Penalized regression approaches to testing for quantitative trait-rare variant association

Sunkyung Kim¹, Wei Pan¹* and Xiaotong Shen²

¹ Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA
² School of Statistics, University of Minnesota, Minneapolis, MN, USA

1. INTRODUCTION

Genome-wide association studies (GWAS) have uncovered many common variants (CVs) associated with complex diseases, but the proportion of variance explained by the identified CVs is often low (Maher, 2008). With the recent advance of sequencing technologies, analysis of rare variants (RVs) has become a feasible alternative. Recent studies have demonstrated that some RVs are associated with complex disease. For example, Kotowski et al. (2006) found that multiple RVs in gene PCSK9 are associated with plasma levels of low-density lipoprotein cholesterol.

In this study, we propose applying some new penalized regression methods to test for association between a quantitative trait and multiple RVs. Differing from the usual application of penalized regression methods to variable selection or risk prediction for high-dimensional data (Kooperberg et al., 2010; Austin et al., 2013), here we focus on their application to hypothesis testing on a quantitative trait in a relatively low-dimensional setting. In such a setting, one commonly used statistical test is the F-test in linear regression. For example, in simple regression, a trait Y is regressed on each of multiple variants sequentially. However, because of the extremely low minor allele frequency (MAF) of a RV, a test to detect the association between a trait and a single RV might be low powered. Also, this approach may be too conservative for a stringent control for multiple testing, e.g., by the Bonferroni correction to control the family-wise error rate. In addition, ultimately, complex diseases are expected to be affected by a combination of multiple genetic variants. Thus an analysis in which a group of variants are tested simultaneously for their joint effects on the trait may be more powerful. In multiple regression, any association between a trait and k RVs, all k RVs are added to a regression model. However, as k increases, the statistical power might decrease due to the cost of large degrees of freedom (DF), k. To avoid the large DF and to aggregate information across multiple RVs, one common strategy is to pool or collapse multiple RVs in a region or gene (Li and Leal, 2008; Madsen and Browning, 2009). One such attempt is the Sum test (Pan, 2009), which was developed to utilize joint effects of multiple variants while reducing the DF. With only 1 DF, the Sum test enhances power under some scenarios (Chapman and Whittaker, 2008; Pan, 2009). However, it is noted that the performance of the Sum test depends on the directions of the variants’ associations with a trait. Thus, in an extreme case where a half of the variants are positively associated with the trait and the other half are negatively associated with similar effect sizes, the positive and negative effects may cancel out, leading to the poor performance of the Sum test and other burden tests (Han and Pan, 2010; Li et al., 2010). In addition, in the Sum test or other pooling-based burden tests, combining or collapsing all variants into just one group ignores the variants’ possibly varying effect sizes, and thus may not work well in those situations. In particular, the Sum test and many burden tests perform poorly if
of subject $j$ with null hypothesis $H_0 : \beta = (\beta_1, \ldots, \beta_k)^T = 0$ by an $F$-test, which is based on the OLS estimates that minimize the residual sum of squares. A potential problem with the test is the power loss due to the large variance of $\hat{\beta}$, since the MAFs of RVs are small.

We also apply other four association tests: the Score, the sum of squared score (SSU), its weighted version SSUw (Pan, 2009), and the univariate minP (UminP) tests. The Score test is popular in general statistics while the UminP test is most widely used for CVs in GWAS; on the other hand, Basu and Pan (2011) showed that the SSU and SSUw tests were powerful in RV association tests for case-control studies. Here, as a secondary contribution, we extend the SSU and SSUw tests to the case with a quantitative trait. All the four tests are based on the score vector $U$ and its covariance matrix $V$ under $H_0$:

\[
U = \sum_{i=1}^{n} (Y_i - \bar{Y})X_i, \\
V = \text{Cov}(U) = \hat{\sigma}^2 \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T,
\]

where $\bar{Y} = \sum_{i=1}^{n} Y_i/n$, $\bar{X} = \sum_{i=1}^{n} X_i/n$, and $\hat{\sigma}^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2/(n-1)$ is the estimate of $\sigma^2$ under $H_0$. The corresponding four test statistics are

\[
T_{\text{Score}} = U^TV^{-1}U, \\
T_{\text{SSU}} = U^TU, \\
T_{\text{SSUw}} = U^TV_d^{-1}U \quad \text{with} \quad V_d = \text{Diag}(V), \\
T_{\text{UminP}} = \max_{j=1}^{k} U_j^2/v_j,
\]

where $U_j$ is the $j$th element of $U$ and $v_j$ is the $(j, j)$th diagonal element of $V$. Under $H_0$, asymptotically $T_{\text{Score}}$ has a $\chi^2_k$ distribution.
each of $T_{SSU}$ and $T_{SSUm}$ has a mixture of chi-squared distributions (Pan, 2009), and the $p$-value of $T_{Unpaired}$ can be numerically obtained (Conneely and Boehnke, 2007).

Next, we extend the Sum test (Pan, 2009) and its modified version, a data-adaptive Sum (aSum) test (Han and Pan, 2010), to the case with a quantitative trait. The Sum test was originated to model multiple variants jointly while inducing a minimum number of DF: while including all the variants in the linear model, it assumes that the variants all have the same effect size (and direction), $\beta_c$, as in the following model:

$$Y_i = \beta_{c,0} + \sum_{j=1}^{k} X_{ij} \beta_c + \epsilon_i$$  \hspace{1cm} (2)

Fitting (Equation 2) is equivalent to conducting a simple regression of $Y$ on a new covariate, the sum of the genotypes over the multiple variants. To address the question of whether any association between the disease and the variants exists, one simply needs to test $H_0: \beta_c = 0$, without the need for multiple testing adjustment. The main advantage of the Sum test is that, because it tests on only one parameter $\beta_c$, there will be no power loss due to the large DF. The common association parameter $\beta_c$ is a weighted average of the individual $\beta_{M,1}, \ldots, \beta_{M,k}$ in the marginal models $Y_i = \beta_{M,0} + X_{ij} \beta_{M,j} + \epsilon_i$ for $j = 1, \ldots, k$ (Pan, 2009). On the other hand, the main problem of the Sum test is its dependence on the signs of $\beta_{M,j}$ or on the coding of each variant (i.e., which allele is chosen as the reference category). If the signs are not the same, the test may have a quite small $\hat{\beta}_c$ and thus low power. To overcome the limitation of the Sum test, Han and Pan (2010) proposed the aSum test for a case-control study design, which can be equally applied to quantitative traits as the following. (1) For each variant $j$, flip its coding to $X_{ij}^* = 2 - X_{ij}$ if $\hat{\beta}_{M,j} < 0$ and its $p$-value $p_{M,j} \leq \alpha_0$ in the marginal model; otherwise use the same coding $X_{ij}^* = X_{ij}$. (2) Fit the model (Equation 2) with the new coding $X^*$. To test $H_0$ in the aSum test, we use a permutation-based log-likelihood ratio test (LRT), which is asymptotically equivalent to the score test. For the choice of $\alpha_0$, we use the same value as recommended by Han and Pan (2010), 0.1, to prevent reduced power when a too small or too large $\alpha_0$ is used.

