Detecting Association of Rare and Common Variants by Testing an Optimally Weighted Combination of Variants

Qiuying Sha,[1] Xuexia Wang,[2] Xinli Wang,[3] and Shuanglin Zhang[1]*

[1] *Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan*
[2] *Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin*
[3] *School of Technology, Michigan Technological University, Houghton, Michigan*

Next-generation sequencing technology will soon allow sequencing the whole genome of large groups of individuals, and thus will make directly testing rare variants possible. Currently, most of existing methods for rare variant association studies are essentially testing the effect of a weighted combination of variants with different weighting schemes. Performance of these methods depends on the weights being used and no optimal weights are available. By putting large weights on rare variants and small weights on common variants, these methods target at rare variants only, although increasing evidence shows that complex diseases are caused by both common and rare variants. In this paper, we analytically derive optimal weights under a certain criterion. Based on the optimal weights, we propose a Variable Weight Test for testing the effect of an Optimally Weighted combination of variants (VW-TOW). VW-TOW aims to test the effects of both rare and common variants. VW-TOW is applicable to both quantitative and qualitative traits, allows covariates, can control for population stratification, and is robust to directions of effects of causal variants. Extensive simulation studies and application to the Genetic Analysis Workshop 17 (GAW17) data show that VW-TOW is more powerful than existing ones either for testing effects of both rare and common variants or for testing effects of rare variants only. *Genet. Epidemiol.* 36:561–571, 2012.

Key words: optimal weights; rare variants; common variants; association studies; next-generation sequencing

Supporting Information is available in the online issue at wileyonlinelibrary.com.

*Correspondence to: Shuanglin Zhang, Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931. E-mail: shuzhang@mtu.edu

Received 11 February 2012; Revised 13 April 2012; Accepted 13 April 2012
Published online 19 June 2012 in Wiley Online Library (wileyonlinelibrary.com/journal/gepi).
DOI: 10.1002/gepi.21649

INTRODUCTION

There is increasing evidence showing that complex diseases are caused by both common and rare variants [Bodmer and Bonilla, 2008; Ng et al., 2009; Pritchard, 2001; Pritchard and Cox, 2002; Stratton and Rahman, 2008; Teer and Mullikin, 2010; Walsh and King, 2007]. The purpose of current genome-wide association studies (GWAS) is to identify common variants that are associated with complex traits. To date, a large number of common variants underlying complex diseases have been identified by GWAS [Heid et al., 2010; Lango Allen et al., 2010; Plenge et al., 2007; Saxena et al., 2007; Thomson et al., 2007; Zeggini et al., 2007]. However, the identified variants account for only a small fraction of disease heritability [Bansal et al., 2010; McCarthy et al., 2008; Schork et al., 2009]. One of potential sources of missing heritability is the contribution of rare variants [Cohen et al., 2006; Ji et al., 2008; Manolio et al., 2009; Marini et al., 2008; Nejentsev et al., 2009; Zhu et al., 2010]. To map common variants, we can use indirect mapping methods based on tagging single-nucleotide polymorphisms (SNPs). However, for rare variant association studies, we need to directly test all rare variants because they are essentially independent of other variants. Next-generation sequencing technology allows sequencing of parts of the genome—or, in the future, the whole genome—of large groups of individuals [Hodges et al., 2007], and thus makes directly testing rare variants feasible [Andre's et al., 2007].

Although statistical methods to detect common variants have been well developed, these methods may not be optimal for detecting rare variants due to allelic heterogeneity as well as the extreme rarity of individual variants [Li and Leal, 2008]. Recently, several statistical methods for detecting associations of rare variants have been developed, including the cohort allelic sums test (CAST) [Morgenthaler and Thilly, 2007], the combined multivariate and collapsing (CMC) method [Li and Leal, 2008], the weighted sum statistic (WSS) [Madsen and Browning, 2009], the variable minor allele frequency (MAF) threshold method [Price et al., 2010], the cumulative minor-allele test (CMAT) [Zawis-TOWski et al., 2010], the adaptive sum test (aSum) [Han and Pan, 2010], and the step-up method [Hoffmann et al., 2010] among others. Let x_{im} denote the genotype (number of minor alleles) of the ith individual at the mth variant. The aforementioned methods for rare variant association studies are essentially testing the effect of a weighted combination of variants, $\sum_m w_m x_{im}$, or its function with different ways to model the weights w_m. All of the CAST, CMC method, variable MAF threshold method, and CMAT set $w_m = 1$. The CMAT tests the effect of $\sum_m w_m x_{im}$, while the CAST, CMC method, and variable MAF threshold method test the effect of $I_{\{\sum_m w_m x_{im} \geq 1\}}$, where $I_{\{\cdot\}}$ is the indicator function. The WSS tests the effect of $\sum_m w_m x_{im}$ with w_m to be the inverse square root of the expected variance based on allele frequencies. The aSum sets $w_m = \text{sign}(\hat{\beta}_m)$, where $\hat{\beta}_m$ is an estimated value of the coefficient of the mth variant based on the marginal logistic linear model. The step-up method models $w_m = a_m s_m v_m$, where a_m is a continuous weight (e.g., to incorporate allele frequencies), s_m determines the direction of the variant effect (deleterious or protective), and v_m is an indicator variable determining whether the variant belongs to the model.

In this paper, we propose a novel Test for testing the effect of an Optimally Weighted combination of variants (TOW). The optimal weights are analytically derived and can be calculated from sampled genotypes and phenotypes. Based on the optimal weights w_m^o, the TOW tests the effect of $\sum_m w_m^o x_{im}$. Furthermore, based on the TOW, we propose a Variable Weight TOW (VW-TOW) to test the effects of both rare and common variants. Both TOW and VW-TOW are applicable to quantitative and qualitative traits, allow covariates, and are robust to directions of effects of causal variants. Extensive simulation studies and applications to the Genetic Analysis Workshop 17 (GAW17) data are used to compare the performance of the two proposed methods with that of three existing methods (CMC, WSS, and SKAT). Results show that VW-TOW demonstrates better performance across a wide range of scenarios. TOW performs better than existing methods when all causal variants are rare.

METHOD

Consider a sample of n individuals. Each individual has been genotyped at M variants in a genomic region (a gene or a pathway). Denote y_i as the trait value of the ith individual for either a quantitative trait or a qualitative trait (1 for cases and 0 for controls for a qualitative trait) and denote $X_i = (x_{i1}, \ldots, x_{iM})^T$ as genotypic score of the ith individual, where $x_{im} \in \{0, 1, 2\}$ is the number of minor alleles the ith individual has at the mth variant. We first describe our methods without considering covariates and then extend our methods to incorporate covariates.

WITHOUT COVARIATES

We use the generalized linear model [Nelder and Wedderburn, 1972]

$$g(E(y_i|X_i)) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_M x_{iM} \tag{1}$$

to model the relationship between trait values and genotypes, where $g(\cdot)$ is a monotone "link" function and $\beta_0, \ldots, \beta_M$ are parameters. Two commonly used models under the generalized linear model framework are the linear model with the identity link for continuous or quantitative traits, and the logistic regression model with the Logit link for a binary trait. Under the generalized linear model, the score test statistic to test the null hypothesis $H_0 : \beta = 0$ is given by [Sha et al., 2011]

$$S = U^T V^{-1} U, \tag{2}$$

where $U = \sum_{i=1}^n (y_i - \bar{y})(X_i - \bar{X})$ and $V = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$. The statistic S asymptotically follows a chi-square distribution with $k = \text{rank}(V)$ degrees of freedom (df). As shown by Sha et al. [2011], the score test includes many commonly used association tests such as the Cochran-Armitage trend test [Armitage, 1955; Cochran, 1954; Zheng et al., 2006], the genotypic chi-square test, the allelic chi-square test [Chapman and Wijsman, 1998], the multimarker genotypic chi-square test, and the haplotypic chi-square test as its special case. For rare variants, however, the score test may lose power due to the sparse data and a large df k.

