SMA (Cai and Sun 2009; Sun and Cai 2007).

The PUMA (Penalized Unified Multiple-locus Association) method and software implements some penalized regression methods for GWAS with an efficient algorithm for Generalized Linear Models and with pre-selection of SNPs based on SMA (Hoffman et al. 2013) to make these methods applicable to large GWAS datasets. It performs heuristic model and SNP selection. PUMA implements MCP with a two-dimensional
model search across its two tuning parameters (2D-MCP), and we chose this method here.

A different type of multivariate method for SNP selection in GWAS was proposed by Zuber et al. (2012). For a continuous phenotype, this method computes correlation-adjusted marginal correlations (CAR scores) as $p_{XY}^{adj} = p_{XX}^{-1/2} p_{XY}$, where $p_{XY}$ is a vector of marginal correlations between each SNP dose (X) and the phenotype (Y), and $p_{XX}$ is the correlation matrix among the SNPs, which is estimated by a computationally efficient shrinkage method. Note that if there are non-SNP study covariates, the phenotype (Y) must be replaced with the residuals from a linear model containing these covariates. For SNP selection, CAR scores are provided as input to the fdrtools R package which computes p-values, FDR and local FDR values based on its empirical null modeling (Strimmer 2008). We refer to this method as CAR.

**Selection of Tuning Parameter Values: Control of the False Discovery Rate**

Here we propose two approaches for combining penalized regression with FDR control, which can be applied to all penalized regression methods considered here and do not require data splitting or sub-sampling. The first approach is based on data permutation and the second on an analytic approximation.

*FDR control using data permutation*

Let $\lambda$ denote the single (Lasso) or the first (MCP, EN) tuning parameter. Penalized regression (with coordinate descent algorithms, Friedman et al., 2007) is efficiently performed by computing the coefficient solutions on a grid of (say 1000) $\lambda$ values,
starting from a maximum value $\lambda_{\text{max}}$ at which all solutions are zero and ending at a minimum value $\lambda_{\text{min}}$ which is zero or produces an excessively large model, with the solutions from any previous $\lambda$ value serving as a warm start for the next $\lambda$. Therefore, the original data set is analyzed on this grid of $\lambda$ values, and the number of non-zero coefficients is determined at each $\lambda$, denoted by $R(\lambda)$. Then $B$ permuted datasets (with random re-orderings of the continuous vector of phenotypes) are analyzed on the same grid of $\lambda$ values, and the number of non-zero coefficients is again determined for each permuted dataset $b$ ($b = 1, \ldots, B$) and $\lambda$ value, denoted by $F(b, \lambda)$. Then for each $\lambda$ value, we compute the permutation FDR as the average number of nonzero solutions in the permuted data over the number of non-zero solutions in the original data,

$$FDR(\lambda) = \frac{1}{B} \sum_{b=1}^{B} F(b, \lambda) / R(\lambda).$$

We note that here we define the FDR as $E(F) / R$, where $F$ denotes the number of false positives and $R$ the number of rejections (recall that in PR a null hypothesis is rejected if the corresponding coefficient solution is nonzero). This definition is different from the original definition $E(F/R)$ and thus does not take into account the dependency between $F$ and $R$, but is easier to work with.

**Analytic FDR control**

We now present an approximate, analytic FDR method which was originally proposed in Breheny (2009). This approach can be applied to all penalized regression methods considered here, but we illustrate it for the Lasso and the MCP. We assume that the predictors have been standardized such that $\sum_{i=1}^{N} x_{ik} = 0$ and $\sum_{i=1}^{N} x_{ik}^2 = N$. We write the
objective function in terms of a single predictor \((k)\), as in Friedman et al. (2007) for the Lasso, and in Breheny and Huang (2011) for the MCP (setting \(\lambda = \lambda_1\) and \(\gamma = \lambda_2\) in [7]).

\[
f_{\text{Lasso}}(\beta_k) = \frac{1}{2}(\beta_k - \beta_k^{\text{OLs}})^2 + \lambda |\beta_k|; \quad f_{\text{MCP}}(\beta_k) = \frac{1}{2}(\beta_k - \beta_k^{\text{OLs}})^2 + \lambda |\beta_k| - \frac{1}{2\gamma} \beta_k^2 \tag{8}
\]

where \(\beta_k^{\text{OLs}} = \frac{1}{N} \sum_{i=1}^{N} x_{ik} r_{i-k}\) and \(r_{i-k} = y_i - \sum_{j=1,j\neq k}^{p} x_{ij} \beta_j\). Differentiating [8] with respect to \(\beta_k\) yields the solutions

\[
\beta_k^{\text{Lasso}} = \begin{cases} 
\beta_k^{\text{OLs}} - \lambda & \text{if } \beta_k^{\text{OLs}} > 0 \text{ and } |\beta_k^{\text{OLs}}| > \lambda \\
\beta_k^{\text{OLs}} + \lambda & \text{if } \beta_k^{\text{OLs}} < 0 \text{ and } |\beta_k^{\text{OLs}}| > \lambda \\
0 & \text{if } |\beta_k^{\text{OLs}}| \leq \lambda
\end{cases}
\]

\[
\beta_k^{\text{MCP}} = \begin{cases} 
\frac{\beta_k^{\text{OLs}} - \lambda}{1 - \frac{1}{\gamma}} & \text{if } \beta_k^{\text{OLs}} > 0 \text{ and } |\beta_k^{\text{OLs}}| > \lambda \\
\frac{\beta_k^{\text{OLs}} + \lambda}{1 - \frac{1}{\gamma}} & \text{if } \beta_k^{\text{OLs}} < 0 \text{ and } |\beta_k^{\text{OLs}}| > \lambda \\
0 & \text{if } |\beta_k^{\text{OLs}}| \leq \lambda \\
\beta_k^{\text{OLs}} & \text{if } |\beta_k^{\text{OLs}}| \geq \gamma \lambda
\end{cases} \tag{9}
\]

Based on [9], we consider the probability of a false positive, or \(Pr(|\beta_k^{\text{OLs}}| > \lambda | \beta_k = 0)\).

