VCSEL: PRIORITIZING SNP-SET BY PENALIZED VARIANCE COMPONENT SELECTION

By Juhyun Kim[1], Judong Shen[2,‡], Anran Wang[2,§],
Devan V. Mehrotra[2,¶], Seyoon Ko[1,†], Jin J. Zhou[3] and Hua Zhou[1,*]

[1]*Department of Biostatistics, University of California Los Angeles, Los Angeles, California, USA, juhkim111@ucla.edu; [*]huazhou@ucla.edu; [†]kos@ucla.edu*

[2]*Biostatistics and Research Decision Sciences, Merck & Co., Inc., Kenilworth, NJ, USA, [‡]judong.shen@merck.com; [§]anran.wang@merck.com; [¶]devan_mehrotra@merck.com*

[3]*Department of Epidemiology and Biostatistics, University of Arizona, Tucson, Arizona, USA, jzhou@arizona.edu*

Single nucleotide polymorphism (SNP) set analysis aggregates both common and rare variants and tests for association between phenotype(s) of interest and a set. However, multiple SNP-sets such as genes, pathways, or sliding windows are usually investigated across the whole genome, in which all groups are tested separately followed by multiple testing adjustments. We propose a novel method to prioritize SNP-sets in a joint multivariate variance component model. Each SNP-set corresponds to a variance component (or kernel), and model selection is achieved by incorporating either convex or non-convex penalties. The uniqueness of this variance component selection framework, which we call VCSEL, is that it naturally encompasses multivariate traits (VCSEL-M) and SNP-set-treatment or -environment interactions (VCSEL-I). We devise an optimization algorithm scalable to many variance components based on the majorization-minimization (MM) principle. Simulation studies demonstrate the superiority of our methods in model selection performance, as measured by the area under the Precision-Recall (PR) curve, compared to the commonly used marginal testing and group penalization methods. Finally, we apply our methods to a real pharmacogenomics study and a real whole exome sequencing study. Some top ranked genes by VCSEL are detected as insignificant by the marginal test methods, which emphasizes formal inference of individual genes with a strict significance threshold. This provides alternative insights for biologists to prioritize follow-up studies and develop polygenic risk score models.

1. Introduction.

The limited success of genome-wide association studies (GWAS) has diverted attention away from common genetic variants, usually denoted by minor allele frequency (MAF) > 0.05. Instead, rare variants (MAF ≤ 0.05) are believed to play an important role in elucidating many common diseases and complex traits (Bodmer and Bonilla, 2008; Manolio et al., 2009; Bansal et al., 2010; Rivas et al., 2011; Gibson, 2012; Gudmundsson et al., 2012; Zuk et al., 2014; Lee et al., 2014). Although association test for common variants in a GWAS analysis is often conducted one variant at a time, this approach results in low statistical power in rare-variant association studies due to their prevalence and extremely low frequency (Li and Leal, 2008; Madsen and Browning, 2009; Zuk et al., 2014). As a remedy, many have proposed single nucleotide polymorphism (SNP) set analysis, also known as gene set, pathway, or region-based analysis (Wu et al., 2010; Dering et al., 2011). In these analyses, variants are binned into a biologically relevant unit such as a gene, pathway, or sliding window, and tested for association with complex traits. Compared to the classical

Keywords and phrases: rare variants, majorization-minimization (MM), penalized estimation, variance components model, restricted maximum likelihood (REML), group selection, nonconvex penalties, multiple phenotypes

single-variant-based approach, SNP-set analysis enjoys increased power as it reduces multiple comparison burden and aggregates weak signals (Rivas and Moutsianas, 2015).

In addition to the high polygenicity—influenced by a large number of genetic variants with small effects—many complex traits are inherently multi-phenotypic. For example, blood pressure is evaluated by both systolic and diastolic pressure measurements. Obesity is determined not only by body mass index but also by waist circumference and body fat percentage. As one indicator may reveal one susceptibility gene over other indicators, it is important to jointly analyze multiple phenotype data in the analysis (Suo et al., 2013). In addition, GWAS have unveiled that many loci affect more than one trait or disease —a phenomenon known as pleiotropy (Sivakumaran et al., 2011; Solovieff et al., 2013). Testing one phenotype at a time, albeit simple and intuitive, fails to exploit the underlying shared genetic architecture of multiple phenotypes and is also subject to multiple testing penalties. On the other hand, multi-trait analyses can increase statistical power to detect association and provide important insights into pathways that certain traits or diseases share (Suo et al., 2013; Hackinger and Zeggini, 2017).

A plethora of marginal test based methods are available to detect associations of a SNP-set with multiple traits, which are termed cross-phenotype associations. For example, Maity, Sullivan and Tzeng (2012); Lee et al. (2017); Wu and Pankow (2016); Broadaway et al. (2016); Zhan et al. (2017); Dutta et al. (2019) take region-based approaches, in which variants are grouped based on pre-specified criteria and tested for cross-phenotype effects. Notably, Multi-SKAT (Dutta et al., 2019) provides a general mixed effect model-based framework for joint analysis of multiple continuous phenotypes, unlike most methods that make specific assumptions about the effects of the variants on multiple phenotypes. However, to our best knowledge, no existing methods investigate sets of genetic variants simultaneously.

Here we propose a method for jointly modeling multiple SNP-sets and selecting groups that are relevant to multiple traits while adjusting for covariates. Suppose we have observations from n individuals with d continuous phenotypes, represented by $n \times d$ matrix, and m SNP-sets. Multivariate response model with $n \times d$ response matrix $\widetilde{\boldsymbol{Y}}$ and $n \times p$ covariate matrix $\boldsymbol{X}$ assumes a multivariate normal model

$$(1) \qquad \operatorname{vec} \widetilde{\boldsymbol{Y}} \sim N(\operatorname{vec}(\boldsymbol{XB}), \boldsymbol{\Sigma}_1 \otimes \widetilde{\boldsymbol{V}}_1 + \cdots + \boldsymbol{\Sigma}_m \otimes \widetilde{\boldsymbol{V}}_m + \boldsymbol{\Sigma}_0 \otimes \boldsymbol{I}_n),$$

where $\boldsymbol{B}$ is the unknown $p \times d$ fixed effects parameters matrix, $\boldsymbol{\Sigma}_i$ are unknown $d \times d$ positive semidefinite variance component matrices, and $\widetilde{\boldsymbol{V}}_i$ are known $n \times n$ kernel matrices for genotypes. The $\operatorname{vec} \widetilde{\boldsymbol{Y}}$ operator in (1) creates an $nd \times 1$ vector from a matrix $\widetilde{\boldsymbol{Y}}$ by stacking its column vectors, and $\otimes$ indicates Kronecker product.

As our interest lies in estimating variance components, we adopt the restricted (or residual) maximum likelihood estimation (REML) approach (Thompson et al., 1962; Patterson and Thompson, 1971; Harville, 1977; Khuri and Sahai, 1985; Robinson, 1987; Searle, Casella and McCulloch, 1992). In the notation of (1), REML first projects $\widetilde{\boldsymbol{Y}}$ to the null space of $\boldsymbol{X}$ and then estimates variance components based on the projected responses. If the columns of the matrix $\boldsymbol{A}$ span the null space of $\boldsymbol{X}^T$ and $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$, then REML estimates parameter $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m)$ by maximizing the log-likelihood of the redefined response matrix $\boldsymbol{Y} = \boldsymbol{A}^T\widetilde{\boldsymbol{Y}}$ whose distribution is as follows:

$$(2) \qquad \operatorname{vec} \boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_1 \otimes \boldsymbol{V}_1 + \cdots + \boldsymbol{\Sigma}_m \otimes \boldsymbol{V}_m + \boldsymbol{\Sigma}_0 \otimes \boldsymbol{I}_{n-p}),$$

where $\boldsymbol{V}_i = \boldsymbol{A}^T\widetilde{\boldsymbol{V}}_i\boldsymbol{A}, i = 1, \ldots, m$. Note that fixed effects have been eliminated.

As there are no closed-form expressions for the REML, we rely on numerical techniques. There are several iterative optimization methods for finding MLE and REML, including Newton's method (Lindstrom and Bates, 1988), Fisher's scoring algorithm, and the expectation-maximization (EM) algorithm (Dempster, Laird and Rubin, 1977; Laird and Ware, 1982;

Laird, Lange and Stram, 1987; Lindstrom and Bates, 1988; Bates and Pinheiro, 1998). Despite their respective advantages, they suffer from either numerical instability, high computational cost or slow convergence. Zhou et al. (2019) address this issue with a minorization-maximization (MM) algorithm that is simple to implement and numerically efficient. Zhai et al. (2018) implements an MM algorithm for penalizing variance components in microbiome data analysis, but it is limited to lasso penalty and a univariate response. The recent paper (Schaid et al., 2020) applies a similar method as Zhai et al. (2018) to the genetic association setting, but still restricted to the univariate response setting.

Since SNPs within a gene/pathway/moving window are treated as a unit, this can be considered a group selection problem with each set being a group and SNP being a variable. Several methods have been proposed to take advantage of grouping structures in variables. Group lasso method (Bakin, 1999; Yuan and Lin, 2006) allows group selection by either including or excluding all variables in the group in the model. Bi-level selection or sparse group method (Huang et al., 2009; Breheny and Huang, 2009; Zhou et al., 2010; Simon et al., 2013) enables both group-wise and within group sparsity. However, these approaches are designed for selecting mean, or fixed effects, hence inappropriate when genetic effects are modeled as random effects.

There exists a considerable body of literature on random effect selection. Lin (1997) proposes score tests to detect the significance of individual variance components. To select important random effects, each component is tested separately, followed by some stepwise procedures. Chen and Dunson (2003), Bondell, Krishna and Ghosh (2010), Fan and Li (2012), and Peng and Lu (2012) consider random effect selection for longitudinal models where observations are divided into independent subjects with a vector of random effects corresponding to each subject. The vectors of random effect are independent and identically distributed with a covariance matrix, which could be a function of one variance component. For these methods, selecting important random effects is essentially limited to within one variance component as it removes rows or columns of covariance matrix or selects components within random effect vectors. No existing method performs a simultaneous selection of random effects at group level to our best knowledge.

Our contributions herein are three-fold. First is developing a novel penalization method for group selection where each group is treated as random effects. Our second contribution is that we devise a general MM-based optimization framework that incorporates both convex and non-convex penalties into variance component models and applies to the analysis of univariate and multivariate traits, respectively. Lastly, we outline an algorithm to incorporate SNP-set-by-treatment or SNP-set-by-environment interaction terms in a univariate trait variance component model, motivated by pharmacogenomic studies.

The remainder of this paper is organized as follows: Section 2 introduces the multivariate response variance component model. In Section 3, we present the VCSEL algorithm that selects variance components in the realm of multivariate response (VCSEL-M). Section 4 extends the algorithm to incorporate interaction terms (VCSEL-I) for a univariate response model. We illustrate the performance of our methods with simulation studies in Section 5 for VCSEL-M and VCSEL-I methods and defer the details for the univariate response VCSEL methods to the Supplementary Materials. In Section 6, the proposed methods are applied to two real datasets: a UK-biobank whole exome sequencing study data and a pharmacogenomic study data. We conclude the paper with a discussion and future research directions in Section 7.