As discussed in the Introduction, a large portion of recently developed methods for rare variant association studies are essentially testing the effect of a weighted combination of variants, $\sum_{m=1}^M w_m x_{im}$. To test the effect of the weighted combination of variants, $x_i = \sum_{m=1}^M w_m x_{im}$, the score test statistic becomes

$$S(w_1, \ldots, w_M) = n\frac{\left(\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})\right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}$$
$$= n\frac{\left(\sum_{m=1}^M w_m \sum_{i=1}^n (y_i - \bar{y})(x_{im} - \bar{x}_m)\right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Because rare variants are essentially independent, we have

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{m=1}^M \sum_{l=1}^M w_m w_l \sum_{i=1}^n (x_{im} - \bar{x}_m)(x_{il} - \bar{x}_l)$$
$$\approx \sum_{m=1}^M w_m^2 \sum_{i=1}^n (x_{im} - \bar{x}_m)^2.$$

Let

$$a_m = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{im} - \bar{x}_m)}{\sqrt{\sum_{i=1}^n (x_{im} - \bar{x}_m)^2}} \quad \text{and}$$

$$u_m = w_m \sqrt{\sum_{i=1}^n (x_{im} - \bar{x}_m)^2}.$$

Then, the score test statistic is approximately equal to

$$S_0(w_1, \ldots, w_M) = n\frac{\left(\sum_{m=1}^M a_m u_m\right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{m=1}^M u_m^2}. \tag{3}$$

As a function of $(u_1, \ldots, u_M)$, $S_0(w_1, \ldots, w_M)$ reaches its maximum when $u_m = a_m$ or $w_m = \sum_{i=1}^n (y_i - \bar{y})(x_{im} - \bar{x}_m)/\sum_{i=1}^n (x_{im} - \bar{x}_m)^2 (m = 1, \ldots, M)$. Thus, the optimal weights, denoted by w_m^O, are given by $w_m^O = \sum_{i=1}^n (y_i - \bar{y})(x_{im} - \bar{x}_m)/\sum_{i=1}^n (x_{im} - \bar{x}_m)^2$.

Let $x_i^o = \sum_{m=1}^M w_m^o x_{im}$. Then,

$$S_0(w_1^o, \ldots, w_M^o) = n \sum_{i=1}^n (y_i - \bar{y})(x_i^o - \bar{x}^o) \Big/ \sum_{i=1}^n (y_i - \bar{y})^2.$$

We define the statistic to Test the effect of the Optimally Weighted combination (TOW) of variants, $\sum_{m=1}^M w_m^o x_{im}$, as

$$T_T = \sum_{i=1}^n (y_i - \bar{y})(x_i^o - \bar{x}^o).$$

Because we use a permutation test to evaluate P-values, we can consider $\sum_{i=1}^n (y_i - \bar{y})^2$ as a constant and thus, T_T is equivalent to $S_0(w_1^o, \ldots, w_M^o)$.

The optimal weight w_m^o is equivalent to $w_m^{o*} = \rho(y, x_m)/\sqrt{\sum_{i=1}^n (x_{im} - \bar{x}_m)^2}$, where $\rho(y, x_m)$ is the correlation coefficient between $y = (y_1, \ldots, y_n)$ and $x_m = (x_{1m}, \ldots, x_{nm})$. Because w_m^{o*} is proportional to $\rho(y, x_m)$, w_m^o will put big weights to the variants that have strong associations with the trait of interests and w_m^o will also adjust the direction of the association. Since w_m^{o*} is proportional to $1/\sqrt{\sum_{i=1}^n (x_{im} - \bar{x}_m)^2}$, w_m^o will put big weights to rare variants. Like most of the existing methods for rare variant association studies, the proposed TOW also targets rare variants and it will lose power when testing the effects of both rare and common variants because it puts small weights on common variants. For testing the effects of both rare and common variants, we propose the following VW-TOW. We divide variants into rare (MAF < the rare variant threshold [RVT]) and common (MAF > RVT) and apply TOW to rare and common variants separately. Let T_r and T_c denote the test statistics of TOW for rare and common variants, respectively. Let $T_\lambda = \lambda \frac{T_r}{\sqrt{\mathrm{var}(T_r)}} + (1 - \lambda) \frac{T_c}{\sqrt{\mathrm{var}(T_c)}}$ and p_λ denote the P-value of T_λ. The test statistic of VW-TOW is defined as

$$T_{VW-T} = \min_{0 \le \lambda \le 1} p_\lambda.$$

In this study, we use a simple method to evaluate the minimization. Divide the interval [0, 1] into K subintervals of equal-length. Let $\lambda_k = k/K$ for $k = 0, 1, \ldots, K$. Then, $\min_{0 \le \lambda \le 1} p_\lambda = \min_{0 \le k \le K} p_{\lambda_k}$.

We use permutation tests to evaluate P-values of both T_T and T_{VW-T}. The standard permutation test can be used to evaluate the P-value of T_T. In the following presentation, we describe the permutation procedure to evaluate the P-value of T_{VW-T}. In each permutation, we randomly shuffle the trait values. Suppose that we perform B times of permutations. Let $T_r^{(b)}$ and $T_c^{(b)}$ denote the values of T_r and T_c, respectively, based on the bth permuted data, where $b = 0$ represents the original data. Based on $T_r^{(b)}$ and $T_c^{(b)}$ $(b = 0, 1, \ldots, B)$, we can calculate $T_{\lambda_k}^{(b)}$ for $b = 0, 1, \ldots, B$ and $k = 0, 1, \ldots, K$, where $\mathrm{var}(T_r)$ and $\mathrm{var}(T_c)$ are estimated using $T_r^{(b)}$ and $T_c^{(b)} (b = 1, \ldots, B)$. Then, we transfer $T_{\lambda_k}^{(b)}$ to $p_{\lambda_k}^{(b)}$ by

$$p_{\lambda_k}^{(b)} = \frac{\#\{T_{\lambda_k}^{(d)} : T_{\lambda_k}^{(d)} > T_{\lambda_k}^{(b)} \text{ for } d = 0, 1, \ldots, B\}}{B}.$$

Let $p^{(b)} = \min_{0 \le k \le K} p_{\lambda_k}^{(b)}$. Then, the P-value of T_{VW-T} is given by

$$\frac{\#\{p^{(b)} : p^{(b)} < p^{(0)} \text{ for } b = 1, 2, \ldots, B\}}{B}.$$

WITH COVARIATES

Suppose that we have p covariates. Let $(z_{i1}, \ldots, z_{ip})^T$ denote covariates of the ith individual. We adjust both trait value y_i and genotypic score x_{im} for the covariates by applying linear regressions. That is,

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 z_{i1} + \cdots + \alpha_p z_{ip} + \varepsilon_i \quad \text{and} \\ x_{im} &= \alpha_{0m} + \alpha_{1m} z_{i1} + \cdots + \alpha_{pm} z_{ip} + \tau_{im}. \end{aligned} \tag{4}$$

Let $\tilde{y}_i$ and $\tilde{x}_{im}$ denote the residuals of y_i and x_{im}, respectively. With covariates, the statistics of TOW and VW-TOW are defined as

$$\begin{aligned} T_{TOW} &= T_T|_{y_i = \tilde{y}_i, x_{im} = \tilde{x}_{im}} \quad \text{and} \\ T_{VW-TOW} &= T_{VW-T}|_{y_i = \tilde{y}_i, x_{im} = \tilde{x}_{im}}, \end{aligned}$$

respectively. Adjusting trait values and genotypic scores for the covariates by applying linear regressions given by (4) is equivalent to modeling the relationship between trait values, covariates, and genotypes by the linear model

