On set-based association tests: insights from a regression using summary statistics

Yanyan Zhao[1] and Lei Sun[1,2]

[1]Department of Statistical Sciences, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada,

yanyan.zhao@utoronto.ca

[2]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, Ontario M5T 3M7, Canada,

sun@utstat.toronto.edu

June 16, 2019

Abstract

Motivated by, but not limited to, association analyses of multiple genetic variants, we propose here a summary statistic-based regression framework. The proposed method requires only variant-specific summary statistics, and it unifies earlier methods based on individual-level data as special cases. The resulting score test statistic, derived from a linear mixed-effect regression model, inherently transforms the variant-specific statistics using the precision matrix to improve power for detecting sparse alternatives. Furthermore, the proposed method can incorporate additional variant-specific information with ease, facilitating omic-data integration. We study the asymptotic properties of the proposed tests under the null and alternatives, and we investigate efficient p-value calculation in finite samples. Finally, we provide supporting empirical evidence from extensive simulation studies and two applications.

Keywords: Correlated test statistics; Regression; Sparse alternatives; Set-based test; Summary statistics; SKAT-O.

1 Introduction

Set-based analyses of multiple variables are increasingly important in many current scientific studies. For example, in modern genome-wide association studies (GWAS) and next-generation sequencing (NGS) analyses, one might be interested in jointly analyzing multiple (rare) genetic variants influencing a complex, heritable trait (also known as gene- or pathway-based studies), identifying one genetic variants influencing multiple traits (also known as pleitrotropy studies), or combining evidence from multiple studies as in the classical meta-analysis.

Without loss of generality, let us focus on set-based analyses of multiple rare genetic variants. In this setting, myriad statistical tests have been proposed, which fit into three general categories (Derkach et al., 2014; Lee et al., 2014): the linear or burden tests (e.g. Morgenthaler and Thilly (2007); Li and Leal (2008); Madsen and Browning (2009)), the quadratic or variance-component tests (e.g. Pan (2009); Neale et al. (2011); Wu and Lin (2011)), and the hybrid tests combining evidence from the linear and quadratic tests (e.g. Lee et al. (2012); Derkach et al. (2013)).

These earlier methods have been studied extensively, however, they can be improved in several aspects. First, these methods may not perform well in the spare-signal setting where only a small proportion of the variants are true signals and variants are highly correlated (Xu et al., 2016). Second, individual/subject-level data may not be available due to logistical challenges or data confidentiality agreement, so it is beneficial to explicitly develop summary statistic-based association tests. In addition, it is important to ensure that summary statistic-based methods can also incorporate additional covariates, such as variant-specific functional annotation information. Finally, earlier work have shown the performance of a simple minimum-p value approach (Derkach et al., 2013) is comparable with that of the optimal sequence kernel association test, SKAT-O (Lee et al., 2012). This suggests that a grid search for the 'optimal' weighting factor may not be necessary, and it is desirable to develop new robust hybrid test statistics with theoretically justified weights.

To this end, we propose a flexible and unifying linear mixed-effect regression model that

requires only variant-specific summary statistics, and we show that earlier methods based on individual-level data are special cases of the proposed testing framework. The statistics derived from the regression model (based on summary statistics) inherently transforms the variant-specific statistics using the precision matrix; this transformation has been proposed to increase signal strength when the signals are sparse in other settings (Fan et al., 2013; Cai et al., 2014). Thus, the proposed regression framework provides an intuitive way to utilize correlations among genetic variants to improve power for detecting sparse alternatives. Furthermore, the proposed method can incorporate additional variant-specific information as covariate(s). For example, the covariate could be the available functional annotation for the set of variants. Both our simulation studies and real data application show that we could have remarkable power gain when the included covariate contains useful information, while power loss is minimal when the covariate is uninformative.

Although the proposed method is motivated by jointly analyzing of multiple rare genetic variants, the general set-based analytical framework can be used for other settings, for example, meta-analysis (Han and Eskin, 2011), PrediXcan incorporating association evidence with gene-expression data (Gamazon et al., 2015), and pleiotropy association study between one genetic variant and multiple phenotypes (Liu and Lin, 2018b). We will provide a detailed discussion on the differences and connections between the proposed method and the earlier work, as well as the additional utilities of the proposed method including enhancing the performance of polygenic risk score (Purcell et al., 2009).

The remainder of the paper is organized as follows. As a proof-of-principle for the proposed set-based regression approach, we first review existing association methods for analyzing a set of rare genetic variants based on individual-level data in Section 2. In Section 3, we outline the proposed regression framework based on summary statistics and derive a catalogue of association test statistics from fixed-, random-, or mixed-regression models. We then demonstrate the analytical equivalency between some of the new statistics and the existing ones for rare variants analyses, and we investigate efficient p-value calculation in finite samples and study asymptotic properties

of the tests. Finally, we discuss covariate adjustment. To provide supporting empirical evidence, we present numerical results from simulation studies in Section 4 and from two application studies in Section 5. We conclude with remarks and discussions in Section 6, and we give theoretical proofs and additional numerical studies in the supplementary material.

2 Existing association tests for jointly analyzing a set of rare genetic variants

2.1 Regression set-up using individual-level data

Let $\boldsymbol{y} = (y_1, \cdots, y_n)'$ denote the phenotype values of n unrelated individuals, and for a set of J genetic variants of interest, let $\boldsymbol{G}_i = (G_{i1}, G_{i2}, \cdots, G_{iJ})'$, $i = 1, \cdots, n$, denote the corresponding genotype data. Assume y_i follows a exponential family distribution with mean μ_i and dispersion parameter $\phi = 1$ (without loss of generality) given $\boldsymbol{G}_i$, $f(y_i|\mu_i) = exp\{\theta(\mu_i)y_i - b(\theta_i) + c(y_i)\}$, the corresponding generalized linear model is

$$g(\mu_i) = \beta_0 + \beta_1 G_{i1} + \cdots + \beta_J G_{iJ}. \tag{1}$$

Individual-level covariate information such as age and sex, if available, should be added to the model but are omitted for the moment. This omission does not change the validity of the methods to be discussed, and for clarify of the presentation we also only consider canonical link functions in this paper.

To evaluate the phenotype-genotype association relationship, we are interested in testing the null hypothesis that $H_0 : \boldsymbol{\beta} = (\beta_1, \cdots, \beta_J)' = \boldsymbol{0}$. Let $\boldsymbol{G}_j = (G_{1j}, G_{2j}, \cdots, G_{nj})'$ be the genotype vector for variant $j, j = 1, \cdots, J$, and $\boldsymbol{G} = (\boldsymbol{G}_1, \cdots, \boldsymbol{G}_J)$, the corresponding score vector is $\boldsymbol{s} = \boldsymbol{G}'(\boldsymbol{y} - \bar{\mu}_y \boldsymbol{1}_n)$, where $\boldsymbol{1}_n$ is a $n \times 1$ unit vector and $\bar{\mu}_y = \frac{1}{n}\sum_{i=1}^{n} y_i$. The j_{th} element of the score

function is $s_j = \sum_i (y_i - \bar{\mu}_y) G_{ij}$, which captures the linear relationship between phenotype $\boldsymbol{y}$ and genotype $\boldsymbol{G}_j$. The variance-covariance matrix of $\boldsymbol{s}$ is

$$\boldsymbol{\Sigma_0} = g_1^{-1}(\bar{\mu}_y)\boldsymbol{G}'(I_n - \mathbf{1_n}\mathbf{1_n}'/n)\boldsymbol{G}, \tag{2}$$

where $g_1(\cdot)$ denotes the first derivative of the link function g, and I_n is a identity matrix of size n.

2.2 Existing methods based on the score vector s

Numerous methods have been proposed to evaluate $H_0 : \boldsymbol{\beta} = (\beta_1, \cdots, \beta_J)' = \mathbf{0}$, among which there are three popular classes, namely the burden or linear (e.g. Madsen and Browning (2009)), variance-component or quadratic (e.g. Wu and Lin (2011)), and hybrid combing the linear and quadratic (e.g. Lee et al. (2012)); see Derkach et al. (2014) for a review.

Although it is not always obvious, most test statistics are functions of $\boldsymbol{s}$. For example, the original burden test (also known as the weighted-sum test) (Madsen and Browning, 2009) first constructs a 'super-variant' for which the genotype is the weighted average across genotypes of the J variants, $G^* = \sum_{j=1}^{J} w_j G_{ij}$, where $\boldsymbol{w} = (w_1, w_2, \cdots, w_J)'$ is a pre-specified weighting factor often associated with minor allele frequency (MAF); Madsen and Browning (2009) chose $w_j = 1/\sqrt{\text{MAF}_j(1-\text{MAF}_j)}$, while Morgenthaler and Thilly (2007) preferred equal weighting using $w_j = 1, j = 1, \cdots, J$. Burden type of tests then performs the phenotype-genotype association analysis via regression, $g(\mu_i) = \beta_0^* + \beta^* G^* = \beta_0^* + \beta^* \sum_{j=1}^{J} w_j G_{ij}$, and testing $H_0 : \beta^* = 0$. However, it is not difficult to show that the score test statistic derived from the above regression using the 'super-variant' G^* is proportional to $T_1 = (\boldsymbol{w}'\boldsymbol{s})^2$, where $(\boldsymbol{w}'\boldsymbol{\Sigma_0}\boldsymbol{w})^{-1} T_1$ is asymptotically chi-square distributed with 1 degrees of freedom (d.f.), χ_1^2.

This T_1 test is also termed as CAST by Morgenthaler and Thilly (2007), the sum test by Pan (2009), and the linear-class test by Derkach et al. (2014), among others. Because T_1 is based on the weighted average of s_j and s_j can be positive or negative depending on the direction of effect

(i.e. sign of β_j in model (1)), T_1 is only powerful when a large proportion of variants are causal *and* effects are in the same direction.