While the $F$-test is based on OLS estimates, in next section we apply some penalized regression methods, the Lasso, glslasso and a recently developed TLP for either only variable selection (TLP-S) or both variable selection and parameter grouping (TLP-SG). In short, both the Lasso and TLP-S consider only variable selection, while the glslasso and TLP-SG pursue parameter grouping along with variable selection to improve power by striking a better balance between goodness-of-fit and reduced DF in the joint model (Equation 1).

### 2.2. PENALIZED REGRESSION BASED TESTS

#### 2.2.1. Parameter estimation from penalized regression

Given a vector of traits $Y = (Y_1, \ldots, Y_n)'$ and a design matrix for $k$ variants $X = (X_1, \ldots, X_k)$, the Lasso estimate of $\beta$ obtained from the penalized least squares function:

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \| Y - X\beta \|^2 + \lambda \sum_{j=1}^{k} |\beta_j|,$$  \hspace{1cm} (3)

where a large $\lambda$ automatically yields some components of $\hat{\beta}$ as 0, realizing variable selection. While Lasso does effective variable selection, its estimates are always biased. To overcome the issue, Shen et al. (2012) proposed a truncated Lasso ($L_1$)-penalty (TLP) $J_1(x) = \min(|x|, \tau)$, which, as $\tau \to 0^+$, tends to the $L_0$-loss, $I(|x| \neq 0)$. The degree of approximation by TLP is controlled by a tuning parameter, $\tau$. See Figure 1 for a display over the different values of $\tau$. Then the TLP-estimate $\hat{\beta}$ is obtained from

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \| Y - X\beta \|^2 + \lambda_1 \sum_{j=1}^{k} \tau \min(|\beta_j|, 1),$$  \hspace{1cm} (4)

and we denote (Equation 4) as TLP-S. The most interesting feature of the TLP is that only smaller $|\beta_j|$’s less than a threshold $\tau$ are penalized, hence realizing variable selection (if some are shrunken to 0) while avoiding penalizing larger $|\beta_j|$’s and thus leading to their almost unbiased estimates.

While both the Lasso and TLP-S consider only variable selection, an alternative way to reduce model complexity is grouping pursuit (Shen and Huang, 2010). To investigate the grouping effects on a test’s power, we apply two recent penalized grouping methods, glslasso and TLP-SG. The $\beta$ estimate from glslasso is based on the following objective function:

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \| Y - X\beta \|^2 + \lambda_1 \sum_{j=1}^{k} \tau \min(|\beta_j|, 1) + \lambda_2 \sum_{j \neq j'} |\beta_j - r(j,j') \beta_{j'}|,$$  \hspace{1cm} (5)

![Figure 1](image-url)
where the first penalty is used for variable selection and the second is to encourage parameter grouping. \( r(j, j') \) is the sign of the correlation between two variants \( X_j \) and \( X_{j'} \), which is used to approximate the target \(|\hat{\beta}_j| \approx |\beta_j|\); this method is denoted \( gflu_{r = cor} \). On the other hand, if \( r(j, j') = 1 \) is used, the penalty targets \( \hat{\beta}_j \approx \beta_j \).

The TLP-SG estimate of \( \beta \) comes from

\[
\hat{\beta} = \arg\min_{\beta} \frac{1}{2} \| Y - X\beta \|^2 + \lambda_1 \sum_{j=1}^{k} I_j(|\beta_j|) + \lambda_2 \sum_{j < j'} I_j(|\beta_j| - |\beta_{j'}|),
\]

where the first penalty is for variable selection while the second shrinks the difference of \(|\beta_j|'s\) if a difference is within the upper bound \( \tau \). The number of the groups of equal parameter estimates is a decreasing function of \( \lambda_2 \). Thus, the tuning parameters \((\lambda_1, \lambda_2, \tau)\) are selected to balance between the model complexity and model goodness-of-fit, which presumably may contribute to enhanced power. As a comparison, in the Sum test all parameters (or variants) are forced to belong to the same single group even if the variants' associations with the trait are quite different both in effect sizes and directions; the TLP-SG method attempts to conduct a more precise grouping over all variants in a data-adaptive way.

To compute \( \hat{\beta} \) in Lasso, gflsso, TLP-S and TLP-SG, we used the Feature Grouping and Selection Over An Undirected Graph (FGSG) package of Yang et al. (2012), which is a C library with interface to MATLAB and is quite fast to run. Its computing efficiency allowed us to estimate separate tuning parameters for each permuted dataset to control the type I error as explained in the next section.

### 2.2.2. Hypothesis testing

To test the null hypothesis \( H_0 : \beta = 0 \) in Equation (1), we conduct a permutation-based test, in which the \( p \)-value is calculated by comparing a test statistic \( T \) applied to the original dataset to the ones \( T_0(b) \) applied to the \( B \) permuted datasets for \( b = 1, \ldots, B \). We use permutation to control the Type I error since the null distribution of a test statistic based on a penalized regression estimate is in general difficult to obtain. The permutation-based testing procedure follows:

Step 1. With the original data \((Y_i, X_i)\), we solve a penalized regression problem to obtain \( \hat{\beta} \) in Equation (1).

Step 2. Calculate a test statistic \( T = T(\hat{\beta}) \).

Step 3. By repeatedly permuting the observed \( Y \) of the original data, we obtain \( B \) sets of permuted data \((Y_i^{(b)}, X_i)\) for \( b = 1, \ldots, B \). For each permuted data set, \((Y_i^{(b)}, X_i)\), we repeat the steps 1 and 2, obtaining the null statistics \( T_0(b) \).

Step 4. The final \( p \)-value is \( \sum_{b=1}^{B} I \left( T < T_0(b) \right) / B \).