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 z_{i1} + \cdots + \alpha_p z_{ip} + \beta_1 x_{i1} + \cdots + \beta_M x_{iM} + \varepsilon_i \\ &= \alpha^T Z_i + \beta^T X_i + \varepsilon_i, \end{aligned} \tag{5}$$

where $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_p)^T$, $\beta = (\beta_1, \ldots, \beta_M)^T$, and $Z_i = (1, z_{i1}, \ldots, z_{ip})^T$. In the Appendix, we show that, under linear model (5), the score test statistic to test the null hypothesis $H_0 : \beta = 0$ is given by

$$SC = \tilde{U}^T \tilde{V}^{-1} \tilde{U}, \tag{6}$$

where $\tilde{U} = \sum_{i=1}^n \tilde{y}_i \tilde{X}_i$, $\tilde{V} = \frac{1}{n} \sum_{i=1}^n \tilde{y}_i^2 \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^T$, and $\tilde{X}_i = (\tilde{x}_{i1}, \ldots, \tilde{x}_{iM})^T$. Thus, the score test statistic to test the effect of the weighted combination of variants, $x_i = \sum_{m=1}^M w_m x_{im}$, is given by $SC(w_1, \ldots, w_M) = n \left(\sum_{i=1}^n \tilde{y}_i \tilde{x}_i\right)^2 / \left(\sum_{i=1}^n \tilde{y}_i^2 \sum_{i=1}^n \tilde{x}_i^2\right)$, where $\tilde{x}_i = \sum_{m=1}^M w_m \tilde{x}_{im}$. Using the same argument as that used in the section Without Covariates, we have that $SC(w_1, \ldots, w_M)$ reaches its maximum when $w_m = \sum_{i=1}^n \tilde{y}_i \tilde{x}_{im} / \sum_{i=1}^n \tilde{x}_{im}^2$ and the maximum of $SC(w_1, \ldots, w_M)$ is equivalent to T_{TOW}.

The R code of the TOW and VW-TOW methods is available at Shuanglin Zhang's homepage http://www.math.mtu.edu/~shuzhang/software.html.

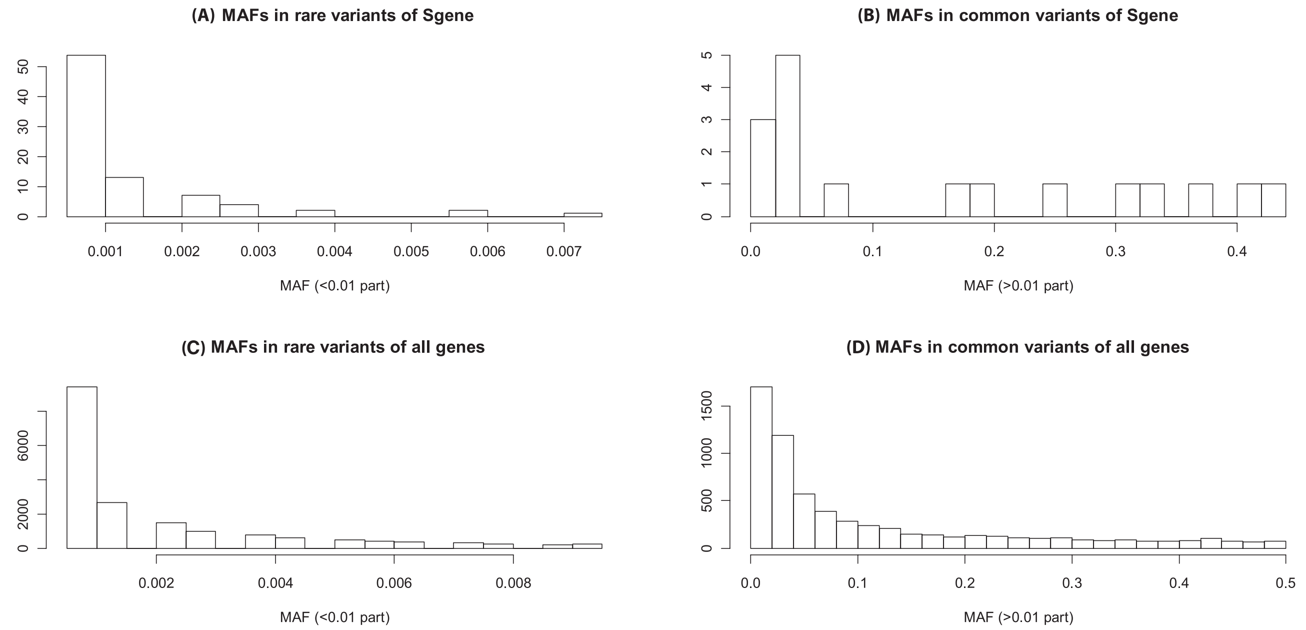

Fig. 1. The distributions of MAFs in the 100 variants in the Sgene and in the 24,487 variants in all the 3,205 genes. (A) The histogram of MAFs in rare variants (MAF $<$ 0.01) in the Sgene and (B) the histogram of MAFs in common variants (MAF $>$ 0.01) in the Sgene. (C) The histogram of MAFs in rare variants (MAF $<$ 0.01) in all the 3,205 genes and (D) the histogram of MAFs in common variants (MAF $>$ 0.01) in all the 3,205 genes.

COMPARISON OF TESTS

We compare the performance of the two proposed tests with that of the WSS [Madsen and Browning, 2009], the CMC method [Li and Leal, 2008], and the sequence kernel association test (SKAT) [Wu et al., 2011]. The rank sum test used by WSS and the T^2 test used by CMC are replaced with the score test based on residuals $\tilde{y}_i$ and $\tilde{x}_{im}$.

SIMULATION

The empirical Mini-Exome genotype data provided by the GAW17 is used for simulation studies. This dataset contains genotypes of 697 unrelated individuals on 3,205 genes. We choose four genes: ELAVL4 (gene1), MSH4 (gene2), PDE4B (gene3), and ADAMTS4 (gene4) with 10, 20, 30, and 40 variants, respectively. We merge the four genes to form a super gene (Sgene) with 100 variants. The distributions of MAFs in the 100 variants in the Sgene and in the 24,487 variants in all the 3,205 genes are given in Figure 1. From this figure, we can see that the distribution of MAFs in the Sgene can represent the distribution of MAFs in all the 3,205 genes. In our simulation studies, we generate genotypes based on the genotypes of 697 individuals in the Sgene. The genotypes are extracted from the sequence alignment files provided by the 1,000 Genomes Project for their pilot3 study (http://www.1000genomes.org). We use the program fastPHASE [Scheet and Stephens, 2006] to infer haplotypic phase for the 697 individuals and calculate haplotype frequencies. To generate the genotype of an individual, we generate two haplotypes according to the haplotype frequencies. To generate a qualitative disease affection status, we use a liability threshold model based on a continuous phenotype (quantitative trait). An individual is defined to be affected if the individual's phenotype is at least one standard deviation larger than the phenotypic mean. This yields a prevalence of 16% for the simulated disease in the general population. In the following, we describe how to generate a quantitative trait.

To evaluate type I error, we generate trait values independent of genotypes by using the model:

$$y = 0.5X_1 + 0.5X_2 + \varepsilon, \tag{7}$$

where X_1 is a continuous covariate generated from a standard normal distribution, X_2 is a binary covariate taking values 0 and 1 with a probability of 0.5, and ε follows a standard normal distribution.