2. Multivariate response variance component model.

Consider the model (2) where $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_m$ are known positive semidefinite matrices. Here $\boldsymbol{V}_i$ is a genotype kernel matrix for the i-th variance component. Different choices of kernels can be readily incorporated in

$\boldsymbol{V}_i$. As defined in Dutta et al. (2019), one popular choice would be $\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{W}_i\boldsymbol{G}_i^T$ where $\boldsymbol{G}_i$ is a genotype matrix corresponding to i-th SNP group and $\boldsymbol{W}_i = \operatorname{diag}(w_1, \ldots, w_q)$ contains the weights of q variants in $\boldsymbol{G}_i$. It corresponds to SKAT and implies that the effects of SNPs in i-th SNP-set are independent. Another choice is $\boldsymbol{G}_i\boldsymbol{W}_i\mathbf{1}\mathbf{1}^T\boldsymbol{W}_i\boldsymbol{G}_i^T$, which corresponds to the Burden test and implies that the effects of SNPs in i-th SNP set are in the same direction. Note that $\mathbf{1}$ denotes a vector of ones. In our simulation studies and real data analysis, we adopt the SKAT genotype kernel and/or the Burden test genotype kernel.

We denote the overall covariance matrix in the model by $\boldsymbol{\Omega}$, i.e.

$$\boldsymbol{\Omega}(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{V}_1 + \cdots + \boldsymbol{\Sigma}_m \otimes \boldsymbol{V}_m + \boldsymbol{\Sigma}_0 \otimes \boldsymbol{I}_{n-p},$$

and assume it to be positive definite. To find estimates of $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m)$, we take a penalization approach by minimizing the penalized negative log-likelihood function

$$\begin{aligned} & - L(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m) + \sum_{i=1}^m P_\lambda(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}_i)}) \\ & = \frac{1}{2}\ln\det\boldsymbol{\Omega} + \frac{1}{2}(\operatorname{vec}\boldsymbol{Y})^T\boldsymbol{\Omega}^{-1}\operatorname{vec}\boldsymbol{Y} + \sum_{i=1}^m P_\lambda(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}_i)}), \end{aligned} \tag{3}$$

where P_λ is a penalty term imposing sparsity on variance components for a given tuning parameter λ. Below we derive iterative procedures for lasso (Tibshirani, 1996) and minimax concave penalty (MCP) (Zhang et al., 2010); only a slight modification is needed to accommodate other penalty functions. In practice, we normalize $\boldsymbol{V}_i$ to have unit Frobenius norm to put the kernel matrices on the equal footing in penalty because the varying number of variants involved in each $\boldsymbol{V}_i$ leads to higher magnitude for sets with a large number of variants compared to those with a small number of variants.

While $\boldsymbol{V}_i$ measures genetic similarity between subjects in the i-th SNP group and is assumed fully known, it is worthwhile noting that no assumptions have been made about $\boldsymbol{\Sigma}_i$, which resides in the phenotype space and reflects how effect sizes of each variant on each phenotype are correlated. Different choices of $\boldsymbol{\Sigma}_i$ have been proposed in Dutta et al. (2019). If one does have *a priori* knowledge about phenotype structure, the algorithm simplifies to the univariate case. For example, if effect sizes of each variant in a SNP-set on different phenotypes are assumed homogeneous, we may write $\boldsymbol{\Sigma}_i = \sigma_i^2\mathbf{1}_d\mathbf{1}_d^T$, where σ_i^2 is a scalar-valued i-th variance component and $\mathbf{1}_d$ is a $d \times 1$ vector of 1's. Then $\boldsymbol{\Omega} = \sum_{i=1}^m \sigma_i^2(\mathbf{1}_d\mathbf{1}_d^T \otimes \boldsymbol{V}_i) + \sigma_0^2(\mathbf{1}_d\mathbf{1}_d^T \otimes \boldsymbol{I}_{n-p})$, where σ_0^2 is a scalar-valued residual variance component. Since $(\mathbf{1}_d\mathbf{1}_d^T \otimes \boldsymbol{V}_i)$ is a known covariance matrix for i-th group, the problem amounts to estimating $\sigma_i^2, i = 0, 1, \ldots, m$.

3. Estimation algorithm.

The MM principle involves majorizing the objective function $f(\boldsymbol{\theta})$ by a surrogate function $g(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ around the current iterate $\boldsymbol{\theta}^{(t)}$ of a search (Lange, Hunter and Yang, 2000; Hunter and Lange, 2004; Lange, 2016). The superscript t indicates the iteration number. Majorization is defined by the following two conditions

$$\begin{aligned} f(\boldsymbol{\theta}^{(t)}) &= g(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}^{(t)}) \\ f(\boldsymbol{\theta}) &\leq g(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}), \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}^{(t)}. \end{aligned}$$

In other words, the surface $\boldsymbol{\theta} \mapsto g(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ lies above the surface $\boldsymbol{\theta} \mapsto f(\boldsymbol{\theta})$ and is tangent to it at the point $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$. Construction of the majorizing function $g(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ constitutes the first M of the MM algorithm. The second M of the algorithm minimizes the surrogate $g(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$

rather than $f(\boldsymbol{\theta})$. If $\boldsymbol{\theta}^{(t+1)}$ denotes the minimizer of $g(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$, then this action forces the descent property $f(\boldsymbol{\theta}^{(t+1)}) \le f(\boldsymbol{\theta}^{(t)})$. This fact follows from the inequalities

$$f(\boldsymbol{\theta}^{(t+1)}) \le g(\boldsymbol{\theta}^{(t+1)} \mid \boldsymbol{\theta}^{(t)}) \le g(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}^{(t)}) = f(\boldsymbol{\theta}^{(t)}),$$

reflecting the definition of $\boldsymbol{\theta}^{(t+1)}$ and the tangency condition. Monotonicity of MM iterates obliterates the need for line search and lends itself to the remarkable numerical stability of the MM algorithm.

We derive a majorizing function of the penalized loss function (3) by working on its three individual terms separately. For the penalty term, we first specialize to the lasso penalty then indicate the generalizations to other penalties.

1. Log-determinant term. The concavity of the map $\boldsymbol{X} \mapsto \ln\det \boldsymbol{X}$ and the supporting hyperplane inequality establish the majorization

$$\ln\det \boldsymbol{\Omega}^{(t)} + \mathrm{tr}[\boldsymbol{\Omega}^{-(t)}(\boldsymbol{\Omega} - \boldsymbol{\Omega}^{(t)})] \ge \ln\det \boldsymbol{\Omega}. \tag{4}$$

2. Quadratic form term. When $\boldsymbol{V}_i$ for all i are positive definite, hence invertible, convexity of the matrix function $(\boldsymbol{X}, \boldsymbol{Y}) \mapsto \boldsymbol{X}^T\boldsymbol{Y}^{-1}\boldsymbol{X}$ where $\boldsymbol{Y} \succ \boldsymbol{0}$ implies

$$\begin{aligned}\boldsymbol{\Omega}^{(t)}\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}^{(t)} &= m\left(\frac{1}{m}\sum_{i=0}^{m}\boldsymbol{\Sigma}_i^{(t)} \otimes \boldsymbol{V}_i\right)\left(\frac{1}{m}\sum_{i=0}^{m}\boldsymbol{\Sigma}_i \otimes \boldsymbol{V}_i\right)^{-1}\left(\frac{1}{m}\sum_{i=0}^{m}\boldsymbol{\Sigma}_i^{(t)} \otimes \boldsymbol{V}_i\right)\\ &\preceq m\sum_{i=0}^{m}\frac{1}{m}(\boldsymbol{\Sigma}_i^{(t)} \otimes \boldsymbol{V}_i)(\boldsymbol{\Sigma}_i \otimes \boldsymbol{V}_i)^{-1}(\boldsymbol{\Sigma}_i^{(t)} \otimes \boldsymbol{V}_i)\\ &= \sum_{i=0}^{m}(\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_i^{(t)}) \otimes \boldsymbol{V}_i, \qquad (5)\end{aligned}$$

or equivalently

$$\boldsymbol{\Omega}^{-1} \preceq \boldsymbol{\Omega}^{-(t)}\left[\sum_{i=0}^{m}(\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_i^{(t)}) \otimes \boldsymbol{V}_i\right]\boldsymbol{\Omega}^{-(t)}. \tag{6}$$

For symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} \preceq \boldsymbol{B}$ means $\boldsymbol{B} - \boldsymbol{A}$ is positive semidefinite. The equality (5) follows from the identities $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$ and $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$. The nonsingularity assumption on $\boldsymbol{V}_i$ can be relaxed by substituting $\boldsymbol{V}_{\epsilon,i} = \boldsymbol{V}_i + \epsilon\boldsymbol{I}_n$ for $\boldsymbol{V}_i$ and sending ϵ to 0.

3. Lasso penalty term. The majorization on the lasso penalty

$$\sqrt{\mathrm{tr}\boldsymbol{\Sigma}_i^{(t)}} + \frac{1}{2\sqrt{\mathrm{tr}\boldsymbol{\Sigma}_i^{(t)}}}(\mathrm{tr}\boldsymbol{\Sigma}_i - \mathrm{tr}\boldsymbol{\Sigma}_i^{(t)}) \ge \sqrt{\mathrm{tr}\boldsymbol{\Sigma}_i} \tag{7}$$

follows from the concavity of the map $x \mapsto \sqrt{x}$ and the support hyperplane inequality.

Merging (4), (6) and (7) generates the overall majorizing function

(8)

$$\begin{aligned}g(\boldsymbol{\Sigma} \mid \boldsymbol{\Sigma}^{(t)}) =& \frac{1}{2}\sum_{i=0}^{m}\left\{\mathrm{tr}\left[\boldsymbol{\Omega}^{-(t)}(\boldsymbol{\Sigma}_i \otimes \boldsymbol{V}_i)\right] + (\mathrm{vec}\boldsymbol{R}^{(t)})^T\left[(\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_i^{(t)}) \otimes \boldsymbol{V}_i\right](\mathrm{vec}\boldsymbol{R}^{(t)})\right\}\\ &+ \frac{1}{2}\sum_{i=1}^{m}\frac{\lambda}{\sqrt{\mathrm{tr}\boldsymbol{\Sigma}_i^{(t)}}}\mathrm{tr}\boldsymbol{\Sigma}_i + c^{(t)},\end{aligned}$$

where vec $\boldsymbol{R}^{(t)} = \boldsymbol{\Omega}^{-(t)}\text{vec}(\boldsymbol{Y})$ with $\boldsymbol{R}^{(t)}$ being a matrix of size $n \times d$ and $c^{(t)}$ is a constant impertinent to the parameters $\boldsymbol{\Sigma}_i$. Parameters $\boldsymbol{\Sigma}_i$ are nicely separated in (8) so we only need to minimize m individual functions

$$
\begin{aligned}
g_i^{(t)}(\boldsymbol{\Sigma}_i) &= \frac{1}{2}\left\{ \text{tr}\left[\boldsymbol{\Omega}^{-(t)}(\boldsymbol{\Sigma}_i \otimes \boldsymbol{V}_i)\right] + \text{tr}(\boldsymbol{R}^{(t)T}\boldsymbol{V}_i\boldsymbol{R}^{(t)}\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_i^{(t)}) + \frac{\lambda}{\sqrt{\text{tr}\boldsymbol{\Sigma}_i^{(t)}}}\text{tr}\boldsymbol{\Sigma}_i \right\} \\
&= \frac{1}{2}\left\{ \text{tr}\left[\boldsymbol{\Omega}^{-(t)}(\boldsymbol{\Sigma}_i \otimes \boldsymbol{V}_i)\right] + \text{tr}(\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{R}^{(t)T}\boldsymbol{V}_i\boldsymbol{R}^{(t)}\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{\Sigma}_i^{-1}) + \frac{\lambda}{\sqrt{\text{tr}\boldsymbol{\Sigma}_i^{(t)}}}\text{tr}\boldsymbol{\Sigma}_i \right\}
\end{aligned} \tag{9}
$$

to update $\boldsymbol{\Sigma}_i$. The first equation follows from the Kronecker identities $(\text{vec}\boldsymbol{A})^T\text{vec}\boldsymbol{B} = \text{tr}(\boldsymbol{A}^T\boldsymbol{B})$ and $\text{vec}(\boldsymbol{CDE}) = (\boldsymbol{E}^T \otimes \boldsymbol{C})\text{vec}(\boldsymbol{D})$. The first trace in the second equation of (9) is linear in $\boldsymbol{\Sigma}_i$ with the coefficient of entry $(\boldsymbol{\Sigma}_i)_{jk}$ equal to

$$
\text{tr}(\boldsymbol{\Omega}_{jk}^{-(t)}\boldsymbol{V}_i) = \mathbf{1}_n^T(\boldsymbol{V}_i \odot \boldsymbol{\Omega}_{jk}^{-(t)})\mathbf{1}_n,
$$

where $\boldsymbol{\Omega}_{jk}^{-(t)}$ is the (j,k)-th $n \times n$ block of $\boldsymbol{\Omega}^{-(t)}$ and $\odot$ is the Hadamard (elementwise) product. The matrix $\boldsymbol{M}_i$ of these coefficients can be written as

$$
\boldsymbol{M}_i = (\boldsymbol{I}_d \otimes \mathbf{1}_n)^T[(\mathbf{1}_d\mathbf{1}_d^T \otimes \boldsymbol{V}_i) \odot \boldsymbol{\Omega}^{-(t)}](\boldsymbol{I}_d \otimes \mathbf{1}_n).
$$