We again define the FDR as \(E(F)/R\) and approximate the numerator with

\[
\hat{E}(F) = \sum_{k=1}^{p} Pr(|\beta_k^{\text{OLs}}| > \lambda | \beta_k = 0) \tag{10}
\]

In general, FDR control based on [10] is expected to be conservative similar to the BH FDR, because in [10] we sum over all variables rather than the (unknown) true null variables. However, this should not matter in the context of GWAS where the proportion of null variables is very close to 1.
We rewrite the probability of a false positive as

$$Pr\left( |\beta_k^{OLS} | > \lambda | \beta_k = 0 \right) = Pr\left( \frac{1}{N} |x_k^T r_{-k} | > \lambda | \beta_k = 0 \right) \quad [11]$$

The distribution of the estimated residuals ($r$) is unknown and complicated but may be approximated by the normal distribution $r_{-k} \sim N(0, V_k)$ or even simpler by $r_{-k} \sim N(0, I \sigma_k^2)$ such that $\frac{1}{N} x_k^T r_{-k} \sim N \left( 0, \frac{1}{N} \sigma_k^2 \right)$. This is an approximation for multiple reasons including the normality assumption, the mean of zero (correct for least squares where the coefficients are estimated unbiasedly but not for PR), and the independence or zero covariances. Then

$$Pr\left( \frac{1}{N} |x_k^T r_{-k} | > \lambda | \beta_k = 0 \right) = 1 - \Phi \left( \frac{\sqrt{N} \lambda}{\sigma_k} \right) + \Phi \left( -\frac{\sqrt{N} \lambda}{\sigma_k} \right) = 2 \Phi \left( -\frac{\sqrt{N} \lambda}{\sigma_k} \right) \quad [12]$$

Using the “natural” estimator $\hat{a}_k^2 = \frac{1}{N} r^T r$, [12] becomes constant for all $k$ and

$$\bar{FDR} = 2p \Phi \left( -\frac{\lambda N}{\sqrt{r^T r}} \right)/R \quad [13]$$

The FDR estimation [13] is noisy when the model becomes saturated; this is not a problem in practice when FDR evaluation starts from a model with all coefficients set to zero ($\lambda_{max}$) and terminates once the desired level of the FDR has been stably exceeded.

When applying the FDR estimation described for the Lasso and the MCP to the EN, equation [13] needs to be modified to

$$\bar{FDR} = 2p \Phi \left( -\frac{\lambda_1 \lambda_2 N}{\sqrt{r^T r}} \right)/R \quad [14]$$
Penalized Regression With Fusion-Type Penalties

It is well-known that the Lasso tends to select a single predictor from a group of predictors which have strong pair-wise correlations (in the extreme the Lasso does not have a unique solution when two variables are perfectly collinear). For this reason, we prefer the EN (see below), but some authors (Kim and Xing 2009; Liu et al. 2011) have suggested that Fused Lasso-type penalties developed for covariates measured on graphs would be more appropriate than the Lasso in such situations. An expectation was that such penalties may increase power and decrease the false positive rate.

Fusion-type penalties impose pairwise similarity on the coefficient solutions of highly correlated predictors or encourage sparsity in the differences among the values of the regression coefficients. Pairwise similarity can be imposed on the effects of correlated (SNP) regressors by adding an extra penalty to the existing penalty term depending on the type of PR method used (here the Lasso). We first consider an added penalty of the following form (referred to as LD2lasso):

\[
\frac{\varphi}{2} \sum_{k=1}^{p} \sum_{m \in S_k} h(r_{km}) \cdot (|\beta_k| - |\beta_m|)^2; \quad \varphi \geq 0
\]

where \( S_k \) is a set of SNPs which are correlated to SNP \( k \), \( r_{km} \) denotes the (Pearson) correlation coefficient among the allelic doses of SNPs \( k \) and \( m \), and \( h(r_{km}) \) is a function of \( r_{km} \) which we specified as \( h(r_{km}) = |r_{km}| \) or \( h(r_{km}) = (r_{km})^2 \). Set \( S_k \) can be limited to adjacent SNPs (Liu et al. 2011), or contain all SNPs whose absolute correlation with SNP \( k \) exceeds a certain threshold (here). Taking absolute values of the coefficients accounts for the fact that two correlated SNPs can have similar effect sizes but opposite
signs due to the arbitrary coding of the SNP alleles. We note that Li and Li (2008; 2010) used a similar penalty for the general case of (genomics) variables whose dependency structure can be represented as a graph. Their penalty is induced by the Laplace matrix of the graph and differs from [15] by dividing the regression coefficients by the square root of the degree of the corresponding variables, which is justified by the argument that variables with more connections should be allowed to have larger coefficients. This argument does not apply in the GWAS context, and we therefore use [15]. Using [3] with the Lasso penalty and the added penalty in [15], the LD2lasso objective function for a single coefficient \( k \) can be written as

\[
 f_{LD2lasso}(\beta_k) = \frac{1}{2} (\beta_k - \beta_k^{PLS})^2 + \lambda \phi |\beta_k| + \lambda (1 - \phi) \frac{1}{2} \sum_{m \in S_k} h(r_{km}) (|\beta_k| - |\beta_m|)^2; \lambda \geq 0, 0 \leq \phi \leq 1 \quad [16]
\]

where

\[
 \beta_k^{PLS} = \frac{1}{N} x_k^T (y_i - \sum_{j=1, j \neq k}^p x_{ij} \beta_j^{LD2lasso})
\]

Then one can show that the LD2lasso solution satisfies

\[
 \beta_k^{LD2lasso} = \text{sign}(\beta_k^{PLS}) \frac{\left| \beta_k^{PLS} - \lambda \left( \phi - (1 - \phi) \sum_{m \in S_k} h(r_{km}) |\beta_m^{LD2lasso}| \right) \right|}{1 + \lambda (1 - \phi) \sum_{m \in S_k} h(r_{km})} \quad [17]
\]

For LD2lasso with analytic FDR control, the probability of a false positive analogous to [12] is
\[
Pr \left( \frac{1}{N} |x_k^T r_{-k}| > \lambda \left( \varphi - (1 - \varphi) \sum_{m \in S_k} h(r_{km}) |\beta_m^{LD2lasso}| \right) | \beta_k = 0 \right)
\]

\[
= 2 \Phi \left( -\frac{\sqrt{N} \kappa}{\sigma_k} \right); \quad \kappa = \lambda \left( \varphi - \lambda (1 - \varphi) \sum_{m \in S_k} h(r_{km}) |\beta_m^{LD2lasso}| \right)
\]  

[18]

where in contrast with [12] the threshold (\(\kappa\)) applied to \(\beta_k^{PLS}\) is now a function of the LD2lasso solutions of other SNPs in strong LD with the current SNP.

An alternative to [15] is a penalty which we refer to as the LD1lasso penalty and which is essentially identical to a penalty term in the General and Graphical Fused Lasso (e.g., Kim and Xing (2009)):

\[
\varphi \sum_{k=1}^p \sum_{m \in S_k} h(r_{km}) \cdot |\beta_k| - |\beta_m|; \quad \varphi \geq 0
\]  

[19]

Computation of the LD2lasso solution ([17]) was implemented with a straightforward path-wise coordinate descent algorithm, because its objective function consist of a convex, continuously differentiable and a separable term (of convex functions) (Friedman \textit{et al}. 2007, Tseng 1988, Tseng, 2001, Chen \textit{et al}. 2010). This does not apply to the LD1lasso, and hence we focused on the LD2lasso which seemed to perform better (results not shown).

Consider a group of SNPs in strong LD and containing a single SNP which is causal for the phenotype. In this situation, the Lasso (and other penalties not including fusion-type penalties) will tend to select a single or few SNPs, while LD2lasso/LD1lasso would have nonzero solutions for several or all SNPs in this group. Hence for PR without fusion-type penalties, a QTL region for subsequent fine-mapping may be
identified as the SNP(s) with non-zero solution plus other SNPs in LD with these SNP(s) above some threshold (e.g., $|r| > 0.5$). For LD2lasso/LD1lasso, a QTL region may be identified as including all (contiguous) SNPs with nonzero coefficients.

