



Published in final edited form as:

*Genet Epidemiol.* 2014 September ; 38(6): 494–501. doi:10.1002/gepi.21834.

## A Rare Variant Association Test Based on Combinations of Single-Variant Tests

Qiuying Sha and Shuanglin Zhang<sup>§</sup>

Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931

### Abstract

Next generation sequencing technologies make directly testing rare variant associations possible. However, the development of powerful statistical methods for rare variant association studies is still underway. Most of existing methods are burden and quadratic tests. Recent studies show that the performance of each of burden and quadratic tests depends strongly upon the underlying assumption and no test demonstrates consistently acceptable power. Thus, combined tests by combining information from the burden and quadratic tests have been proposed recently. However, results from recent studies (including this study) show that there exist tests that can outperform both burden and quadratic tests. In this article, we propose three classes of tests that include tests outperforming both burden and quadratic tests. Then, we propose the optimal combination of single-variant tests (OCST) by combining information from tests of the three classes. We use extensive simulation studies to compare the performance of OCST with that of burden, quadratic and optimal single-variant tests. Our results show that OCST either is the most powerful test or has similar power with the most powerful test. We also compare the performance of OCST with that of the two existing combined tests. Our results show that OCST has better power than the two combined tests.

### Keywords

rare variant; association study; next generation sequencing

### Introduction

Recent studies show that complex diseases are caused by both common and rare variants [Pritchard, 2001; Pritchard and Cox, 2002; Walsh and King, 2007; Stratton and Rahman, 2008; Bodmer and Bonilia, 2008; Ng *et al.*, 2009; Teer and Mullikin, 2010]. To detect disease associated common variants, indirect mapping methods based on tagging SNPs can be used. However, to detect disease associated rare variants, direct association mapping methods in which all variants must be identified should be used because rare variants are essentially independent of other variants. Next-generation sequencing technology allows sequencing of the whole genome of large groups of individuals, and thus makes direct association mapping feasible [Andre's *et al.*, 2007; Metzker, 2010].

<sup>§</sup>Corresponding author Shuzhang Zhang: shuzhang@mtu.edu. Qiuying Sha: qsha@mtu.edu.

The authors have no conflict of interests to declare.

Statistical methods for common variant association studies have been well developed. However, the variant by variant methods for common variant association studies may not be optimal for rare variant association studies due to allelic heterogeneity as well as the extreme rarity of individual variants [Li and Leal, 2008]. Recently, statistical methods for rare variant association studies by summarizing genotype information from multiple variants have been developed. These methods can be roughly divided into three groups: burden tests, quadratic tests, and combined tests.

Burden tests include the cohort allelic sums test (CAST) [Morgenthaler and Thilly, 2007], the combined multivariate and collapsing (CMC) method [Li and Leal, 2008], the weighted sum (WS) method [Madsen and Browning, 2009], the variable minor allele frequency threshold (VT) method [Price *et al.*, 2010], and the cumulative minor-allele test (CMAT) [Zawistowski *et al.*, 2010], among others. Burden tests collapse rare variants in a genomic region into a single burden variable and then regress the phenotype on the burden variable to test for the cumulative effects of rare variants in the region [Lee *et al.*, 2012]. Let  $x_{im}$  denote the genotype (number of minor alleles) of the  $i^{th}$  individual at the  $m^{th}$  variant. As shown by Sha *et al.* [2012], the burden variables of the aforementioned methods are all the weighted combination of variants,  $\sum_m w_m x_{im}$ , or its function with different ways to model the weights  $w_m$ . Let  $s_m$  denote the score test statistic from a linear model or a logistic model for the  $m^{th}$  variant. Linear test statistics with the form  $\sum_m W_m s_m$  are also based on the burden variable  $\sum_m w_m x_{im}$ . Thus, from the way of collapsing genotypes, burden tests and linear tests are equivalent. So, burden tests are also called linear tests [Derkach *et al.*, 2012].

Quadratic tests with test statistics in the form  $\sum_m W_m s_m^2$  include C-alpha test [Neale *et al.*, 2011], sequence kernel association test (SKAT) [Wu *et al.*, 2011], and the test for testing the effects of the optimally weighted combination of variants TOW [Sha *et al.*, 2012]. Recently developed adaptive weighting methods for rare variant association studies [Han and Pan, 2010; Hoffmann *et al.*, 2010; Lin and Tang, 2011; Yi and Zhi, 2011; Sha *et al.*, 2013], as pointed out by Derkach *et al.*, [2012], are operationally similar to quadratic tests. Combined tests include the test using Fisher's method to combine information from the linear and quadratic statistics (Fisher-CT) [Derkach *et al.*, 2012] and the optimal linear combination of the burden test and SKAT (SKAT-O) [Lee *et al.*, 2012].

Burden tests and quadratic tests perform quite differently. Burden tests or linear tests implicitly assume that all the rare variants are causal and directions of effects are all the same. If these assumptions are true, burden tests can outperform quadratic tests; otherwise, burden tests can perform poorly and quadratic tests can outperform burden tests [Wu *et al.*, 2011; Lee *et al.*, 2012; Sha *et al.*, 2012; Derkach *et al.*, 2012]. Ladouceur *et al.* [2012] showed that the performance of each of burden and quadratic tests depends strongly upon the underlying assumption and no test demonstrates consistently acceptable power despite the large sample size. To increase the robustness of the test, both SKAT-O and Fisher-CT combine a burden and a quadratic test aiming to have advantages of both burden and quadratic tests. However, burden and quadratic tests cannot cover all situations. Kinnamon *et al.* [2012] demonstrated that the single-variant test with statistic  $\max_m s_m^2$  can outperform both burden and quadratic tests when there are a large number of neutral variants and small

number of causal variants. Results of this study show that the tests with statistics

$\sum_m |s_m|^p$  ( $p \geq 4$ ),  $\sum_m \left( s_m I_{\{s_m \geq 0\}} \right)^2$ , or  $\sum_m \left( -s_m I_{\{s_m \leq 0\}} \right)^2$  can outperform both burden and quadratic tests in some situations.