We apply each of several test statistics in Step 2. First, across all penalized methods, we use a 1-df \( F \)-statistic (1-df) to test the association between \( Y \) and \( X\hat{\beta} \), where \( \hat{\beta} \) is the penalized estimate of \( \beta \) in Step 1. Specifically, we fit a linear model

\[
Y_i = \alpha_0 + (X_i\hat{\beta})' \alpha + \epsilon_i,
\]

and test \( H'_0 : \alpha = 0 \). This 1-df test uses variable selection and possibly parameter grouping result from the corresponding penalized method, while allowing testing with only 1 DF. Second, for TLP-SG, we also apply the corresponding SSU and SSUw tests, where the test statistics \( T_{SSU} \) and \( T_{SSUw} \) are both based on the selected variables from the corresponding penalized estimates. Specifically,

\[
T_{SSU} = U^* U^*, \quad T_{SSUw} = U^* (V_d^*)^{-1} U^* \quad \text{with} \quad V_d^* = \text{Diag}(V^*),
\]

where \( U^* \) is a sub-component vector of the score vector \( U \) corresponding to \(|\hat{\beta}_j| \neq 0 \), and \(|\hat{\beta}_j| > 0.001 \) is considered as non-zero. Similarly, \( V^* \) is the corresponding sub-matrix of the covariance matrix \( V \). Note that the grouping information is not used.

#### 2.2.3. Selection of tuning parameters

To select the suitable tuning parameters, we apply a grid-search with Akaike's information criterion (AIC) (Akaike, 1974):

\[
\text{AIC} = -2 \log L + 2p,
\]

where \( \log L = \frac{-n \log (\hat{\sigma}^2) - n - p - 1}{2} \) is the log-likelihood with the penalized estimate plugged-into model (Equation 1), and \( \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (Y_i - \beta_0 - X_i\hat{\beta})^2}{n - p - 1} \). The effective number of the parameters, \( p \), in AIC is computed as the number of non-zero \(|\hat{\beta}_j|'s\) for Lasso and TLP-S, as the number of non-zero unique \( \hat{\beta}_j \)'s for \( gflu_{r = cor} \); and as the number of non-zero unique \( |\hat{\beta}_j|'s \) for \( gflu_{r = cor} \) and TLP-SG, respectively.

For \( \lambda \) in Lasso, the one resulting in the smallest AIC out of 50 equally spaced points in \([0.001, 10]\) is selected. Similarly, the values of each of \( \lambda_1 \), \( \lambda_2 \) and \( \tau \) in other methods are searched over five equally spaced grid points of \([0.001, 1], [0.001, 0.5], \) and \([0.001, 0.5]\), respectively. For each permuted dataset \((Y_i^{(b)}, X_i)\) for \( b = 1, \ldots, B \), we also estimate its own \((\lambda_1^{(b)}, \lambda_2^{(b)}, \tau^{(b)})\) to properly control the type I error.

### 3. RESULTS

#### 3.1. SIMULATIONS

We consider two simulation schemes. In the first scheme, we generate only RVs with a total of 200 replicates and \( n = 400 \) in each replicate. The permutation size is set as \( B = 100 \). For each replicate, to generate \( k \) variants including six causal ones in linkage disequilibrium (LD), as in Wang et al. (2007), two latent vectors from multivariate normal distribution MVN(0, \( R \)) are simulated, where \( R \) has a first order auto regressive (AR1) structure; the association between any two elements of the latent vector decreases by \( \rho = 0.8 \) times as 1 lag increases. Then, the vector is dichotomized by comparing a test statistic \( T \) applied to the original dataset to the one resulting in the smallest AIC out of 50 equally spaced points in \([0.001, 10]\) is selected. Similarly, the values of each of \( \lambda_1 \), \( \lambda_2 \) and \( \tau \) in other methods are searched over five equally spaced grid points of \([0.001, 1], [0.001, 0.5], \) and \([0.001, 0.5]\), respectively. For each permuted dataset \((Y_i^{(b)}, X_i)\) for \( b = 1, \ldots, B \), we also estimate its own \((\lambda_1^{(b)}, \lambda_2^{(b)}, \tau^{(b)})\) to properly control the type I error.
the randomly located six causal variants with \( \sigma^2 = 2 \) in model (Equation 1), where the intercept \( \beta_0 \) is set as 0.3 throughout the simulations. The considered three cases are:

- **Case 1:** \( \beta = (0.9, 0.9, 0.9, 0.9, 0.9, 0, \ldots, 0)' \)
- **Case 2:** \( \beta = (1.2, 1.2, 1.2, -1.2, -1.2, -1.2, 0, \ldots, 0)' \)
- **Case 3:** \( \beta = (1.4, 1.3, -1.2, 1.2, -1.3, 1.4, 0, \ldots, 0)' \).

In each case, we vary the number of non-causal RVs \( k \) from 0 to 24 so that the total number of RVs, \( k \), ranges from 6 to 30. The Type I error is computed from the \( Y \) under \( H_0 : \beta = (0, \ldots, 0)' \).

In the second scheme, multiple RVs and two CVs are generated to mimic the GAW17 data we use later. The frequency of one allele for each CV is randomly distributed between 0.2 and 0.7, and CVs may or may not be chosen as a causal variant in each replicate. When a CV is randomly selected as a causal variant, its effect size \( \beta_j \) is scaled down to \( \beta_j/10 \) in the following cases to prevent its dominating association with the outcome. The considered three cases for mixed RVs and CVs are:

- **Case 1:** \( \beta = (1, 1, 1, 1, 1, 1, 0, \ldots, 0)' \)
- **Case 2:** \( \beta = (1.5, 1.5, 1.5, -1.5, -1.5, -1.5, 0, \ldots, 0)' \)
- **Case 3:** \( \beta = (1.1, 1.3, -1.2, 1.2, -1.3, 1.1, 0, \ldots, 0)' \).