**Analyses of Multiple Chromosomes**

The analysis of multiple chromosomes may be regarded as a multi-group analysis, which we first discuss in the context of SMA. For a given biomedical trait of interest, many chromosomes may show no signal, while other chromosomes may show sparse signals or even extended stronger signals. To control the GWER such grouping structure is irrelevant. However, for FDR control there is no unique way to proceed in the presence of a group structure. As argued by Efron (2008), the usual pooled FDR analysis, where the group structure is simply ignored, can be overly conservative or overly liberal for any particular group. Efron also established that a separate analysis, where FDR control (at the same level) is performed separately for each group and subsequently the results are simply combined, is valid in terms of achieving overall (i.e. across groups) FDR control (at that same level). Efron’s concern about joint FDR analysis of all groups (here chromosomes) within an experiment or study applies directly to SMA with FDR control in GWAS. While this also applies to PR with FDR control, here a separate analysis of each chromosome implies that QTL on other chromosomes are not included in the PR multi-marker model, potentially lowering detection power and diminishing the postulated advantage of an all-markers over single marker analysis. Lastly, for most GWAS, one expects that the proportions of true null hypotheses are close to 1 and very similar across groups (chromosomes), so there may
be little difference between joint and separate analyses and group-based analyses described below.

**SMA**

Besides these two basic strategies of pooled and separate analyses, some specialized methods that take the group structure into account have recently been proposed in the context of univariate analysis (SMA) (Cai and Sun 2009; Hu *et al.* 2010). Cai and Sun estimated the local FDR within each group and then applied the (approximate oracle) thresholding procedure to the combined set of local FDR values. For SMA of multiple chromosomes, we compared (i) pooled analysis based on the BH method or the local FDR thresholding procedure, (ii) separate analysis based on BH or the local FDR procedure, and (iii) the local FDR based grouping procedure of Cai and Sun (2009).

Sun and Cai (2009) extended their earlier method to dependent test statistics using a Hidden Markov Model (HMM). Wei *et al.* (2009) combined this method with the group-based method of Cai and Sun (2009) in the context of GWAS and implemented it in the *R* package *PLIS*. We tested this method on the single chromosome 21 simulation with 8 true QTLs, where it produced a very high number of false positives (irrespective of whether the alternative component was fit with a mixture or a single normal distribution) and hence we did not further pursue the *PLIS* method.

**Penalized Regression**

For PR with FDR control, we also compared separate and pooled analyses. For the pooled analysis, the SNPs from all three chromosomes were fitted jointly with the Lasso penalty, and the penalty parameter value was chosen based on controlling the FDR at level 0.05. For the separate analysis, the Lasso analysis was performed separately for
each chromosome by omitting the SNPs on the other chromosomes and choosing a separate penalty parameter value to control the FDR at level 0.05 for each chromosome.

**Comparison Among Methods**

We define a Causal True Positive (CTP) as a QTL (defined earlier as a functional or causal SNP affecting the phenotype) which is significant (significant means identified by any of the methods considered). We define a Linked True Positive (LTP) as a SNP which is significant and has an absolute correlation with a QTL above a certain threshold (T) (note that T is a threshold on the absolute correlation coefficient, not the squared coefficient). We define a CLTP as being either a CTP or a LTP. Methods are compared, across the replicate simulations, based on (1) True Positive Rate $\text{TPR}_1 = \frac{\text{number of CTPs}}{\text{number of QTLs}}$, (2) $\text{TPR}_2 = \frac{\text{number of QTLs detected by at least one CLTP (CTP or LTP)}}{\text{number of QTLs}}$, and (3) threshold-specific empirical tFDR = $\frac{\text{number of significant SNPs which are not a CLTP at a given T}}{\text{number of significant SNPs}}$. For criterion (2), we used $T = 0.9, 0.7$ and 0.5 for an increasingly relaxed definition of TP. For criterion (3), we used $T = 0.5, 0.3$ and 0.25. The first value ($T = 0.5$) was chosen because our simulation was designed such that a QTL had a power of 0.5 to be detected, computed using Quanto for a given heritability, sample size and chromosome-wide p-value threshold, and this power decreased to $< 0.01$ for a SNP correlated with the QTL at 0.5 (assuming that SNP and QTL have the same allele frequency). The second value ($T = 0.3$) is just below the threshold of “useful LD” ($r = 0.316$ or $r^2 = 0.1$) as defined by Kruglyak (1999) and Pritchard and Przeworski (2001) as
the value at which the sample size is increased at most 10-fold. The last value (T = 0.25) is close to the threshold used in a previous comparison study (Ayers and Cordell 2010). For completeness, we also provide tFDR values at the remaining thresholds used for TPR2 (0.9, 0.7).

It is not straightforward to evaluate the empirical FDR in GWAS simulations due to LD. We experimented with alternative ways of defining the tFDR, such as partitioning a chromosome into 100kb windows and computing tFDR (for a given T) as the number of windows in which the significant SNPs are all false positives divided by the number of windows that contain at least one significant SNP. Such an alternative approach did not change any results significantly, so we report tFDR for the first definition.

The results are presented in figures while more detailed results in the form of tables are available in Supplement A; the tables include standard errors for all average (across replicates) tFDR, TPR1 and TPR2 values which were used for computing p-values for differences in these values between methods or for a one-sided test of average tFDR to exceed level 0.05.

RESULTS AND DISCUSSION

Single chromosome simulation with two or eight QTLs: Single marker analysis

For two isolated QTLs, SMA results are summarized in Figure 1 (and Table S1 in Supplement A). For SMA with GWER control (referred to as Genome-Wide Threshold GWT in Figure 1) which controlled the FWER conservatively, the TPR was low (0.2) as expected. For SMA with FDR control, BH and BY controlled the FDR at level 0.05
based on two (three) of the three empirical threshold-specific FDRs (tFDRs), while the local FDR based approximate oracle procedures LFDR1 and LFDR2 had tFDR values all above 0.05 although not significantly (at nominal p-value of 0.05). The TPR was somewhat higher (0.31) for BY compared with the GWT, while it was substantially higher for BH, LFDR1 and LFDR2 (above 0.5).

(Figure 1 about here)

For the eight-QTL scenario (Figure 2 and Table S2 in Supplement A), TPR was increased for the FDR based SMA methods relative to the two-QTL case (0.7 and above for BH, LFDR1 and LFDR2; around 0.6 for BY compared with 0.4 or below for GWT). All FDR methods controlled the FDR in the sense of producing at least one tFDR value which was not significantly higher than 0.05 (nominal p-value). The smallest p-value for exceeding level 0.05 by a tFDR value at T=0.3 was p=0.034 for SMA-BH.