In this article, through the optimal combination of single-variant tests under different criteria, we first obtain three classes of tests that are well beyond burden and quadratic tests. Then, we propose the optimal combination of single-variant tests (OCST) by combining information from tests of the three classes. Using extensive simulation studies, we compare the performance of OCST with that of the burden, quadratic, and the optimal single-variant tests. Our results show that, in a wide range of scenarios, OCST either is the most powerful test or has similar power with the most powerful test. We also compare power of OCST with that of the two existing combined tests: Fisher-CT and SKAT-O. We are able to demonstrate that OCST has better power than both Fisher-CT and SKAT-O.

## 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  $i^{\text{th}}$  individual for either a quantitative trait or a qualitative trait (1 for cases and 0 for controls for a qualitative trait) and denote  $x_{im}$  as the genotypic score of the  $i^{\text{th}}$  individual at the  $m^{\text{th}}$  variant, where  $x_{im} \in \{0, 1, 2\}$  is the number of minor alleles. If there are no covariates, we use the generalized linear model [Nelder and Wedderburn, 1972]

$$g(E(y_i | x_{im})) = \beta_0 + \beta_1 x_{im}$$

to model the relationship between trait values and genotypes at the  $m^{\text{th}}$  variant, where  $g(\cdot)$  is a monotone “link” function. Under the generalized linear model, the score test statistic to test the null hypothesis  $H_0: \beta_1 = 0$  is given by [Sha *et al.*, 2011]

$$s_m = \frac{U_m}{\sqrt{V_m}}, \quad (1)$$

where  $U_m = \sum_{i=1}^n (y_i - \bar{y}) (x_{im} - \bar{x}_m)$  and  $V_m = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_{im} - \bar{x}_m)^2$ . The statistic  $s_m$  asymptotically follows the standard normal distribution. If there are covariates, we use the method proposed by Sha *et al.* [2012] to adjust the effect of the covariates. Let  $(z_{i1}, \dots, z_{ip})^T$  denote covariates of the  $i^{\text{th}}$  individual. We adjust both trait value  $y_i$  and genotypic score  $x_{im}$  for the covariates by applying linear regressions. That is,

$$y_i = \alpha_0 + \alpha_1 z_{i1} + \dots + \alpha_p z_{ip} + \varepsilon_i \quad \text{and} \quad x_{im} = \alpha_{0m} + \alpha_{1m} z_{i1} + \dots + \alpha_{pm} z_{ip} + \tau_{im}. \quad (2)$$

Let  $\tilde{y}_i$  and  $\tilde{x}_{im}$  denote the residuals of  $y_i$  and  $x_{im}$ , respectively. With covariates, we replace  $y_i$  and  $x_{im}$  by  $\tilde{y}_i$  and  $\tilde{x}_{im}$  in  $s_m$ .

Let  $S_m = s_m^2$ . Current quadratic tests for rare variant association studies are combinations of  $S_m$ . The statistic of TOW [Sha *et al.*, 2012]  $T_{TOW} = \sum_{m=1}^M S_m$  and the statistic of SKAT [Wu

et al., 2011]  $T_{SKAT} = \sum_{m=1}^M w_m S_m$ , where  $w_m = V_m W_m$  and  $W_m$  is the weight used by SKAT. Since  $\sum_{m=1}^M s_m = \sum_{i=1}^n \left( y_i - \bar{y} \right) x_i$ , where  $x_i = \sum_{m=1}^M w_m x_{im}$  and  $w_m = 1/\sqrt{V_m}$  is asymptotically equivalent to the weights used by Weighted Sum (WS) method [Madsen and Browning, 2009],  $\sum_{m=1}^M s_m$  is a burden test and is similar to WS method. These observations motivate us to consider combinations of  $S_m$  and combinations of  $s_m$ .

First, we consider the optimal combinations of  $S_1, \dots, S_M$  under different criteria, that is,

$$T_a^{a(p)} = \max_{u_1, \dots, u_M} \sum_{m=1}^M u_m S_m \quad \text{under the condition } \sum_{m=1}^M u_m^p = 1 \text{ for } p \in (1, \infty). \text{ By solving the}$$

maximization problem, we have  $T_a^{a(p)} = \left( \sum_{m=1}^M S_m^{p/(p-1)} \right)^{(p-1)/p}$ . The class of tests

$\left\{ T_a^{a(p): p \in (1, \infty)} \right\}$  is equivalent to  $\{ T_a(z): z \in (1, \infty) \}$ , where  $T_a(z) = \left( \sum_{m=1}^M S_m^z \right)^{1/z}$ . We further extend  $\{ T_a(z): z \in (1, \infty) \}$  to  $A_a = \{ T_a(z): z \in (1, \infty) \}$ , where we define

$T_a(\infty) = \lim_{z \rightarrow \infty} T_a(z) = \max_{1 \leq m \leq M} S_m$  and  $T_a(1) = \sum_{m=1}^M S_m$ . Each test in  $A_a$  can be more powerful than other tests in  $A_a$  in some scenarios. No test can be consistently more powerful than other tests in  $A_a$  (see Figures S1 and S2). Note that TOW ( $T_a(2)$ ) belongs to  $A_a$ . In most cases, there is another test in  $A_a$  that is more powerful than TOW (Figures S1 and S2).