To evaluate power, we consider two cases: (1) rare causal variants in which causal variants are all rare (MAF < RVT) and (2) both causal variants in which causal variants contain both rare and common variants. In the case of rare causal variants, we randomly choose n_c rare variants as causal variants, where n_c is determined by the percentage of causal variants among rare variants. In the case of both causal variants, we randomly select n_c rare and one common variant (MAF > RVT) as causal variants. For power comparison, we consider three different values of RVT (0.005, 0.01, and 0.03). Denote n_r and n_p as the number of risk rare variants and protective rare variants, respectively, where $n_r + n_p = n_c$. For an individual, let x_i^r, x_j^p, and x_c denote the genotypic scores of the ith risk rare variant, the jth protective rare variant, and the common causal variant, respectively. We assume that all the n_c rare causal variants have the same heritability such that rarer variants have larger effects. Under this assumption, disease model is given by

$$y = 0.5X_1 + 0.5X_2 + \sum_{i=1}^{n_r} \beta_i^r x_i^r - \sum_{j=1}^{n_p} \beta_j^p x_j^p + \beta_c x_c + \varepsilon,$$

TABLE I. The estimated type I error rates of the two proposed tests

			$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.001$	
Trait	Gene	Sample size	TOW	VW-TOW	TOW	VW-TOW	TOW	VW-TOW
Quan	4	1,000	0.0496	0.0507	0.0110	0.0110	0.0008	0.0014
		2,000	0.0486	0.0483	0.0101	0.0098	0.0008	0.0012
		3,000	0.0489	0.0475	0.0106	0.0105	0.0015	0.0010
	Sgene	1,000	0.0513	0.0509	0.0104	0.0092	0.0011	0.0013
		2,000	0.0472	0.0483	0.0112	0.0097	0.0008	0.0012
		3,000	0.0485	0.0497	0.0097	0.0080	0.0016	0.0009
Qual	4	1,000	0.0490	0.0497	0.0108	0.0110	0.0011	0.0013
		2,000	0.0482	0.0492	0.0106	0.0108	0.0013	0.0011
		3,000	0.0496	0.0506	0.0109	0.0103	0.0015	0.0016
	Sgene	1,000	0.0502	0.0496	0.0101	0.0081	0.0011	0.0011
		2,000	0.0492	0.0476	0.0114	0.0103	0.0013	0.0011
		3,000	0.0465	0.0462	0.0102	0.0094	0.0014	0.0008

Note: Quan represents quantitative traits; Qual represents qualitative traits.

where X_1, X_2, and ε are the same as those in Equation (7); β_i^r, β_j^p, and β_c are constants and their values depend on the total heritability and the ratio of the heritability of rare causal variants to the heritability of the common causal variant. In the case of rare causal variants, $\beta_c = 0$.

SIMULATION RESULTS

For type I error evaluation, we consider different kinds of traits, different sample sizes, different haplotype structures (different genes), and different significance levels. In each simulation scenario, P-values are estimated by 10,000 permutations and type I error rates are evaluated using 10,000 replicated samples. For 10,000 replicated samples, the 95% confidence intervals (CIs) for type I error rates of nominal levels 0.05, 0.01, and 0.001 are (0.046, 0.054), (0.008, 0.012), and (0.0004, 0.0016), respectively. The estimated type I error rates of the two proposed tests are summarized in Table I. From this table, we can see that all the estimated type I error rates are within the 95% CIs, which indicates that the estimated type I error rates are not significantly different from the nominal levels. Thus, the two proposed tests are all valid tests.

For power comparisons, we consider two different cases: (1) rare causal variants in which all causal variants are rare (MAF < RVT) and (2) both causal variants in which causal variants contain both rare and common (one common variant) and the heritability of the common variant is as big as twice of the heritability of all the rare causal variants. The distributions of MAFs in rare causal variants and in common causal variants for different values of RVT are given in Figure 2. In each of the two cases, we consider different values of heritability, different values of RVT, different kinds of traits, different percentages of protective variants, and different percentages of neutral variants. In each of the simulation scenarios, P-values are estimated using 10,000 permutations and power is evaluated using 200 replicated samples at a significance level of 0.001. In all cases, we use $RVT = 0.01$ in VW-TOW and CMC, although different values of RVT are used to generate data.

Power comparisons of the five tests (VW-TOW, TOW, CMC, SKAT, and WSS) for different values of heritability based on a quantitative trait are given in Figure 3. As shown in Figure 3, in the case of both causal variants, VW-TOW and CMC have similar power and are more powerful than the other three tests. Among the other three tests (TOW, SKAT, and WSS), WSS is the least powerful one. TOW and SKAT have similar power when $RVT \leq 0.01$ and TOW is much more powerful than SKAT when $RVT = 0.03$. WSS loses power because it gives common variants very small weights. CMC has high power because it gives common variants big weights (as big as that for rare variants). Comparing to the case of $RVT \leq 0.01$, SKAT loses power when $RVT = 0.03$. The reason is that when $RVT \leq 0.01$, MAFs of a large portion of common causal variants are within interval (0.01, 0.05) (Figure 2) and SKAT puts decent nonzero weights for variants with MAF in (0.01, 0.05); when $RVT = 0.03$, MAFs of a large portion of common causal variants are ≥ 0.15 (Figure 2) and SKAT puts almost zero weights for variants with MAF ≥ 0.15. In the case of rare causal variants, VW-TOW and TOW have similar power and are more powerful than the other three tests. Among the other three tests (CMC, SKAT, and WSS), WSS is more powerful than SKAT and SKAT is more powerful than CMC. CMC loses power also because it gives common variants big weights and thus, common neutral variants will introduce large noises.

Power comparisons of the five tests for different values of heritability based on a qualitative trait are given in Figure 4. By comparing Figure 3 with Figure 4, we can see that patterns of power comparisons based on a qualitative trait are very similar to that based on a quantitative trait. However, the power improvement of CMC and VW-TOW over the other three tests in the case of both causal variants and the power improvement of TOW and VW-TOW over the other three tests in the case of rare causal variants are smaller based on a qualitative trait than that based on a quantitative trait.

Comparisons of power as a function of percentage of protective variants are given in Figure 5. This figure shows that, for a quantitative trait, TOW and VW-TOW are much more powerful than other three tests. TOW, VW-TOW, and SKAT are robust to the percentage of protective variants, whereas CMC and WSS suffer substantial loss of power when both risk and protective variants are present. Power comparisons based on a qualitative trait have similar patterns to those

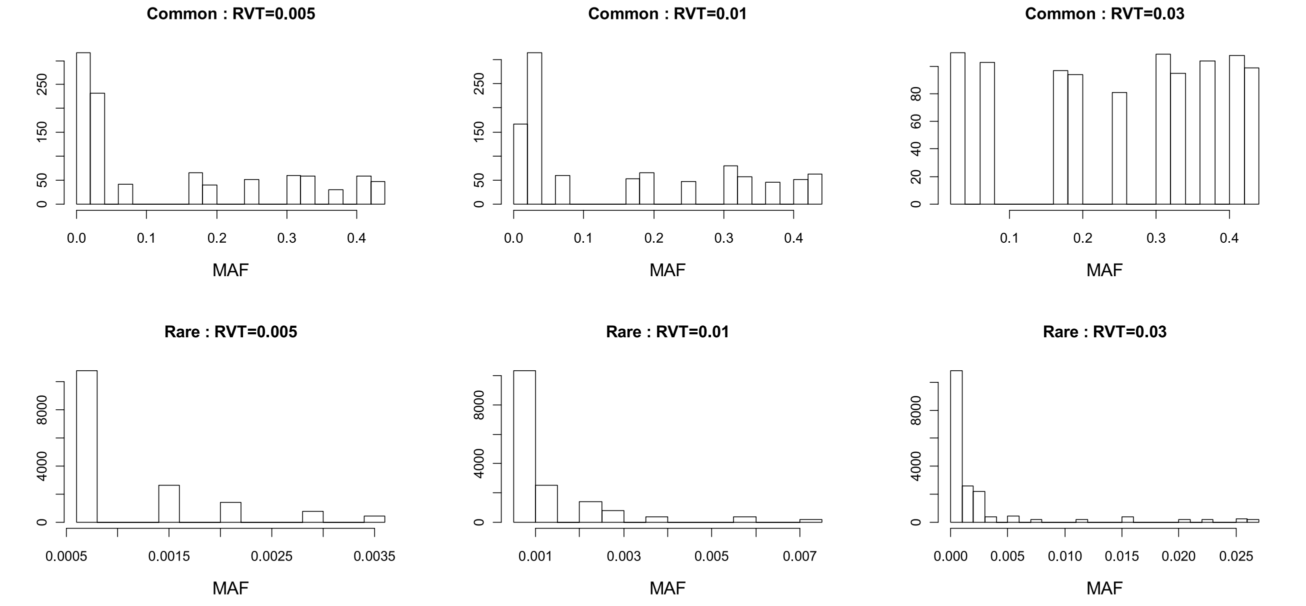

Fig. 2. Distributions of MAFs in causal variants. The figure is based on 1,000 replications. In each replication, the percentage of causal variants among rare variants is 20%. The top channel gives the histograms of MAFs in common (MAF > RVT) causal variants and the bottom channel gives the histograms of MAFs in rare (MAF < RVT) causal variants. RVT represents the rare variant threshold.