Variance-component tests, such as SKAT (Wu and Lin, 2011), SSU (Pan, 2009), and C-alpha (Neale et al., 2011), offer an alternative approach that belongs to the quadratic class of tests (Derkach et al., 2014). Again, although most of the original tests started with the regression model (1), this class of tests can be formulated as $T_2 = \boldsymbol{s}'\boldsymbol{A}\boldsymbol{s}$, where $\boldsymbol{A}$ is a positive or semi-definite symmetric matrix, and T_2 asymptotically follows a weighted chi-square distribution. For example, Derkach et al. (2014) has noted that $\boldsymbol{A} = \boldsymbol{I}$ leads to C-alpha of Neale et al. (2011) and SSU of Pan (2009), while $\boldsymbol{A} = diag\{a_1, \cdots, a_J\}$ leads to SKAT of Wu and Lin (2011), where a_j depends on the MAF of variant j; see Table 1 of Derkach et al. (2014) for a summary. These quadratic tests are robust to heterogenous effect directions, but they are less powerful than linear tests when most variants are causal and with the same direction of effects.

Since the true genetic model is unknown, omnibus hybrid tests combining T_1 and T_2 tests have been been proposed. For example, Lee et al. (2012) proposed SKAT-O, a weighted linear combination of a burden-type of test and SKAT, $Q_\rho = \rho(\boldsymbol{w}'\boldsymbol{s})^2 + (1-\rho)\boldsymbol{s}'\boldsymbol{A}\boldsymbol{s} = \rho T_1 + (1-\rho)T_2$, where ρ can be interpreted as the unknown pairwise correlation between β_j under the alternative, $\boldsymbol{A} = \text{diag}\{w_1^2, \cdots, w_J^2\}$, and w_j depends on the MAF of variant $j, j = 1, \cdots, J$. A grid search for the 'optimal' ρ is then performed, $0 = \rho_1 < \rho_2 < \cdots < \rho_m = 1$. Let p_ρ be the corresponding p-value based on Q_ρ, the test statistic for SKAT-O is $T_{skato} = \min\{p_{\rho_1}, \cdots, p_{\rho_m}\}$. The asymptotic p-value of T_{skato} can be calculated with one-dimensional numerical integration.

Instead of considering data-driven 'optimal' ρ then adjusting for the inherent selection bias, Derkach et al. (2013) proposed two simpler yet competitive hybrid test statistics, T_{Fisher} and T_{Minp}. Let p_{T_1} and p_{T_2} be the p-values corresponding to T_1 and T_2, respectively, the Fisher and Minp statistics take the form of $T_{Fisher} = -2\log(p_{T_1}) - 2\log(p_{T_2})$, $T_{Minp} = \min\{p_{T_1}, p_{T_2}\}$. If p_{T_1} and p_{T_2} are asymptotically independent under the null hypothesis, T_{Fisher} has an asymptotic distribution of χ^2_4, and T_{Minp} has an asymptotic distribution of $Beta(1,2)$.

Previous work have shown that T_{Minp} and T_{skato} perform similarly, and they are slightly more powerful than T_{Fisher} when T_1 has no power (Derkach et al., 2013). In contrast, T_{Fisher} has better power than T_{Minp} and T_{skato} when both T_1 and T_2 have some power. However, we expect all three hybrid tests to have little power under sparse alternatives (Donoho and Jin, 2004; Barnett et al., 2017) when only a small proportion of variants in the set is causal.

To improve performance, we first note that if variants are correlated with each other, we can consider for example $\mathbf{\Sigma_0}^{-1}\boldsymbol{s}$ instead of $\boldsymbol{s}$ in constructing a more powerful test under sparse alternatives. Second, it is clear that we only need variant-specific summary statistics $\boldsymbol{s} = (s_1, \cdots, s_J)'$ for jointly analyzing the J variants of interest. Further, the fact that T_{Minp} and T_{skato} having similar performance suggests that a grid search for ρ might not be necessary, and an easy-to-compute yet theoretically justified 'optimal' ρ could exist. Lastly, when additional variant-specific information z_j (e.g. variant j being non-synonymous or not) is available, we can improve power by incorporating $z_j, j = 1, \cdots, J$. Intuitively we can achieve this by modifying w_j proportional to z_j, but a less add-hoc approach is desirable. To this end, we will consider a flexible and unifying regression framework that (i) requires only s_j and $\mathbf{\Sigma_0}$, (ii) yields T_1 and T_2, and hybrid statistics similar to T_{skato} as special cases, and (iii) provides new test statistics that incorporate the precision matrix $\mathbf{\Sigma_0}^{-1}$ if desired, as well as account for covariate information $z_j, j = 1, \cdots, J$, if available.

3 A regression framework based on summary statistics

Here we assume $\boldsymbol{s} = (s_1, \cdots, s_J)'$ is available, summarizing the association relationship between the phenotype of interest and a set of J genetic variants as detailed in Section 2. We also assume that $\mathbf{\Sigma_0}$ is known or estimated accurately from a reference panel. We let $\boldsymbol{z} = (z_1, \cdots, z_J)'$ represent variant-specific information available (Ionita-Laza et al., 2016), which can be multi-dimensional but assumed to be one single covariate for notation simplicity.

3.1 Fixed-effect and random-effect models

We first consider a fixed-effect (FE) model that models the common effect present among $s_j, j = 1, \cdots, J$,

$$\boldsymbol{s} = \mu \boldsymbol{w} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\boldsymbol{w} = (w_1, \cdots, w_J)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_J)'$, and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma_0})$. Based on this model, we aim to test $H_0 : \mu = 0$, and the corresponding score test statistic is

$$T_{FE} = \left(\boldsymbol{w}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{w}\right)^{-1} (\boldsymbol{w}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s})^2. \tag{4}$$

The equivalence between $(\boldsymbol{w}'\boldsymbol{\Sigma_0}\boldsymbol{w})^{-1} T_1$ $(= (\boldsymbol{w}'\boldsymbol{\Sigma_0}\boldsymbol{w})^{-1} (\boldsymbol{w}'\boldsymbol{s})^2)$ in Section 2 and T_{FE} above is apparent, if we let w in T_{FE} to be $\boldsymbol{\Sigma_0}\boldsymbol{w}$ (Table 1).

Alternatively, we can consider the following random effect (RE) model,

$$\boldsymbol{s} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\boldsymbol{\eta} \sim N(\mathbf{0}, \tau^2 \boldsymbol{R})$, $\boldsymbol{R}$ is a predefined positive or semi-definite symmetric matrix, and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma_0})$. If we test the following hypothesis, $H_0 : \tau^2 = 0$, the corresponding score test statistic is

$$T_{RE} = \left(2tr\left(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R}\right)^2\right)^{-1/2} \left(\boldsymbol{s}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R}\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s} - tr(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R})\right) = c_2^{-1/2}(Q(\boldsymbol{s}) - c_1), \tag{6}$$

where $c_1 = tr(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R})$ is the mean of $Q(\boldsymbol{s})$ and $c_2 = 2tr(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R})^2$ is the variance of $Q(\boldsymbol{s})$. The analytical equivalence between $T_2 = \boldsymbol{s}'\boldsymbol{A}\boldsymbol{s}$ and $Q(\boldsymbol{s}) = \boldsymbol{s}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R}\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s}$ (the key element of T_{RE}) is also apparent, if we let $\boldsymbol{R} = \boldsymbol{\Sigma_0}\boldsymbol{A}\boldsymbol{\Sigma_0}$ (Table 1). In addition, if we let $\boldsymbol{R} = \boldsymbol{\Sigma_0}\boldsymbol{W}\boldsymbol{R}_\rho\boldsymbol{W}\boldsymbol{\Sigma_0}$, where $\boldsymbol{W} = \text{diag}\{w_j\}$ and $\boldsymbol{R}_\rho = \rho \mathbf{1_J}\mathbf{1_J}' + (1-\rho)\boldsymbol{I_J}$, then $Q(\boldsymbol{s}) = \rho T_1 + (1-\rho)T_2 = Q_\rho$, the key element of T_{skato}.

3.2 Mixed-effect model

The fixed-effect model (3) captures the common underlying effect, while the random-effect model (5) accounts for potential heterogeneity. A logical next step to consider a mixed-effect (ME) modelling framework that includes models (3) and (5) as special cases,

$$\boldsymbol{s} = \mu \boldsymbol{w} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{7}$$

where $\boldsymbol{\eta} \sim N(0, \tau^2 \boldsymbol{R})$ and $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Sigma_0})$. If we test the following null hypothesis,

$$H_0 : \mu = 0, \tau^2 = 0, \tag{8}$$

the corresponding score vector is $\left(\boldsymbol{w}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s},\ \frac{1}{2}\boldsymbol{s}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R}\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s} - \frac{1}{2}tr(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R})\right)'$, and the test statistic is

$$T_{ME} = \left(\boldsymbol{w}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{w}\right)^{-1}\left(\boldsymbol{w}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s}\right)^2 + c_2^{-1}\left(\boldsymbol{s}'\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R}\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s} - c_1\right)^2, \tag{9}$$

where $c_1 = tr(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R})$ and $c_2 = 2tr(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{R})^2$ as before for T_{RE}.