Figure 2 displays the TLP-S and TLP-SG solution paths of \( |\hat{\beta}| \) over a tuning parameter given other(s), where two horizontal lines at 1.2 and 0 give the true parameter values for Case 2 set-up with only RVs. In contrast to piece-wise linear solution paths of the Lasso estimates, the TLP solution paths are like step functions with only RVs. In contrast to piece-wise linear solution paths of the Lasso estimates, the TLP solution paths are like step functions.
FIGURE 2 | Solution paths of $|\hat{\beta}_j|$'s in a simulated dataset of Case 2 with $k = 22$ RVs for TLP-S and TLP-SG over the values of a tuning parameter given other(s). The true values of $|\hat{\beta}_j|$'s at 1.2 and 0 are given by two horizontal lines. (A) TLP-S: $\tau = 0.15$. (B) TLP-S: $\lambda_1 = 0.1$. (C) TLP-SG: $(\lambda_2, \tau) = (0.01, 0.15)$. (D) TLP-SG: $(\lambda_1, \tau) = (0.02, 0.15)$. (E) TLP-SG: $(\lambda_1, \lambda_2) = (0.1, 0.01)$.

to 1.2; all variants are positively associated with the trait Q2 but in differential magnitudes. In this study, we test on each of all causal genes (PLAT, SREBF1, SIRT1, VLDLR, VNN3, PDGFD, BCHE, INSIG1, LPL, RARB, VNN1, and VWF) except GCKR, which contains just one SNP. The number of causal variants ($nC$) in each gene affecting Q2, and some summary statistics of their MAFs and pairwise correlations (COR) are listed in Table 5. Within each gene, most variants are RVs, but a few are CVs with their MAFs larger than 5%. First, we test for any association between Q2 and all variants gene by gene as shown in Table 6, and then test on each gene without its CVs as shown in Table 7.
Table 1 | Empirical Type I error and Power at the nominal level $\alpha = 0.05$ based on 200 replicates for the RVs only set-ups with six causal RVs and a varying number of non-causal RVs.

| Model fitting | Test statistics | # of non-causal RVs | # of non-causal RVs |
|---------------|----------------|--------------------|--------------------|
|               |                | 0  | 8  | 16 | 24 | 0  | 8  | 16 | 24 |
|               |                | Null | | | | Case 1 | | | |
| OLS           | $F$-test     | 0.030 | 0.080 | 0.040 | 0.060 | 0.715 | 0.480 | 0.340 | 0.260 |
| OLS           | Score        | 0.030 | 0.080 | 0.035 | 0.055 | 0.710 | 0.470 | 0.320 | 0.245 |
| OLS           | SSU          | 0.030 | 0.060 | 0.045 | 0.045 | 0.830 | 0.660 | 0.510 | 0.405 |
| OLS           | SSUw         | 0.035 | 0.080 | 0.055 | 0.060 | 0.810 | 0.625 | 0.500 | 0.380 |
| OLS           | UminP        | 0.045 | 0.070 | 0.050 | 0.035 | 0.675 | 0.445 | 0.360 | 0.310 |
| OLS           | Sum          | 0.055 | 0.075 | 0.040 | 0.075 | 0.915 | 0.685 | 0.525 | 0.460 |
| OLS           | aSum         | 0.035 | 0.065 | 0.035 | 0.060 | 0.910 | 0.715 | 0.575 | 0.520 |
| Lasso         | 1df          | 0.055 | 0.075 | 0.050 | 0.080 | 0.710 | 0.415 | 0.325 | 0.270 |
| gflasso$_{r = \text{cor}}$ | 1df        | 0.035 | 0.080 | 0.050 | 0.090 | 0.690 | 0.415 | 0.240 | 0.295 |
| gflasso$_{r = 1}$ | 1df        | 0.035 | 0.070 | 0.050 | 0.075 | 0.685 | 0.375 | 0.225 | 0.275 |
| TLP-S         | 1df          | 0.050 | 0.085 | 0.050 | 0.075 | 0.720 | 0.450 | 0.305 | 0.255 |
| TLP-SG        | 1df          | 0.055 | 0.085 | 0.055 | 0.070 | 0.700 | 0.450 | 0.290 | 0.250 |
| TLP-SG        | SSU          | 0.055 | 0.080 | 0.040 | 0.060 | 0.700 | 0.520 | 0.440 | 0.390 |
| TLP-SG        | SSUw         | 0.040 | 0.075 | 0.045 | 0.070 | 0.790 | 0.500 | 0.365 | 0.320 |

Case 2

|                |                | 0.635 | 0.515 | 0.440 | 0.455 | 0.745 | 0.640 | 0.550 | 0.490 |
|                |                | 0.625 | 0.500 | 0.425 | 0.395 | 0.745 | 0.635 | 0.525 | 0.470 |
|                |                | 0.590 | 0.530 | 0.505 | 0.445 | 0.710 | 0.645 | 0.595 | 0.555 |
|                |                | 0.570 | 0.505 | 0.475 | 0.445 | 0.715 | 0.660 | 0.570 | 0.525 |
|                |                | 0.450 | 0.410 | 0.400 | 0.310 | 0.665 | 0.595 | 0.425 | 0.425 |
|                |                | 0.145 | 0.125 | 0.145 | 0.100 | 0.485 | 0.310 | 0.260 | 0.215 |
|                |                | 0.450 | 0.430 | 0.355 | 0.340 | 0.665 | 0.590 | 0.535 | 0.500 |
| Lasso         | 1df          | 0.615 | 0.465 | 0.405 | 0.390 | 0.765 | 0.585 | 0.465 | 0.435 |
| gflasso$_{r = \text{cor}}$ | 1df      | 0.620 | 0.530 | 0.435 | 0.480 | 0.765 | 0.600 | 0.480 | 0.520 |
| gflasso$_{r = 1}$ | 1df      | 0.615 | 0.535 | 0.435 | 0.425 | 0.750 | 0.585 | 0.475 | 0.495 |
| TLP-S         | 1df          | 0.615 | 0.505 | 0.455 | 0.425 | 0.760 | 0.600 | 0.530 | 0.475 |
| TLP-SG        | 1df          | 0.615 | 0.485 | 0.445 | 0.415 | 0.755 | 0.605 | 0.450 | 0.450 |
| TLP-SG        | SSU          | 0.565 | 0.470 | 0.460 | 0.445 | 0.705 | 0.605 | 0.510 | 0.525 |
| TLP-SG        | SSUw         | 0.585 | 0.505 | 0.460 | 0.415 | 0.745 | 0.585 | 0.485 | 0.475 |

Maximum power in bold.

In Table 6, when both RVs and CVs within a gene are included, the identity of the most powerful test differs across the genes: the $F$-test is the winner for the genes VLDLR, VNN3, PDGFD, and LPL; however, for the genes VLDLR, BCHE, VNN1, and VWF, the SSU or SSUw test is the best. The two gflasso-based tests work quite similarly over all genes. The TLP based tests perform best for the genes SREBF1, RARAB, VNN1, and INSIG1. After removing a few CVs in each gene (Table 7), the SSU test recovers good power for the genes PDGFD, BCHE and LPL. The Sum test is the winner for gene BCHE, while the $F$-test based on the OLS estimates perform best for genes VNN3, SREBF1, and PDGFD. For gene VNN1, the TLP-SG with the SSU statistic has the highest power.