All tests in  $A_a$  are robust to the directions of the effects of causal variants. From the literature [Sha et al., 2012; Wu et al., 2011], we learn that tests being robust to directions of the effects of causal variants are less powerful than burden tests when directions of the effects of causal variants are all the same and there are not many neutral variants. This observation leads us to consider the optimal combination of  $s_1, \dots, s_M$  besides the class of tests  $A_a$ . To consider the optimal combination of  $s_1, \dots, s_M$ , we propose to use either

$$T_b^{a(p)} = \max_{w_1, \dots, w_M} \sum_{m=1}^M w_m s_m \quad \text{under the condition } \sum_{m=1}^M w_m^p = 1 \text{ and } w_m \geq 0 \text{ for } m = 1, \dots, M \text{ or}$$

$$T_c^{a(p)} = \max_{w_1, \dots, w_M} \sum_{m=1}^M w_m s_m \quad \text{under the condition } \sum_{m=1}^M (-w_m)^p = 1 \text{ and } w_m \geq 0 \text{ for } m = 1, \dots, M.$$

Using the same argument for  $A_a$ , we have that  $T_b^{a(p)}$  lead to the class of tests  $A_b = \{ T_b(z): z \in$

$[1, \infty) \}$ , where  $T_b(z) = \left( \sum_{m=1}^M \left( s_m I_{\{s_m \geq 0\}} \right)^z \right)^{\frac{1}{z}}$  and  $T_c^{a(p)}$  lead to the class of tests

$A_c = \{ T_c(z): z \in [1, \infty) \}$ , where  $T_c(z) = \left( \sum_{m=1}^M \left( -s_m I_{\{s_m \leq 0\}} \right)^z \right)^{\frac{1}{z}}$ .

Each of the three test classes  $A_a, A_b,$  and  $A_c$  has its own favorite scenario. The favorite scenario of  $A_a$  is that both risk and protective variants are present. The favorite scenario of  $A_b$  is that all causal variants are risk variants while the favorite scenario of  $A_c$  is that all causal variants are protective variants (see Figures S3 and S4). Let  $P_a(z), P_b(z),$  and  $P_c(z)$  denote the p-values of  $T_a(z), T_b(z)$  and  $T_c(z)$ , respectively. Our proposed Optimal Combination of Single-variant Tests (OCST) is defined as

$$T_{OCST} = \min_{z \in [1, \infty]} (P_a(z), P_b(z), P_c(z)).$$

$T_{OCST}$  can be obtained by a simple grid search across a range of  $z$ . For a given grid  $1 < z_1 < \dots < z_k < \infty$ , the test statistic  $T_{OCST} = \min_{1 \leq k \leq K} (P_a(z_k), P_b(z_k), P_c(z_k))$ .

We use a permutation test to evaluate the p-value of OCST. In each permutation, we randomly shuffle the trait values. Suppose that we perform  $B$  times of permutations. Let  $s_m^{(b)}$  denote the values of  $s_m$  based on the  $b^{th}$  permuted data, where  $b=0$  represents the original data. Based on  $s_m^{(b)}$  ( $b=0, 1, \dots, B$ ), we can calculate  $T_s^{(b)}(z_k)$  for  $s = a, b$ , or  $c$ . Then, we transfer  $T_s^{(b)}(z_k)$  to  $P_s^{(b)}(z_k)$  by

$$P_s^{(b)}(z_k) = \frac{1}{B+1} \sum_{d=0}^B I(T_s^{(d)}(z_k) \geq T_s^{(b)}(z_k)), \quad \text{for } s=a, b, \text{ and } c.$$

Let  $P^{(b)} = \min_{1 \leq k \leq K} (P_a^{(b)}(z_k), P_b^{(b)}(z_k), P_c^{(b)}(z_k))$ . Then, the p-value of OCST is given by

$$\frac{1}{B} \sum_{b=1}^B I(P^{(b)} < P^{(0)}).$$

For a simulation study with  $R$  replicates, the above procedure will be rather computationally expensive. In our simulation studies, we use the pooling permutation method proposed by Guo and Lin [2009] to evaluate p-values. In the pooling permutation method, permuted samples from all the replicates are pooled together to form a joint sample from the null distribution. Suppose that we have  $R$  replicates and we perform  $B$  permutations for each replicate. Let  $T_s^{(b,r)}(z_k)$  denote the value of  $T_s(z_k)$  based on the  $b^{th}$  permuted data in the  $r^{th}$  replicate for  $s=a, b$ , or  $c$ , where  $b=0$  represents original data. Then, we transfer  $T_s^{(b,r)}(z_k)$  to the corresponding p-value  $P_s^{(b,r)}(z_k)$  by

$$P_s^{(b,r)}(z_k) = \frac{1}{(B+1)R} \sum_{l=1}^R \sum_{d=0}^B I(T_s^{(d,l)}(z_k) \geq T_s^{(b,r)}(z_k)), \quad \text{for } s=a, b, \text{ and } c.$$

Let  $P^{(b,r)} = \min_{1 \leq k \leq K} (P_a^{(b,r)}(z_k), P_b^{(b,r)}(z_k), P_c^{(b,r)}(z_k))$ . Then, the p-value of OCST in the  $r^{th}$  replicate is given by

$$\frac{1}{BR} \sum_{l=1}^R \sum_{b=1}^B I(P^{(b,l)} < P^{(0,r)}).$$

Since the permutation samples are pooled across all replicates to form a sample from the null,  $B$  can be set to be much smaller than the situation when only one sample is analyzed.