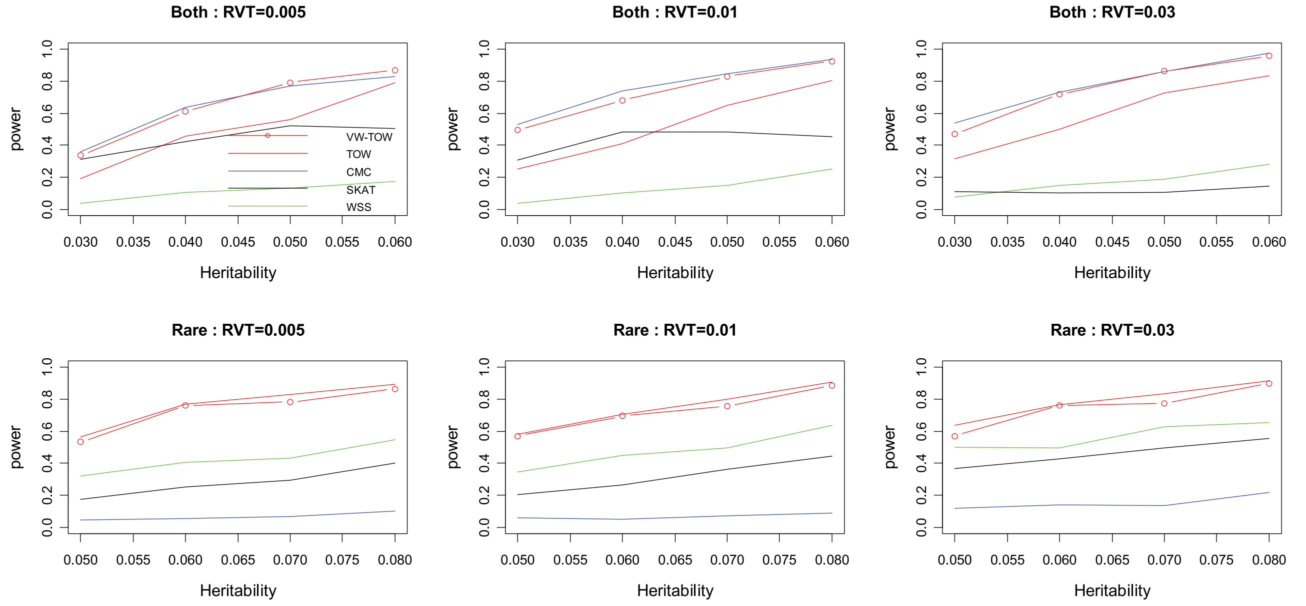

Fig. 3. Power comparisons of five tests for different values of heritability based on a quantitative trait. RVT represents the rare variants threshold. Rare means that all causal variants are rare (MAF < RVT). Both means that causal variants contain both rare and common (one common variant) and the heritability of the common variant is as big as twice the heritability of all the rare causal variants. *x*-axis represents the total heritability of all causal variants. Sample size is 1,000. In this set of simulations, all causal variants are risk variants and 20% of rare variants are causal.

based on a quantitative trait. However, the power of TOW, VW-TOW, and SKAT decreases with the increase of the percentage of protective variants, although decreases not as fast as that of WSS and CMC. As pointed out by Wu et al. [2011], decrease in power of TOW, VW-TOW, and SKAT in the presence of both risk and protective variants is due to the fact that protective variants lower MAFs in cases and thus make observing rare variants in cases more difficult. The larger decrease in power of WSS and CMC is additionally driven by sensitivity to direction of effect due to aggregation of genotypes.

Comparisons of power as a function of percentage of neutral variants are given in Figure 6. As shown by this figure, patterns of power comparisons based on a quantitative trait

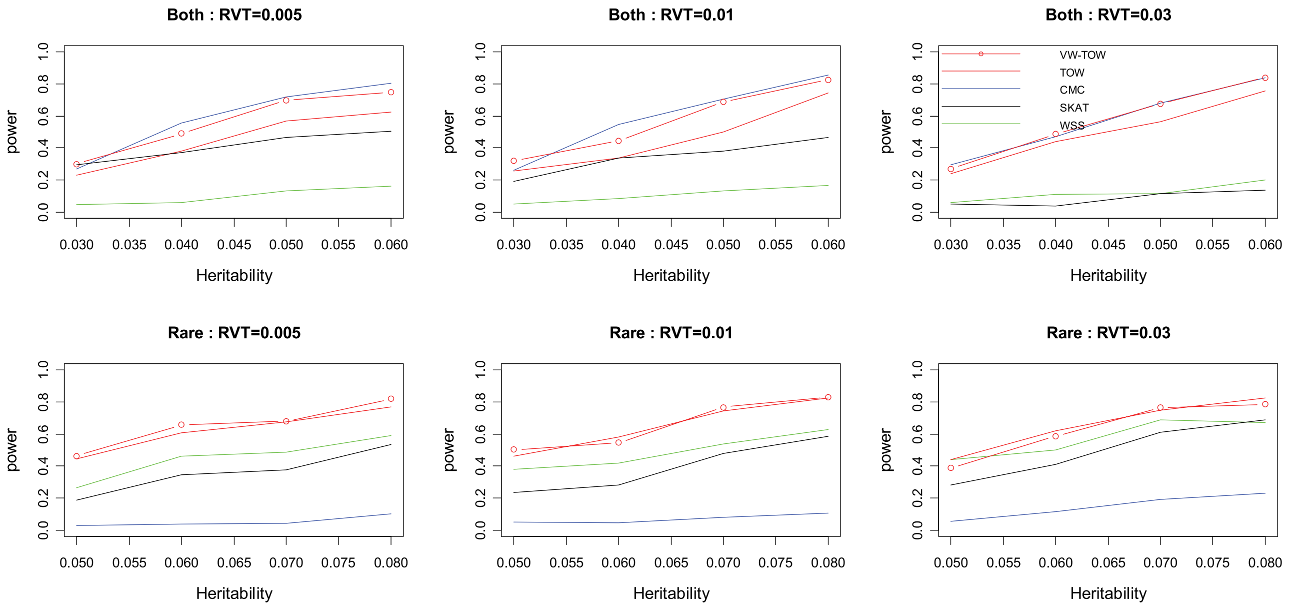

Fig. 4. Power comparisons of five tests for different values of heritability based on a qualitative trait. RVT represents the rare variants threshold. Rare means that all causal variants are rare (MAF < RVT). Both means that causal variants contain both rare and common (one common variant) and the heritability of the common variant is as big as twice the heritability of all the rare causal variants. *x*-axis represents the total heritability of all causal variants. Sample size is 1,000. In this set of simulations, all causal variants are risk variants and 20% of rare variants are causal.