A potential advantage of penalized regression is variable selection, which is missing from existing global tests. Table 8 shows the results of causal variant selection by the penalized methods. Overall, each penalized method could eliminate some non-associated variants at the cost of omitting some causal ones. In general, in agreement with simulated data, the Lasso and TLP-SG seem to select fewest variants, including both TPs and FPs, while TLP-S and gflasso give higher numbers of both TPs and FPs.

4. DISCUSSION

In this study we have conducted hypothesis testing to detect the association between a quantitative trait and multiple RVs based on some new penalized regression methods. In addition to the traditional use of penalized regression for variant selection, we have also considered several state-of-the-art grouping pursuit methods that smooth the effect sizes of the variants, either $\beta_i$ or $|\beta_i|$, in a data-adaptive way, which can be considered as a generalization of the Sum and other genotype pooling/collapsing-based burden tests. In particular, our proposed TLP-SG overcomes several limitations of the Sum and other burden tests. First, by variable selection, the result of TLP-SG is presumably less influenced
Table 2 | Empirical Type I error and Power at the nominal level $\alpha = 0.05$ based on 200 replicates for the RVs + CVs set-ups with six causal variants and a varying number of non-causal ones.

| Model fitting | Test statistics | # of non-causal variants | # of non-causal variants |
|---------------|----------------|--------------------------|--------------------------|
|               |                | 0 | 8 | 16 | 24 | 0 | 8 | 16 | 24 |
|                |                | Null |                |                |                | Case 1 |                |                |                |
| OLS           | F-test         | 0.025 | 0.045 | 0.065 | 0.050 | 0.760 | 0.520 | 0.355 | 0.385 |
| OLS           | Score          | 0.020 | 0.045 | 0.065 | 0.035 | 0.760 | 0.515 | 0.345 | 0.350 |
| OLS           | SSU            | 0.060 | 0.050 | 0.090 | 0.030 | 0.490 | 0.210 | 0.125 | 0.110 |
| OLS           | SSUw           | 0.040 | 0.035 | 0.060 | 0.035 | 0.845 | 0.695 | 0.510 | 0.510 |
| OLS           | UminP          | 0.030 | 0.055 | 0.060 | 0.025 | 0.715 | 0.540 | 0.380 | 0.410 |
| OLS           | Sum            | 0.055 | 0.060 | 0.075 | 0.045 | 0.695 | 0.450 | 0.315 | 0.315 |
| OLS           | aSum           | 0.050 | 0.060 | 0.065 | 0.045 | 0.665 | 0.435 | 0.325 | 0.340 |
| Lasso         | 1df            | 0.030 | 0.045 | 0.060 | 0.045 | 0.750 | 0.515 | 0.360 | 0.375 |
| gflasso$_{r=\text{cor}}$ | 1df        | 0.030 | 0.030 | 0.070 | 0.015 | 0.760 | 0.450 | 0.275 | 0.415 |
| gflasso$_{r=1}$ | 1df          | 0.030 | 0.030 | 0.070 | 0.015 | 0.765 | 0.455 | 0.290 | 0.385 |
| TLP-S         | 1df            | 0.035 | 0.050 | 0.050 | 0.030 | 0.750 | 0.540 | 0.360 | 0.370 |
| TLP-SG        | 1df            | 0.035 | 0.035 | 0.065 | 0.045 | 0.750 | 0.515 | 0.335 | 0.315 |
| TLP-SG        | SSU            | 0.075 | 0.060 | 0.055 | 0.065 | 0.495 | 0.230 | 0.140 | 0.105 |
| TLP-SG        | SSUw           | 0.030 | 0.055 | 0.055 | 0.045 | 0.845 | 0.675 | 0.435 | 0.375 |

Table 3 | Mean numbers of TP(sd)/FP(sd) of the methods in Case 2 with both RVs and CVs.

| Method               | # of non-causal variants |
|----------------------|--------------------------|
|                     | 0 | 8 | 16 | 24 |
| OLS                  | 5.9(0.2)/. | 5.9(0.3)/79(0.4) | 5.9(0.3)/15.7(0.5) | 6.0(0.2)/23.5(0.7) |
| Lasso                | 4.4(1.7)/. | 3.7(1.6)/2.9(2.1) | 3.5(1.7)/4.7(3.3) | 3.2(1.7)/5.9(4.2) |
| gflasso$_{r=\text{cor}}$ | 5.4(1.0)/. | 4.8(1.5)/5.1(2.4) | 4.1(2.0)/8.0(5.0) | 3.5(2.2)/9.3(7.5) |
| gflasso$_{r=1}$       | 5.2(1.1)/. | 4.5(1.6)/4.8(2.4) | 4.3(1.9)/8.9(5.4) | 4.1(2.1)/12.3(8.7) |
| TLP-S                | 5.4(0.9)/. | 4.7(1.1)/4.3(1.5) | 4.4(1.1)/7.5(2.1) | 4.3(1.1)/11.0(2.9) |
| TLP-SG               | 4.7(2.0)/. | 4.3(1.8)/4.2(3.2) | 3.6(1.6)/5.3(4.5) | 3.5(1.5)/6.3(5.0) |

Maximum power in bold.

by the presence of many non-associated variants to be tested. Second, rather than pooling all the variants into a single group or two groups, TLP-SG automatically determines the number of groups to be formed based on the given data. Furthermore, since TLP-SG shrinks the effects sizes $|\beta_i|$, not $\beta_i$, toward each other, it is robust to varying association directions of the causal variants. However, based on our studies on both simulated and real sequence data, we found that TLP-SG and other penalized.
methods sometimes might be more powerful than some existing global tests, though they do not always outperform the SSU or SSUw test. The discovery of no uniform gain of penalized methods over existing global tests is interesting and even surprising, which as TLP aims to reduce the biases of large coefficient estimates when fitting with many permuted datasets, to save computing time, we only searched relatively few grid points for the tuning parameters, which might not have covered some suitable tuning parameter values. These are all issues to be addressed in the future.