(Figure 2 about here)

**Single chromosome simulation with two isolated QTLs: PR with FDR control**

We verified the expectation that Lasso with standard tuning parameter value selection based on CV produces an unacceptably high empirical FDR (above 0.5; results not shown). The results for PR with FDR control for this scenario are presented in Figure 3 and Tables S3 and S4 in Supplement A. Differences in empirical tFDR and TPR among all PR methods with analytic FDR control were small; there was a very slight advantage for the Elastic Net with the mixing parameter set to 0.5 (EN50); Lasso and MCP with different values for the second tuning parameter (from $\lambda_2=3$ to $\lambda_2=100$) produced almost
identical results. All PR methods were able to control the FDR by producing tFDR values below 0.05 for all T at or below 0.5. In terms of power and using the EN50 to represent PR, in comparison with SMA-BH, the average TPR1 and TPR2 values of EN50 were lower with p-values for the differences ranging from 0.004 to 0.013. Comparing the EN50 with SMA-BY, the average TPR1 and TPR2 values of EN50 were higher with all p-values around 0.002. The Lasso with permutation FDR control had slightly higher tFDR values compared with the Lasso with analytic FDR control; TPR1 and TPR2 of the permutation Lasso was similar to that of SMA-BH and higher than that of the Lasso with analytic FDR control (smallest p-value for the difference equal to 0.01). The results for Lasso with permutation FDR were stable across different sets of permutations (100 to 400 permutations).

(Figure 3 about here)

**Single chromosome simulation with eight QTLs: PR with FDR control**

The results for this scenario are presented in Figure 4 and Tables S5 and S6 in Supplement A. All PR methods were able to control the FDR by producing tFDR values below (or not significantly above) 0.05 for all T at or below 0.5. Most of the differences in empirical tFDR and TPR among different PR methods with analytic FDR control remained small but some were more pronounced for the eight-QTL data relative to the two-QTL data: EN50 had a higher TRP1 and TRP2 than the Lasso, with a p-value of $2 \times 10^{-5}$ for the difference in TPR1 and $p = 0.001$ for TPR2 at $T = 0.9$; MCP with low values for the second tuning parameter ($\lambda_2 = 3, 10$) had lower TPR1 (p-values $1.3 \times 10^{-14}$ and 0.003, respectively) and TPR2 (p-values $2 \times 10^{-15}$ to 0.03 and 0.003 to 0.33,
respectively, with the upper end of each range representing TPR2 at $T = 0.5$) than the Lasso. Power was expectedly higher relative to the 2-QTL scenario (for EN50, TPR2 at $T = 0.5$ increased from 0.43 to 0.73). Comparing the EN50 with SMA-BH, the average TPR1 and TPR2 values of EN50 were lower with p-values for the differences ranging from $5 \times 10^{-22}$ to 0.001. Comparing the EN50 with SMA-BY, EN50 had a significantly higher value for TPR2 at $T = 0.5$ (p-value 0.0005) while EN50 had a lower value for TPR1 (p-value 0.005). In general, differences among PR methods were largest for TPR1 and TPR2 at $T = 0.9$ and were diminished for TPR2 with $T = 0.5$. Comparing Lasso with permutation versus analytic FDR estimation, the permutation Lasso had lower TPR1 and TPR2 with p-values for the differences at or below 0.0006. It is surprising that the permutation method, which was the most powerful approach in the 2-QTL scenario, is so conservative in the 8-QTL scenario. This appears to be due to the fact that permuting the phenotypes, while justified under the global null hypothesis of no association between any SNPs and the phenotypes, is conservative when evaluating the significance of a QTL in the presence of substantial signal from other QTLs. By permuting the phenotype, which consists of both fixed signal and random error, the contribution of the random errors is overestimated. This does not affect the validity of the permutation approach in terms of properly controlling the FDR but does diminish power in cases like the 8-QTL scenario.

(Figure 4 about here)

The eight-QTL simulation included two groups of two QTLs each with within-group $r^2 \approx 0.5$ (absolute correlation among SNP doses of 0.71). Because of the Lasso’s
tendency to only select a single SNP per group and the EN’s ability to deal with co-
linearity and potential co-selection of correlated SNPs (SNPs in LD), we took a closer 
look at the behavior of Lasso and EN in this situation. Averaged over both groups and 
the 200 replicate simulations and including LTPs (T = 0.5), the Lasso selected 0 SNPs 
in 22%, 1 SNP in 43%, 2 SNPs in 26% and >2 SNPs in 9% of cases. By comparison, 
the EN selected 0 SNPs in 19%, 1 SNP in 37%, 2 SNPs in 31% and >2 SNPs in 13% of 
cases. Expectedly the EN selected 2 and >2 SNPs more frequently but the difference 
was not substantial.

**Single chromosome simulation with two/eight QTLs: Fusion-Type Penalties**

Here we focused on the LD2lasso with \( \phi \) (the relative weight given to the Lasso penalty) 
fixed at two different values (0.9, 0.5), with the two LD functions \( h(r) = |r| \) and \( h(r) = r^2 \), 
and with \( h(r) = 0 \) if \( |r| < 0.85 \) (for this threshold, the median number of SNP ‘neighbors’ 
was 2 with interquartile range (IQR) 0 to 5 and range 0 to 39) or \( |r| < 0.50 \) (median 
number of SNP ‘neighbors’ equal to 14 with IQR 6 to 25 and range 0 to 102). Because 
we did not optimize the implementation of the LD2lasso for computational speed, we 
compared the LD2lasso with the Lasso and EN50 based on only 20 data replicates 
which were randomly selected from the 200 replicates. For the chromosome 21 data 
with two isolated QTLs, the LD2lasso achieved higher power than Lasso and EN50 
while maintaining FDR control (Figure 5 and Table S7 in Supplement A). The largest 
difference was the increase in TPR1 from 0.35 for the Lasso to 0.5 for the best 
LD2lasso analysis with \( \phi = 0.5 \) and \( h(r) = r^2 \) (p-value < 0.05), while all other differences 
in TPR1 and TPR2 were smaller.
For the chromosome 21 data with eight QTLs (Figure 6 and Table S8 in Supplement A), however, all LD2lasso variants were unable to control the FDR with average tFDR values at the lowest threshold of \( T = 0.25 \) near 0.2 (p-value for the differences to FDR level 0.05 below \( 4 \times 10^{-4} \)). We recall that the eight-QTL data contained a group of four QTLs with weak pairwise LD (\( 0.01 \leq r^2 \leq 0.1 \)); the high tFDR values (0.2 to 0.26) for LD2lasso were mostly due to SNPs in LD with the group of four causal SNPs at small absolute correlation values. In terms of TPR, the difference in TPR1 between the best LD2lasso analysis with \( \varphi = 0.5, h(r) = r^2 \) and \( |r| < 0.85 \) (\( |r| < 0.50 \)) and the Lasso had a p-value of \( 5.4 \times 10^{-5} \) \( (1.7 \times 10^{-5}) \), while the difference in TPR1 between the best LD2lasso and the EN50 had a larger p-value of 0.10 (0.008). TPR2 values were also higher, but most p-values for the differences in TPR2 between the same methods were above 0.05 due to smaller differences (less than half) and the small number of replicates.