Intuitively, following the construction of T_{ME}, we can also consider another hybrid test statistic, T_{12}, by combining $T_1 = (\boldsymbol{w}'\boldsymbol{s})^2$ and $T_2 = \boldsymbol{s}'\boldsymbol{A}\boldsymbol{s}$ weighted by the corresponding standard deviations,

$$T_{12} = \left(\boldsymbol{w}'\boldsymbol{\Sigma_0}\boldsymbol{w}\right)^{-1}\left(\boldsymbol{w}'\boldsymbol{s}\right)^2 + \check{c}_2^{-1}\left(\boldsymbol{s}'\boldsymbol{A}\boldsymbol{s} - \check{c}_1\right)^2 \tag{10}$$

where $\check{c}_1 = tr(\boldsymbol{\Sigma_0}\boldsymbol{A})$ is the mean of T_2 under the null and $\check{c}_2 = 2tr(\boldsymbol{\Sigma_0}\boldsymbol{A})^2$ is the variance of T_2. The connection between T_{12} with T_{skato} is also immediate. However, there are two key differences. First, given a ρ, T_{skato} relies on $Q_\rho = \rho(\boldsymbol{w}'\boldsymbol{s})^2 + (1-\rho)\boldsymbol{s}'\boldsymbol{A}\boldsymbol{s} = \rho T_1 + (1-\rho)T_2$. In contrast, T_{12} combines T_1 and the *square* of *centralized* T_2 (not T_2 itself). That is, $T_{12} = \sqrt{2}\mathrm{sd}_{T_1}^{-1} T_1 + (\mathrm{sd}_{T_2}^{-1}(T_2 - \mathrm{mean}_{T_2}))^2$. Secondly, T_{skato} searches for the 'optimal' ρ that minimizes the p-value associated with Q_ρ then adjusts for selection bias. In contrast, T_{12} uses the corresponding variances. Lastly, we

note that when J is large, centralized T_2 is approximately normally distributed with mean zero under the null. Thus $(\mathrm{sd}_{T_2}^{-1}(T_2 - \mathrm{mean}_{T_2}))^2$ is χ_1^2 distributed asymptotically, and it is on the same scale as $\mathrm{sd}_{T_1}^{-1} T_1$.

The score test statistic derived directly from the mixed-effect model (7) is T_{ME}, which uses $\boldsymbol{\Sigma_0}^{-1}\boldsymbol{s}$, instead of $\boldsymbol{s}$, to account for the correlation between the tested variants. Transforming $\boldsymbol{s}$ by the precision matrix has been considered previously in other settings, for example, by Cai et al. (2014) for a two-sample high-dimensional means test. As illustrated in Cai et al. (2014), the transformation could improve power under sparse alternatives in the presence of high correlation.

Inspired by Hall and Jin (2010), we can also consider Cholesky decomposition of $\boldsymbol{\Sigma_0}$, $\boldsymbol{C\Sigma_0 C'} = \boldsymbol{I_J}$, and use $\boldsymbol{Cs}$ instead of $\boldsymbol{s}$. This will result in another new hybrid test statistic,

$$\tilde{T}_{ME} = \left(\boldsymbol{w'w}\right)^{-1}\left(\boldsymbol{w'Cs}\right)^2 + \left(2tr\left(\boldsymbol{R}^2\right)\right)^{-1}\left(\boldsymbol{s'C'RCs} - tr(\boldsymbol{R})\right)^2. \tag{11}$$

In the two-class classification context, Fan et al. (2013) has shown that $\boldsymbol{\Sigma_0}^{-1}$ leads to better performance than $\boldsymbol{C}$. In our setting, we will study both T_{ME} and $\tilde{T}_{ME}$ for completeness.

3.3 Additional covariate adjustment

Increasingly, additional variant-specific information, $\boldsymbol{z} = (z_1, \cdots, z_J)$, such as functional annotation or gene-expression evidence are available and should be utilized to increase power (Finucane et al., 2015). One could consider directly modifying $\boldsymbol{w}$ to reflect the additional information, but a principled approach is lacking. The proposed regression framework, however, can naturally incorporate $\boldsymbol{z}$ as a covariate into the mixed-effect model (7),

$$\boldsymbol{s} = \mu\boldsymbol{w} + \theta\boldsymbol{z} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}. \tag{12}$$

If we are interested in testing, $H_0 : \mu = 0, \theta = 0, \tau^2 = 0$, the corresponding score test statistic

is

$$T_{ME,cov} = \boldsymbol{s}'\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{u}(\boldsymbol{u}'\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{u})^{-1}\boldsymbol{u}'\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{s} + c_2^{-1}\left(\boldsymbol{s}'\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{R}\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{s} - c_1\right)^2, \tag{13}$$

where $\boldsymbol{u} = (\boldsymbol{w}, \boldsymbol{z})$, $c_1 = tr(\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{R})$, and $c_2 = 2tr(\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{R})^2$. Similarly as before, we can construct $T_{12,cov}$ and $\tilde{T}_{ME,cov}$ versions of T_{12} and $\tilde{T}_{ME}$ that account for covariate information, but for a more focused study we will not examine the difference between these tests here. Table 1 summarizes all the tests discussed so far.

3.4 Asymptotic distributions of the proposed test statistics

The asymptotic distributions of the existing tests, T_1, T_2, T_{skato}, T_{Fisher}, and T_{Minp}, have been previously established as $n \to \infty$, assuming that individual-level data are available. Here, we are interested in the asymptotic distribution of the various tests when $J \to \infty$. In the next section, we will study finite sample behaviour of the tests. We begin with some mild conditions needed for Theorem 1. For two sequences of real numbers $\{a_{1J}\}$ and $\{a_{2J}\}$, denote $a_{1J} = o(a_{2J})$ if $\lim_{J\to\infty}(a_{1J}/a_{2J}) = 0$, and $\boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{0}}\boldsymbol{C}' = \boldsymbol{I}_{\boldsymbol{J}}$.

Condition 1. *Let b_j be the j_{th} element of the vector $\boldsymbol{b}$, $j = 1, \cdots, J$, where $\boldsymbol{b} = \boldsymbol{w}'\boldsymbol{C}'$, and let λ_j be the j_{th} eigenvalue of the matrix $\boldsymbol{C}\boldsymbol{R}\boldsymbol{C}'$, $j = 1, \cdots, r$. Denote $b_{max} = \max_j b_j$ and $\lambda_{max} = \max_j \lambda_j$. $b_{max} = o\left(\sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1}\boldsymbol{w}}\right)$ and $\lambda_{max} = o\left(\sqrt{tr(\boldsymbol{R}\boldsymbol{\Sigma}_{\boldsymbol{0}}^{-1})^2}\right)$.*

Condition 2. *Let $\check{b}_j$ be the j_{th} element of the vector $\check{\boldsymbol{b}}$, $j = 1, \cdots, J$, where $\check{\boldsymbol{b}} = \boldsymbol{w}'\boldsymbol{C}^{-1}$, and let $\check{\lambda}_j$ be the j_{th} eigenvalue of the matrix $(\boldsymbol{C}^{-1})'\boldsymbol{R}\boldsymbol{C}^{-1}$, $j = 1, \cdots, r$. Denote $\check{b}_{max} = \max_j \check{b}_j$ and $\check{\lambda}_{max} = \max_j \check{\lambda}_j$. $\check{b}_{max} = o\left(\sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}_{\boldsymbol{0}}\boldsymbol{w}}\right)$ and $\check{\lambda}_{max} = o\left(\sqrt{tr(\boldsymbol{R}\boldsymbol{\Sigma}_{\boldsymbol{0}})^2}\right)$.*

Condition 3. *Let $\tilde{\lambda}_j$ be the j_{th} eigenvalue of the matrix $\boldsymbol{R}$, $j = 1, \cdots, r$. Denote $w_{max} = \max_j w_j$ and $\tilde{\lambda}_{max} = \max_j \tilde{\lambda}_j$. $w_{max} = o(\sqrt{\boldsymbol{w}'\boldsymbol{w}})$ and $\tilde{\lambda}_{max} = o(\sqrt{tr(\boldsymbol{R}^2)})$.*

Theorem 1. *Under the null hypothesis of (8),*

Table 1: **Summary of different test statistics for analyzing a set of genetic variants.** $\boldsymbol{s} = (s_1, \cdots, s_J)$ is a vector of summary association statistics for J genetic variants, $\boldsymbol{w}$ is a vector of weights, $\boldsymbol{\Sigma_0}$ is a known or estimated variance-covariance matrix for $\boldsymbol{s}$, and $\boldsymbol{R}$ and $\boldsymbol{A}$ are positive definite (or semi-definite) symmetric matrices. [a]The equivalence can be analytical or numerical; see corresponding sessions for details. See Table 1 of Derkach et al. (2014) for a more detailed summary of the previous methods.

Proposed Approach		Equivalence[a] with previous methods based on individual-level data	
Regression Models	Test statistics	Methods	Test statistics, e.g.
Fixed-effect Model, Section 3.1		Linear T_1	
$\boldsymbol{s} = \mu \boldsymbol{w} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma_0})$ (3) $H_0 : \mu = 0$	T_{FE} (4)	let $\boldsymbol{w}$ be $\boldsymbol{\Sigma_0 w}$	CAST, Morgenthaler and Thilly (2007) Sum-test/burden, Pan (2009) Weighted-sum, Madsen and Browning (2009)
Random-effect Model, Section 3.1		Quadratic T_2	
$\boldsymbol{s} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \boldsymbol{\eta} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{R})$ (5) $H_0 : \tau^2 = 0$	T_{RE} (6)	let $\boldsymbol{R}$ be $\boldsymbol{\Sigma_0 A \Sigma_0}$	SSU, Pan (2009) C-alpha, Neale et al. (2011) SKAT, Wu and Lin (2011)
Mixed-effect Model, Section 3.2		Hybrid T_1 '+' T_2	
$\boldsymbol{s} = \mu \boldsymbol{w} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$ (7) $H_0 : \mu = 0, \tau^2 = 0$	T_{ME} (9) T_{12} (10) $\tilde{T}_{ME}$ (11)		T_{skato}, Lee et al. (2012) T_{Fisher}, T_{Minp}, Derkach et al. (2013)
Covariate Adjustment, Section 3.3			
$\boldsymbol{s} = \mu \boldsymbol{w} + \theta \boldsymbol{z} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$ (12) $H_0 : \mu = 0, \theta = 0, \tau^2 = 0$	$T_{ME,cov}$ (13)		

(a) Assume Condition 1 holds, then, $T_{ME} \xrightarrow{d} \chi^2_2$, *as* $J \to \infty$.

(b) Assume Condition 2 holds, then, $T_{12} \xrightarrow{d} \chi^2_2$, *as* $J \to \infty$.

(c) Assume Condition 3 holds, then, $\tilde{T}_{ME} \xrightarrow{d} \chi^2_2$, *as* $J \to \infty$.