Another non-convex penalty is SCAD (Fan and Li, 2001), methods; in particular, it is unclear how to count the effective number of parameters in the AIC. Alternatively, one may want to try a more popular model selection method, multi-fold cross-validation. However, for RVs as considered here, if we divide the data into multiple folds, the training data may contain several monomorphic variants, causing non-identifiability of their corresponding effect sizes. Second, due to the repeated model-fitting with many permuted datasets, to save computing time, we only searched relatively few grid points for the tuning parameters, which might not have covered some suitable tuning parameter values. These are all issues to be addressed in the future.

### Table 4 | Means, sd’s and MSEs of some causal ($\beta_{cs}$) and non-causal ($\beta_{ncs}$) variants’ regression coefficient estimates when $k = 30$ in Case 2 with both RVs and CVs.

| Methods     | $\beta_{cs} = 1.5$ |       | $\beta_{cs} = 1.5$ |       | $\beta_{ncs} = 0$ |       |
|-------------|-------------------|-------|-------------------|-------|-------------------|-------|
|             | Mean        | sd     | MSE       | Mean        | sd     | MSE       |
| OLS         | 1.59        | 1.26   | 3.16      | 1.54        | 1.53   | 4.69      |
| Lasso       | 0.93        | 0.87   | 1.82      | 0.84        | 0.81   | 1.74      |
| gflasso$_{cor}$ | 0.88    | 0.93   | 2.11      | 0.80        | 0.86   | 1.96      |
| gflasso$_{A1}$ | 0.83   | 0.92   | 2.15      | 0.74        | 0.86   | 2.05      |
| TLP-S       | 1.35        | 1.14   | 2.60      | 1.25        | 1.15   | 2.70      |
| TLP-SG      | 1.28        | 1.16   | 2.72      | 1.29        | 1.15   | 2.70      |

### Table 5 | MAFs (%) and pair-wise correlations (COR) in the values of (min, mean, max) for the 12 genes influencing the quantitative trait Q2 in the GAW17 data.

| Gene       | All       | Causal     | Non-causal |
|------------|-----------|------------|------------|
|           | MAF       | COR        |            |
| PLAT       | (0.072,0.098,45.12) | (0.072,0.206,0.574) | (0.072,2.855,45.12) |
| SREBF1     | (0.072,0.699,7.747)  | (0.072,0.222,0.43)  | (0.072,1.04,7.747)  |
| SIRT1      | (0.072,0.856,16.71)  | (0.072,0.12,0.215)  | (0.072,1.33,16.71)  |
| VLDLR      | (0.072,1.047,9.469)  | (0.072,0.126,0.287) | (0.072,1.435,9.469) |
| VNN3       | (0.072,4.429,40.53)  | (0.072,2.065,14.56) | (0.072,6.303,40.53) |
| PDGFD      | (0.072,4.115,31.56)  | (0.072,2.287,0.861) | (0.072,6.501,31.56) |
| BCHE       | (0.072,0.625,14.56)  | (0.072,0.155,0.287) | (0.072,1.076,14.56) |
| INSIG1     | (0.072,0.775,3.587)  | (0.072,0.072,0.072) | (0.072,1.829,3.587) |
| LPL        | (0.072,1.854,14.490) | (0.072,0.598,1.578) | (0.072,2.076,14.490) |
| RARB       | (0.072,0.352,1.363)  | (0.072,0.287,0.502) | (0.072,0.367,1.363) |
| VNN1       | (0.072,2.675,17.070) | (0.574,8.824,17.070) | (0.072,0.215,0.359) |
| VWF        | (0.072,0.944,2.080)  | (0.072,0.323,0.574)  | (0.359,1.265,2.080)  |
| PLAT       | (−0.143,0.002,0.753) | (−0.088,−0.003,−0.001) | (−0.143,0.007,0.753) |
| SREBF1     | (−0.038,0.007,0.635) | (−0.009,−0.004,−0.001) | (−0.038,0.024,0.635) |
| SIRT1      | (−0.044,0.004,0.707) | (−0.004,0.003,0.33)  | (−0.044,0.002,0.499) |
| VLDLR      | (−0.135,−0.001,0.331) | (−0.003,−0.002,−0.001) | (−0.135,0.001,0.331) |
| VNN3       | (−0.422,−0.002,0.59)  | (−0.104,−0.01,0.072)  | (−0.422,−0.001,0.341) |
| PDGFD      | (−0.156,−0.007,0.276) | (−0.007,−0.004,−0.001) | (−0.156,−0.007,0.276) |
| BCHE       | (−0.044,0.001,0.499)  | (−0.005,0.004,0.499)  | (−0.044,−0.002,0.075) |
| INSIG1     | (−0.010,0.009,0.128)  | (−0.001,−0.001,−0.001) | (0.128,0.128,0.128)  |
| LPL        | (−0.138,−0.002,0.215) | (−0.010,−0.006,−0.002) | (−0.138,−0.002,0.215) |
| RARB       | (−0.025,−0.003,0.073) | (−0.004,−0.004,−0.004) | (−0.025,−0.005,−0.001) |
| VNN1       | (−0.046,0.038,0.945)  | (0.055,0.055,0.555)  | (−0.005,0.091,0.945)  |
| VWF        | (0.113,0.316,0.564)   | (0.265,0.265,0.265)   | (0.127,0.246,0.466)   |

www.frontiersin.org
estimates resulting from the Lasso or $L_1$ penalty. Although SCAD can be equally applied and compared here, we chose the TLP as a representative of non-convex penalties for its good properties: as shown by Shen et al. (2012), $L_q$ regularization is optimal in variable selection, and its computational surrogate, TLP, shares the same property for sufficiently small tau; furthermore, the variable selection consistency of TLP regularization also led to enhanced parameter estimation and prediction in numerical studies with finite sample sizes. Nevertheless, we note that, penalized regression methods have been intensively studied for high-dimensional data, but not for the type of data considered here, which are low dimensional but with RVs as sparse predictors.