Single chromosome simulation with eight QTLs: PR versus SMA and CAR

Because all previous comparisons focused on differences among methods in empirical TPR and tFDR for a single FDR cut-off of 0.05, here we present figures depicting TPR1 versus tFDR at \( T = 0.25 \) and \( T = 0.50 \) (Figure 7) and TPR2 (\( T = 0.50 \)) versus tFDR at \( T = 0.25 \) and \( T = 0.50 \) (Figure 8). These figures were generated by varying the p-value cut-off for SMA and CAR and varying the penalty parameter for PR. In terms of power to
detect causal SNPs (TPR1) and when using the most relaxed definition of tFDR (with T = 0.25), SMA dominated all other methods with CAR being second followed by LD2lasso and EN50, with Lasso and MCP ($\lambda_2 = 10$) performing worst (Figure 7). When using a more stringent definition of false positives with tFDR at T=0.50, SMA/CAR dominated in terms of TPR1 only for tFDR values above 0.05. (Figure 7 about here)

In terms of power to detect QTLs with causal or linked SNPs (TPR2 at T=0.5) (Figure 8) and when using the most relaxed definition of tFDR (with T = 0.25), EN50 and for higher tFDR(T=0.25) values SMA slightly dominated the other methods, with the LD2lasso now performing worst for the most relevant range of tFDR(T=0.25) from 0.01 to 0.10. When replacing tFDR(T=0.25) with tFDR(T=0.50), EN50 performed best (slightly better than Lasso and MCP), with SMA, CAR and LD2lasso now separated as the worst performers. (Figure 8 about here)

**Single chromosome simulation with two/eight QTLs: Previous approaches**

The multi-split analysis of Meinshausen *et al.* (2009) was evaluated based on a random subset of 20 replicates of the chromosome 21 2-QTL data and on another random subset of 20 replicates of the chromosome 21 8-QTL data (Table S9 in Supplement A). It controlled the FDR quite conservatively (all tFDR values equal to zero) and had low power relative to the single-step PR methods with FDR control, as expected. Power (TPR) was below 0.18 for the two-QTL data and below 0.21 for the eight-QTL data. P-
values for the differences in TPR between the multi-split analysis and the Lasso ranged from 0.01 to 0.04 for the two-QTL data and from $2.4 \times 10^{-7}$ to $4 \times 10^{-4}$ for the eight-QTL data.

The Adaptive Lasso with local FDR control of Sampson et al. (2013) was evaluated using all 200 replicates of the chromosome 21 two-QTL and eight-QTL data (Table S10 in Supplement A). R code for executing this method was generously provided to us by the first author. For our data, the local FDR Adaptive Lasso controlled the FDR at the local FDR threshold of 0.1 (tFDR values below 0.05), but its power (TPR) was much lower than that of the Lasso for both the two-QTL and the eight-QTL data (p-values for the differences in TPR to the Lasso ranged from $1 \times 10^{-12}$ to $2 \times 10^{-13}$ and from $2.4 \times 10^{-67}$ to $2.5 \times 10^{-83}$, respectively). When raising the local FDR threshold to 0.5, the FDR was not controlled for the two-QTL data (tFDR values highly significantly above 0.05) but was controlled for the eight-QTL data (tFDR values below 0.05) where the TPR values however remained significantly below the values of the Lasso.

PUMA was run on the eight-QTL data with 200 replicates, choosing the 2D-MCP method, omitting the SNP pre-selection step and otherwise using default parameter values. The results are in Table S11 (Supplement A) for four different p-value thresholds for SNP selection ranging from $1 \times 10^{-07}$ to 0.05, and these results need to be compared with our MCP with analytic FDR control (and four fixed values of the second tuning parameter $\lambda_2$ in [9]) in Table S6. The authors recommended a p-value threshold of $1 \times 10^{-07}$ for 2D-MCP, which controlled the FDR in terms of the tFDR values in Table S11; the next lower threshold of $1 \times 10^{-06}$ had tFDR values above 0.06 whose one-sided test for
exceeding level 0.05 had p-values from 0.055 to 0.11. Overall, TPR1 and TPR2 values for 2D-MCP were significantly below those for MCP with analytic FDR control (p-values for differences from $9 \times 10^{-05}$ to $8 \times 10^{-29}$ for 2D-MCP SNP selection threshold $1 \times 10^{-07}$, and from 0.032 to $2 \times 10^{-22}$ for 2D-MCP SNP selection threshold $1 \times 10^{-06}$). To explain these differences, we looked at the values for $\lambda_2$ for the models chosen in PUMA by AIC over the 200 datasets, which surprisingly had a median of 3.46, with almost all values well below 10. For our own MCP analysis with FDR control we had chosen four values for the second tuning parameter (3, 10, 30 and 100), of which the lowest value performed the worst. We then replaced the heuristic AIC model selection followed by SNP selection of PUMA with our analytic FDR tuning parameter selection, which significantly improved TPR1 and TPR2 (last column of Table S11). The analytic FDR criterion selected models with larger values for $\lambda_2$ with a median of 33.6 (as well as larger values for $\lambda_1$, median 0.166 versus 0.310) This improvement still underperformed the results in Table S6 because PUMA uses an (internally determined) grid of values for $\lambda_1$ and $\lambda_2$ in 2D-MCP, which was too sparse for optimal performance of the analytic FDR approach.

Analyses of multiple chromosomes

Here we analyzed the simulated data including chromosomes 19 (no QTL), 21 (8 QTLs) and 22 (2 QTLs) with 200 replicates. For single marker analysis (SMA), the average tFDR values for pooled and separate BH analyses and for the local FDR grouping procedure were all above 0.05 for $T = 0.25$ and 0.3 but most were insignificantly higher (largest p-value of 0.001 for separate BH). Differences in TPR between SMA based pooled and separate analyses and the grouping procedure of Cai and Sun (2009) were
small and insignificant, although there was a trend for increasing TPR from pooled analysis to separate analysis to the grouping procedure (see Table S12 in Supplement A).

For penalized regression using Lasso PR with analytic FDR control, pooled and separate analyses controlled the FDR in terms of their average tFDR values, and differences in TPR between the pooled and the separate Lasso were small and insignificant (Table S13 in Supplement A). We therefore compared the separate and pooled analyses for individual chromosomes. From Table S13, we can see that for chromosome 21 (with eight QTLs), TPR1 and TPR2 decreased somewhat from separate to pooled analysis with p-values for the differences ranging from 0.02 to 0.05. For chromosome 22 (with only two QTLs), in contrast, TPR1 and TPR2 increased significantly as expected from separate to pooled analysis with all p-values below 4×10^{-10}.