When J is small, significance evaluation based on the above asymptotic distributions may not be adequate. In Theorem 2 we provide an approximation for the finite-sample distribution of T_{ME}; results for T_{12}, and $\tilde{T}_{ME}$ are similar. Note that the main computational cost involved in Theorem 2 is the calculation of eigenvalues $\lambda_j, j = 1, \cdots, J$.

Theorem 2. *Let* λ_j *be the* j_{th} *eigenvalue of the matrix* $\boldsymbol{CRC'}$, $j = 1, \cdots, r$, *then*

$$T_{ME} \overset{\mathscr{D}}{=} u_1^2 + \left(2\sum_{j=1}^{r}\lambda_j^2\right)^{-1}\left(\sum_{j=1}^{r}\lambda_j(v_j^2-1)\right)^2,$$

where u_1 *and* v_j, $j = 1, \cdots, r$, *are independent* $N(0,1)$, *and* $\overset{\mathscr{D}}{=}$ *denotes equality in distribution.*

We note that the above finite and asymptotic results are with respect to J. The validities of T_{ME}, T_{12}, and $\tilde{T}_{ME}$ do not require $n \to \infty$ explicitly, as long as $\boldsymbol{s}$ is multivariate normal. The distributions of the existing tests, T_1, T_2, T_{skato}, T_{Fisher}, and T_{Minp} for finite J have been established. As $J \to \infty$, it is easy to show that both T_1 and T_2 are asymptotically χ^2_1 distributed under some mild conditions that are similar to the ones specified above. The constructions of T_{skato}, T_{Fisher}, and T_{Minp} depend on p-values of Q_ρ, T_1, or T_2, thus we do not pursue their asymptotic distributions with respect to J.

3.5 Power comparison

We first establish the asymptotic distributions of T_{ME}, T_{12}, and $\tilde{T}_{ME}$ (as $J \to \infty$) under the alternative, $H_1 : \mu = \mu_1, \tau^2 = \tau_1^2$, requiring some mild conditions on the $\boldsymbol{w}$, $\boldsymbol{R}$, $\boldsymbol{\Sigma_0}$ as specified in Theorem 3.

Condition 4. *Let λ_j be the j_{th} eigenvalue of the matrix $\boldsymbol{CRC}'$, $j = 1, \cdots, r$. $c^{-1} \leq \lambda_{min} \leq \lambda_{max} \leq c$ for some constant $c > 0$.*

Theorem 3. *Under the alternative hypothesis H_1,*

(a) Assume Conditions 1 and 4 hold, then

$$T_{ME} \xrightarrow{d} \pi_0^2 \chi_1^2 \left(\varphi_0^2/\pi_0^2\right) + \pi_1^2 \chi_1^2 \left((\varphi_1+\varphi_2)^2/\pi_1^2\right), \text{ as } J \to \infty,$$

where $\varphi_0 = \mu_1 \sqrt{\boldsymbol{w'\Sigma_0^{-1}w}}$, $\varphi_1 = 2^{-1/2}\tau_1^2 \sqrt{tr(\boldsymbol{\Sigma_0^{-1}R})^2}$, $\varphi_2 = \mu_1^2 \boldsymbol{w'\Sigma_0^{-1}R\Sigma_0^{-1}w} \Big/ \sqrt{2tr(\boldsymbol{\Sigma_0^{-1}R})^2}$, $\pi_0^2 = 1 + \tau_1^2 \boldsymbol{w'\Sigma_0^{-1}R\Sigma_0^{-1}w}/\boldsymbol{w'\Sigma_0^{-1}w}$, *and* $\pi_1^2 = 1 + \left(\tau_1^4 tr(\boldsymbol{\Sigma_0^{-1}R})^4 + 2\tau_1^2 tr(\boldsymbol{\Sigma_0^{-1}R})^3\right)/tr(\boldsymbol{\Sigma_0^{-1}R})^2$.

(b) Assume Conditions 2 and 4 hold, then

$$T_{12} \xrightarrow{d} \breve{\pi}_0^2 \chi_1^2 \left(\breve{\varphi}_0^2/\breve{\pi}_0^2\right) + \breve{\pi}_1^2 \chi_1^2 \left((\breve{\varphi}_1+\breve{\varphi}_2)^2/\breve{\pi}_1^2\right), \text{ as } J \to \infty,$$

where $\breve{\varphi}_0 = \mu_1 \boldsymbol{w'w}/\sqrt{\boldsymbol{w'\Sigma_0 w}}$, $\breve{\varphi}_1 = \tau_1^2 tr(\boldsymbol{R}^2)/\sqrt{2tr(\boldsymbol{\Sigma_0 R})^2}$, $\breve{\varphi}_2 = \mu_1^2 \boldsymbol{w'Rw}/\sqrt{2tr(\boldsymbol{\Sigma_0 R})^2}$, $\breve{\pi}_0^2 = 1 + \tau_1^2 \boldsymbol{w'Rw}/\boldsymbol{w'\Sigma_0 w}$, *and* $\breve{\pi}_1^2 = 1 + \left(\tau_1^4 tr(\boldsymbol{R}^4) + 2\tau_1^2 tr(\boldsymbol{\Sigma_0 R^3})\right)/tr(\boldsymbol{\Sigma_0 R})^2$.

(c) Assume Conditions 3 and 4 hold, then

$$\tilde{T}_{ME} \xrightarrow{d} \tilde{\pi}_0^2 \chi_1^2 \left(\tilde{\varphi}_0^2/\tilde{\pi}_0^2\right) + \tilde{\pi}_1^2 \chi_1^2 \left((\tilde{\varphi}_1+\tilde{\varphi}_2)^2/\tilde{\pi}_1^2\right), \text{ as } J \to \infty,$$

where $\tilde{\varphi}_0 = \mu_1 \boldsymbol{w'Cw}/\sqrt{\boldsymbol{w'w}}$, $\tilde{\varphi}_1 = \tau_1^2 tr(\boldsymbol{CRC'R})/\sqrt{2tr(\boldsymbol{R}^2)}$, $\tilde{\varphi}_2 = \mu_1^2 \boldsymbol{w'C'RCw}/\sqrt{2tr(\boldsymbol{R}^2)}$, $\tilde{\pi}_0^2 = 1 + \tau_1^2 \boldsymbol{w'CRC'w}/\boldsymbol{w'w}$, *and* $\tilde{\pi}_1^2 = 1 + \left(\tau_1^4 tr(\boldsymbol{RCRC'})^2 + 2\tau_1^2 tr(\boldsymbol{RCRC'R})\right)/tr(\boldsymbol{R}^2)$.

To compare the asymptotic power between T_{ME} and T_{12}, first let us consider the simple case of no random effect, i.e. $\tau_1 = 0$. In that case, $\pi_0 = \breve{\pi}_0 = \pi_1 = \breve{\pi}_1 = 1$. Thus T_{ME} is reduced to T_{FE}, and T_{12} is reduced to $(\boldsymbol{w'\Sigma_0 w})^{-1} T_1$ based on Theorem 3. Further, T_{ME} is at least as powerful as T_{12} provided that $\varphi_0^2/\breve{\varphi}_0^2 = \boldsymbol{w'\Sigma_0 w w'\Sigma_0^{-1} w}/(\boldsymbol{w'w})^2 \geqslant 1$. In fact, the above inequality always holds as

long as $\boldsymbol{\Sigma_0}$ is a positive definite symmetric matrix. This reveals that if the true underlying model for $\boldsymbol{s}$ is a fixed-effect model, then T_{ME} is more powerful than T_{12}. Our analytical conclusion here is consistent with that observed by Liu and Lin (2018a) for jointly analyzing multiple phenotypes.

Second, if we consider a local alternative assuming $\tau_1^2\lambda_{max} = o(1)$, then $\pi_0 = 1+o(1), \breve{\pi}_0 = 1+o(1), \pi_1 = 1+o(1)$, and $\breve{\pi}_1 = 1+o(1)$. As a result, T_{ME} is at least as powerful as T_{12} provided that $\varphi_0^2/\breve{\varphi}_0^2 = \boldsymbol{w'\Sigma_0 w w'\Sigma_0^{-1} w}/(\boldsymbol{w'w})^2 \geqslant 1$, $\varphi_1^2/\breve{\varphi}_1^2 = tr(\boldsymbol{\Sigma_0 R})^2 tr(\boldsymbol{\Sigma_0^{-1} R})^2/tr^2(\boldsymbol{R}^2) \geqslant 1$, $\varphi_2/\breve{\varphi}_2 = \frac{\boldsymbol{w'\Sigma_0^{-1} R \Sigma_0^{-1} w}}{\boldsymbol{w'Rw}} \cdot \frac{\sqrt{tr(\boldsymbol{\Sigma_0 R})^2}}{\sqrt{tr(\boldsymbol{\Sigma_0^{-1} R})^2}} \geqslant 1$. The first two inequalities always hold as long as $\boldsymbol{R}$ and $\boldsymbol{\Sigma_0}$ are positive definite symmetric matrices. The last one depends on the specific structures of $\boldsymbol{w}, \boldsymbol{R}$, and $\boldsymbol{\Sigma_0}$ which we will exam by means of simulation.

4 Simulation studies

To compare the finite-sample performance of T_{ME}, $\tilde{T}_{ME}$, and T_{12}, with T_{skato}, T_{Minp}, and T_{Fisher}, we conduct extensive simulation studies, examining the effects of different correlation structures and signal sparsities. For the purpose of mimicking the rare variants association study scenario to obtain the summary statistics, $\boldsymbol{s}$ and $\boldsymbol{\Sigma_0}$, we follow the individual-level data generating framework used in Derkach et al. (2014). We assume $E(y_i|\boldsymbol{G}_i) = \beta_0 + \beta_1 G_{i1} + \cdots + \beta_J G_{iJ}$, where y_i is normally distributed with variance $\sigma^2 = 1$, and G_{ij} is Bernoulli with $Pr(G_{ij} = 1) = p_j$, $i = 1, \cdots, n$ and $j = 1, \cdots, J$, and p_j is approximately twice the minor allele frequency of variant j. Given this set-up, $\boldsymbol{s} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma_0})$, where $\boldsymbol{\mu} = (np_1(1-p_1)\beta_1, \cdots, np_J(1-p_J)\beta_J)'$, $\boldsymbol{\Sigma}_0 = \{\sigma_{jk}^2\}_{J\times J}$, $\sigma_{jk}^2 = np_j(1-p_j)$ for $j = k$ and $\sigma_{jk}^2 = n(p_{jk} - p_j p_k)$ for $j \neq k$, and $p_{jk} = P(G_{ij} = 1, G_{ik} = 1)$.