Setting the derivative of (9) to zeros yields the stationarity condition

$$
\boldsymbol{M}_i + \frac{\lambda}{\sqrt{\text{tr}\boldsymbol{\Sigma}_i^{(t)}}}\boldsymbol{I}_d = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{R}^{(t)T}\boldsymbol{V}_i\boldsymbol{R}^{(t)}\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{\Sigma}_i^{-1}, \tag{10}
$$

which is a Riccati equation admitting the explicit solution

$$
\boldsymbol{\Sigma}_i^{(t+1)} = \boldsymbol{L}_i^{-(t)T}[\boldsymbol{L}_i^{(t)T}(\boldsymbol{\Sigma}_i^{(t)}\boldsymbol{R}^{(t)T}\boldsymbol{V}_i\boldsymbol{R}^{(t)}\boldsymbol{\Sigma}_i^{(t)})\boldsymbol{L}_i^{(t)}]^{1/2}\boldsymbol{L}_i^{-(t)}
$$

in terms of the Cholesky factor $\boldsymbol{L}_i^{(t)}$ of the matrix on the left hand side of (10).

Algorithm 1 summarizes the MM algorithm for lasso penalized multivariate variance components model (VCSEL-M-lasso). Each iteration computes $m+1$ Cholesky factorizations and symmetric square roots of $d \times d$ positive semidefinite matrices. In most applications, d is a small number. Our convergence criteria are based on the change in objective function (3) (the penalized negative log-likelihood function) values. The procedure is repeated until the relative change in the objective function value is less than a tolerance value ($10^{-6}\times$[|objective function value at the current iterate| $+ 1$] by default). For tuning parameters, we first locate the tuning parameter λ value, after which all the variance component estimates turn zero—which we denote the maximum λ. Then we create a solution path using a set number of equidistant tuning parameter values from 0 to the maximum λ.

Nonconvex penalties reduce the bias by applying less shrinkage to the large nonzero components. As an example, we illustrate with the MCP. An extra tuning parameter $\gamma > 1$ controls the concavity of the penalty. In our case where $\sqrt{\text{tr}(\boldsymbol{\Sigma}_i)}$ is nonnegative, MCP is defined as

$$
P_\gamma(\sqrt{\text{tr}(\boldsymbol{\Sigma}_i)}; \lambda) = \begin{cases} \lambda\sqrt{\text{tr}(\boldsymbol{\Sigma}_i)} - \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{2\gamma}, & \text{if } \sqrt{\text{tr}(\boldsymbol{\Sigma}_i)} \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2, & \text{if } \sqrt{\text{tr}(\boldsymbol{\Sigma}_i)} > \gamma\lambda \end{cases}. \tag{11}
$$

MCP converges to lasso penalty as $\gamma \to \infty$. Derivation of the majorization for MCP is described in detail in the Supplementary Material S.1. Algorithm 2 summarizes the MM algorithm for MCP penalized multivariate response variance component model (VCSEL-M-MCP).

```
Input  : $\boldsymbol{Y}, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_m, \lambda$
Output: $\hat{\boldsymbol{\Sigma}}_0, \hat{\boldsymbol{\Sigma}}_1, \ldots, \hat{\boldsymbol{\Sigma}}_m$
Initialize $\boldsymbol{\Sigma}_i^{(0)}$ positive definite, $i = 1, \ldots, m$
repeat
    $\boldsymbol{\Omega}^{(t)} \leftarrow \sum_{i=1}^m \boldsymbol{\Sigma}_i^{(t)} \otimes \boldsymbol{V}_i + \boldsymbol{\Sigma}_0^{(t)} \otimes \boldsymbol{I}$
    $\boldsymbol{R}^{(t)} \leftarrow \text{reshape}(\boldsymbol{\Omega}^{-(t)} \text{vec} \boldsymbol{Y}, n, d)$
    for $i = 1, \ldots, m$ do
        Cholesky $\boldsymbol{L}_i^{(t)} \boldsymbol{L}_i^{(t)T} \leftarrow (\boldsymbol{I}_d \otimes \boldsymbol{1}_n)^T [(\boldsymbol{1}_d \boldsymbol{1}_d^T \otimes \boldsymbol{V}_i) \odot \boldsymbol{\Omega}^{-(t)}](\boldsymbol{I}_d \otimes \boldsymbol{1}_n) + \frac{\lambda}{\sqrt{\text{tr} \boldsymbol{\Sigma}_i^{(t)}}} \boldsymbol{I}_d$
        $\boldsymbol{\Sigma}_i^{(t+1)} \leftarrow \boldsymbol{L}_i^{-(t)T} [\boldsymbol{L}_i^{(t)T} (\boldsymbol{\Sigma}_i^{(t)} \boldsymbol{R}^{(t)T} \boldsymbol{V}_i \boldsymbol{R}^{(t)} \boldsymbol{\Sigma}_i^{(t)}) \boldsymbol{L}_i^{(t)}]^{1/2} \boldsymbol{L}_i^{-(t)}$
    end
    Cholesky $\boldsymbol{L}_0^{(t)} \boldsymbol{L}_0^{(t)T} \leftarrow (\boldsymbol{I}_d \otimes \boldsymbol{1}_n)^T [(\boldsymbol{1}_d \boldsymbol{1}_d^T \otimes \boldsymbol{I}_n) \odot \boldsymbol{\Omega}^{-(t)}](\boldsymbol{I}_d \otimes \boldsymbol{1}_n)$
    $\boldsymbol{\Sigma}_0^{(t+1)} \leftarrow \boldsymbol{L}_0^{-(t)T} [\boldsymbol{L}_0^{(t)T} (\boldsymbol{\Sigma}_0^{(t)} \boldsymbol{R}^{(t)T} \boldsymbol{R}^{(t)} \boldsymbol{\Sigma}_0^{(t)}) \boldsymbol{L}_0^{(t)}]^{1/2} \boldsymbol{L}_0^{-(t)}$
until objective value converges;
```

Algorithm 1: VCSEL algorithm for lasso penalized multivariate response variance component model (3) (VCSEL-M-lasso).

```
Input  : $\boldsymbol{Y}, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_m, \lambda, \gamma$
Output: $\widehat{\boldsymbol{\Sigma}}_0, \widehat{\boldsymbol{\Sigma}}_1, \ldots, \widehat{\boldsymbol{\Sigma}}_m$
Initialize $\boldsymbol{\Sigma}_i^{(0)}$ positive definite, $i = 1, \ldots, m$ ;
repeat
    $\boldsymbol{\Omega}^{(t)} \leftarrow \sum_{i=1}^m \boldsymbol{\Sigma}_i^{(t)} \otimes \boldsymbol{V}_i + \boldsymbol{\Sigma}_0^{(t)} \otimes \boldsymbol{I}$ ;
    $\boldsymbol{R}^{(t)} \leftarrow \text{reshape}(\boldsymbol{\Omega}^{-(t)} \text{vec} \boldsymbol{Y}, n, d)$ ;
    for $i = 1, \ldots, m$ do
        if $\sqrt{\text{tr}(\boldsymbol{\Sigma}_i^{(t)})} \leq \gamma\lambda$ then
            Cholesky
              $\boldsymbol{L}_i^{(t)} \boldsymbol{L}_i^{(t)T} \leftarrow (\boldsymbol{I}_d \otimes \boldsymbol{1}_n)^T [(\boldsymbol{1}_d \boldsymbol{1}_d^T \otimes \boldsymbol{V}_i) \odot \boldsymbol{\Omega}^{-(t)}](\boldsymbol{I}_d \otimes \boldsymbol{1}_n) + \left( \frac{\lambda}{\sqrt{\text{tr} \boldsymbol{\Sigma}_i^{(t)}}} - \frac{1}{\gamma} \right) \boldsymbol{I}_d$
        else
            Cholesky $\boldsymbol{L}_i^{(t)} \boldsymbol{L}_i^{(t)T} \leftarrow (\boldsymbol{I}_d \otimes \boldsymbol{1}_n)^T [(\boldsymbol{1}_d \boldsymbol{1}_d^T \otimes \boldsymbol{V}_i) \odot \boldsymbol{\Omega}^{-(t)}](\boldsymbol{I}_d \otimes \boldsymbol{1}_n)$;
        end
        $\boldsymbol{\Sigma}_i^{(t+1)} \leftarrow \boldsymbol{L}_i^{-(t)T} [\boldsymbol{L}_i^{(t)T} (\boldsymbol{\Sigma}_i^{(t)} \boldsymbol{R}^{(t)T} \boldsymbol{V}_i \boldsymbol{R}^{(t)} \boldsymbol{\Sigma}_i^{(t)}) \boldsymbol{L}_i^{(t)}]^{1/2} \boldsymbol{L}_i^{-(t)}$
    end
    Cholesky $\boldsymbol{L}_0^{(t)} \boldsymbol{L}_0^{(t)T} \leftarrow (\boldsymbol{I}_d \otimes \boldsymbol{1}_n)^T [(\boldsymbol{1}_d \boldsymbol{1}_d^T \otimes \boldsymbol{I}_n) \odot \boldsymbol{\Omega}^{-(t)}](\boldsymbol{I}_d \otimes \boldsymbol{1}_n)$ ;
    $\boldsymbol{\Sigma}_0^{(t+1)} \leftarrow \boldsymbol{L}_0^{-(t)T} [\boldsymbol{L}_0^{(t)T} (\boldsymbol{\Sigma}_0^{(t)} \boldsymbol{R}^{(t)T} \boldsymbol{R}^{(t)} \boldsymbol{\Sigma}_0^{(t)}) \boldsymbol{L}_0^{(t)}]^{1/2} \boldsymbol{L}_0^{-(t)}$
until objective value converges;
```

Algorithm 2: VCSEL algorithm for MCP penalized multivariate response variance component model (3) (VCSEL-M-MCP).

4. Interaction model. Genomic differences among people place some individuals at grave risk of harm from certain medications while others may benefit from the same drug. For that reason, detecting those genetic variants that contribute to variability in treatment responses is the main objective in pharmacogenetic (PGx) studies. Several methods have been proposed to test the interaction effect or jointly test the genetic main effect and the interaction effect (Broadaway et al., 2015; Chen, Meigs and Dupuis, 2014; Zhao et al., 2019; Yang et al., 2019; Zhang et al., 2020). However, they are limited to testing a single SNP-set. Hence, in this section we illustrate the VCSEL method that incorporates interaction terms between gene and treatment in the univariate response setting $(d = 1)$.