**Analysis of Real Data**

We applied PR and SMA methods to data from the Health ABC GWAS of interleukin 6 soluble receptor (IL-6 SR). The analysis included 786 Health ABC Caucasians who had both GWAS and serum IL-6 SR measurements (a continuous phenotype, approximately normally distributed) available (see Supplement B for details). The genotype data included 750,424 SNPs (after standard edits) from chromosomes 1, 3, 4, 6 and 19. Covariates included in the analysis models were age, gender, site of data collection, and one principal component score obtained using Eigenstrat. We analyzed these data by SMA with GWER threshold (SMA-GWT), SMA with BH FDR control (SMA-BH), and
with the Elastic Net with 50% Lasso penalty (EN50) and analytic FDR control. SMA-GWT is commonly used in practice, and SMA-BH performed best among all SMA methods with FDR control on the simulated data. The Elastic Net was chosen as the single PR method because it performed best on the simulated data among all PR methods combined with our novel method for FDR control, and because differences among PR methods were generally small. For SMA-BH and EN50 (EN50-sep), chromosomes were analyzed separately, and for EN50, a joint analysis of all chromosomes was also performed (EN50-joint).

Chromosome 1 represents a special situation in that it contains a region with an extremely strong QTL signal. The very stringent SMA-GWT threshold identified 74 SNPs for this chromosome, many more than the 15 EN50-sep SNPs (see Table 1). Of these 74 SNPs, 61 were in LD above the absolute correlation threshold of 0.25 with EN50-sep SNPs, and the remaining 13 SNPs had largest absolute correlations with EN50-sep SNPs between 0.20 and 0.24 and therefore likely still overlap with EN50-sep SNPs (given the very strong signal) rather than represent independent signals. The EN50-sep selected a small subset of the peak area SNPs which had the smallest SMA p-values (Figure 9). While the SMA-GWT SNPs were all located in the area of the major peak, several EN50-sep SNPs were located in regions clearly separated from the peak, which mostly overlapped with SNPs detected by SMA-BH except for one EN50-sep SNP (see Figure 9). Expectedly, SMA-BH identified the largest number of SNPs (161), of which 101 were in LD above the absolute correlation threshold of 0.25 with EN50-sep SNPs. The remaining 60 SMA-BH SNPs had largest absolute correlations with EN50-sep SNPs between 0.07 and 0.24, indicating together
with Figure 9 that SMA-BH selected some SNPs physically well separated from and likely not in LD with any EN50 SNPs. To check on the interpretation of the SNP-SNP correlation values, we simulated independent SNPs for \( N = 786 \) representing 11 different MAF values between 0.01 and 0.5 (assuming Hardy-Weinberg equilibrium) and computed all pairwise absolute correlations over 100,000 replications, yielding 75th, 95th and 99th percentiles just below 0.025, 0.07, and 0.13, respectively, with a maximum value of 0.31. Based on our simulation studies (in particular the eight-QTL simulation), SMA-BH may have higher power than EN50-sep (by about 10%) but also may not completely control the FDR, so the 60 SMA-BH SNPs with very small correlations to EN50-sep SNPs may represent some additional true signals and some false positives. We also re-ran EN50-sep on chromosome 1 by excluding all peak area SNPs but including a single peak area SNP (with the smallest p-value) in the model without shrinkage, but this analysis did not identify any additional SNPs (such as any of the 60 SMA-BH SNPs) compared to the previous EN50-sep analysis.

(Figure 9 about here)

On chromosome 4, SMA-GWT, SMA-BH and EN50-sep identified similar numbers of SNPs, 8 for SMA-GWT (all correlated with EN50 SNPs), 10 for SMA-BH (all but 1 correlated with EN50 SNPs), and 8 for EN50 (6 of which were correlated with SMA-GWT or SMA-BH SNPs).

On chromosomes 3 and 19, no QTLs were identified with the stringent SMA-GWT, while SMA-BH and EN50-sep identified the same QTLs (Table 1 and Figure 9). For chromosome 3, the 30 SMA-BH SNPs were all in LD (absolute correlation threshold
of 0.25) with the five EN50-sep SNPs, and for chromosome 19, the three SMA-BH SNPs were all in LD with the one EN50-sep SNP. No SNPs were identified by any method on chromosome 6. Overall, these results show that SMA-BH and EN50-sep identify very similar candidate regions in a genome scan, that EN50 (and other PR methods) will select a subset of SNPs in a cluster of correlated SNPs, here those with the smallest SMA p-values, and that for follow-up (fine-mapping) studies, all EN50 (or other PR) identified SNPs plus a set of SNPs correlated with the identified SNPs (above a certain threshold) should be considered.

Lastly, the joint analysis of all chromosomes by EN50 with FDR control (EN50-joint), compared to the separated analyses (EN50-sep), identified fewer SNPs (20 SNPs for EN50-joint compared with 29 SNPs for EN50-sep). EN50-joint identified one SNP on chromosome 3 and another on chromosome 4 which were not in LD with EN50-sep SNPs. The EN50-joint SNP on chromosome 3 is likely a false positive due to its -$\log_{10}(p$-value) being much smaller than those of all other identified (by any method) SNPs (Figure 10). The EN50-joint SNP on chromosome 4 however was in LD (absolute correlation above 0.9) with a SMA-BH SNP.

(Figure 10 about here)

**Concluding Remarks**

The goals of this study were to provide a review and an independent comparison of penalized regression methods with single marker analysis for variable (SNP) selection in GWAS, and to implement and evaluate penalty/tuning parameter value selection by
FDR control. A recent application of PR to GWAS (Waldmann et al., 2013) arrived at this conclusion: “Hence, we can conclude that it is important to analyze GWAS data with both the lasso and the elastic net and an alternative tuning criterion to minimum MSE is needed for variable selection.” Here we provided such an alternative criterion.

We provided both a simple, approximate analytic method and a permutation-based method for FDR control in PR. The analytic method performed consistently well although being somewhat conservative. The permutation method expectedly had higher power (was less conservative) than the analytic method in the two-QTL scenario but was more conservative than the analytic method in the eight-QTL scenario. This initially unexpected behavior of the permutation method appears to be due to the fact that permuting the phenotypes means permuting some signal along with the random error and thus overestimates the amount of random error in the data. This causes the permutation method to become increasingly conservative with increasing numbers and effect sizes of causal SNPs (QTLs). Because our simulated data had small sample sizes (near 200) and contained QTLs jointly explaining 20 or 48 percent of the phenotypic variance (for the two- or eight-QTL data, respectively), this problem of the permutation method is more pronounced in our simulation than in real GWAS data characterized by very large sample sizes (thousands to ten thousands) and small SNP effects (explaining 1% or less of the overall variation). However, the problem may manifest itself also in real data, when there are larger numbers of causal SNPs or causal SNPs with larger effect sizes as in our Health ABC GWAS. This is unfortunate because the main advantage of multi-marker over single-marker approaches is that they are able to use the signal from other markers in determining significance. It may be
possible to overcome this limitation by estimating the residuals using the fitted model and then permuting the residuals instead of the phenotype; we are currently investigating the performance of this approach.

Based on our comparisons with previous strategies for SNP selection in PR, it appears that our analytic FDR criterion is currently the best approach to SNP selection when using PR for GWAS, and this approach should be appealing to practitioners who desire a measure of error, in particularly in terms of FDR, associated with the selected SNPs. Lastly we note that our simple analytic FDR method can also be applied to penalized logistic regression.