In summary, the established benefit of penalized regression for variable selection and risk prediction for high-dimensional data (Kooperberg et al., 2010) did not seem to directly translate into substantial power gains in genetic association testing. In addition to the current work, there exist three recent reports (Croiseau and Cordell, 2009; Martinez et al., 2010; Basu et al., 2011) questioning the effectiveness of the Lasso penalized regression in hypothesis testing, while Basu et al. (2011) showed that several variable selection approaches did not outperform some global tests (e.g., the SSU or SSUw test) for association analysis of CVs. Due to the limitations mentioned above, we cannot conclude here that any penalized regression method would not outperform exiting global association tests; rather, further investigation on enhanced tuning parameter selection and better choice of the test statistic is warranted. Finally, we note that the capability of variable selection by penalized regression can be useful, e.g., in narrowing down causal variants.

| Model fitting Test stats | Gene(k. $nC$) | BCHE | INSIG1 | LPL | RARB | VNN1 | VWF |
|--------------------------|---------------|------|--------|-----|------|------|-----|
|                          |               | (28,13) | (5,3) | (20,3) | (11,2) | (7,2) | (6,2) |
| OLS F-test               | 0.375         | 0.065  | 0.305  | 0.135 | 0.750 | 0.110 |
| OLS Score                | 0.365         | 0.065  | 0.295  | 0.135 | 0.740 | 0.110 |
| OLS SSU                  | 0.400         | 0.090  | 0.050  | 0.100 | 0.945 | 0.170 |
| OLS SSUw                 | 0.405         | 0.055  | 0.300  | 0.130 | 0.715 | 0.210 |
| OLS UminP                | 0.300         | 0.060  | 0.285  | 0.110 | 0.820 | 0.170 |
| OLS Sum                  | 0.180         | 0.080  | 0.030  | 0.145 | 0.925 | 0.210 |
| OLS aSum                 | 0.120         | 0.100  | 0.090  | 0.145 | 0.935 | 0.210 |
| Lasso 1df                | 0.315         | 0.050  | 0.205  | 0.135 | 0.655 | 0.090 |
| gfllasso1 = cor          | 0.300         | 0.055  | 0.220  | 0.120 | 0.720 | 0.110 |
| gfllasso0 = 1            | 0.300         | 0.055  | 0.215  | 0.125 | 0.695 | 0.110 |
| TLP-S 1df                | 0.355         | 0.060  | 0.270  | 0.160 | 0.720 | 0.110 |
| TLP-SG 1df               | 0.135         | 0.080  | 0.115  | 0.999 | 0.665 | 0.080 |
| TLP-SG SSU               | 0.045         | 0.110  | 0.040  | 0.070 | 0.945 | 0.140 |
| TLP-SG SSUw              | 0.155         | 0.075  | 0.135  | 0.085 | 0.675 | 0.145 |

Maximum power in bold.
Table 8 | Mean numbers of TP(sd)/FP(sd) in the GAW17 data, where q1 and q0 denote the numbers of the causal and non-causal variants in each gene.

| Gene(q1/q0) | OLS | Lasso | gfllasso<sub>r = cor</sub> | gfllasso<sub>r = 1</sub> | TLP-S | TLP-SG |
|-------------|-----|-------|--------------------------|--------------------------|-------|-------|
| PLAT(8/20)  | 8.0(0.0)/20.0(0.2) | 0.8(1.4)/21.2(7.7) | 0.6(1.5)/19.2(6.2) | 1.5(2.8)/38.6(6.4) | 5.0(1.6)/11.4(2.9) | 1.7(2.1)/3.9(4.5) |
| SREBF1(10/14) | 10.0(0.0)/14.0(0.1) | 2.1(2.6)/2.6(3.0) | 2.4(3.1)/3.5(3.9) | 6.5(4.4)/8.6(1.6) | 6.6(1.6)/8.5(2.4) | 2.6(1.2)/3.1(2.2) |
| SIRT1(9/14)  | 9.0(0.0)/14.0(0.1) | 2.0(1.9)/2.5(2.2) | 1.9(2.3)/3.1(2.3) | 4.9(3.7)/6.9(5.8) | 5.0(1.6)/7.2(2.5) | 2.0(1.2)/2.3(1.3) |
| VLDLR(8/19)  | 8.0(0.0)/19.0(0.0) | 0.8(1.5)/2.7(3.0) | 0.8(1.7)/2.6(3.3) | 2.1(2.2)/4.5(7.1) | 5.0(1.5)/11.7(2.6) | 1.5(1.7)/3.8(7.3) |
| VNN3(7/8)   | 7.0(0.0)/8.0(0.2) | 2.7(1.3)/2.5(1.7) | 3.5(1.6)/3.9(2.1) | 4.4(2.1)/4.6(2.6) | 5.0(1.2)/5.6(1.4) | 2.8(1.4)/2.3(1.9) |
| PDGFDG(4/7)  | 4.0(0.0)/7.0(0.2) | 1.2(1.1)/2.3(1.9) | 2.1(1.3)/4.2(2.0) | 2.5(1.4)/4.4(2.2) | 2.9(1.0)/5.3(1.3) | 1.2(1.1)/2.0(2.0) |
| BCHE(13/15)  | 13.0(0.1)/15.0(0.1) | 3.4(2.0)/2.8(2.8) | 3.0(3.5)/1.3(2.5) | 7.1(5.2)/7.7(6.0) | 7.2(1.7)/7.5(2.3) | 4.1(3.1)/3.4(3.4) |
| INSIG1(3/2)  | 3.0(0.0)/2.0(0.0) | 0.2(0.6)/0.4(0.6) | 1.0(1.1)/1.0(1.7) | 0.8(1.2)/1.0(1.7) | 1.6(0.8)/1.4(0.5) | 0.7(1.1)/0.7(0.8) |
| LPL(3/17)   | 3.0(0.0)/16.0(2.0) | 1.0(0.8)/2.9(3.0) | 1.1(0.9)/4.0(4.1) | 1.4(1.0)/5.4(5.7) | 2.5(0.7)/10.8(2.4) | 1.2(0.8)/3.2(3.4) |
| RARB(2/9)   | 2.0(0.0)/9.0(0.1) | 0.7(0.7)/1.2(1.8) | 0.8(0.7)/2.1(2.7) | 1.0(0.8)/3.2(3.6) | 1.6(0.5)/5.1(1.7) | 0.8(0.7)/1.4(1.8) |
| VNN1(2/5)   | 2.0(0.0)/5.0(0.0) | 1.5(0.5)/0.6(1.0) | 1.8(0.4)/2.4(1.7) | 1.7(0.5)/1.7(1.8) | 1.9(0.3)/2.8(1.3) | 1.5(0.5)/1.1(1.7) |
| VWF(2/4)    | 2.0(0.0)/4.0(0.1) | 0.2(0.5)/1.0(1.1) | 1.0(0.8)/2.7(1.2) | 1.0(0.8)/2.7(1.2) | 1.5(0.6)/3.5(0.7) | 0.4(0.6)/1.2(1.2) |

**ACKNOWLEDGMENTS**

We thank the editor and reviewers for many helpful and constructive comments and suggestions. This research was supported by NIH grants R01HL65462, R01HL105397, R01GM081535, and R01HL116720, and by the Minnesota Supercomputing Institute.