In the subsequent studies, we consider $J = 10,\ 50,\ 100,\ 500$, and 1000, p_j randomly drawn from Unif $(0.005, 0.02)$, $j = 1, \cdots, J$, and $\boldsymbol{\beta} = \boldsymbol{0}$ under the null and following various different structures under alternatives. For $\text{diag}\{\sigma_{jj}^{-1}\}\boldsymbol{\Sigma_0}\,\text{diag}\{\sigma_{jj}^{-1}\}$, we consider an AR(1) pattern with correlation $\tilde{\rho}$, and $\tilde{\rho} = 0.2$, 0.5, and 0.8. For $\boldsymbol{w}$ and $\boldsymbol{A}$ in T_{skato}, T_{Minp}, T_{Fisher}, and T_{12}, we choose the commonly used $w_j = 1/\sqrt{p_j(1-p_j)}$ and $\boldsymbol{A} = \text{diag}\{w_j^2\}$ without loss of generality. For $\boldsymbol{w}$ and

$\boldsymbol{R}$ in T_{ME} and $\tilde{T}_{ME}$, we choose the same w_j and let $\boldsymbol{R} = \boldsymbol{A}$ for a fair comparison.

4.1 Type I error

To examine the validity of the proposed tests, T_{ME}, $\tilde{T}_{ME}$, and T_{12}, we generate $\boldsymbol{s}$ from $N(\boldsymbol{0}, \boldsymbol{\Sigma_0})$, independently, 10^6 times for each J and $\tilde{\rho}$ combination. Table 2 provides the empirical type I error rates for $\alpha = 5\%, 1\%, 0.1\%$, and 0.01% estimated based on the 10^6 replications for $\tilde{\rho} = 0.5$; results similar for other $\tilde{\rho}$ values (see Supplementary Material). $T_{.,asy}$ represents the asymptotic results based on Theorem 1. For $T_{.,apr}$ based on Theorem 2, 10^7 independent random variables u_1 and v_j are generated for each replication. The results in Table 2 show that for small J, in combination with stringent α level, p-value evaluation based on the asymptotic distributions in Theorem 1 is not adequate. In that case, the approximate solution in Theorem 2 should be used.

For the existing methods, T_{skato}, T_{Minp}, and T_{Fisher}, we observed in our simulation studies that T_{skato} is slightly conservative for the α levels considered regardless of the size of J, T_{Minp} is also slightly conservative for small J but has correct test size when $J > 50$, and T_{Fisher} has inflated type I error when correlation is strong; see supplementary material for detailed simulation results.

4.2 Power without covariates

We consider two different simulation designs to evaluate power. For both designs, P_c, the proportion of causal variants for a given set of J variants, is randomly drawn from $\text{Unif}(0.01, 0.1)$ for the case of sparse signal, and $P_c \sim \text{Unif}\,(0.1, 0.5)$ for moderately sparse case. Among the causal variants, the proportion of deleterious variants with $\beta_j > 0$, $P_d \sim \text{Unif}\,(0.5, 0.75)$. We note that for each P_c and P_d combination, the locations of the signals ($\beta_j \neq 0$) are randomly drawn from $\{1, 2, \cdots, J\}$, without replacement. This randomness helps us to comprehensively explore the effect of different correlation structures between causal variants, between non-causal variants, as well as between causal and non-causal variants.

Table 2: **Type I error evaluation.** Empirical test sizes for $\alpha = 5\%$, 1%, 0.1%, and 0.01%, estimated based on 10^6 replications independently simulated under the null. For p-value evaluation, $T_{\cdot,asy}$ represents using the asymptotic distributions in Theorem 1, and $T_{\cdot,apr}$ represents using the approximate solution in Theorem 2 with 10^7 independent simulated $N(0,1)$ variables. Results here are for $\tilde{\rho} = 0.5$ in $\mathbf{\Sigma}_0$; results for other $\tilde{\rho}$ values are provided in the Supplementary Material.

J	α	$T_{12,asy}$	$\tilde{T}_{me,asy}$	$T_{me,asy}$	$T_{12,apr}$	$\tilde{T}_{me,apr}$	$T_{me,apr}$
10	5%	6.1558	5.4401	5.6556	5.7470	5.1676	5.2226
	1%	3.1219	2.2454	2.4281	1.7649	1.3830	1.2969
	0.1%	1.4531	0.7937	0.9427	0.2079	0.1575	0.1316
	0.01%	0.7725	0.3260	0.4294	0.0196	0.0158	0.0141
20	5%	5.6845	5.0952	5.5795	5.4308	4.9646	5.2198
	1%	2.5849	1.8642	2.4312	1.6450	1.2781	1.3667
	0.1%	1.0219	0.5893	0.9643	0.2212	0.1565	0.1476
	0.01%	0.4703	0.2167	0.4550	0.0246	0.0162	0.0149
50	5%	5.1085	4.9573	5.1317	5.0333	4.9307	4.9875
	1%	1.9088	1.3816	1.7082	1.4498	1.1470	1.1475
	0.1%	0.5952	0.3178	0.5002	0.2103	0.1446	0.1307
	0.01%	0.2148	0.0852	0.1794	0.0230	0.0165	0.0132
100	5%	4.9508	4.9621	5.0064	4.923	4.9672	4.9664
	1%	1.5279	1.2223	1.3854	1.2788	1.0938	1.1030
	0.1%	0.3796	0.2207	0.3082	0.1846	0.1428	0.1273
	0.01%	0.1149	0.0464	0.0831	0.025	0.0162	0.0138
500	5%	4.9335	4.9757	5.0020	4.9307	4.9759	4.9948
	1%	1.1265	1.0426	1.0917	1.0720	1.0112	1.0238
	0.1%	0.1662	0.1273	0.1547	0.1288	0.1094	0.1130
	0.01%	0.0332	0.0176	0.0281	0.0202	0.0123	0.0139
1000	5%	4.9841	4.9887	4.9552	4.9743	4.9889	4.9436
	1%	1.0475	1.0045	1.0254	1.0151	0.9886	0.9864
	0.1%	0.1365	0.1127	0.1190	0.1183	0.1065	0.0951
	0.01%	0.0205	0.0135	0.0162	0.0163	0.0116	0.0109

Design one follows the approach of Derkach et al. (2014). That is, $\boldsymbol{s} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\mu} = (np_1(1-p_1)\beta_1, \cdots, np_J(1-p_J)\beta_J)'$, $n = 500$, $\text{diag}\{\sigma_{jj}^{-1}\}\boldsymbol{\Sigma_0}\,\text{diag}\{\sigma_{jj}^{-1}\}$ follows the AR(1) model with varying $\tilde{\rho}$, and $|\beta_j| \sim \text{Unif}(0.5, 1.5)$ for both the sparse and moderately sparse cases. Design two assumes that $\boldsymbol{s}$ is drawn from the mixed-effect model (7) with varying magnitudes of μ and τ^2. Specifically, $\boldsymbol{s} \sim N(\boldsymbol{w}^*\beta, \tau^2\boldsymbol{I}_J + \boldsymbol{\Sigma}_0)$, where $\boldsymbol{w}^* = (np_1(1-p_1)\text{sign}(\beta_1), \cdots, np_J(1-p_J)\text{sign}(\beta_J))'$, β represents the common fixed effect of $\boldsymbol{\beta}$, and τ^2 captures the random effect. For this design, $\beta \sim \text{Unif}(0.5, 1.5)$ for both the sparse and moderately sparse cases as in design one, and $\tau^2 \sim \text{Unif}(1, 2)$.

For both designs, to compare power between methods we focus on $J = 100$ and $\alpha = 0.05$ without loss of generality; results for other parameter values are characteristically similar. The empirical power for $\alpha = 0.05$ are estimated from 10^3 independently simulated replicates, and using the empirical critical values obtained from 10^4 corresponding null replicates. As the performances under sparse and moderate sparse cases share similar patterns, we present the results for the sparse case here and the other results in the supplementary material.

Under simulation design one and the sparse signal case, Figures 1-2 show the empirical power of the proposed test statistics T_{12}, $\tilde{T}_{ME}$, and T_{ME} as compared to T_{skato}, T_{Minp}, and T_{Fisher}, respectively, for $\tilde{\rho} = 0.5$ and $\tilde{\rho} = 0.2$; see supplementary material for $\tilde{\rho} = 0.8$. In each figure, there are 200 randomly simulated models where $p_j \sim \text{Unif}(0.005, 0.02)$, $P_c \sim \text{Unif}(0.01, 0.1)$, $P_d \sim \text{Unif}(0.5, 0.75)$, and $|\beta_j| \sim \text{Unif}(0.5, 1.5)$; see supplementary material for the moderately sparse case where $P_c \sim \text{Unif}(0.1, 0.5)$.

Based on the results in Figures 1-2, first we note that when the correlation among variants is relatively strong (e.g. $\tilde{\rho} = 0.5$ in Figures 1 and $\tilde{\rho} = 0.8$ in supplementary material), T_{ME} derived from the proposed mixed-effect regression model, using $\boldsymbol{\Sigma_0^{-1}}\boldsymbol{s}$ instead of $\boldsymbol{s}$, increases power. However, this approach may not be advantageous when there is only weak correlation in conjunction with sparse signal (e.g. $\tilde{\rho} = 0.2$ in Figures 2) as discussed in Section 3.2. Interestingly, the new test T_{12} (which has the same structure as T_{ME} but uses $\boldsymbol{s}$) and test $\tilde{T}_{ME}$ (uses $\boldsymbol{Cs}$, where $\boldsymbol{C}'\boldsymbol{\Sigma}\boldsymbol{C} = \boldsymbol{I}_J$)

have comparable power with T_{skato}, but without the need to search for the 'optimal' ρ. Our simulation results also confirm that the three existing hybrid tests, T_{skato}, T_{Minp}, and T_{Fisher}, largely have similar performance, where T_{skato} and T_{Minp} perform more similar with each other than with T_{Fisher}.