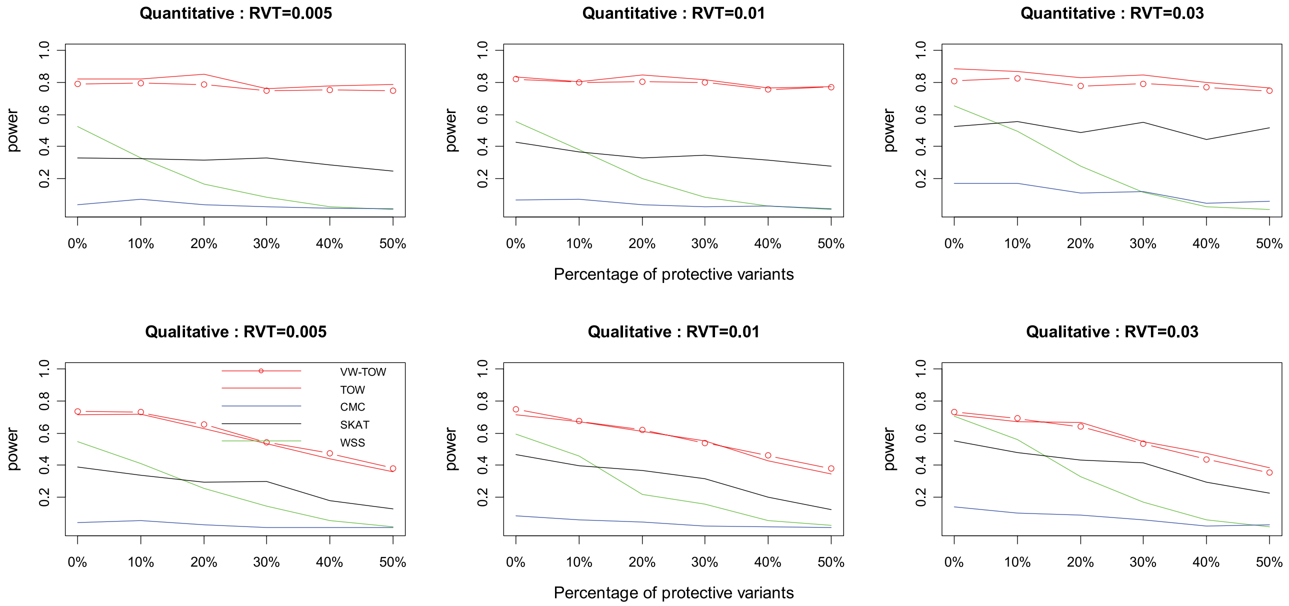

Fig. 5. Power comparisons of five tests for different percentages of protective variants. RVT represents the rare variants threshold. *x*-axis represents the percentage of protective variants. Sample size is 1,000. In this set of simulations, all causal variants are rare variants, 20% of rare variants are causal, and heritability of all the causal variants is 0.07.

are similar to those based on a qualitative trait. The power of TOW, VW-TOW, and SKAT is relatively robust to the increase of neutral variants, while the power of WSS and CMC decreases rapidly with the increase of neutral variants. TOW and VW-TOW have similar power in all the cases. TOW and VW-TOW are more powerful than the other three tests when the percentage of neutral variants is large (>50%~70%), while WSS is the most powerful one when the percentage of neutral variants is small (<50%~70%).

In summary, except the case that the percentage of neutral variants is small and all causal variants are risk variants, VW-TOW is either the most powerful test or has similar power to the most powerful test in the case of either rare or both causal variants and TOW is the most powerful test

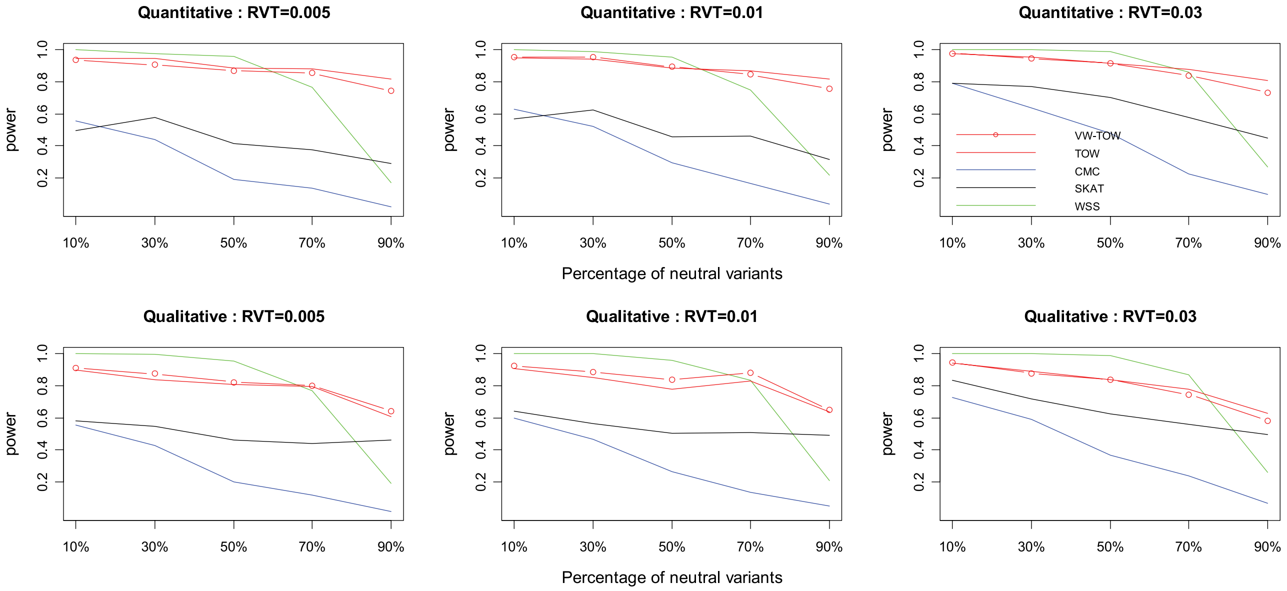

Fig. 6. Power comparisons of five tests for different percentages of neutral variants among all rare variants. RVT represents the rare variants threshold. *x*-axis represents the percentage of neutral variants among all rare variants. Sample size is 1,000. In this set of simulations, all causal variants are rare variants, all causal variants are risk variants, and heritability of all the causal variants is 0.07.

in the case of rare causal variants. The power of TOW and VW-TOW is robust to the increase of protective variants and is also relatively robust to the increase of neutral variants. Power simulation results based on other genes (gene3 and gene4) yield the same conclusions (Figures S1–S8 in the Supporting Information).

ANALYSIS OF THE GAW17 DATASET

The GAW17 dataset consists of a collection of 697 unrelated individuals and their genotypes and phenotypes. SNP genotypes of the 697 individuals are obtained from the sequence alignment files provided by the 1,000 Genomes Project for their pilot3 study. There is a total of 24,487 SNPs in 3,205 genes. The genotypes are held fixed for all 200 simulation replicates. A total of 200 replicates of three quantitative traits Q1, Q2, and Q4 are simulated. Covariates include age, sex, and smoking status. Q4 has a heritability of 0.70, but none of this genetic component is due to genes in this dataset. Thus, we do not consider Q4 for the purpose of power comparisons.

We perform power comparisons using quantitative traits Q1 and Q2. All powers are estimated at a significance level of 0.001 and every two replicates are merged to increase the sample size. Q1 is influenced by nine genes, whereas Q2 is influenced by 13 genes. There are 1–13 causal variants per gene and MAFs in causal variants range from 0.07% to 17.1%. In all cases, the minor allele is associated with higher means of the two quantitative traits, which means that there are no protective variants. For the purpose of power comparison, we omit causal genes that have one variant, causal genes in which all of the five tests have 100% power, and causal genes in which all of the five tests have a power less than 10%.