## Comparison of Tests

We compare the performance of the proposed test with that of (1) the weighted sum (WS) method [Madsen and Browning, 2009], (2) the sequence kernel association test (SKAT)

[Wu *et al.*, 2011], (3)  $T_a(\infty) = \max_{1 \leq m \leq M} S_m$  that is called maximum single-variant test

(MAXST), and (4)  $T_a(2) = \sum_{m=1}^M s_m^2$  that is the same as TOW [Sha *et al.*, 2012]. The rank sum test used by WS is replaced with the score test based on residuals  $\tilde{y}_i$  and  $x_{im}$ . We also compare the performance of the proposed method with two combined tests: Fisher-CT and SKAT-O [Derkach *et al.*, 2012; Lee *et al.*, 2012].

## Simulation

The empirical Mini-Exome genotype data provided by the 17<sup>th</sup> genetic analysis workshops (GAW17) are used for simulation studies. This dataset contains genotypes of 697 unrelated individuals on 3205 genes. We choose six genes: AHNAK (gene1), AKAP13 (gene2), COL6A3 (gene3), FREM2 (gene4), MDN1 (gene5), and TG (gene6) with 231, 163, 187, 143, 187, and 146 variants, respectively. We merge the six genes to form a super gene (Sgene) with 1057 variants. We use Sgene because the distributions of the minor allele frequencies (MAFs) in the 1057 variants in the Sgene and in the 24487 variants in all the 3205 genes are very similar (Figure S5). In our simulation studies, we generate genotypes based on the genotypes of 697 individuals in the Sgene. The genotypes of the GAW17 data set are extracted from the sequence alignment files provided by the 1000 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:

$$y = 0.5X_1 + 0.5X_2 + \varepsilon, \quad (3)$$

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  $\varepsilon$  follows a standard normal distribution.

To evaluate power, we assume that there are  $M$  variants in total and there are  $n_{cau}$  causal variants, where  $M$  is determined by  $n_{cau}$  and the percentage of neutral variants. When  $M$  and

$n_{cau}$  are given, we randomly choose  $M$  variants from 1057 variants of Sgene as total variants and randomly choose  $n_{cau}$  rare variants (MAF<0.01) from  $M$  variants as causal variants. Denote  $n_r$  and  $n_p$  as the number of risk variants and protective variants, respectively, where  $n_r + n_p = n_{cau}$ . For an individual, let  $x_i^r$  and  $x_j^p$  denote the genotypic scores of the  $i^{th}$  risk variant and the  $j^{th}$  protective variant, respectively. The disease model is given by

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

where  $X_1$ ,  $X_2$ , and  $\varepsilon$  are the same as those in equation (3);  $\beta_i^r$  and  $\beta_j^p$  are constants and their values depend on the heritability of each causal variant. We have two models to determine the heritability of each causal variant. Let  $h_i$  denote the heritability of the  $i^{th}$  causal variant and let  $h_T = \sum_{i=1}^{n_{cau}} h_i$  denote the total heritability. In Model 1, let  $r_1, \dots, r_{n_{casu}}$  be random numbers between 0 and 1, then,  $h_i = h_T r_i / \sum_{j=1}^{n_{casu}} r_j$  for  $i=1, \dots, n_{casu}$ . In model 2,  $h_1 = 0.5h_T$ . Let  $r_2, \dots, r_{n_{casu}}$  be random numbers between 0 and 1, then,

$h_i = h_T r_i / \left( 2 \sum_{j=2}^{n_{casu}} r_j \right)$  for  $i=2, \dots, n_{casu}$ . Under Model 1, all causal variants have the same expected heritability. Under Model 2, the heritability of one of the causal variants is much larger than that of other causal variants.

## Results

In simulation studies, p-values are estimated using a pooling permutation method [Guo and Lin, 2009] in which permuted samples from all the replicates are pooled together to form a joint sample from the null distribution. In each replicate, we perform 20 permutations. Type I error rates are evaluated using 10,000 replicated samples, while powers are evaluated using 1,000 replicated samples.

For type I error evaluation, we consider different kinds of traits, different haplotype structures (different genes), and different significance levels. 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 five tests are summarized in Table 1. As shown in this table, more than 95% 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, all the five tests are valid tests.

For power comparisons, we conduct two sets of simulations. In simulation set 1, we compare the power of OCST with that of burden (WS), quadratic (SKAT and TOW), and optimal single-variant (MAXST) tests. In simulation set 2, we compare the power of OCST with that of two combined tests (SKAT-O and Fisher-CT). For simulation set 1, we compare the power of the five tests for power as a function of the percentage of neutral variants (Figures 1, 2, S6, S7) and as a function of the percentage of protective variants (Figures 3, S8). The power of TOW and the power of SKAT have similar patterns in all the simulation

scenarios, but TOW is consistently more powerful than SKAT. In the following discussion of power comparisons, we omit SKAT.

As shown by the power comparisons for power as a function of the percentage of neutral variants (Figures 1, 2), in all the cases, OCST either is the most powerful test or has similar power with the most powerful test. WS is the most powerful test and OCST has similar power with WS when there are no protective variants and the percentage of neutral variants is small; TOW is the most powerful test and OCST has similar power with TOW under model 1 when both protective and risk variants are present; MAXST is the most powerful test and OCST has similar power with MAXST under model 2 when both protective and risk variants are present and the percentage of neutral variants is large; OCST is the most powerful test otherwise. With the increase of neutral variants, power of all the tests decreases while the power of WS decreases the fastest and power of MAXST decreases the slowest. With the decrease of number of causal variants, power of all the tests increases while the power of WS increase the slowest and power of MAXST increases the fastest. The reason that the power of MAXST decreases the slowest with the increase of neutral variants and increases the fastest with the decrease of causal variants is that MAXST essentially only depends on the variant with the largest heritability. This reason can also explain why the power of MAXST is higher under model 2 than that under model 1.