The results for simulation design two and under the sparse case are shown in Figures 3-4, respectively, for $\tilde{\rho} = 0.5$ and $\tilde{\rho} = 0.2$; see supplementary material for $\tilde{\rho} = 0.8$. In each figure, there are 200 randomly simulated models where p_j, P_c, and P_d are simulated as in design one, while $\beta \sim \text{Unif}(0.5, 1.5)$ and $\tau^2 \sim \text{Unif}\,(1,2)$; see supplementary material for additional results using other parameter values. As expected, when $\boldsymbol{s}$ follows a mixed-effect model (7), the advantage of the proposed T_{ME} is enhanced. In this case, power of T_{ME} is considerably higher than the other tests even when $\tilde{\rho} = 0.2$ for most of the 200 models simulated (Figures 4). We also note that in our power studies, we used the sub-optimal $w_j = 1/\sqrt{p_j(1-p_j)}$ and $\boldsymbol{R} = \text{diag}\{w_j^2\}$. Performance of T_{ME} can be further improved by using the oracle $w_j = np_j(1-p_j)sign(\beta_j)$ and $\boldsymbol{R} = \boldsymbol{I}$.

4.3 Power with covariates

We now briefly study the effect of incorporating variant-specific additional information $\boldsymbol{z} = (z_1, \cdots, z_J)'$. As discussed before, although one may revise w_j to be proportional to z_j, in additional to MAF, it is not immediately clear how to choose an 'optimal' weighting function. Thus, we only study the proposed $T_{ME,cov}$, derived directly from the regression model (12), and we consider simulation design two only. Without loss of generality, we assume z_j to be an indicator variable, for example the variant being non-synonymous ($z_j = 1$) or synonymous ($z_j = 0$). For causal variants we let $Pr(z_j = 1) = 0.5$, and for non-causal variants $Pr(z_j = 1) = 0$. We consider both the case of informative $\boldsymbol{z}$ ($\theta \neq 0$ in model (12)) and the case of uninformative $\boldsymbol{z}$ ($\theta = 0$).

Figure 5 shows the results for $J = 100$ and $\tilde{\rho} = 0.5$ under the sparse signal case where $P_c = 0.1$; $P_d \sim \text{Unif}\,(0.5, 1)$ and $p_j \sim \text{Unif}\,(0.005, 0.02)$. Because of the additional information available from $\boldsymbol{z}$, we decrease β to be drawn from $\text{Unif}\,(0.1, 1)$ and choose $\tau^2 = 0$. When $\boldsymbol{z}$ is informative,

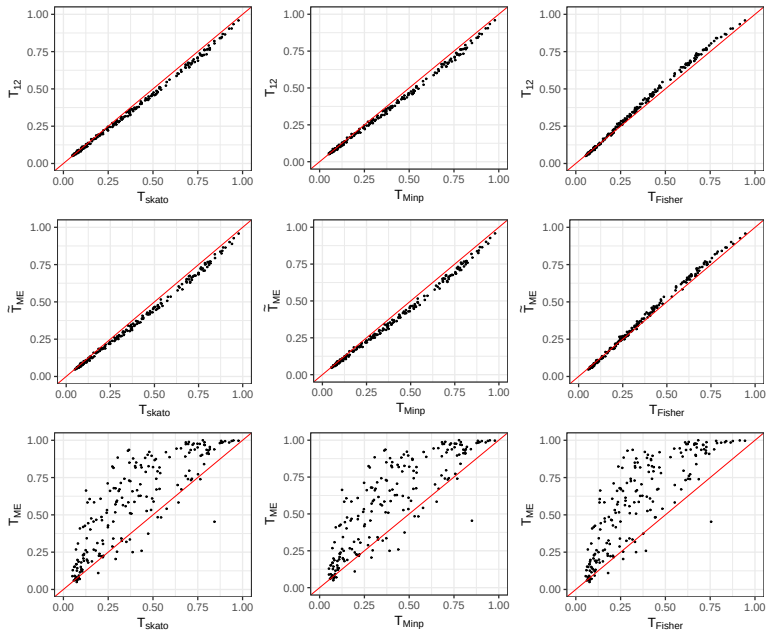

Figure 1: **Comparison of power for sparse signals and $\tilde{\rho} = 0.5$ based on design ONE.** Compare T_{12}, $\tilde{T}_{ME}$, T_{ME}, T_{Fisher}, T_{Minp} and T_{skato} using 200 alternative models with sparse signals simulated based on design one. $J = 100$ and the proportion of the causal variants varies from 1% to 10%. Sample size $n = 500$ and $\alpha = 0.05$.

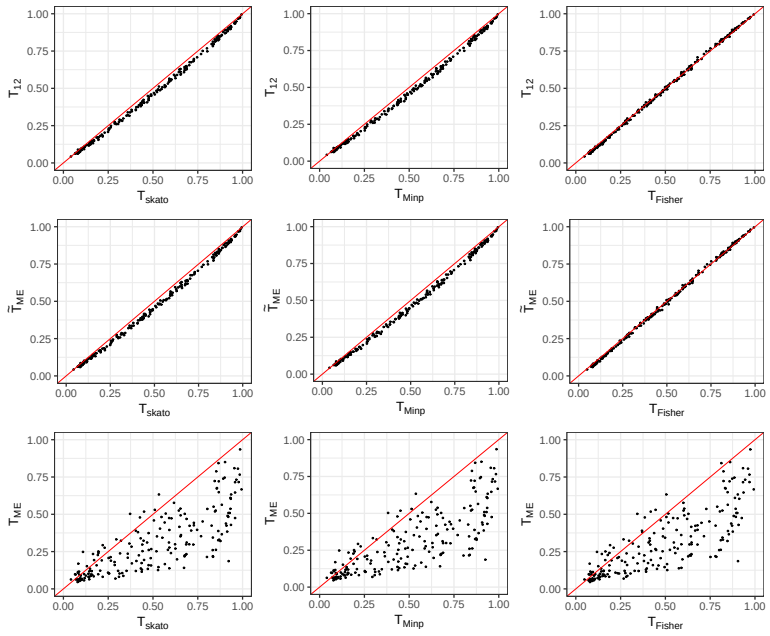

Figure 2: **Comparison of power for sparse signals and $\tilde{\rho} = 0.2$ based on design ONE.** Compare T_{12}, $\tilde{T}_{ME}$, T_{ME}, T_{Fisher}, T_{Minp} and T_{skato} using 200 alternative models with sparse signals simulated based on design one. $J = 100$ and the proportion of the causal variants varies from 1% to 10%. Sample size $n = 500$ and $\alpha = 0.05$.

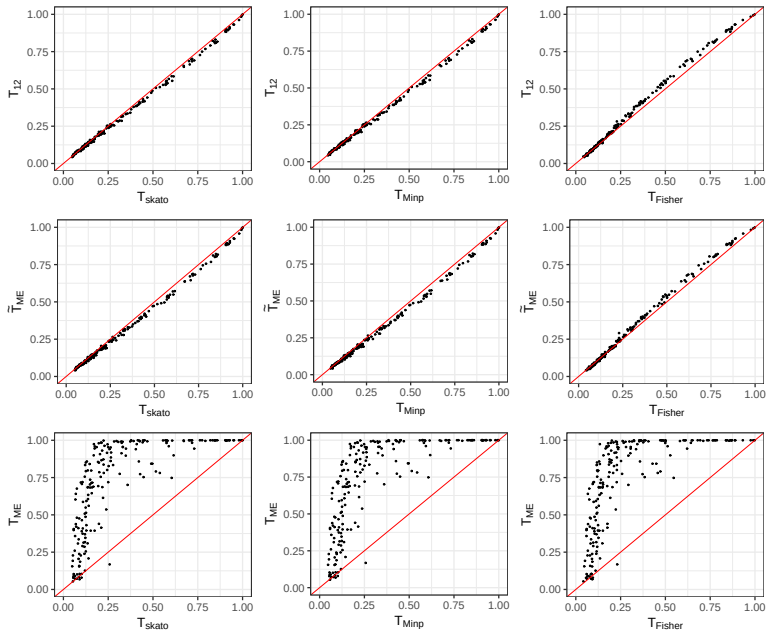

Figure 3: **Comparison of power for sparse signals and $\tilde{\rho} = 0.5$ based on design TWO.** Compare T_{12}, $\tilde{T}_{ME}$, T_{ME}, T_{Fisher}, T_{Minp} and T_{skato} using 200 alternative models with sparse signals simulated based on design two. $J = 100$ and the proportion of the causal variants varies from 1% to 10%. Sample size $n = 500$ and $\alpha = 0.05$.