If there are m genes under consideration, we have $2m+1$ variance components in total, including the residual variance component, because each gene is associated with two variance components, one for the gene itself and the other for the interaction between gene and treatment. For the i-th SNP-set, σ_{i1} and σ_{i2} denote the genetic effect and interaction effect variance components, respectively. Let $\boldsymbol{G}_i$ be the corresponding genotype matrix and $\boldsymbol{T} = \operatorname{diag}(t_1, \ldots, t_n)$ be a diagonal matrix where $t_i \in \{0, 1\}$ indicates treatment status. Then linear weighted kernels associated with σ_{i1} and σ_{i2} are $\boldsymbol{V}_{i1} = \boldsymbol{G}_i \boldsymbol{W}_i \boldsymbol{G}_i^T$ and $\boldsymbol{V}_{i2} = \boldsymbol{T}\boldsymbol{G}_i \boldsymbol{W}_i \boldsymbol{G}_i^T \boldsymbol{T}^T$ respectively. The matrix $\boldsymbol{W}_i = \operatorname{diag}(w_1, \ldots, w_q)$ contains the weights of the q variants in the i-th SNP-set. We remind readers that linear weighted kernels can be readily replaced by other choices of kernels. Note that $\boldsymbol{T}$ matrices are not limited to binary values. For example, one can swap diagonal entries in $\boldsymbol{T}$ matrix with environmental variable values, which are often continuous. Simulation studies 5.2 demonstrate this option of continuous values.

For a given response vector $\boldsymbol{y}$, the penalized loglikelihood augmented by group penalty on two variance components of each gene can be written as

$$f(\boldsymbol{\sigma}) = \frac{1}{2} \log\det \boldsymbol{\Omega}(\boldsymbol{\sigma}) + \frac{1}{2} \boldsymbol{y}^T [\boldsymbol{\Omega}(\boldsymbol{\sigma})]^{-1} \boldsymbol{y} + \sum_{i=1}^{m} P_\lambda(\sigma_{i1}, \sigma_{i2}), \tag{12}$$

where $\boldsymbol{\Omega}(\boldsymbol{\sigma}) = \sum_{i=1}^m \left(\sigma_{i1}^2 \boldsymbol{V}_{i1} + \sigma_{i2}^2 \boldsymbol{V}_{i2}\right) + \sigma_0^2 \boldsymbol{I}_n$ and $\boldsymbol{\sigma} = (\sigma_0, \sigma_{i1}, \sigma_{i2}, i = 1, \ldots, m)$ collects all $2m+1$ variance components. We introduce two routes to constructing interaction models: 1) include/exclude main effects and interaction term together as a pair (VCSEL-I) and 2) enforce hierarchy restriction that only allows interaction term into the model when the corresponding main effect is included (VCSEL-Ih).

4.1. *all-in/all-out (VCSEL-I).* Often in the discovery phase, genetic main effect and gene-treatment interaction effect are jointly tested. This approach examines the association between the trait of interest and genetic marker while accounting for gene-treatment interaction. To majorize the group lasso penalty on a pair of variance components, we apply the support hyperplane inequality to the concave map $x \mapsto \sqrt{x}$

$$P_\lambda(\sigma_{i1}, \sigma_{i2}) = \lambda \sqrt{\sigma_{i1}^2 + \sigma_{i2}^2} \le \frac{\lambda}{2} \frac{\sigma_{i1}^2 + \sigma_{i2}^2}{\sqrt{\sigma_{i1}^{(t)2} + \sigma_{i2}^{(t)2}}} + c^{(t)},$$

where $c^{(t)}$ is an irrelevant constant. Combining with the univariate case of inequalities (4) and (6), the surrogate function given t-th iterate $\boldsymbol{\sigma}^{(t)}$ is

$$g(\boldsymbol{\sigma}|\boldsymbol{\sigma}^{(t)}) = \sum_{i=1}^{m} \sum_{j=1}^{2} \left[\frac{\sigma_{ij}^2}{2} \operatorname{tr}(\boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{ij}) + \frac{1}{2} \frac{\sigma_{ij}^{4(t)}}{\sigma_{ij}^2} \boldsymbol{y}^T \boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{ij} \boldsymbol{\Omega}^{-(t)} \boldsymbol{y} + \lambda \frac{\sigma_{ij}^2}{2\sqrt{\sigma_{i1}^{(t)2} + \sigma_{i2}^{(t)2}}} \right] + \frac{\sigma_0^2}{2} \operatorname{tr}(\boldsymbol{\Omega}^{-(t)}) + \frac{1}{2} \frac{\sigma_0^{4(t)}}{\sigma_0^2} \boldsymbol{y}^T \boldsymbol{\Omega}^{-2(t)} \boldsymbol{y}.$$

Then the update $\sigma_{ij}^{(t+1)}$ for $i = 1, \ldots, m$ and $j = 1, 2$ is

$$\sigma_{ij}^{(t+1)} = \sigma_{ij}^{(t)} \left[\frac{\boldsymbol{y}^T \boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{ij} \boldsymbol{\Omega}^{-(t)} \boldsymbol{y}}{\operatorname{tr}(\boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{ij}) + \lambda / \sqrt{\sigma_{i1}^{(t)2} + \sigma_{i2}^{(t)2}}} \right]^{1/4}.$$

Algorithm 3 summarizes the VCSEL algorithm for the all-in/all-out interaction with lasso penalty (VCSEL-I-lasso). A similar algorithm for MCP penalty (VCSEL-I-MCP) is summarised in Supplementary Materials S.2.

```
Input  : $\boldsymbol{y}, \boldsymbol{V}_{11}, \boldsymbol{V}_{12}, \ldots, \boldsymbol{V}_{m1}, \boldsymbol{V}_{m2}, \lambda$
Output: $\hat{\sigma}_0^2, \hat{\sigma}_{11}^2, \hat{\sigma}_{12}^2, \ldots, \hat{\sigma}_{m1}^2, \hat{\sigma}_{m2}^2$
Initialize $\sigma_0^{(0)}, \sigma_{ij}^{(0)} > 0$, $i = 1, \ldots, m$, $j = 1, 2$;
repeat
    $\boldsymbol{\Omega}^{(t)} \leftarrow \sum_{i=1}^m (\sigma_{i1}^{2(t)} \boldsymbol{V}_{i1} + \sigma_{i2}^{2(t)} \boldsymbol{V}_{i2}) + \sigma_0^{2(t)} \boldsymbol{I}$ ;
    $\sigma_{ij}^{(t+1)} \leftarrow \sigma_{ij}^{(t)} \left( \frac{\boldsymbol{y}^T \boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{ij} \boldsymbol{\Omega}^{-(t)} \boldsymbol{y}}{\operatorname{tr}(\boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{ij}) + \lambda / \sqrt{\sigma_{i1}^{(t)2} + \sigma_{i2}^{(t)2}}} \right)^{1/4}$ , $i = 1, \ldots, m$, $j = 1, 2$ ;
    $\sigma_0^{(t+1)} \leftarrow \sigma_0^{(t)} \left( \frac{\boldsymbol{y}^T \boldsymbol{\Omega}^{-2(t)} \boldsymbol{y}}{\operatorname{tr}(\boldsymbol{\Omega}^{-(t)})} \right)^{1/4}$
until objective value converges;
```

Algorithm 3: VCSEL algorithm with lasso penalty for selecting main effect and interaction effect variance components as a pair (VCSEL-I-lasso).

4.2. *Hierarchical interactions (VCSEL-Ih).* In the confirmation phase of gene-drug testing, interest lies in detecting gene-treatment interaction. Choi, Li and Zhu (2010) argue that for easier interpretability, interaction terms should be included only if all corresponding main effects are in the model. We integrate this idea by assuming interaction effect variance component to be a constant multiple of genetic effect counterpart, i.e. $\sigma_{i2}^2 = \gamma_i \sigma_{i1}^2$. Whenever the variance component for i-th gene σ_{i1} is equal to 0, the interaction variance component σ_{i2} is automatically set to 0. Following Choi, Li and Zhu (2010), we penalize both variance component σ_{i1} and interaction parameter γ_i. Then our objective function with lasso penalty becomes

$$f(\boldsymbol{\sigma}) = \frac{1}{2} \log \det \boldsymbol{\Omega} + \frac{1}{2} \boldsymbol{y}^T \boldsymbol{\Omega}^{-1} \boldsymbol{y} + \lambda_1 \sum_{i=1}^m \sigma_{i1} + \lambda_2 \sum_{i=1}^m \gamma_i,$$

where $\boldsymbol{\Omega} = \sum_{i=1}^m (\sigma_{i1}^2 \boldsymbol{V}_{i1} + \sigma_{i2}^2 \boldsymbol{V}_{i2}) + \sigma_0^2 \boldsymbol{I} = \sum_{i=1}^m (\sigma_{i1}^2 \boldsymbol{V}_{i1} + \gamma_i \sigma_{i1}^2 \boldsymbol{V}_{i2}) + \sigma_0^2 \boldsymbol{I}$. Both λ_1 and λ_2 are tuning parameters controlling the strength of the penalty terms.

The already familiar majorizations (4), (6) and (7) yields the surrogate function

$$\begin{aligned} g(\boldsymbol{\sigma} \mid \boldsymbol{\sigma}^{(t)}) = \sum_{i=1}^m \Bigg[& \frac{\sigma_{i1}^2}{2} \operatorname{tr}(\boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i1}) + \frac{\gamma_i \sigma_{i1}^2}{2} \operatorname{tr}(\boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i2}) + \frac{1}{2} \frac{\sigma_{i1}^{4(t)}}{\sigma_{i1}^2} \boldsymbol{y}^T \boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i1} \boldsymbol{\Omega}^{-(t)} \boldsymbol{y} \\ & + \frac{1}{2} \frac{\gamma_i^{2(t)} \sigma_{i1}^{4(t)}}{\gamma_i \sigma_{i1}^2} \boldsymbol{y}^T \boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i2} \boldsymbol{\Omega}^{-(t)} \boldsymbol{y} + \frac{\lambda_1}{2 \sigma_{i1}^{(t)}} \sigma_{i1}^2 + \lambda_2 \gamma_i \Bigg] \\ & + \frac{\sigma_0^2}{2} \operatorname{tr}(\boldsymbol{\Omega}^{-(t)}) + \frac{1}{2} \frac{\sigma_0^{4(t)}}{\sigma_0^2} \boldsymbol{y}^T \boldsymbol{\Omega}^{-2(t)} \boldsymbol{y}. \end{aligned}$$

We adopt the block update strategy to decrease the objective value of $g(\boldsymbol{\sigma} \mid \boldsymbol{\sigma}^{(t)})$. Given $\gamma_i = \gamma_i^{(t)}$, we update σ_{i1} by

$$\sigma_{i1}^{2(t+1)} = \sigma_{i1}^{2(t)} \sqrt{\frac{\boldsymbol{y}^T \boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i1} \boldsymbol{\Omega}^{-(t)} \boldsymbol{y} + \gamma_i^{(t)} \boldsymbol{y}^T \boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i2} \boldsymbol{\Omega}^{-(t)} \boldsymbol{y}}{\operatorname{tr}(\boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i1}) + \gamma_i^{(t)} \operatorname{tr}(\boldsymbol{\Omega}^{-(t)} \boldsymbol{V}_{i2}) + \lambda_1 / \sigma_{i1}^{(t)}}}, \quad i = 1, \ldots, m.$$

Given $\sigma_{i1} = \sigma_{i1}^{(t+1)}$, we first update the covariance matrix

$$\widetilde{\boldsymbol{\Omega}}^{(t)} = \sum_{i=1}^m \left(\sigma_{i1}^{2(t+1)} \boldsymbol{V}_{i1} + \gamma_i^{(t)} \sigma_{i1}^{2(t+1)} \boldsymbol{V}_{i2} \right) + \sigma_0^{2(t+1)} \boldsymbol{I},$$

then update the i-th interaction parameter by

$$\gamma_i^{(t+1)} = \gamma_i^{(t)} \sqrt{\frac{\boldsymbol{y}^T \widetilde{\boldsymbol{\Omega}}^{-(t)} \boldsymbol{V}_{i2} \widetilde{\boldsymbol{\Omega}}^{-(t)} \boldsymbol{y}}{\operatorname{tr}(\widetilde{\boldsymbol{\Omega}}^{-(t)} \boldsymbol{V}_{i2}) + 2\lambda_2 / \sigma_{i1}^{2(t+1)}}}.$$

Summary of the algorithm for this hierarchical interaction selection method with lasso penalty (VCSEL-Ih-lasso) is left to Supplementary Materials S.2.