Powers of the five tests to detect association between each of the five remaining causal genes and Q1 and between each of the seven remaining causal genes and Q2 are given in Table II. As shown in Table II, VW-TOW, TOW, CMC, SKAT, and WSS are the most powerful test in 5, 2, 2, 2, and 1 of 12 genes, respectively. Causal variants in the five genes in which VW-TOW is the most powerful test are in a wide range of MAF (0.07%–1.22%). Causal variants in the genes in which either TOW or WSS is the most powerful test are all rare (MAF < 0.3%). Each of the two genes in which SKAT is the most powerful test contains causal variants with MAF in (0.01, 0.05). Each of the two genes in which CMC is the most powerful test contains common causal variants with MAF > 0.09. Results from analysis of the GAW17 dataset are consistent with those from simulation studies.

DISCUSSION

Most of the recently developed methods for rare variant association studies are essentially testing the effect of a weighted combination of variants. Thus, choosing appropriate weights is critical to the performance of these methods. In this paper, we analytically derived the optimal weights. Based on the optimal weights, we proposed TOW that tests the effect of the optimally weighted combination of variants. We further developed VW-TOW to test the effects of both rare and common variants. We used extensive simulation studies and application to the GAW17 dataset to compare the performance of TOW and VW-TOW with that of the existing methods. Our results show that, in most cases, TOW is the most powerful test for testing rare variants. VW-TOW is the most powerful test or has similar power with the most powerful test in either testing effects of both rare and common variants or testing effects of rare variants only.

For testing rare variants, most of the recently developed methods put large weights on rare variants and small weights on common variants. By putting small weights on

TABLE II. Power of the five tests to test the association between each of the five causal genes and quantitative trait Q1 and between each of the seven causal genes and quantitative trait Q2

Traits	Gene name	No. of variants, no. of causal variants	Min, max, mean MAF	WSS	CMC	TOW	VW-TOW	SKAT
Q1	ARNT	18, 5	0.07, 1.15, 0.33	0.18	0.98	0.68	0.96	**0.99**
	ELAVL4	10, 2	0.07, 0.07, 0.07	0.03	0.28	0.26	**0.31**	0.00
	FLT4	10, 2	0.07, 0.14, 0.11	**0.49**	0.28	0.25	0.28	0.13
	HIF1A	8, 4	0.07, 1.22, 0.39	0.09	0.59	0.22	**0.63**	0.60
	VEGFA	6, 1	0.22, 0.22, 0.22	0.12	0.12	0.22	**0.24**	0.1
Q2	BCHE	29, 13	0.07, 0.29, 0.10	0.19	0.18	**0.35**	0.26	0.11
	LPL	20, 3	0.07, 1.58, 0.60	0.01	0.22	0.17	0.27	**0.39**
	PDGFD	11, 4	0.07, 0.86, 0.29	0.07	0.20	0.22	**0.29**	0.15
	SIRT1	24, 9	0.07, 0.22, 0.12	0.51	0.30	**0.64**	0.63	0.57
	SREBF1	24, 10	0.07, 0.43, 0.22	0.25	0.30	0.19	**0.33**	0.07
	VNN1	7, 2	0.57, 17.1, 8.82	0.06	**0.93**	0.67	0.90	0.02
	VNN3	15, 7	0.07, 9.83, 2.06	0.33	**0.84**	0.57	0.51	0.41

Note: Min, max, mean MAF: the minimum, maximum, and mean MAF (in percentage) at causal variants. In each row, the boldfaced number represents the highest power in the row.

common variants, these methods will lose power when testing the effects of both rare and common variants. To test the effects of both rare and common variants, CMC puts same weights on rare and common variants. By putting large weights on common variants, CMC loses power when testing rare variants only because putting large weights on common neutral variants will introduce large noises. Our proposed VW-TOW, by choosing weights adaptively, has good performance in testing the effects of both rare and common variants and in testing the effects of rare variants only.

In case-control studies, it has been long recognized that population stratification can confound association results. In association studies of common variants, several methods have been developed to control for population stratification by using a set of unlinked genetic markers genotyped in the same samples [Devlin and Roeder, 1999; Price et al., 2006; Pritchard et al., 2000; Zhang et al., 2003]. Theoretically, our proposed methods can be easily modified to be robust to population stratification through principal component (PC) approach [Price et al., 2006; Zhang et al., 2003]. Let $T_i = (t_{i1}, t_{i2}, \ldots, t_{iK})^T$ denote the first K PCs of genotypes at genomic markers of the ith individual. We can put T_i as covariates in our proposed methods to adjust for population effects, which is equivalent to the method of Price et al. [2006] to adjust both the trait y_i and genotypic score x_{im} for the first K PCs, T_i, by applying a linear regression. Although the PC approach performs well in association studies of common variants, its performance in association studies of rare variants may need more investigation.

Our proposed TOW is related to SKAT. Using the notations given in the Method section, when there are no covariates, the test statistics of both TOW and SKAT can be written as $T = \sum_{m=1}^{M} \frac{U_m^2}{V_m}$, where $U_m = \sum_{i=1}^{n}(y_i - \bar{y})(x_{im} - \bar{x}_m)$. In TOW, $V_m = \sum_{i=1}^{n}(x_{im} - \bar{x}_m)^2$, while in SKAT, $\sqrt{\frac{1}{V_m}} = \mathrm{Beta}(MAF_m; a_1, a_2)$, the beta distribution density function with prespecified parameters a_1 and a_2 evaluated at the sample MAF for the mth variant in the data. When there are covariates, TOW and SKAT use different methods to adjust for the effects of covariates. TOW adjusts both trait value y_i and genotypic score x_{im} for the covariates by applying linear regressions given by (4). SKAT adjusts only trait value y_i (not genotypic score x_{im}) for the covariates by applying a linear regression for a quantitative trait and applying a logistic regression for a qualitative trait.

TOW is derived for independent variants. Since common variants within a gene are usually correlated, we may wonder how the performance of TOW for common variants is. We may learn the performance of TOW for common variants from the relationship between TOW and some existing methods for common variants. TOW is related to the weighted sum of squared score (SSUw) proposed by Pan [2009]. In fact, when there are no covariates, TOW and SSUw are the same. Pan [2009] has shown that, to test association between multiple correlated common variants and the trait of interests, SSUw is more powerful than existing standard methods such as the score test given by Equation (2) in most cases. Thus, our proposed TOW, though derived for independent rare variants, has the good performance for correlated common variants.

ACKNOWLEDGMENTS

Research reported in this publication was partially supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number R03HG006155 for Q.S. and S.Z. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Genetic Analysis workshops are supported by NIH grant R01 GM031575 from the National Institute of General Medical Sciences. Preparation of the Genetic Analysis Workshop 17 Simulated Exome Dataset was supported in part by NIH R01 MH059490 and used sequencing data from the 1,000 Genomes Project (http://www.1000genomes.org). The authors have no conflict of interest to declare.

WEB RESOURCES

The R code of the TOW and VW-TOW methods is available at Shuanglin Zhang's homepage http://www.math.mtu.edu/∼shuzhang/software.html.

REFERENCES

Andre's A, Clark A, Shimmin L, Boerwinkle E, Sing C, Hixson J. 2007. Understanding the accuracy of statistical haplotype inference with sequence data of known phase. Genet Epidemiol 31:659–671.

Armitage P. 1955. Tests for linear trends in proportions and frequencies. Biometrics 11:375–386.

Bansal V, Libiger O, Torkamani A, Schork NJ. 2010. Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet 11:773–785.

Bodmer W, Bonilla, C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40(6):695–701.

Chapman NH, Wijsman EM. 1998. Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. Am J Hum Genet 63:1872–1885.

Cochran WG. 1954. Some methods for strengthening the common x2 tests. Biometrics 10:417–451.

Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low density lipoprotein levels. Proc Natl Acad Sci USA 103:1810–1815.

Devlin B, Roeder K. 1999. Genomic control for association studies. Biometrics 55:997–1004.

Han F, Pan W. 2010. A data-adaptive sum test for disease association with multiple common or rare variants. Hum Hered 70:42–54.

Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Magi R, Workalemahu T, White C, Bouatia-Naji N, Harris T, Berndt S, Ingelsson E, Willer C, Weedon M, Luan J, Vedantam S, Esko T, Kilpeläinen T, Kutalik Z, Li S, Monda K. 2010. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 42:949–960.

Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. 2007. Genome-wide in situ exon capture for selective resequencing. Nat Genet 39:1522–1527.

Hoffmann TJ, Marini NJ, Witte JS. 2010. Comprehensive approach to analyzing rare genetic variants. PLoS One 5(11):e13584. doi:10.1371/journal.pone.0013584.

Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. 2008. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 40:592–599.

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segrè AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Mägi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M,Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Zhao JH, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI,Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, König IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N,Haiqing S, Hartikainen AL, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpeläinen TO, Koiranen M, Kolcic I,Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Paré G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietiläinen KH, Pouta A, Ridderstråle M, Rotter JI, Sambrook JG, Sanders AR, Schmidt CO, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL,Nelis M, Peters MJ, Ripatti S, van Meurs JB, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kähænen M, Kaprio J, Kathiresan S,Kiemeney L, Kocher T, Launer LJ, Lehtimäki T, Melander O, Mosley TH Jr, Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tænjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD,Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Grænberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Lathrop GM, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Cupples LA, Erdmann J, Fox CS, Gudnason V, Gyllensten U, Harris TB, Hayes RB, Jarvelin MR, Mooser V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Vælzke H, Watkins H, Wilson JF, Groop LC,Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832–838.

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

Madsen BE, Browning SR. 2009. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 5:e1000384.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF,McCarroll SA, Visscher PM. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753.

Marini NJ, Gin J, Ziegle J, Keho KH, Ginzinger D, Gilbert DA, Rine J. 2008. The prevalence of folate-remedial MTHFR enzyme variants in humans. Proc Natl Acad Sci USA 105:8055–8060.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9:356–369.

Morgenthaler S, Thilly WG. 2007. A strategy to discover genes that carry multiallelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutat Res 615:28–56.

Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. 2009. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324:387–389.

Nelder J, Wedderburn R. 1972. Generalized linear models. J R Stat Soc Ser A 135:370–384.

Ng SB, Turner EH, Robertson PD. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nat Lett 461:272–276.

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

Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Coblyn J, Weinblatt ME, Gabriel SB, Purcell S, Klareskog L, Gregersen PK, Shadick NA, Daly MJ, Altshuler D. 2007. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 39:1477–1482.

Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Lee-Jen Wei LJ, Sunyaev SR. 2010. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 86:832–838.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. PCs analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909.

Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69:124–137.

Pritchard JK, Cox NJ. 2002. The allelic architecture of human disease genes: common disease-common variant . . . or not? Hum Mol Genet 11:2417–2423.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000. Association mapping in structured populations. Am J Hum Genet 67:170–181.

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson BK, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Råstam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjögren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell. 2007. Genomewide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336.

Scheet, P, Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78(4):629–644.

Schork NJ, Murray SS, Frazer KA, Topol EJ. 2009. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19:212–219.

Sha Q, Zhang Z, Zhang S. 2011. An improved score test for genetic association studies. Genet Epidemiol 35:350–359.

Stratton MR, Rahman N. 2008. The emerging landscape of breast cancer susceptibility. Nat Genet 40:17–22.

Teer JK, Mullikin JC. 2010. Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet. 19(R2): R145–R151.

Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, Donn R, Symmons D, Hider S, Bruce IN, Wellcome Trust Case Control Consortium, Wilson AG, Marinou I, Morgan A, Emery P; YEAR Consortium, Carter A, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Strachan D, Worthington J. 2007. Rheumatoid arthritis association at 6q23. Nat Genet 39:1431–1433.

Walsh T, King MC. 2007. Ten genes for inherited breast cancer. Cancer Cell 11:103–105.

Wu M, Lee S, Cai T, Li Y, Boehnke M, Lin X. 2011. Rare variant association testing for sequencing data using the sequence kernel association test (SKAT). Am J Hum Genet 89:82–93.

Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S. 2010. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. Am J Hum Genet 87: 604–617.

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney ASF, McCarthy MI, Hattersley AT. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341.

Zhang S, Zhu X, Zhao H. 2003. On a semi-parametric test to detect associations between quantitative traits and candidate genes using unrelated individuals. Genet Epidemiol 24:44–56.

Zheng G, Freidlin B, Gastwirth JL. 2006. Robust genomic control for association studies. Am J Hum Genet 78:350–356.

Zhu X, Feng T, Li Y, Lu Q, Elston RC. 2010. Detecting rare variants for complex traits using family and unrelated data. Genet Epidemiol 34:171–187.

APPENDIX

Use the notation in the Method section. Let $X = (X_1, \ldots, X_n)^T$, $Z = (Z_1, \ldots, Z_n)^T$, and $Y = (y_1, \ldots, y_n)^T$. Under the linear model

$$y_i = \alpha^T Z_i + \beta^T X_i + \varepsilon_i,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent and $\varepsilon_i \sim N(0, \sigma^2)$, the log-likelihood (up to a constant) is given by

$$\log l = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(Y - Z\alpha - X\beta)^T(Y - Z\alpha - X\beta).$$

Then,

$$\frac{\partial \log l}{\partial \beta} = \frac{1}{\sigma^2}(Y - Z\alpha - X\beta)^T X,$$

$$\frac{\partial \log l}{\partial \alpha} = \frac{1}{\sigma^2}(Y - Z\alpha - X\beta)^T Z$$

$$\frac{\partial^2 \log l}{\partial \beta \beta^T} = -\frac{1}{\sigma^2} X^T X, \; \frac{\partial^2 \log l}{\partial \alpha \alpha^T} = -\frac{1}{\sigma^2} Z^T Z, \quad \text{and}$$

$$\frac{\partial^2 \log l}{\partial \alpha \beta^T} = -\frac{1}{\sigma^2} Z^T X.$$

Let $\hat{\alpha}$ and $\hat{\sigma}^2$ denote the maximum likelihood estimates of α and σ^2 under null hypothesis $H_0 : \beta = 0$. Then, $\hat{\alpha} = (Z^T Z)^{-1} Z^T Y$ and $\hat{\sigma}^2 = \frac{1}{n} Y^T (I - P) Y = \frac{1}{n}\tilde{Y}^T\tilde{Y}$, where $P = Z(Z^T Z)^{-1} Z^T$, $\tilde{Y} = (\tilde{y}_1, \ldots, \tilde{y}_n)^T$, and $\tilde{y}_i$ is the residual of y_i under the linear regression $y_i = \alpha^T Z_i + \varepsilon_i$. Let $\theta = (\alpha^T, \beta^T)^T$. The score and information matrix are

$$S = \frac{\partial \log l}{\partial \theta}_{|\alpha=\hat{\alpha},\beta=0} = \frac{1}{\hat{\sigma}^2}(0, U^T)^T \quad \text{and}$$

$$I = -E\frac{\partial^2 \log l}{\partial \theta \theta^T}_{|\alpha=\hat{\alpha},\beta=0} = \frac{1}{\hat{\sigma}^2}\begin{pmatrix} Z^T Z & Z^T X \\ X^T Z & X^T X \end{pmatrix},$$

where $U = \tilde{Y}^T X$. Note that $(I - P)^2 = I - P$. We have $U = \tilde{Y}^T X = Y^T (I - P) X = \tilde{Y}^T \tilde{X}$ and $X^T (I - P) X = \tilde{X}^T \tilde{X}$, where $\tilde{X} = (\tilde{x}_{im})$ and $\tilde{x}_{im}$ is the residual of x_{im} under the linear regression (4). The score test statistic is given by

$$T_{\text{linear}} = \frac{1}{\hat{\sigma}^2} U^T V^{-1} U,$$

where $U = \tilde{Y}^T \tilde{X} = \sum_{i=1}^n \tilde{y}_i \tilde{X}_i$ and $V = \tilde{X}^T \tilde{X} = \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^T$.