**REFERENCES**

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723. doi: 10.1109/TAC.1974.110705

Almasy, L. A., Dyer, T. D., Peralta, J. M., Kent, J. W. Jr., Charlesworth, J. C., Curran, J. E., et al. (2011). Genetic analysis workshop p 17 mini-exome simulation. *BMC Proc.* 5(suppl 9):S2. doi: 10.1186/1753-6561-5-S9-S2

Austin, E., Pan, W., and Shen, X. (2013). Penalized regression and risk prediction in Genome-Wide association studies. *Stat. Anal. Data Min.* 6, 315–328. doi: 10.1002/sam.11183

Basu, S., and Pan, W. (2011). Comparison of statistical tests for disease association with rare variants. *Genet. Epidemiol.* 35, 606–619. doi: 10.1002/gepi.20609

Basu, S., Pan, W., Shen, X., and Oetting, W. (2011). Multi-locus association testing with penalized regression. *Genet. Epidemiol.* 35, 755–765. doi: 10.1002/gepi.20625

Chapman, J. M., and Whittaker, J. (2008). Analysis of multiple SNPs in a candidate gene or region. *Genet. Epidemiol.* 32, 560–566. doi: 10.1002/gepi.20330

Conneely, K. N., and Boehnke, M. (2007). So many correlated tests, so little time! *Genet. Epidemiol.* 32, 560–566. doi: 10.1002/gepi.20625

Kim, S., and Xing, E. P. (2009). Generalized linear modeling with regularization for detecting common disease rare haplotype association. *Genet. Epidemiol.* 33, 308–316. doi: 10.1002/gepi.20382

Han, E., and Pan, W. (2010). A data-adaptive sum test for disease association with multiple common or rare variants. *Hum. Hered.* 70, 42–54. doi: 10.1159/000288704

Kim, S., and Xing, E. P. (2009). Statistical estimation of correlated genome associations to a quantitative trait network. *PLoS Genet.* 5:e1000587. doi: 10.1371/journal.pgen.1000587

Kooperberg, C., LeBlanc, M. L., and Obenchain, V. (2010). Risk prediction using genome-wide association studies. *Genet. Epidemiol.* 34, 643–652. doi: 10.1002/gepi.20509

Kotowski, I. K., Pertsemidis, A., Luke, A., Cooper, R. S., Vega, G. L., Cohen, J. C., et al. (2006). A spectrum of PCSK9 alleles contributes to plasma levels of low density lipoprotein cholesterol. *Am. J. Hum. Genet.* 78, 410–422. doi: 10.1086/500615

Kim et al. Penalized regression for rare variant

Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* 96, 1348–1360. doi: 10.1198/016214501753382273

Guo, W., and Lin, S. (2009). Generalized linear modeling with regularization for detecting common disease rare haplotype association. *Genet. Epidemiol.* 33, 308–316. doi: 10.1002/gepi.20382

Madsen, B. E., and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 5:e1000384. doi: 10.1371/journal.pgen.1000384

Maher, B. (2008). Personal genomes: the case of the missing heritability. *Nature* 456, 18–21. doi: 10.1038/456018a

www.frontiersin.org May 2014 | Volume 5 | Article 121 | 11
Malo, N., Libiger, O., and Schork, N. J. (2008). Accommodating linkage disequilibrium in genetic-association analyses via ridge regression. *Am. J. Hum. Genet.* 82, 375–385. doi: 10.1016/j.ajhg.2007.10.012

Martinez, J. G., Carroll, R. J., Muller, S., Sampson, J. N., and Chatterjee, N. (2010). A note on the effect on power of score tests via dimension reduction by penalized regression under the null. *Int. J. Biostat.* 6, 12. doi: 10.2202/1557-4679.1231

Pan, W. (2009). Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet. Epidemiol.* 33, 497–507. doi: 10.1002/gepi.20402

Roeder, K., Bacanu, S. A., Song, V., Zhang, X., and Devlin, B. (2005). Analysis of single-locus tests to detect gene/disease associations. *Genet. Epidemiol.* 28, 207–219. doi: 10.1002/gepi.20050

Shen, X., and Huang, H.-C. (2010). Grouping pursuit through a regularization solution surface. *J. Am. Stat. Assoc.* 105, 727–739. doi: 10.1198/jasa.2010. tm09380

Shen, X., Pan, W., and Zhu, Y. (2012). Likelihood-based selection and sharp parameter estimation. *J. Am. Stat. Assoc.* 107, 223–232. doi: 10.1080/01621459.2011.645783

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

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. B* 67, 91–108. doi: 10.1111/j.1467-9868.2005.00490.x

Tzeng, J. Y., and Bondell, H. D. (2010). A comprehensive approach to haplotype specific analysis via penalized likelihood. *Eur. J. Hum. Genet.* 18, 95–103. doi: 10.1038/ejhg.2009.118

Wang, T., and Elston, R. C. (2007). Improved power by use of a weighted score test for linkage disequilibrium mapping. *Am. J. Hum. Genet.* 80, 353–360. doi: 10.1086/511312

Yang, S., Yuan, L., Lai, Y.-C., Shen, X., Wonka, P., and Ye, J. (2012). “Feature grouping and selection over an undirected graph,” in *KDD’12 Proceedings of the 18th ACM SIGKDD International Conference On Knowledge Discovery and Data Mining* (SIGKDD 2012) (New York, NY), 922–930. doi: 10.1145/233950.30239675

Zhou, H., Alexander, D. H., Sehl, M. E., Sinsheimer, J. S., Sobel, E. M., and Lange, K. (2011). Penalized regression for genome-wide association screening of sequence data. *Pac. Symp. Biocomput.* 2011, 106–117. doi: 10.1142/9789814335058_0012

Zhu, Y., Shen, X., and Pan, W. (2013). Simultaneous grouping pursuit and feature selection in regression over an undirected graph. *J. Am. Statist. Assoc.* 108, 713–725. doi: 10.1080/01621459.2013.770704

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2014; accepted: 18 April 2014; published online: 13 May 2014.

Citation: Kim S, Pan W and Shen X (2014) Penalized regression approaches to testing for quantitative trait-rare variant association. *Front. Genet.* 5:121. doi: 10.3389/fgene.2014.00121

This article was submitted to Statistical Genetics and Methodology, a section of the journal Frontiers in Genetics.

Copyright © 2014 Kim, Pan and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.