5. Simulation studies. We conduct simulation studies to examine the selection performance of the proposed methods. We compare with R packages Multi-SKAT (Dutta et al., 2019) and rareGE (Chen, Meigs and Dupuis, 2014) for multivariate response and interaction model, respectively. Both Multi-SKAT and rareGE are marginal approaches that test one SNP-set at a time and make a formal inference. This contrasts with our method that encompasses multiple SNP-sets in a joint model and provides rankings. For readers interested in the results on a univariate response, we summarize the results in Supplementary Materials S.4, in which we compare the selection performance of VCSEL to the group lasso. The group lasso is a group selection method designed for selecting fixed effects. Interestingly, the proposed penalized variance component model outperforms group lasso even when the data is generated from a fixed effects model, not to mention under a variance component model (see Supplementary Materials S.4).

Both the lasso and MCP penalties are demonstrated for multivariate trait and interaction models. Unless otherwise specified, $\gamma = 2.0$ is used for the MCP penalty. We use the area under Precision-Recall curve (auPRC) to evaluate performance. Similar to Receiver Operator Characteristic (ROC) curves, Precision-Recall (PR) curves (recall on the x-axis and precision on the y-axis) illustrate the tradeoff between precision and recall for varying cutoff values (Manning and Schütze, 1999; Raghavan, Bollmann and Jung, 1989). *Precision* is defined as the number of true positives over the total number of declared positives, while *recall* is defined as the number of true positives over the number of true positives plus the number of false negatives. A PR curve closer to the upper right corner, which corresponds to 100% precision and 100% recall, generally represents a better classifier. Since we want to take the influence of all cutoff values into account, we report auPRC, which is an aggregate measure of performance across all tuning parameter values and has a range of [0, 1]. An auPRC close to 1 indicates that the classifier returns accurate results (high precision) and most of all positive results (high recall).

Although ROC curves are the most popular metric for binary classifiers, PR curves are more suitable when the class distribution is highly skewed, usually negative instances outnumbering positive instances (Davis and Goadrich, 2006; Saito and Rehmsmeier, 2015). In fact, PR curves have been cited as an alternative in unbalanced datasets (Craven and Bockhorst, 2005; Bunescu et al., 2005; Davis et al., 2005; Goadrich, Oliphant and Shavlik, 2004; Kok and Domingos, 2005; Singla and Domingos, 2005). As we expect the number of positive variance components to be greatly exceeded by that of zero variance components, we deem auPRC to be an appropriate metric.

For the marginal testing methods—Multi-SKAT and rareGE—we calculate the auPRC by ranking all genes by their p-values and assuming that each gene enters the solution path from the smallest to largest. For example, the gene with the smallest p-value enters the solution path first, and the gene with the largest value would be the last one to enter the solution path.

For a sample of size n, we form genotype matrix $\boldsymbol{G}$ by randomly pairing $2n$ haplotypes drawn from a haplotype pool (SKAT.haplotypes in the SKAT R-package). The genotype values in matrix G are coded as 0, 1 and 2, representing the number of minor alleles while an additive genetic model is assumed. Assuming that there are m SNP-sets, we partition $\boldsymbol{G}$ into m submatrices of pre-specified window length:

$$\boldsymbol{G} = \left[\boldsymbol{G}_1 \middle| \boldsymbol{G}_2 \middle| \cdots \middle| \boldsymbol{G}_m\right],$$

where $\boldsymbol{G}_i \in \mathbb{R}^{n \times q_i}, i = 1, \ldots, m$, represents the i-th SNP-set.

We fix the number of positive variance components, excluding the residual variance component, to be 5. We calculate each auPRC over 100 tuning parameter values and report the average auPRCs along with their standard errors across 20 replicates.

5.1. *Simulation studies for multiple traits.* Here we compare selection performance of Algorithm 1 and MultiSKAT (Dutta et al., 2019) package in R. We generate three phenotypes ($n = 2000$, $d = 3$) from the following:

$$\text{vec}(\boldsymbol{Y}) = \text{vec}(\boldsymbol{X}\boldsymbol{B}) + \boldsymbol{L}_{\boldsymbol{\Omega}}\boldsymbol{\epsilon}, \ \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}_{nd}), \tag{13}$$

where $\boldsymbol{L}_{\boldsymbol{\Omega}}$ is the lower triangular Cholesky factor of $\boldsymbol{\Omega} = \sum_{i=1}^{m} \boldsymbol{\Sigma}_i \otimes \boldsymbol{V}_i + \boldsymbol{\Sigma}_0 \otimes \frac{1}{\sqrt{n}}\boldsymbol{I}_n$. Depending on the genotype kernel, $\boldsymbol{V}_i$ equals to $\frac{1}{||\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{W}_i\boldsymbol{G}_i^T||_F}\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{W}_i\boldsymbol{G}_i^T$ (SKAT genotype kernel) or $\frac{1}{||\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{W}_i\boldsymbol{G}_i^T||_F}\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{W}_i\boldsymbol{G}_i^T$ (Burden test genotype kernel) where $\boldsymbol{W}_i$ is diagonal matrix whose entry equals to the weights $w_k = Beta(\text{MAF}_k; 1, 25)$ with MAF_k being the minor allele frequency of the k-th genetic variant (Wu et al., 2011). We use this weight since it is the default version in MultiSKAT package. We set $\boldsymbol{X}$ to be a $n \times 1$ matrix of 1s and $\boldsymbol{B}$ to be a $1 \times d$ matrix of 0.5s. For non-zero variance components $\boldsymbol{\Sigma}_i$, we incorporate two structures proposed in Dutta et al. (2019). The first choice is $\boldsymbol{\Sigma}_i = \boldsymbol{1}_d\boldsymbol{1}_d^T$, which implies that effect sizes of a variant on d different phenotypes are homogeneous. Hence it is called homogeneous kernel. The second structure is $\boldsymbol{\Sigma}_i = \boldsymbol{I}_d$, also known as heterogeneous kernel, which assumes that effect sizes of a variant on different phenotypes are heterogeneous or independent. Non-zero variance component matrices are spread across all m groups to create a scenario of low linkage disequilibrium (LD) between causal SNP-sets or variance components:

$$\boldsymbol{\Sigma}_i = \begin{cases} \boldsymbol{1}_d\boldsymbol{1}_d^T \text{ or } \boldsymbol{I}_d & \text{if } i = 1, 10, 20, 30, 40 \ (m = 40) \\ & \text{if } i = 1, 25, 50, 75, 100 \ (m = 100) \\ \boldsymbol{I}_d & \text{if } i = 0 \\ \boldsymbol{0} & \text{else.} \end{cases}$$

In this case, causal genes, or signal variance components are dispersed, hence there is little correlation among causal genes. One notable difference between VCSEL-M and Multi-SKAT is that Multi-SKAT does not estimate $\boldsymbol{\Sigma}_i$ while VCSEL-M estimates $\boldsymbol{\Sigma}_i$. In fact, Multi-SKAT requires one to provide phenotype kernel structure, which is $\boldsymbol{\Sigma}_i$ in our notation, for testing association between a SNP-set and multiple phenotypes. In our simulations, we supply the ground truth $\boldsymbol{\Sigma}_i$, whether it be $\boldsymbol{1}_d\boldsymbol{1}_d^T$ or $\boldsymbol{I}_d$, when calling Multi-SKAT, hence giving an advantage to the Multi-SKAT method.

Figure 1 and Table 1 describe simulation results. Overall, our methods perform as well as Multi-SKAT, if not better. Despite having the ground truth $\boldsymbol{\Sigma}_i$ as an input argument, Multi-SKAT does not perform well when phenotype kernel has a homogeneous structure, as seen in the left panel of Figure 1.

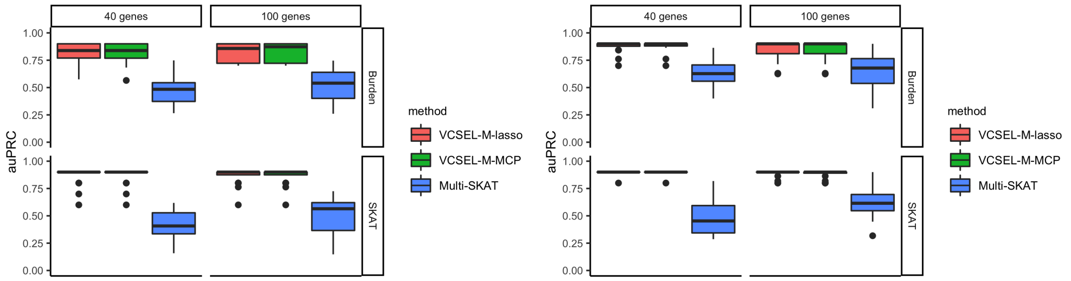

Fig 1: The auPRCs of VCSEL-M-lasso, VCSEL-M-MCP and Multi-SKAT under 40 and 100 genes and different genotype kernels for models with 6 non-zero variance components and 3 simulated traits ($d = 3$), using haplotype data from the SKAT R-package. The left and right panels assume $\boldsymbol{\Sigma}_i = \mathbf{1}_d\mathbf{1}_d^T$ and $\boldsymbol{\Sigma}_i = \boldsymbol{I}_d$, respectively, for non-zero variance components.

5.2. *Simulation studies for interaction models.* Here we compare selection performance of Algorithm 3, S.2.1 and rareGE (Chen, Meigs and Dupuis, 2014) package in R. We generate a phenotype from

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{L}_{\boldsymbol{\Omega}}\boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{I}_n)$$

where $n = 500$. Here covariate matrix $\boldsymbol{X}$ is a 500×3 matrix whose first column is a vector of 1's, second column is generated from $N(50, 5^2)$, and third column from $N(25, 4^2)$, which mimic covariate matrix in simulation studies of Chen, Meigs and Dupuis (2014). $\boldsymbol{L}_{\boldsymbol{\Omega}}$ is the lower triangular Cholesky factor of $\boldsymbol{\Omega} = \sum_{j=1}^{2}\sum_{i=1}^{m}\sigma_{ij}^2\boldsymbol{V}_{ij} + \frac{\sigma_0^2}{\sqrt{n}}\boldsymbol{I}_n$. Following the default option of rareGE package, we set

$$\boldsymbol{V}_{i1} = \frac{1}{||\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{G}_i^T||_F}\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{G}_i^T$$
$$\boldsymbol{V}_{i2} = \frac{1}{||\boldsymbol{E}\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{G}_i^T\boldsymbol{E}_i||_F}\boldsymbol{E}\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{G}_i^T\boldsymbol{E},$$

where $\boldsymbol{W}_i$ diagonal matrix whose entry equals to to the commonly used weights $\sqrt{w_k} = Beta(\text{MAF}_k; 1, 25)$ with MAF_k being the MAF of the k-th genetic variant (Wu et al., 2011). $\boldsymbol{E}$ is a diagonal matrix whose entries coincide with that of the second column in $\boldsymbol{X}$. $\boldsymbol{G}_i$ is a submatrix of genotype matrix we form from haplotypes data in the SKAT R-package, as explained in the beginning of Section 5. We restrict $\boldsymbol{G}_i$ to only include SNPs with MAF less than 0.05 for fair comparison with rareGE method. This constraint leads to the number of SNPs ranging from 18 to 51 with a median of 33 for groups with window length of 5kb

TABLE 1
The auPRCs of VCSEL-M-lasso, VCSEL-M-MCP, and Multi-SKAT across varying size and number of genes, using SKAT.haplotypes data from the SKAT R-package. In parentheses are standard deviation/$\sqrt{\text{no. replicates}}$.