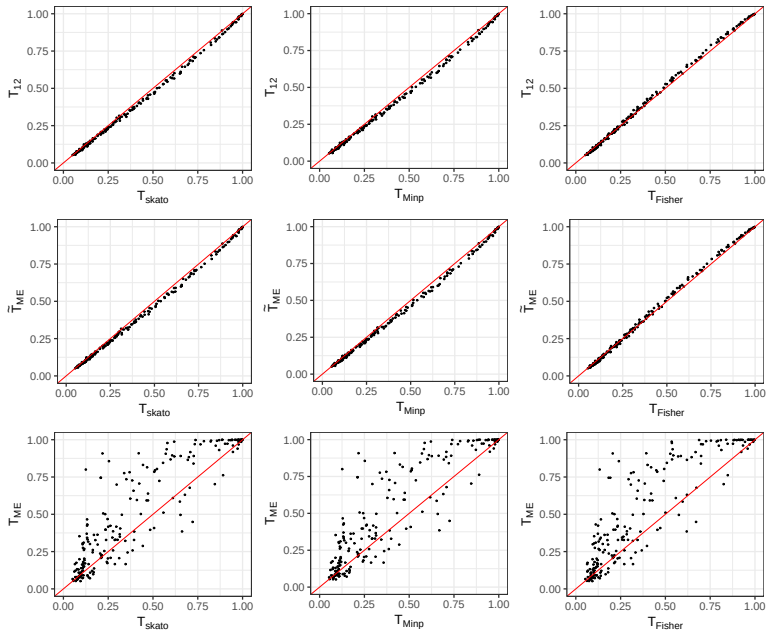

Figure 4: **Comparison of power for sparse signals and $\tilde{\rho} = 0.2$ based on design TWO.** Compare T_{12}, $\tilde{T}_{ME}$, T_{ME}, T_{Fisher}, T_{Minp} and T_{skato} using 200 alternative models with sparse signals simulated based on design two. $J = 100$ and the proportion of the causal variants varies from 1% to 10%. Sample size $n = 500$ and $\alpha = 0.05$.

$\theta \sim \text{Unif}(1,4)$. As expected, there can be substantial power gain when incorporating informative covariate information (left plot in Figure 5), at the cost of slightly reduced power when $\boldsymbol{z}$ is uninformative (right plot in Figure 5).

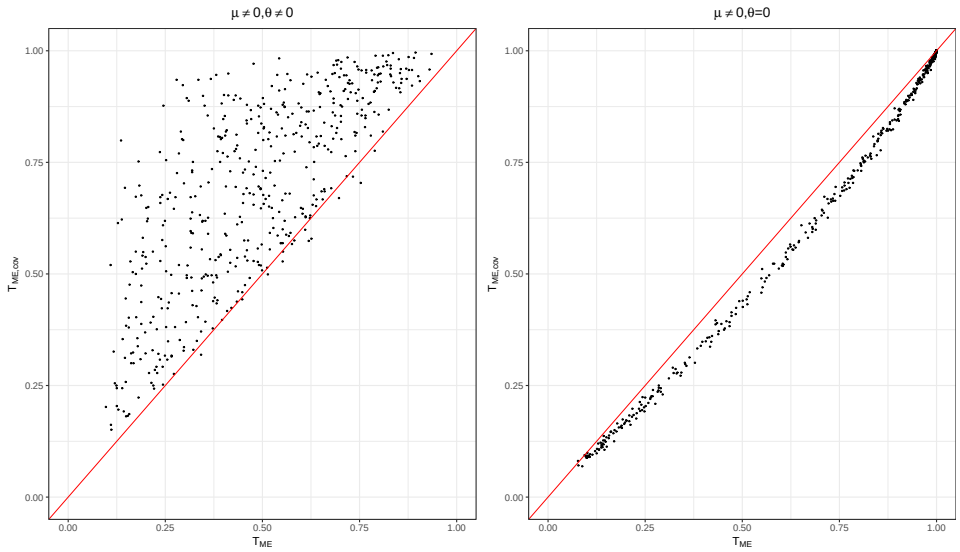

Figure 5: **Power with and without covariates**. Compare $T_{ME,cov}$ and T_{ME} using 500 alternative models with sparse signals simulated based on design two. $J = 100$, $P_c = 0.1$, $P_d \sim U(0.5,1)$, $\beta \sim U(0.1,1)$, $\tilde{\rho} = 0.5$. Sample size $n = 500$ and $\alpha = 0.05$.

5 Applications

In this section, we examine nine test statistics through two data applications. The nine tests examined include the four new methods, $T_{ME,cov}$, T_{ME}, $\tilde{T}_{ME}$, and T_{12}, and the existing methods, T_{skato}, T_{Minp}, T_{Fisher}, as well as T_1 and T_2 for completeness; see Table 1 for a summary of the different tests. In the implementation of $T_{ME,cov}$, we use variants being non-synonymous or synonymous, annotated using the UCSC genome browser at https://genome.ucsc.edu/, as the variant-specific information.

The first application highlights the advantage of the proposed T_{ME} in the presence of high or moderately high correlation between variants, and it also demonstrates that the method is not

limited to analyses of rare variants. The second application revisits the genetic analysis workshop 17 (GAW17) rare variants data previously studied by Derkach et al. (2014). This application reveals the benefit of incorporating additional variant-specific information using $T_{ME,cov}$, derived from the proposed summary statistic-based mixed-effect regression model (12).

5.1 Cystic Fibrosis (CF) data - common variants

Cystic Fibrosis is a life-limiting genetic condition for which lung function is a primary co-morbidity of interest. To indirectly study gene-environment interactions, Soave et al. (2015) proposed a joint location-scale (JLS) test and applied it to lung function measures in CF individuals, $n = 1,409$ from a Canadian sample and $n = 1,232$ from a French sample. They discovered and replicated the significance of the SLC9A3 complex set (35 common variants from four genes) based on the JLS test. However, the signal appears to come from the scale (interaction) component. For the traditional location (mean) test based on T_2, the SLC9A3 complex set was only significant in the Canadian sample but not replicated in the French sample. Here we exam the performance of the nine tests applied to the French sample.

To implement all the tests we use $\boldsymbol{w} = \mathbf{1}$ and $\boldsymbol{A} = \boldsymbol{I}$, and $\boldsymbol{R} = \boldsymbol{I}$ for fair comparison, since we did not find that MAF-dependent weighting enhance the performance of the tests. Because the number of variants of interest here is not large enough for using the asymptotic distributions, we obtain empirical p-values for all tests based on 10^4 permutation replicates.

Results in Table 3 show that only some of the genes appear to be truly associated with lung function in CF. For SLC9A3, all tests have suggestive evidence with T_1 having p-value < 0.05. For SLC9A3R1, benefiting from the correlation structure (Figure S9), the proposed T_{ME} and $T_{ME,cov}$ (and $\tilde{T}_{ME}$), which use $\boldsymbol{\Sigma}_{\mathbf{0}}^{-1}\boldsymbol{s}$ (and $\boldsymbol{Cs}$) instead of $\boldsymbol{s}$, are significant. When jointly analyzing all four genes in the SLC9A3 complex set, none of the tests is statistically significant but $\tilde{T}_{ME}$ has the smallest p-value. A larger sample is needed to make a definitive conclusion of true association. The covariate information (non-synonymous vs. synonymous) appear not to be informative here,

but the performance of $T_{ME,cov}$ is similar to that of T_{ME}.

Table 3: **Empirical p-values of the tests in the CF data application.**

		Empirical p-value								
Gene	J	T_1	T_2	T_{skato}	T_{Minp}	T_{Fisher}	T_{12}	$\tilde{T}_{ME}$	T_{ME}	$T_{ME,cov}$
SLC9A3	7	0.0443	0.1198	0.079	0.0783	0.0541	0.0718	0.0586	0.0886	0.0860
EZR	10	0.2192	0.6856	0.3241	0.3209	0.3897	0.239	0.3602	0.3473	0.3474
SLC9A3R2	10	0.804	0.5951	0.6965	0.6984	0.709	0.9037	0.7999	0.9683	0.8791
SLC9A3R1	8	0.0999	0.1103	0.1471	0.1656	0.0846	0.1042	0.0250	0.0243	0.0243
4-gene jointly	35	0.8372	0.3079	0.4749	0.4738	0.5671	0.9261	0.1142	0.1710	0.1709

5.2 The Genetic Analysis Workshop 17 (GAW17) data - rare variants

Here we apply the method to the GAW17 data provided by the 1000 Genomes Project (Consortium et al., 2010), focusing on the simulated quantitative trait Q2. The phenotype Q2 is influenced by 72 variants in 13 genes but not by environmental factors, and the genotypes of these variants are obtained from a 'mini-exome' next-generation sequencing experiment. Available to us are 200 replicates, simulated based on a true phenotype-genotype association model determined by the GAW17 study group but blinded to this analysis. We consider $n = 321$ unrelated Asian samples (Han Chinese, Denver Chinese, and Japanese) and use only variants with MAF less than 0.05. The description of the variants is provided in Table 4. Among the 13 genes, GCKR is not analyzed since only one variant remained after variant screening. VNN1 does not have any causal rare variants but is kept for negative control.

For each of the 200 alternative replicates, we calculate the empirical p-values (based on 10^4 permutation replicates) for the nine test statistics. For each test, the power for $\alpha = 0.05$ is estimated as the proportion of the 200 replicates for which the empirical p-values ≤ 0.05. We separate the 11 genes into three categories based on power as in Derkach et al. (2014) (which examined T_1 and T_2)

and Derkach et al. (2013) (which examined T_1, T_2, T_{Minp}, and T_{Fisher}), and we also jointly analyze all genes within each category.

In this application, because the correlation is weak among variants (Figure S10), we anticipate that methods relies on $\boldsymbol{s}$ will have better power than those based on $\boldsymbol{\Sigma_0^{-1}s}$ or $\boldsymbol{Cs}$. Indeed, results in Table 4 show that T_{12} has better performance than T_{ME} and $\tilde{T}_{ME}$. However, this application clearly demonstrate the potential of incorporating informative covariates. For example, the power of analyzing RARB, PLAT and VLDLR is significantly improved using $T_{ME,cov}$, at the cost of slightly reduced performance if the included covariate is not (detectably) informative. Interestingly, T_{Fisher} has comparable performance as T_{12}, and both outperform T_{skato} in almost all cases. Although the individual T_1 and T_2 tests may have the highest power for certain genes, the robustness of the hybrid tests is evident based on the overall performance exhibited in Table 4.

6 Discussion

In this paper, we considered a summary statistic-based regression framework to analyze a set of J variants simultaneously. As delineated in Table 1, the proposed approach is flexible and adaptive. The score test derived from the fixed-effect model, T_{FE}, unifies the linear class of tests (also known as the burden tests), T_1, derived from models requiring individual-level data, while T_{RE} from the random-effect model connects the quadratic class or variance component tests, T_2. Further, the score test derived from the random-effect model offers a new hybrid test, T_{ME}, that naturally aggregates information from T_{FE} and T_{RE}.