The power comparisons for power as a function of the percentage of protective variants are given in figure 3. As shown by figure 3, again, OCST either is the most powerful test or has similar power with the most powerful test. TOW is the most powerful test and OCST has similar power with TOW when the percentage of neutral variants is small; MAXST is the most powerful test and OCST has similar power with MAXST under model 2 when the percentage of neutral variants is large; OCST is the most powerful test otherwise. When both protective and risk variants are present, the power of WS decreases dramatically, while the power of OCST decreases slightly and the power of TOW and MAXST doesn't decrease at all.

Power comparisons based on a qualitative trait have similar patterns to those based on a quantitative trait (Figures S6-S8). However, the power of TOW and MAXST decreases in the presence of both risk and protective variants, although decreases not as fast as that of WS (Figure S8). As pointed out by Wu *et al.* [2011] and Sha *et al.* [2012], decrease in power of TOW and MAXST 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.

In simulation set 2, we compare the power of OCST, Fisher-CT, and SKAT-O for power as a function of percentage of protective variants. Results are summarized in Figure 4. This figure shows that OCST is consistently more powerful than Fisher-CT and Fisher-CT is consistently more powerful than SKAT-O. Power simulation results based on a qualitative trait yield the same conclusions, but differences in power between the three tests are smaller than those based on a quantitative trait (Figure S9).



We also perform simulation studies to compare the power of the proposed test (OCST) with the Adaptive Weighting test (AW2) proposed by Sha *et al.* [2013]. The power comparisons of these two tests for power as a function of the heritability for quantitative traits are given in Figure 5. This figure shows that OCST is consistently more powerful than AW2.

## Discussion

There is increasing interest to detect associations between rare variants and complex traits. Reasons are that (1) the common variants identified through genome-wide association studies (GWAS) account for only a small portion of the presumed phenotypic variation and (2) the development of next-generation sequencing technology has made directly testing all rare variants feasible. Several statistical methods for rare variant association studies have been developed recently. However, recent studies show that the performance of each of these methods depends strongly upon the underlying assumption and no method demonstrates consistently acceptable power [Ladouceur *et al.* 2012]. More recently, Derkach *et al.*, [2012] and Lee *et al.*, [2012] proposed combined tests by combining information from the burden and quadratic tests. However, results from this study and from Kinnamon *et al.* [2012] show that there exist tests that can outperform both burden and quadratic tests in some situations. In this article, we propose a novel combined test OCST by combining information from tests of the three classes that are well beyond burden and quadratic tests. Our results show that, comparing with burden and quadratic tests, OCST either is the most powerful test or has similar power with the most powerful test. Our results also show that OCST has better power than the two combined tests: Fisher-CT and SKAT-O.

All the existing methods discussed in this article are for unrelated individuals only. Although our proposed method is also described using unrelated individuals, our method can be applied to family-based data as long as there is a single-variant test. As an example, we consider the within-family test  $T_{WFT}$  and admixture between-family test  $T_{adBFT}$  proposed by Fang *et al.* [2012] for family-based rare variant association studies. We can use either  $T_{WFT}$  or  $T_{adBFT}$  as the single-variant test  $s_m$  and then our method can be applied to family-based data through this  $s_m$ .

Using formula (2) to adjust for the effect of covariates for binary traits may look strange. Previous researches showed that using formula (2) to adjust for the effect of covariates for binary traits works well. To control for population stratification, Price *et al.* [2006] used formula (2) to adjust for the effect of covariates (eigenvectors) for binary traits and they showed that this method works fine. In rare variant association studies, Sha *et al.* [2012] used formula (2) to adjust for the effect of covariates for binary traits and their results also showed that this method works very well.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

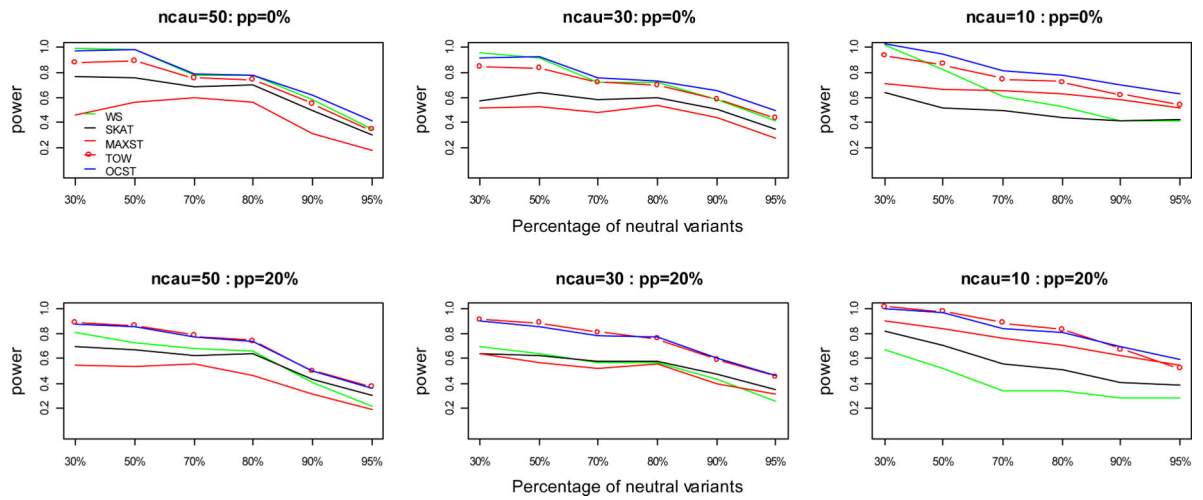
Research reported in this article was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number R03 HG006155. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes 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 Data Set was supported in part by NIH R01 MH059490 and used sequencing data from the 1000 Genomes Project ([www.1000genomes.org](http://www.1000genomes.org)).