Genotype kernel	Phenotype kernel	No. genes	VCSEL-M-lasso	VCSEL-M-MCP	MultiSKAT
$\boldsymbol{G}_i\boldsymbol{W}_i\mathbf{1}_{q_i}\mathbf{1}_{q_i}^T\boldsymbol{W}_i\boldsymbol{G}_i^T$ (Burden)	$\boldsymbol{\Sigma}_i = \mathbf{1}_d\mathbf{1}_d^T$	100 (2kb/gene)	0.82 (0.019)	0.82 (0.020)	0.52 (0.032)
		40 (5kb/gene)	0.82 (0.020)	0.82 (0.021)	0.48 (0.034)
	$\boldsymbol{\Sigma}_i = \boldsymbol{I}_d$	100 (2kb/gene)	0.84 (0.021)	0.84 (0.021)	0.65 (0.035)
		40 (5kb/gene)	0.87 (0.012)	0.88 (0.012)	0.63 (0.029)
$\boldsymbol{G}_i\boldsymbol{W}_i\boldsymbol{I}_{q_i}\boldsymbol{W}_i\boldsymbol{G}_i^T$ (SKAT)	$\boldsymbol{\Sigma}_i = \mathbf{1}_d\mathbf{1}_d^T$	100 (2kb/gene)	0.86 (0.017)	0.86 (0.017)	0.48 (0.041)
		40 (5kb/gene)	0.87 (0.018)	0.87 (0.018)	0.42 (0.030)
	$\boldsymbol{\Sigma}_i = \boldsymbol{I}_d$	100 (2kb/gene)	0.88 (0.008)	0.88 (0.009)	0.62 (0.031)
		40 (5kb/gene)	0.90 (0.005)	0.90 (0.005)	0.48 (0.038)

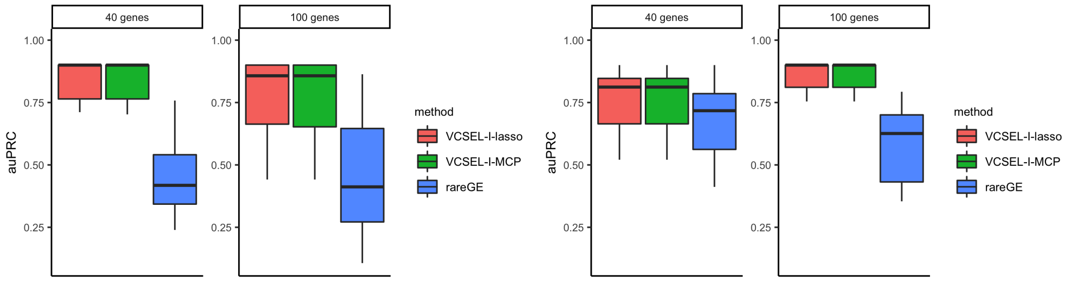

Fig 2: The auPRCs of VCSEL-I-lasso, VCSEL-I-MCP, and rareGE under 40 and 100 genes for models with 6 non-zero variance components, using haplotype data from the SKAT R-package. True variance component values in the left panel mimic low LD scenario (14) while those in the right panel mimic high LD scenario (15).

and that ranging from 3 to 29 with a median of 13 for groups with window length of 2kb. We set the effect strength of non-zero variance components to be 2.236. Two scenarios are simulated. The first is low LD setting:

$$\sigma_{i1} = \sigma_{i2} = \begin{cases} 2.236 & i = 1, 11, 20, 30, 40 \; (m = 40) \\ & i = 1, 26, 50, 75, 100 \; (m = 100) \\ 1.0 & i = 0 \\ 0.0 & \text{else.} \end{cases} \tag{14}$$

The second is high LD setting, where the first 5 variance components are set to be non-zero:

$$\sigma_{i1} = \sigma_{i2} = \begin{cases} 2.236 & i = 1, 2, 3, 4, 5 \\ 1.0 & i = 0 \\ 0.0 & \text{else.} \end{cases} \tag{15}$$

In Supplementary Materials S.5, we quantify the correlations between SNP-sets in these high/low LD settings via the canonical correlation analysis. The true fixed effects parameter values are set to be $\boldsymbol{\beta} = (0.5, 0.1, 0.05)^T$.

As seen in Figure 2, VCSEL-I method is competitive against rareGE. The outperformance of VCSEL-I method is more dramatic under the low LD scenario, probably because the marginal test rareGE is not able to jointly model the multiple SNP-sets.

6. Real data analysis.

To test the multivariate response model, we apply our methods to the genetic data from the UK Biobank exome sequencing study (Sudlow et al., 2015). By doing so, we aim to identify genes associated with two quantitative lipid traits: high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C). For the analysis, we only use measurements from the initial assessment visit. We regress each phenotype separately on age, age^2, sex, and the top five principal components and inverse normal transform respective residuals. The transformed residuals are used as our response variables. For our samples, we extract self-reported white British individuals (data field 21000: Ethnic background) with no genetic kinship to other participants (data field 22021: Genetic kinship to other participants) and without any medication for cholesterol, blood pressure, diabetes or exogenous hormones at baseline (data field 6153 and 6177: Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones). After removing individuals with missing values, we have 18,020 samples and genotype information of 8,959,608 variants, which are

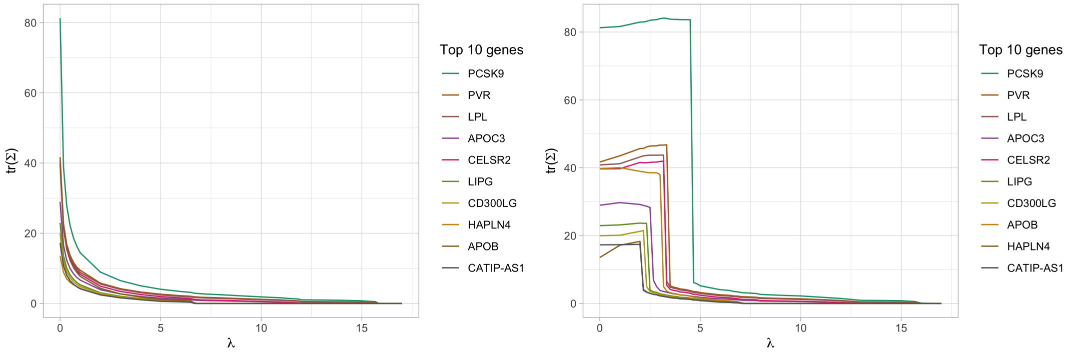

Fig 3: Solution paths of VCSEL-M-lasso (left) and VCSEL-M-MCP (right) methods in the analysis of 200 genes and two lipid measurements (HDL-C, LDL-C).

grouped into 26,395 genes based on the annotation information from SnpEff software (Cingolani et al., 2012) with GRCh38 human reference genome. We then remove monoallelic variants, common variants with MAF > 0.05, variants in sex chromosome from the analysis. Finally, we have the data of 18,020 individuals and genotype information of 4,312,036 low-frequency/rare (MAF ≤ 0.05) variants in 25,460 genes with at least three of those variants in each gene. Because the number of genes is too large, we first screen 25,460 genes down to 200 genes according to their p-values from Multi-SKAT omnibus approach that combines results across three pre-specified phenotype kernels (homogeneous, heterogeneous, and phenotype covariance kernels). Then we carry out a penalized estimation of the 200 variance components in the joint model (1) using the Burden test genotype kernel. This is akin to the sure independence screening strategy by Fan and Lv (2008), which entails large-scale screening accompanied by moderate-scale variable selection. Genes are ranked according to the order they appear in the solution path. Figure 3 illustrates the solution paths obtained from VCSEL-M-lasso and VCSEL-M-MCP methods, along with their corresponding lists of the top ten genes in the order they appear in the solution path. Table 2 lists the top 10 genes together with their marginal p-values from Multi-SKAT. Most genes that are highly ranked by VCSEL methods—*PCSK9*, *PVR*, *LPL*, *APOC3*, *CELSR2*, *LIPG*, *CD300LG*, and *APOB* in the top 10 list—have their marginal test p-values under the false discovery rate (FDR) $< 5\%$ threshold and/or are known to play a role in modulating lipid levels (Benn et al., 2005; Cohen et al., 2005; Heid et al., 2008; Abifadel et al., 2009; Wallace et al., 2008; Tachmazidou et al., 2013; Lange et al., 2014; Holmen et al., 2014; Surakka et al., 2015). VCSEL methods identify genes that are not deemed significant by marginal testing but have association evidence in the literature. *HAPLN4* has been shown significant association with LDL-C and total cholesterol levels (Southam et al., 2017) and *APOC4* with HDL-C, LDL-C (Hoffmann et al., 2018; Wojcik et al., 2019).

Next, we apply our methods to the GWAS of Ezetimibe response in IMPROVE-IT (IMProved Reduction of Outcomes: Vytroin Efficacy International Trial), which is a phase 3b, multicenter, double-blind, randomized study to establish the clinical benefit and safety of Vytorin (Ezetimibe/Simvastatin tablet) versus Simvastatin mono-therapy in high-risk subjects (Cannon et al., 2015). In this PGx study using IMPROVE-IT clinical data, we are interested in discovering genes associated with 1) the efficacy of Vytorin treatment for 2,808 European patients who receive a greater benefit compared with the Simvastatin mono-therapy and 2) the joint efficacy of Ezetimibe/Simvastatin treatment and the Simvastatin mono-therapy treatment for 5,661 European patients. The endpoint for this gene-based variance component selection analysis is LDL-C fold-change at 1-month. The standard GWAS quality control and

Lasso Rank	MCP Rank	Gene	Marginal p-value	# Variants
1	1	*PCSK9*	3.37×10^{-20}	353
2	2	*PVR*	3.56×10^{-20}	111
3	4	*LPL*	5.73×10^{-18}	198
4	3	*APOC3*	2.04×10^{-7}	61
5	5	*CELSR2*	4.05×10^{-13}	986
6	6	*LIPG*	2.36×10^{-13}	225
7	7	*CD300LG*	6.56×10^{-10}	189
8	9	*HAPLN4*	2.86×10^{-3}	141
9	8	*APOB*	5.33×10^{-11}	947
10	10	*CATIP-AS1*	2.81×10^{-3}	16
11	11	*APOC4*	1.34×10^{-4}	74

TABLE 2
Top genes selected by the lasso and MCP penalized variance component model are tallied with their marginal p-values from the Multi-SKAT omnibus test in an association study of 200 genes and bivariate trait: HDL-C and LDL-C.