When the focus is on variable selection (here the identification of a causal SNP by itself or a linked SNP) and not on estimation or prediction, our results indicate that among all PR methods investigated here, a version of the Elastic Net (EN50) performs better than the other PR methods in most situations, although differences were often small. Based on our analyses of simulated and real data, the EN50 should identify QTL regions that are very similar to those identified by SMA combined with BH FDR control. In contrast with SMA-BH, EN50 identifies a QTL region with a single SNP or few SNPs, hence subsequent fine-mapping should include the EN50 SNPs plus additional SNPs in LD with the EN50 SNPs above a certain threshold. Expectedly, EN50 and SMA-BH have substantially more power than SMA with the genome-wide type-I error threshold.

Incorporating fusion-type penalties developed for covariates measured on graphs may improve power but can also generate more false positives, or rather wide QTL regions for follow-up, in situations when there are multiple, moderately correlated causal
SNPs located on the same chromosome, and false positives are defined as any significant SNPs correlated with a causal SNP below an absolute value of 0.25.

When applying PR to large genome-wide datasets for joint (pooled) analysis of all chromosomes, pre-screening SNPs based on SMA p-values is still necessary due to the high memory requirements for holding the entire SNP genotype matrix in memory (see Hoffman et al. (2013) for more details). We note that if pre-screening is performed, then for tuning parameter value selection by our analytic FDR method the number of markers ($p$) in [13] should be set to the total number of markers (prior to pre-selection) to maintain control of the FDR (which should produce the same result as what would be obtained without pre-selection). It would however be prudent to perform PR with FDR control both on the entire set of SNPs from all chromosomes (pooled analysis) as well as separately on each chromosome (separate analysis). Our results from the simulated and real data representing several chromosomes indicate that the pooled analysis does not necessarily provide better power for all chromosomes.

We note that currently there is substantial interest in extending GWAS of single phenotypes to high-dimensional phenotypes (e.g., Marttinen et al. (2012)). While high-dimensional phenotypes allow aspects of modeling that are not feasible with a single phenotype, the results of the current study still provide useful information for the design of analysis methods for GWAS of high-dimensional phenotypes.

Lastly, code used for data simulation and code for EN50 analysis with analytic FDR control are provided in Supplement C. An example simulated dataset with eight QTLs on chromosome 21 is available at the Dryad Digital Repository
ACKNOWLEDGEMENTS

This work was supported by the National Institute of General Medical Sciences grant R01HG005254 to I.H., by the National Institute on Aging grant 1R01AG032098-01A1 and the National Heart, Lung and Blood Institute grant R01HL101250 to Y.L., and by a Virginia Tech Genetics, Bioinformatics and Computational Biology Ph.D. program fellowship to H.Y.. The HABC study was supported by the National Institute on Aging contracts N01AG62101, N01AG62103, and N01AG62106. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C.
LITERATURE CITED


Efron, B., 2005 Local false discovery rates. Division of Biostatistics, Stanford University.


Gauderman, W. J. and J. Morrison, 2001 QUANTO documentation (Technical report no. 157). Department of Preventive Medicine, University of Southern California.


Kohavi, R., 1995 A study of cross-validation and bootstrap for accuracy estimation and model selection, pp. 1137-1145 in IJCAI.


Tseng, P., 1988 Coordinate ascent for maximizing nondifferentiable concave functions. Massachusetts Institute of Technology, Laboratory for Information and Decision Systems.


Table 1: Analysis of the real data with Single Marker Analysis using the Genome-Wide Threshold p-value < $5.5\times10^{-8}$ (SMA-GWT) or Benjamini-Hochberg FDR control (SMA-BH) applied separately to the data on each chromosome, Elastic Net penalized regression with analytic FDR control applied separately to each chromosome (EN50s) or jointly across all five chromosomes (EN50j). No. sig denotes the number of significant SNPs, $\cap$ denotes overlap with. The three numbers for overlap (e.g., of SMA-GWT with EN50s in row 3) represent the number of significant SMA-GWT SNPs that are: identical to a significant EN50s SNP / correlated at or above 0.8 with a significant EN50s SNP / correlated at or above 0.5 with a significant EN50s SNP / correlated at or above 0.25 with a significant EN50s SNP.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SMA-GWT</strong></td>
<td>No. sig</td>
<td>74</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>$\cap$EN50s</td>
<td>9/17/42/61</td>
<td>0/0/0/0</td>
<td>6/8/8/8</td>
<td>0/0/0/0</td>
<td>0/0/0/0</td>
</tr>
<tr>
<td><strong>SMA-BH</strong></td>
<td>No. sig</td>
<td>161</td>
<td>30</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>$\cap$EN50s</td>
<td>13/23/49/101</td>
<td>5/30/30/30</td>
<td>6/9/9/9</td>
<td>0/0/0/0</td>
<td>1/1/3/3</td>
</tr>
<tr>
<td><strong>EN50j</strong></td>
<td>No. sig</td>
<td>14</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>$\cap$EN50s</td>
<td>14/14/14/14</td>
<td>2/2/2/2</td>
<td>2/2/2/2</td>
<td>0/0/0/0</td>
<td>0/0/0/0</td>
</tr>
<tr>
<td><strong>EN50s</strong></td>
<td>No. sig</td>
<td>15</td>
<td>5</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>$\cap$SMA-GWT</td>
<td>9/9/9/9</td>
<td>0/0/0/0</td>
<td>6/6/6/6</td>
<td>0/0/0/0</td>
<td>0/0/0/0</td>
</tr>
<tr>
<td>$\cap$SMA-BH</td>
<td>13/13/13/14</td>
<td>5/5/5/5</td>
<td>6/6/6/6</td>
<td>0/0/0/0</td>
<td>1/1/1/1</td>
</tr>
</tbody>
</table>
Figure 1: Single Marker Analysis (SMA) of chromosome 21 data with two isolated QTLs. 200 replicates, sample size N = 201. Empirical power (defined as True Positive Rate TPR) on the left: Dark blue represents TPR1 as defined in the main text (TPR including only the significant QTLs), and light blue represents TPR2 (TPR also including significant SNPs linked to QTLs at absolute correlation threshold 0.5, 0.7 or 0.9). Empirical thresholded False Discovery Rate (tFDR) with absolute correlation thresholds 0.25, 0.3, 0.5, 0.7 and 0.9 on the right: red horizontal line represents the tFDR value of 0.05. GWT: Genome-Wide Threshold which represents the p-value threshold $5.5 \times 10^{-8}$; BH: Benjamini-Hochberg; BY: Benjamini-Yekutieli; LFDR1 and LFDR2: local FDR based thresholding using different estimators of the proportion of true null hypotheses (see main text for details).