In contrast to the well-known SKAT-O, it is worth emphasizing two notable differences. First, the proposed framework aggregates evidence across J variants based on $\boldsymbol{\Sigma_0^{-1}s}$, a precision matrix-based transformation of the score vector, that can increase power for sparse alternatives (Fan et al., 2013; Cai et al., 2014). Secondly, when additional variant-specific information is available, it is straightforward to derive $T_{ME,cov}$ that accounts for covariate effects. We have demonstrated these

Table 4: **Empirical power of the tests in the GAW17 data application.**

Gene	J_C	J_N	Empirical Power								
			T_1	T_2	T_{skato}	T_{Minp}	T_{Fisher}	T_{12}	$\tilde{T}_{ME}$	T_{ME}	$T_{ME,cov}$
7 genes for which the maximum power is 10% or more											
SIRT1	4	7	0.44	0.385	0.455	0.43	0.495	0.5	0.315	0.285	0.285
BCHE	6	9	0.29	0.39	0.405	0.39	0.435	0.445	0.45	0.46	0.41
PDGFD	3	5	0.295	0.385	0.385	0.38	0.425	0.45	0.37	0.31	0.26
SREBF1	4	5	0.495	0.25	0.440	0.440	0.445	0.400	0.440	0.405	0.355
RARB	1	5	0.06	0.135	0.095	0.110	0.085	0.090	0.090	0.100	0.215
PLAT	4	7	0.13	0.125	0.105	0.100	0.135	0.12	0.115	0.13	0.165
VLDLR	4	3	0.11	0.085	0.100	0.095	0.115	0.115	0.110	0.110	0.16
7-gene jointly	26	41	0.935	0.765	0.94	0.935	0.945	0.94	0.79	0.765	0.765
4 genes for which the maximum power is 10% or less											
VNN3	2	2	0.035	0.04	0.035	0.035	0.04	0.04	0.04	0.04	0.035
INSIG1	3	1	0.05	0.03	0.03	0.03	0.035	0.035	0.025	0.025	0.015
LPL	1	3	0.03	0.065	0.035	0.035	0.04	0.035	0.025	0.035	0.050
VWF	1	3	0.025	0.01	0.01	0.01	0.015	0.01	0.035	0.04	0.045
4-gene jointly	7	9	0.075	0.01	0.06	0.06	0.055	0.045	0.055	0.055	0.045
1 gene for which there is no rare causal variants, used as a negative control											
VNN1	0	3	0.015	0.045	0.045	0.045	0.035	0.04	0.04	0.04	0.045

features of the proposed method using both analytical results and empirical studies.

To exploit the assumption of signal sparsity, various supremum-type tests have been proposed including the generalized higher criticism (Barnett et al., 2017) for sparse signals, and most recently the generalized Berk-Jones statistic (Sun and Lin, 2017) for moderate sparse signals. These methods, tailored for common variants, are not easy to adjust for additional variant-specific information when individual-level data are not available. See Lin and Zeng (2010) for a general discussion of the relative efficiency between mega- and meta-analysis.

The proposed set-based testing framework is a general one, and it can be used for other settings such as pleiotropy studies of multiple phenotypes, where the analtyical unit is each of the phenotypes. In that context, Liu and Lin (2018b) also proposed a summary statistic-based linear mixed-effect regression model, but they focused on the special case of $\boldsymbol{w} = \mathbf{1}$ and $\boldsymbol{R} = \boldsymbol{I}$. In addition, Liu and Lin (2018b) derived two score test statistics, respectively, for testing $\mu = 0$ and $\tau^2 = 0$ *separately*, then considered different ways to combine the evidence including SKAT-O type of statistics. In contrast, we derive T_{ME} from testing $\mu = 0$ and $\tau^2 = 0$ *jointly*, and the weighting factors are inherently justified. We also study the asymptotic properties of the proposed tests under the null and alternatives, in addition to the study of covariate adjustments when only variant-specific information is available.

The proposed method can also be used for the study of polygenic risk score (PRS)(Purcell et al., 2009), and the connection between PRS and burden type of tests (T_1) has been noted by Pan et al. (2015). In principle, T_{ME} can overcome the poor statistical efficiency of T_1 as adopted in the PRS test. However, the estimation of large precision matrices can be challenges and requires special considerations (Fan et al., 2016). The link between T_1 and existing PrediXcan (Gamazon et al., 2015) for association and tissue-specific gene-expression data integration has also been noted (Xu et al., 2017). The performance of T_{ME} in this setting and comparison with other concurrently developed newer methods are of our future research interest.

Fix-, random-, and mixed-effect models for summary statistics have been studied for meta-

441 analysis of GWAS (Han and Eskin, 2011). In that context, a likelihood ratio test was implemented
442 for the mixed-effect model, and the resulting test is also known as the new random-effect meta-
443 analysis. The original test of Han and Eskin (2011) was designed for meta-analysis of independent
444 studies, and a modified procedure has since been developed by Lee et al. (2017) to account for
445 correlations between studies but not covariate effects. Comparison between the two approaches
446 for meta-analysis and other studies warrants future investigations.

447 Acknowledgements

448 We thank Dr. Lisa J. Strug and her lab, and Dr. Harriet Corvol for providing the cystic fibrosis
449 application data, and we thank the Genetic Analysis Workshop 17 (GAW17) committee and the
450 1000 Genomes Project for providing the GAW17 application data. We would also like to thank Dr.
451 Andriy Derkach for helpful discussions. YZ is a trainee of the CIHR STAGE (Strategic Training
452 in Advanced Genetic Epidemiology) training program at the University of Toronto. This research
453 is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-
454 04934 and RGPAS-522594), the Canadian Institutes of Health Research (CIHR, MOP-310732),
455 and the University of Toronto McLaughlin Centre Accelerator Grants in Genomic Medicine (MC-
456 2019-15).

References

Barnett, I., R. Mukherjee, and X. H. Lin, 2017: The generalized higher criticism for testing snp-set effects in genetic association studies. *J. Amer. Statist. Assoc.*, **112**, 64–76.

Cai, T. T., W. Liu, and Y. Xia, 2014: Two-sample test of high dimensional means under dependence. *J. R. Statist. Soc. B*, **76**, 349–372.

Consortium, . G. P., and Coauthors, 2010: A map of human genome variation from population-scale sequencing. *Nature*, **467 (7319)**, 1061.

Derkach, A., J. F. Lawless, and L. Sun, 2013: Robust and powerful tests for rare variants using fisher's method to combine evidence of association from two or more complementary tests. *Genet. Epidemiol.*, **37**, 110–121.

Derkach, A., J. F. Lawless, and L. Sun, 2014: Pooled association tests for rare genetic variants: a review and some new results. *Statistical Science*, **29**, 302–321.

Donoho, D., and J. Jin, 2004: Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, **32**, 962–994.

Fan, J., Y. Liao, and H. Liu, 2016: An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, **19**, 1–32.

Fan, Y. Y., J. S. Jin, and Z. G. Yao, 2013: Optimal classification in sparse gaussian graphic model. *Ann. Statist.*, **41**, 2537–2571.

Finucane, H. K., and Coauthors, 2015: Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*, **47 (11)**, 1228–1235.

Gamazon, E. R., and Coauthors, 2015: A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics*, **47**, 1091–1098.

Hall, P., and J. Jin, 2010: Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.*, **38**, 1686–1732.

Han, B., and E. Eskin, 2011: Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 586–598.

Ionita-Laza, I., K. McCallum, B. Xu, and J. Buxhaum, 2016: A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nature Genetics*, **48**, 214–220.

Lee, C. H., E. Eskin, and B. Han, 2017: Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects. *Bioinformatics*, **33**, 379–388.

Lee, S., G. Abecasis, M. Boehnke, and X. Lin, 2014: Rare variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.*, **95**, 5–23.

Lee, S., M. C. Wu, and X. Lin, 2012: Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*, **13**, 762–775.

Li, B., and S. M. Leal, 2008: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.*, **83**, 311–321.

Lin, D., and D. Zeng, 2010: On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika*, **97**, 321–332.

Liu, Z. H., and X. H. Lin, 2018a: A geometric perspective on the power of principal component association tests in multiple phenotype studies. *J. Amer. Statist. Assoc.*

Liu, Z. H., and X. H. Lin, 2018b: Multiple phenotype association tests using summary statistics in genome-wide association studies. *Biometrics*, **74**, 165–175.

Madsen, B., and S. Browning, 2009: A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genetics*, **5**, e1000 384.

Morgenthaler, S., and W. Thilly, 2007: A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (cast). *Mutation Research*, **615**, 28–56.

Neale, B. M., and Coauthors, 2011: Testing for an unusual distribution of rare variants. *PLoS Genetics.*, **7**, e1001 322.

Pan, W., 2009: Asymptotic tests of association with multiple snps in linkage disequilibrium. *Genet. Epidemiol.*, **33**, 497–507.

Pan, W., Y. M. Chen, and P. Wei, 2015: Testing for polygenic effects in genome-wide association studies. *Genet. Epidemiol.*, **39**, 306–316.

Purcell, S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'donovan, P. F. Sullivan, and P. Sklar, 2009: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460 (7256)**, 748–752.

Soave, D., and Coauthors, 2015: A joint location-scale test improves power to detect associated snps, gene sets, and pathways. *Am. J. Hum. Genet.*, **97**, 125–138.

Sun, R., and X. H. Lin, 2017: Set-based tests for genetic association using the generalized berk-jones statistic. *arXiv:1710.02469*.

Wu, S. C. T. L. Y. B. M., M. C.and Lee, and X. Lin, 2011: Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.*, **89**, 82–93.

Xu, G. J., L. F. Lin, P. Wei, and W. Pan, 2016: An adaptive two-sample test for high-dimensional. *Biometrika*, **103**, 609–624.

522 Xu, Z., C. Wu, P. Wei, and W. Pan, 2017: A powerful framework for integrating eqtl and gwas
523 summary data. *Genetics*, **207**, 893–902.