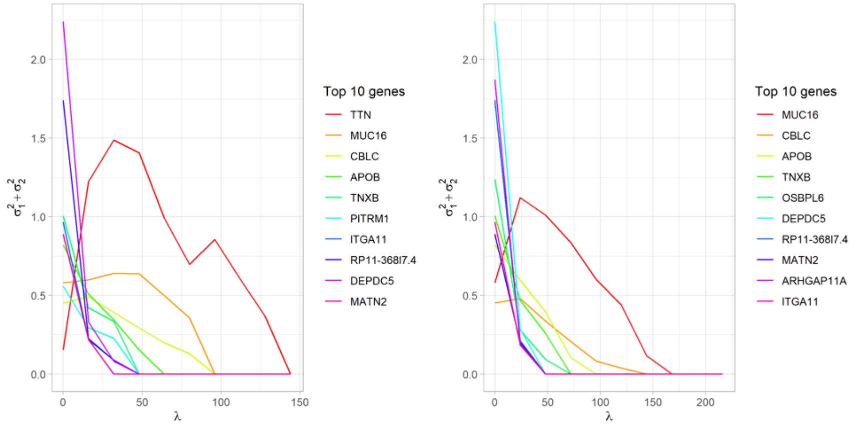

Fig 4: Solution paths of VCSEL-I-lasso (left) and VCSEL-I-MCP (right) methods in the analysis of 200 genes and the LDL-C response of all the patients receiving the Vytorin (Ezetimibe/Simvastatin tablet) treatment and Simvastatin mono-therapy in the IMPROVE-IT PGx study.

SNP imputation are conducted. We focus on the low frequency variants ($0.01 \leq \text{MAF} \leq 0.05$) after imputation (with imputation quality scores $r^2 > 0.5$) and putatively functional variants with consequences as non-synonymous, splice-site, non-sense, and frameshift variants annotated from the GEMINI software (Paila et al., 2013). Missing genotypes are imputed by their column mean. In total, there are 208,123 low frequency variants in 2,572 genes with at least two low frequency variants in each gene. The covariate matrix includes age, gender, prior lipid lowering therapy, early Acute Coronary Syndrome (ACS) trial, high risk ACS diagnosis, and the top five principal components calculated from the GWAS data to adjust for population structure. Because the number of genes is too large, we first screen the 2,572 genes down to 200 genes according to their marginal p-values from SKAT-O (Lee, Wu and Lin, 2012) for the analysis of Vytorin treatment effect and the other 200 genes according to their marginal p-values from the Composite Kernel Association Test (CKAT) (Zhang et al., 2020) for the analysis of Ezetimibe/Simvastatin treatment and the Simvastatin mono-therapy treatment joint effects. Then we analyze the two sets of the 200 genes by penalized estimation of the 200 variance components respectively.

Figure 4 illustrates the solution paths from VCSEL-I-lasso and VCSEL-I-MCP methods, along with their corresponding lists of the top ten genes in the order they appear in the solution path for the analysis of Ezetimibe/Simvastatin treatment and the Simvastatin monotherapy treatment joint effects. The top five genes selected by the VCSEL-I-lasso method are *TTN*, *MUC16*, *CBLC*, *APOB* and *TNXB*, and those selected by the VCSEL-I-MCP method are *MUC16*, *CBLC*, *APOB* and *TNXB* and *OSBPL6*. *CBLC* and *TNXB* are selected by both methods and have been shown to associate with statins response in literature. More specifically, similar as the *BCAM* gene, *CBLC* gene, close to *BCAM* gene, has been shown to associate with the response to statins (LDL-C change) and multiple non-drug-response LDL-C related traits as well (Postmus et al., 2014; Deshmukh et al., 2012; Supplementary Table S1). In addition, *TNXB* gene also shows a significant association with the non-drug-response LDL-C trait in the literature (Supplementary Table S1). We defer the analysis results for studying the efficacy of Vytorian treatment to Supplementary Materials S.6.

The above analyses demonstrate that VCSEL methods can provide well-known and potentially new association evidence between genes and the drug response LDL-C in the IMPROVE-IT PGx study and the lipid phenotypes in the UK Biobank whole exome sequencing study. More work is needed to further interpret both top ranked genes with some association evidence and without any literature support to identify causal genes.

7. Discussion. This article provides a variance component selection framework for identifying SNP-sets associated with quantitative traits, particularly for multivariate traits and SNP-set-treatment interactions. Simulation studies and real data analyses have testified to the competitiveness of the proposed methods, compared to the traditional marginal tests.

Additionally, our methods can adjust for sample relatedness by augmenting the model with a kinship matrix. More precisely, borrowing the notation of (2), the model becomes

$$\operatorname{vec} \boldsymbol{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_1 \otimes \boldsymbol{V}_1 + \cdots + \boldsymbol{\Sigma}_m \otimes \boldsymbol{V}_m + \boldsymbol{\Sigma}_g \otimes \boldsymbol{\Phi} + \boldsymbol{\Sigma}_0 \otimes \boldsymbol{I}_{n-p}),$$

where $\boldsymbol{\Phi}$ is the kinship matrix, and $\boldsymbol{\Sigma}_g$ is a matrix describing the shared heritability between the phenotypes. Along with the residual variance component $\boldsymbol{\Sigma}_0$, coheritability variance component $\boldsymbol{\Sigma}_g$ would remain in the model without any regularization.

While chiefly motivated by association testing in genetics, we envision the analysis to be applicable beyond genetics. For instance, in random effects ANOVA with many factors, each represented by a variance component, one may wish to select factors that are relevant to the response. This ANOVA scenario has been alluded in Supplementary Materials S.4.

There are some limitations to the proposed methods. First, it is difficult to conduct formal inference on the selected SNP-sets. Second, it does not apply to biobank-scale data. We recommend this method for datasets of size up to $n \times d = 50{,}000$ where n is the number of samples and d is the number of traits. This is because VCSEL methods involve inverting the covariance matrix $\boldsymbol{\Omega}$ in each iteration, which is computationally expensive. Additionally, we do not suggest jointly fitting all 20,000-25,000 genes in the human genome using our method. We recommend that the number of genes is reduced before fitting the model by the sure independence screening strategy, which has been extensively studied and investigated (Fan and Lv, 2008).

In this paper, we focus on the ranking of genes and report the overall selection performance by auPRC. In practice, the tuning parameters can be chosen according to the extended Bayesian information criteria (Chen and Chen, 2008). Future research should entail post-selection inference and investigation of the algorithms' theoretical properties and address limitations mentioned above.

8. Implementation. All our methods are implemented in the open source, high-performance technical computer language Julia (Bezanson et al., 2017), and the software is freely available at https://github.com/juhkim111/VCSEL.jl.

Acknowledgements. This work is partially supported by National Institutes of Health (NIH) grants T32 HG02536, R01 HG006139 and R35 GM141798 and National Science Foundation grant DMS-2054253.

SUPPLEMENTARY MATERIAL

Supplement to "VCSEL: Prioritizing SNP-set by penalized variance component selection". We provide information about data sets used in this paper and the supplemental results (e.g., extra simulations, tables, figures and result summaries).
().

Supplementary Table S1. We provide additional results from real data analysis.
().

REFERENCES

ABIFADEL, M., RABÈS, J.-P., DEVILLERS, M., MUNNICH, A., ERLICH, D., JUNIEN, C., VARRET, M. and BOILEAU, C. (2009). Mutations and polymorphisms in the proprotein convertase subtilisin kexin 9 (PCSK9) gene in cholesterol metabolism and disease. *Human Mutation* **30** 520–529.

BAKIN, S. (1999). Adaptive Regression and Model Selection in Data Mining Problems, PhD thesis.

BANSAL, V., LIBIGER, O., TORKAMANI, A. and SCHORK, N. J. (2010). Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics* **11** 773.

BATES, D. M. and PINHEIRO, J. C. (1998). Computational methods for multilevel modelling. *University of Wisconsin, Madison, WI* 1–29.

BENN, M., NORDESTGAARD, B. G., JENSEN, J. S., GRANDE, P., SILLESEN, H. and TYBJÆRG-HANSEN, A. (2005). Polymorphism in APOB associated with increased low-density lipoprotein levels in both genders in the general population. *The Journal of Clinical Endocrinology & Metabolism* **90** 5797–5803.

BEZANSON, J., EDELMAN, A., KARPINSKI, S. and SHAH, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review* **59** 65-98.

BODMER, W. and BONILLA, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics* **40** 695.

BONDELL, H. D., KRISHNA, A. and GHOSH, S. K. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics* **66** 1069–1077.

BREHENY, P. and HUANG, J. (2009). Penalized methods for bi-level variable selection. *Statistics and Its Interface* **2** 369.

BROADAWAY, K. A., DUNCAN, R., CONNEELY, K. N., ALMLI, L. M., BRADLEY, B., RESSLER, K. J. and EPSTEIN, M. P. (2015). Kernel approach for modeling interaction effects in genetic association studies of complex quantitative traits. *Genetic Epidemiology* **39** 366–375.

BROADAWAY, K. A., CUTLER, D. J., DUNCAN, R., MOORE, J. L., WARE, E. B., JHUN, M. A., BIELAK, L. F., ZHAO, W., SMITH, J. A., PEYSER, P. A. et al. (2016). A statistical approach for testing cross-phenotype effects of rare variants. *The American Journal of Human Genetics* **98** 525–540.

BUNESCU, R., GE, R., KATE, R. J., MARCOTTE, E. M., MOONEY, R. J., RAMANI, A. K. and WONG, Y. W. (2005). Comparative experiments on learning information extractors for proteins and their interactions. *Artificial Intelligence in Medicine* **33** 139–155.

CANNON, C. P., BLAZING, M. A., GIUGLIANO, R. P., MCCAGG, A., WHITE, J. A., THEROUX, P., DARIUS, H., LEWIS, B. S., OPHUIS, T. O., JUKEMA, J. W. et al. (2015). Ezetimibe added to statin therapy after acute coronary syndromes. *New England Journal of Medicine* **372** 2387–2397.

CHEN, J. and CHEN, Z. (2008). Extended Bayesian information criterion for model selection with large model space. *Biometrika* **95** 759-771.

CHEN, Z. and DUNSON, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59** 762–769.

CHEN, H., MEIGS, J. B. and DUPUIS, J. (2014). Incorporating gene-environment interaction in testing for association with rare genetic variants. *Human Heredity* **78** 81–90.

CHOI, N. H., LI, W. and ZHU, J. (2010). Variable selection with the strong heredity constraint and its oracle property. *Journal of the American Statistical Association* **105** 354–364.

CINGOLANI, P., PLATTS, A., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. and RUDEN, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6** 80-92.

COHEN, J., PERTSEMLIDIS, A., KOTOWSKI, I. K., GRAHAM, R., GARCIA, C. K. and HOBBS, H. H. (2005). Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature Genetics* **37** 161–165.

CRAVEN, M. and BOCKHORST, J. (2005). Markov networks for detecting overalpping elements in sequence data. In *Advances in Neural Information Processing Systems* 193–200.

DAVIS, J. and GOADRICH, M. (2006). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* 233–240. ACM.

DAVIS, J., BURNSIDE, E. S., DE CASTRO DUTRA, I., PAGE, D., RAMAKRISHNAN, R., COSTA, V. S. and SHAVLIK, J. W. (2005). View Learning for Statistical Relational Learning: With an Application to Mammography. In *IJCAI* 677–683. Citeseer.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)* 1–38.

DERING, C., HEMMELMANN, C., PUGH, E. and ZIEGLER, A. (2011). Statistical analysis of rare sequence variants: an overview of collapsing methods. *Genetic Epidemiology* **35** S12–S17.

DESHMUKH, H. A., COLHOUN, H. M., JOHNSON, T., MCKEIGUE, P. M., BETTERIDGE, D. J., DURRINGTON, P. N., FULLER, J. H., LIVINGSTONE, S., CHARLTON-MENYS, V., NEIL, A. et al. (2012). Genome-wide association study of genetic determinants of LDL-c response to atorvastatin therapy: importance of Lp (a). *Journal of Lipid Research* **53** 1000–1011.