## References

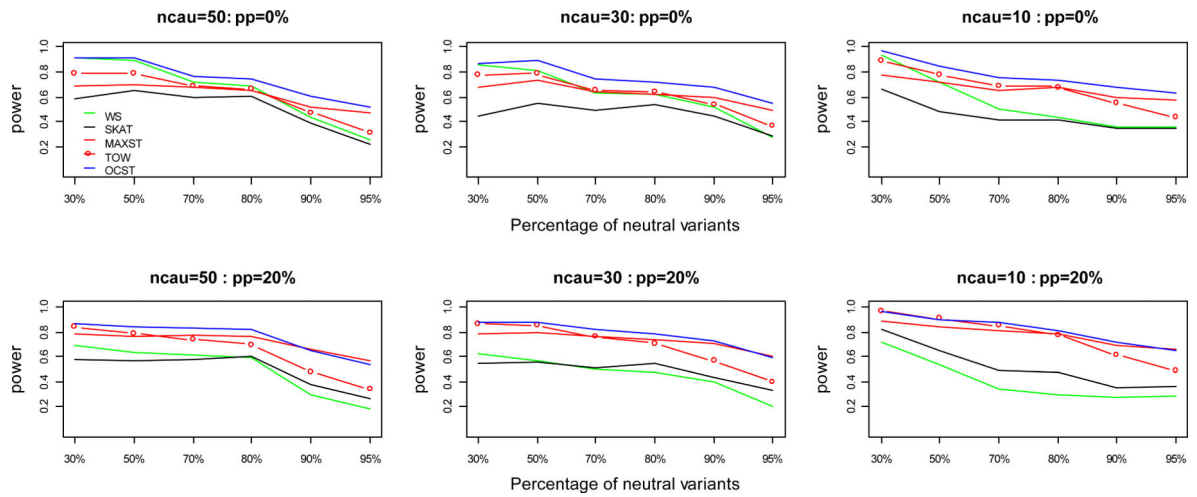
- Andrés AM, Clark AG, Shimmin L, Boerwinkle E, Sing CF, Hixson JE. Understanding the accuracy of statistical haplotype inference with sequence data of known phase. *Genetic Epi.* 2007; 31:659–671.
- Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet.* 2008; 40(6):695–701. [PubMed: 18509313]
- Derkach A, Lawless J, Sun L. Robust and powerful tests for rare variants using fisher’s method to combine evidence of association from two or more complementary tests. *Genetic Epi.* 2012; 37(1): 110–121.
- Guo W, Lin S. Generalized linear modeling with regularization for detecting common disease rare haplotype association. *Genet Epi.* 2009; 33:308–316.
- Han F, Pan W. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered.* 2010; 70:42–54. [PubMed: 20413981]
- Hoffmann TJ, Marini NJ, Witte JS. Comprehensive approach to analyzing rare genetic variants. *PLoS ONE.* 2010; 5(11):e13584. [PubMed: 21072163]
- Kinnamon DD, Hershberger RE, Martin ER. Reconsidering association testing methods using single-variant test statistics as alternatives to pooling tests for sequence data with rare variants. *PLoS ONE.* 2012; 7(2):e30238. [PubMed: 22363423]
- Ladouceur M, Dastani Z, Aulchenko YS, Greenwood CMT, Richards JB. The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals. *PLoS Genet.* 2012; 8:e1002496. [PubMed: 22319458]
- Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet.* 2012; 91:224–237. [PubMed: 22863193]
- Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet.* 2008; 83:311–321. [PubMed: 18691683]
- Lin D-Y, Tang Z-Z. A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet.* 2011; 89:354–367. [PubMed: 21885029]
- Madsen BE, Browning SR. A group-wise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 2009; 5:e1000384. [PubMed: 19214210]
- Metzker ML. Sequencing technologies – the next generation. *Nature Reviews Genetics.* 2010; 11:31–46.
- Morgenthaler S, Thilly WG. A strategy to discover genes that carry multiallelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res.* 2007; 615:28–56. [PubMed: 17101154]
- Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. Testing for an unusual distribution of rare variants. *PLoS Genet.* 2011; 7:e1001322. [PubMed: 21408211]
- Nelder J, Wedderburn R. Generalized linear models. *J R Stat Soc Ser A.* 1972; 135:370–384.
- Ng SB, Turner EH, Robertson PD. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature Letters.* 2009; 461:272–276.

- Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet.* 2010; 86:832–838. [PubMed: 20471002]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. PCs analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38:904–909. [PubMed: 16862161]
- Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet.* 2002; 11:2417–2423. [PubMed: 12351577]
- Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet.* 2001; 69:124–137. [PubMed: 11404818]
- Sha Q, Wang X, Wang X, Zhang S. Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genet Epi.* 2012; 36(6):561–71.
- Sha Q, Wang S, Zhang S. Adaptive clustering and adaptive weighting methods to detect disease associated rare variants. *Eur J Hum Genet.* 2013; 21(3):332–7. [PubMed: 22781093]
- Sha Q, Zhang Z, Zhang S. An improved score test for genetic association studies. *Genetic Epi.* 2011; 35:350–359.
- Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 2006; 78(4): 629–644. [PubMed: 16532393]
- Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet.* 2008; 40:17–22. [PubMed: 18163131]
- Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet.* 2010 doi: 10.1093/hmg/ddq333.
- Walsh T, King MC. Ten genes for inherited breast cancer. *Cancer Cell.* 2007; 11:103–105. [PubMed: 17292821]
- Wu M, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare variant association testing for sequencing data using the sequence kernel association test (SKAT). *Am J Hum Genet.* 2011; 89:82–93. [PubMed: 21737059]
- Yi N, Zhi D. Bayesian analysis of rare variants in genetic association studies. *Genet Epi.* 2011; 35:57–69.
- Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. *Am J Hum Genet.* 2010; 87:604–617. [PubMed: 21070896]