Figure 2: Single Marker Analysis (SMA) of chromosome 21 data with eight QTLs. 200 replicates, sample size N = 201. Empirical power (defined as True Positive Rate TPR) on the left: Dark blue represents TPR1 as defined in the main text (TPR including only the significant QTLs), and light blue represents TPR2 (TPR also including significant SNPs linked to QTLs at absolute correlation threshold 0.5, 0.7 or 0.9). Empirical thresholded False Discovery Rate (tFDR) with absolute correlation thresholds 0.25, 0.3, 0.5, 0.7 and 0.9 on the right: red horizontal line represents the tFDR value of 0.05. GWT: Genome-Wide Threshold which represents the p-value threshold $5.5 \times 10^{-8}$; BH: Benjamini-Hochberg; BY: Benjamini-Yekutieli; LFDR1 and LFDR2: local FDR based thresholding using different estimators of the proportion of true null hypotheses (see main text for details).

Figure 3: Penalized Regression (PR) analyses of chromosome 21 data with two isolated QTLs. 200 replicates, sample size N = 201. Empirical power (defined as True Positive Rate TPR) on the top: Dark blue represents TPR1 as defined in the main text (TPR including only the significant QTLs), and light blue represents TPR2 (TPR also including significant SNPs linked to QTLs at absolute correlation threshold 0.5, 0.7 or 0.9). Empirical thresholded False Discovery Rate (tFDR) with absolute correlation thresholds 0.25, 0.3, 0.5, 0.7 and 0.9 on the bottom: red horizontal line represents the tFDR value of 0.05. The following penalized regression methods (in the order shown) are represented. AdaLasso: Adaptive Lasso where all three initial estimators (SMA, Lasso with CV, Ridge Regression with CV) produced the same result; EN_Lambda2 = x: Elastic Net with Lasso portion of $\lambda_2 = x$ (x = 0.3, 0.5, 0.9); Lasso_anly: Lasso with analytical FDR control; Lasso_perm: Lasso with permutation FDR control; MCP_lambda2 = x: Minimax Concave Penalty with $\lambda_2 = x$ (x = 3, 10, 30, 100).

Figure 4: Penalized Regression (PR) analyses of chromosome 21 data with eight QTLs. 200 replicates, sample size N = 201. Empirical power (defined as True Positive
Rate TPR) on the top: Dark blue represents TPR1 as defined in the main text (TPR including only the significant QTLs), and light blue represents TPR2 (TPR also including significant SNPs linked to QTLs at absolute correlation threshold 0.5, 0.7 or 0.9). Empirical thresholded False Discovery Rate (tFDR) with absolute correlation thresholds 0.25, 0.3, 0.5, 0.7 and 0.9 on the bottom: red horizontal line represents the tFDR value of 0.05. The following penalized regression methods (in the order shown) are represented. AdaLasso: Adaptive Lasso where all three initial estimators (SMA, Lasso with CV, Ridge Regression with CV) produced the same result; EN_Lambda2 = x: Elastic Net with Lasso portion of $\lambda_2 = x$ ($x = 0.3, 0.5, 0.9$); Lasso_anly: Lasso with analytical FDR control; Lasso_perm: Lasso with permutation FDR control; MCP_lambda2 = x: Minimax Concave Penalty with $\lambda_2 = x$ ($x = 3, 10, 30, 100$).

Figure 5: Empirical power (True Positive Rate TPR) and thresholded False Discovery Rate (tFDR) for LD2lasso versus Lasso and Elastic Net (EN) Penalized Regression (PR) analyses of chromosome 21 data with two isolated QTLs. 20 randomly selected replicates, sample size N = 201. Dark blue represents TPR1 as defined in the main text (TPR including only the significant QTLs), and light blue represents TPR2 (TPR also including significant SNPs linked to QTLs at absolute correlation threshold 0.5, 0.7 or 0.9). The empirical tFDR was computed at five absolute correlation thresholds between causal and linked SNPs (T = 0.25, 0.3, 0.5, 0.7, 0.9). The red horizontal line represents the tFDR value of 0.05. EN_Lambda2 = 0.5 denotes the Elastic Net with Lasso portion of $\lambda_2 = 0.5$; LD2lasso_0.9_r^2 (LD2lasso_0.9_\|r\|) represents the LD2lasso with weight on the Lasso penalty equal to $\varphi = 0.9$ and LD function $h(r) = r^2$ ($h(r) = |r|$). The LD function was set to zero when the absolute correlation between two SNPs was $|r| < 0.85$.

Figure 6: Empirical power (True Positive Rate TPR) and thresholded False Discovery Rate (tFDR) for LD2lasso versus Lasso and Elastic Net (EN) Penalized Regression (PR) analyses of chromosome 21 data with eight QTLs. 20 randomly selected replicates, sample size N = 201. Dark blue represents TPR1 as defined in the main text (TPR including only the significant QTLs), and light blue represents TPR2 (TPR also including significant SNPs linked to QTLs at absolute correlation threshold 0.5, 0.7 or 0.9). The empirical tFDR was computed at five absolute correlation thresholds between causal and linked SNPs (T = 0.25, 0.3, 0.5, 0.7, 0.9). The red horizontal line represents the tFDR value of 0.05. EN_Lambda2 = 0.5 denotes the Elastic Net with Lasso portion of $\lambda_2 = 0.5$; LD2lasso_0.9_r^2 (LD2lasso_0.9_\|r\|) represents the LD2lasso with weight on the Lasso penalty equal to $\varphi = 0.9$ and LD function $h(r) = r^2$ ($h(r) = |r|$). The LD function was set to zero when the absolute correlation between two SNPs was $|r| < 0.85$. 
Figure 7: Plot of True Positive Rate TPR1 versus thresholded False Discovery Rate (tFDR): TPR1 versus tFDR (T=0.25) and TPR1 versus tFDR (T=0.50). TPR1 and tFDR are defined in the main text. MCP was run with $\lambda_2 = 10$. CAR represents correlation-adjusted marginal correlations.

Figure 8: Plot of True Positive Rate TPR2 versus thresholded False Discovery Rate (tFDR): TPR2 (T=0.50) versus tFDR (T=0.25) and TPR2 (T = 0.50) versus tFDR (T=0.50). TPR2 and tFDR are defined in the main text. MCP was run with $\lambda_2 = 10$. CAR represents correlation-adjusted marginal correlations.

Figure 9: Plot of – log$_{10}$(p-value) for SNPs on chromosomes 1 (left), 3, 4, 6, and 19 (right) for the real data. The purple, black, and red circles represent significant SNPs selected by both SMA with BH FDR control (SMA-BH) and EN50 (Elastic Net with 50% weight on L1), by SMA-BH only, and by EN50 only, respectively. EN50-sep denotes EN50 applied to each chromosome separately. The grey line is the genome-wide threshold of $5.5 \times 10^{-8}$, above which SNPs were selected by SMA-GWT. The y-axis represents raw p-values from SMA.

Figure 10: Plot of -log$_{10}$(p-value) for SNPs on chromosomes 1 (left), 3, 4, 6, and 19 (right) for the real data. EN50-sep (EN50-joint) denotes the Elastic-Net with 50% weight on L1 applied to each chromosome separately (to all chromosomes jointly). The purple, red and black circles represent significant SNPs selected by both EN50-sep and EN50-joint, by EN50-sep only and by EN50-joint only, respectively. The y-axis represents raw p-values from SMA.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 6
Figure 7
Figure 8
Figure 9
Figure 10