DUTTA, D., SCOTT, L., BOEHNKE, M. and LEE, S. (2019). Multi-SKAT: General framework to test for rare-variant association with multiple phenotypes. *Genetic Epidemiology* **43** 4–23.

FAN, Y. and LI, R. (2012). Variable selection in linear mixed effects models. *Annals of Statistics* **40** 2043.

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 849–911.

GIBSON, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics* **13** 135.

GOADRICH, M., OLIPHANT, L. and SHAVLIK, J. (2004). Learning ensembles of first-order clauses for recall-precision curves: A case study in biomedical information extraction. In *International Conference on Inductive Logic Programming* 98–115. Springer.

GUDMUNDSSON, J., SULEM, P., GUDBJARTSSON, D. F., MASSON, G., AGNARSSON, B. A., BENEDIKTSDOTTIR, K. R., SIGURDSSON, A., MAGNUSSON, O. T., GUDJONSSON, S. A., MAGNUSDOTTIR, D. N. et al. (2012). A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nature Genetics* **44** 1326–1329.

HACKINGER, S. and ZEGGINI, E. (2017). Statistical methods to detect pleiotropy in human complex traits. *Open Biology* **7** 170125.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72** 320–338.

HEID, I. M., BOES, E., MÜLLER, M., KOLLERITS, B., LAMINA, C., COASSIN, S., GIEGER, C., DÖRING, A., KLOPP, N., FRIKKE-SCHMIDT, R. et al. (2008). Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circulation: Cardiovascular Genetics* **1** 10–20.

HOFFMANN, T. J., THEUSCH, E., HALDAR, T., RANATUNGA, D. K., JORGENSON, E., MEDINA, M. W., KVALE, M. N., KWOK, P.-Y., SCHAEFER, C., KRAUSS, R. M. et al. (2018). A large electronic-health-record-based genome-wide study of serum lipids. *Nature Genetics* **50** 401–413.

HOLMEN, O. L., ZHANG, H., FAN, Y., HOVELSON, D. H., SCHMIDT, E. M., ZHOU, W., GUO, Y., ZHANG, J., LANGHAMMER, A., LØCHEN, M.-L. et al. (2014). Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nature genetics* **46** 345–351.

HUANG, J., MA, S., XIE, H. and ZHANG, C.-H. (2009). A group bridge approach for variable selection. *Biometrika* **96** 339–355.

HUNTER, D. R. and LANGE, K. (2004). A tutorial on MM algorithms. *The American Statistician* **58** 30–37.

KHURI, A. I. and SAHAI, H. (1985). Variance components analysis: a selective literature survey. *International Statistical Review/Revue Internationale de Statistique* 279–300.

KOK, S. and DOMINGOS, P. (2005). Learning the structure of Markov logic networks. In *Proceedings of the 22nd International Conference on Machine Learning* 441–448. ACM.

LAIRD, N., LANGE, N. and STRAM, D. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association* **82** 97–105.

LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* 963–974.

LANGE, K. (2016). *MM Optimization Algorithms*. SIAM.

LANGE, K., HUNTER, D. R. and YANG, I. (2000). Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics* **9** 1–20.

LANGE, L. A., HU, Y., ZHANG, H., XUE, C., SCHMIDT, E. M., TANG, Z.-Z., BIZON, C., LANGE, E. M., SMITH, J. D., TURNER, E. H. et al. (2014). Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *The American Journal of Human Genetics* **94** 233–245.

LEE, S., WU, M. C. and LIN, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13** 762–775.

LEE, S., ABECASIS, G. R., BOEHNKE, M. and LIN, X. (2014). Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics* **95** 5–23.

LEE, S., WON, S., KIM, Y. J., KIM, Y., CONSORTIUM, T.-G., KIM, B.-J. and PARK, T. (2017). Rare variant association test with multiple phenotypes. *Genetic Epidemiology* **41** 198–209.

LI, B. and LEAL, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics* **83** 311–321.

LIN, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84** 309–326.

LINDSTROM, M. J. and BATES, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* **83** 1014–1022.

MADSEN, B. E. and BROWNING, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genetics* **5** e1000384.

MAITY, A., SULLIVAN, P. F. and TZENG, J.-I. (2012). Multivariate Phenotype Association Analysis by Marker-Set Kernel Machine Regression. *Genetic Epidemiology* **36** 686–695.

MANNING, C. D. and SCHÜTZE, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R., CHAKRAVARTI, A. et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461** 747.

PAILA, U., CHAPMAN, B. A., KIRCHNER, R. and QUINLAN, A. R. (2013). GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Computational Biology* **9**.

PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58** 545–554.

PENG, H. and LU, Y. (2012). Model selection in linear mixed effect models. *Journal of Multivariate Analysis* **109** 109–129.

POSTMUS, I., TROMPET, S., DESHMUKH, H. A., BARNES, M. R., LI, X., WARREN, H. R., CHASMAN, D. I., ZHOU, K., ARSENAULT, B. J., DONNELLY, L. A. et al. (2014). Pharmacogenetic meta-analysis of genome-wide association studies of LDL cholesterol response to statins. *Nature Communications* **5** 5068.

RAGHAVAN, V., BOLLMANN, P. and JUNG, G. S. (1989). A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems (TOIS)* **7** 205–229.

RIVAS, M. A. and MOUTSIANAS, L. (2015). Power of Rare Variant Aggregate Tests. In *Assessing Rare Variation in Complex Traits* 185–199. Springer.

RIVAS, M. A., BEAUDOIN, M., GARDET, A., STEVENS, C., SHARMA, Y., ZHANG, C. K., BOUCHER, G., RIPKE, S., ELLINGHAUS, D., BURTT, N. et al. (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genetics* **43** 1066–1073.

ROBINSON, D. L. (1987). Estimation and use of variance components. *The Statistician* 3–14.

SAITO, T. and REHMSMEIER, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One* **10** e0118432.

SCHAID, D. J., SINNWELL, J. P., LARSON, N. B. and CHEN, J. (2020). Penalized variance components for association of multiple genes with traits. *Genetic Epidemiology* **n/a**.

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. New York: Wiley.

SIMON, N., FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2013). A sparse-group lasso. *Journal of Computational and Graphical Statistics* **22** 231–245.

SINGLA, P. and DOMINGOS, P. (2005). Discriminative training of Markov logic networks. In *Proceedings of the 20th National Conference on Artificial Intelligene (AAAI)* **5** 868–873. AAAI Press.

SIVAKUMARAN, S., AGAKOV, F., THEODORATOU, E., PRENDERGAST, J. G., ZGAGA, L., MANOLIO, T., RUDAN, I., MCKEIGUE, P., WILSON, J. F. and CAMPBELL, H. (2011). Abundant pleiotropy in human complex diseases and traits. *The American Journal of Human Genetics* **89** 607–618.

SOLOVIEFF, N., COTSAPAS, C., LEE, P. H., PURCELL, S. M. and SMOLLER, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* **14** 483.

SOUTHAM, L., GILLY, A., SÜVEGES, D., FARMAKI, A.-E., SCHWARTZENTRUBER, J., TACHMAZIDOU, I., MATCHAN, A., RAYNER, N. W., TSAFANTAKIS, E., KARALEFTHERI, M. et al. (2017). Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. *Nature communications* **8** 1–11.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J., LANDRAY, M. et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**.

SUO, C., TOULOPOULOU, T., BRAMON, E., WALSHE, M., PICCHIONI, M., MURRAY, R. and OTT, J. (2013). Analysis of multiple phenotypes in genome-wide genetic mapping studies. *BMC Bioinformatics* **14** 151.

SURAKKA, I., HORIKOSHI, M., MÄGI, R., SARIN, A.-P., MAHAJAN, A., LAGOU, V., MARULLO, L., FERREIRA, T., MIRAGLIO, B., TIMONEN, S. et al. (2015). The impact of low-frequency and rare variants on lipid levels. *Nature Genetics* **47** 589–597.

TACHMAZIDOU, I., DEDOUSSIS, G., SOUTHAM, L., FARMAKI, A.-E., RITCHIE, G. R., XIFARA, D. K., MATCHAN, A., HATZIKOTOULAS, K., RAYNER, N. W., CHEN, Y. et al. (2013). A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. *Nature communications* **4** 1–6.

THOMPSON, W. A. et al. (1962). The problem of negative estimates of variance components. *The Annals of Mathematical Statistics* **33** 273–289.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

WALLACE, C., NEWHOUSE, S. J., BRAUND, P., ZHANG, F., TOBIN, M., FALCHI, M., AHMADI, K., DOBSON, R. J., MARÇANO, A. C. B., HAJAT, C. et al. (2008). Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *The American Journal of Human genetics* **82** 139–149.

WOJCIK, G. L., GRAFF, M., NISHIMURA, K. K., TAO, R., HAESSLER, J., GIGNOUX, C. R., HIGHLAND, H. M., PATEL, Y. M., SOROKIN, E. P., AVERY, C. L. et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570** 514–518.

WU, B. and PANKOW, J. S. (2016). Sequence kernel association test of multiple continuous phenotypes. *Genetic Epidemiology* **40** 91–100.

WU, M. C., KRAFT, P., EPSTEIN, M. P., TAYLOR, D. M., CHANOCK, S. J., HUNTER, D. J. and LIN, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *The American Journal of Human Genetics* **86** 929–942.

WU, M. C., LEE, S., CAI, T., LI, Y., BOEHNKE, M. and LIN, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* **89** 82–93.

YANG, T., CHEN, H., TANG, H., LI, D. and WEI, P. (2019). A powerful and data-adaptive test for rare-variant–based gene-environment interaction analysis. *Statistics in Medicine* **38** 1230–1244.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 49–67.

ZHAI, J., KIM, J., KNOX, K. S., TWIGG III, H. L., ZHOU, H. and ZHOU, J. J. (2018). Variance component selection with applications to microbiome taxonomic data. *Frontiers in Microbiology* **9** 509.

ZHAN, X., ZHAO, N., PLANTINGA, A., THORNTON, T. A., CONNEELY, K. N., EPSTEIN, M. P. and WU, M. C. (2017). Powerful genetic association analysis for common or rare variants with high-dimensional structured traits. *Genetics* **206** 1779–1790.

ZHANG, C.-H. et al. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38** 894–942.

ZHANG, H., ZHAO, N., MEHROTRA, D. V. and SHEN, J. (2020). Composite Kernel Association Test (CKAT) for SNP-set Joint Assessment of Genotype and Genotype-by-treatment Interaction in Pharmacogenetics Studies. *Bioinformatics*. btaa125.

ZHAO, N., ZHANG, H., CLARK, J. J., MAITY, A. and WU, M. C. (2019). Composite kernel machine regression based on likelihood ratio test for joint testing of genetic and gene–environment interaction effect. *Biometrics* **75** 625–637.

ZHOU, H., SEHL, M. E., SINSHEIMER, J. S. and LANGE, K. (2010). Association screening of common and rare genetic variants by penalized regression. *Bioinformatics* **26** 2375.

ZHOU, H., HU, L., ZHOU, J. and LANGE, K. (2019). MM algorithms for variance components models. *Journal of Computational and Graphical Statistics* **28** 350–361.

ZUK, O., SCHAFFNER, S. F., SAMOCHA, K., DO, R., HECHTER, E., KATHIRESAN, S., DALY, M. J., NEALE, B. M., SUNYAEV, S. R. and LANDER, E. S. (2014). Searching for missing heritability: designing rare variant association studies. *Proceedings of the National Academy of Sciences* **111** E455–E464.