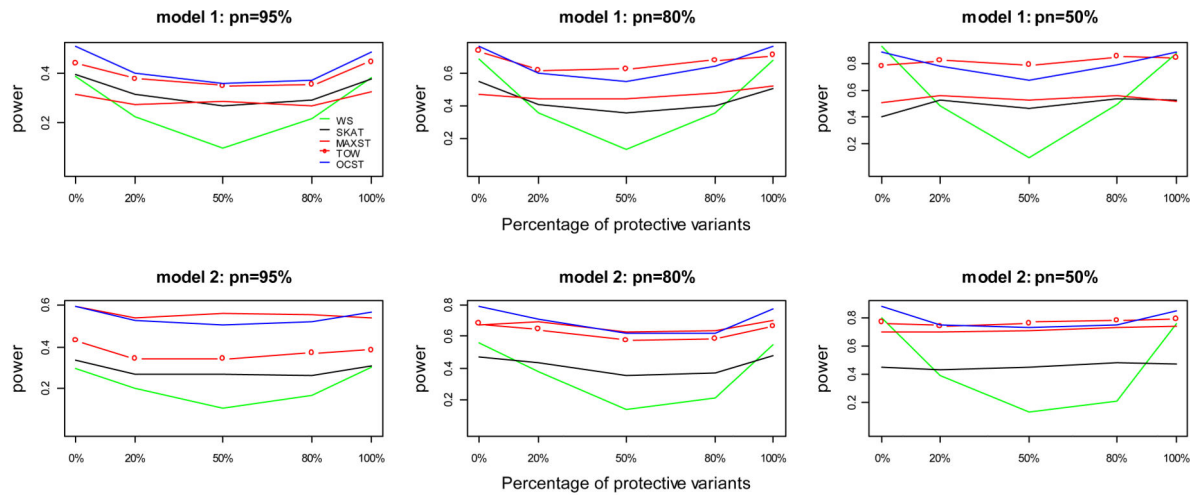
**Figure 1.**

Power comparisons of the five tests for power as a function of the percentage of neutral variants under model 1 for quantitative traits.  $ncau$  represents the number of causal variants.  $pp$  represents the percentage of protective variants. In this set of simulations, sample size is 1000; significance level is 0.001; total heritability is 0.05.  $Tn = ncau / (1 - pn)$ , where  $Tn$  represents the total number of variants and  $pn$  represents the percentage of neutral variants, and the  $Tn$  variants are randomly chosen from the 1057 variants of  $S_{gene}$ .



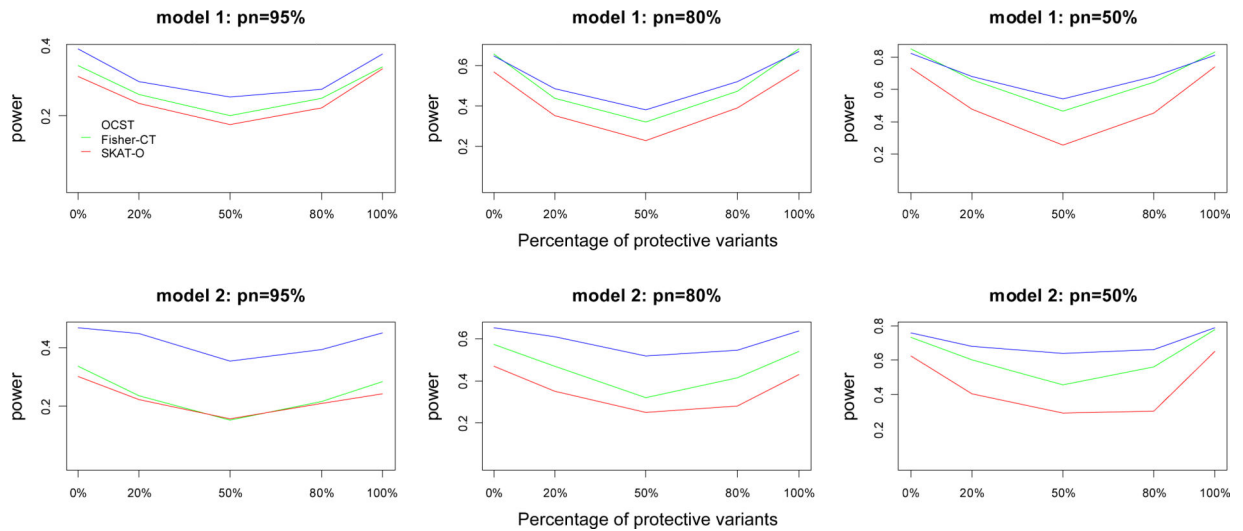
**Figure 2.**

Power comparisons of the five tests for power as a function of the percentage of neutral variants under model 2 for quantitative traits.  $ncau$  represents the number of causal variants.  $pp$  represents the percentage of protective variants. In this set of simulations, sample size is 1000; significance level is 0.001; total heritability is 0.05.  $Tn = ncau / (1 - pn)$ , where  $Tn$  represents the total number of variants and  $pn$  represents the percentage of neutral variants, and the  $Tn$  variants are randomly chosen from the 1057 variants of *Sgene*.



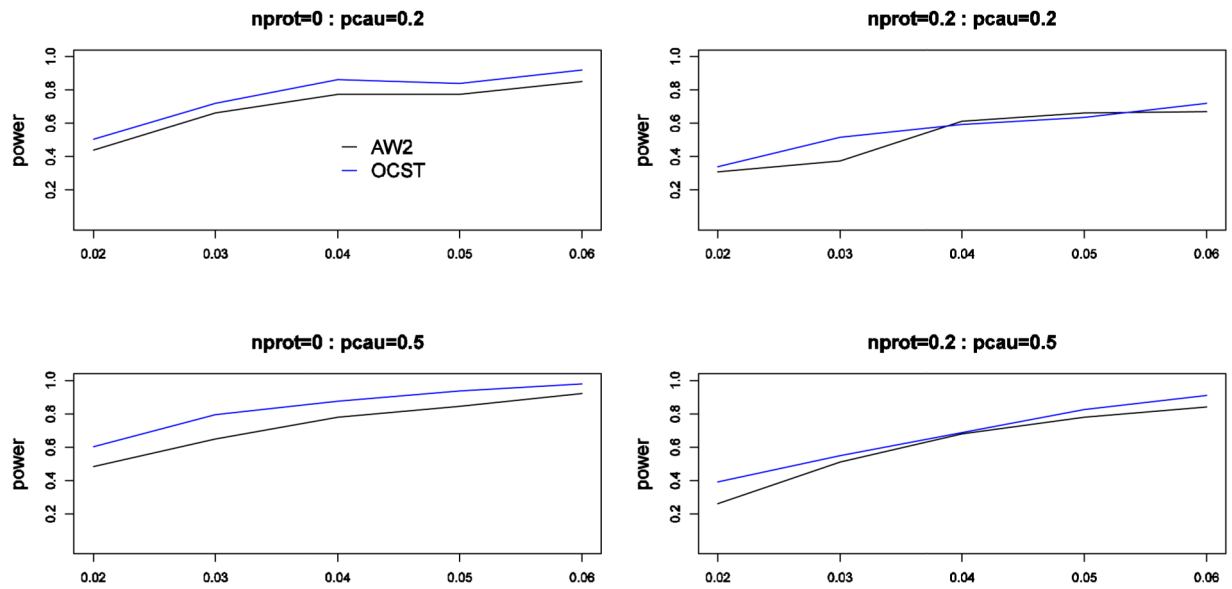
**Figure 3.**

Power comparisons of five tests for power as a function of the percentage of protective variants for quantitative traits. The number of causal variants is 20.  $pn$  represents the percentage of neutral variants. In this set of simulations, sample size is 1000; significance level is 0.001; total heritability is 0.05.  $Tn = ncau / (1 - pn)$ , where  $Tn$  represents the total number of variants and  $ncau$  represents the number of causal variants, and the  $Tn$  variants are randomly chosen from the 1057 variants of *Sgene*.



**Figure 4.**

Power comparisons of three tests (OCST, Fisher-CT, and SKAT-O) for power as a function of the percentage of protective variants for quantitative traits. The number of causal variants is 20.  $pn$  represents the percentage of neutral variants. In this set of simulations, sample size is 1000; significance level is 0.001; total heritability is 0.05.  $Tn = ncau / (1 - pn)$ , where  $Tn$  represents the total number of variants and  $ncau$  represents the number of causal variants, and the  $Tn$  variants are randomly chosen from the 1057 variants of *Sgene*.



**Figure 5.**

Power comparisons of the proposed test (OCST) and AW2 proposed by Sha et al. [2013] for power as a function of the heritability for quantitative traits. The number of causal variants is 30. nprot represents the percentage of protective variants. pcau represents the percentage of causal variants. In this set of simulations, sample size is 1000; significance level is 0.001.



**Table 1**

Estimated type I error rates (in percentage) of the five tests.

$\alpha$	Gene	Quantitative Traits					Qualitative Traits				
		WS	MAXST	TOW	SKAT	OCST	WS	MAXST	TOW	SKAT	OCST
5%	Gene1	5.21	5.11	4.82	5.18	4.96	5.04	5.26	5.04	5.04	5.20
	Gene3	5.06	5.06	5.12	5.10	5.14	4.72	5.37	4.97	4.93	5.07
	Gene5	4.92	5.37	4.88	4.54	5.08	4.98	5.37	5.14	5.03	4.97
	Sgene	5.23	5.09	4.74	5.18	4.73	4.77	4.98	4.68	4.90	5.03
	Gene1	0.96	0.86	0.89	1.08	0.77	1.09	1.05	1.09	0.91	0.97
1%	Gene3	1.09	1.09	1.09	1.06	1.14	0.99	1.06	0.97	1.06	0.90
	Gene5	0.93	1.12	0.98	0.84	0.93	0.85	1.22	1.01	1.15	1.04
	Sgene	0.89	1.04	1.12	1.02	1.02	0.98	1.08	0.92	0.94	0.99
	Gene1	0.11	0.06	0.10	0.05	0.13	0.13	0.08	0.09	0.10	0.09
	Gene3	0.07	0.11	0.12	0.08	0.07	0.07	0.08	0.11	0.10	0.08
0.1%	Gene5	0.12	0.15	0.06	0.03	0.12	0.10	0.10	0.12	0.14	0.16
	Sgene	0.13	0.13	0.13	0.07	0.18	0.06	0.11	0.11	0.14	0.13

Note:  $\alpha$  denotes the significance level. In this set of simulations, sample size is 1000.