

# A Data-Adaptive Sum Test for Disease Association with Multiple Common or Rare Variants

Fang Han Wei Pan

Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minn., USA

## Key Words

Genome-wide association study • Logistic regression •  
Multimarker analysis • Single nucleotide polymorphism

## Abstract

Since associations between complex diseases and common variants are typically weak, and approaches to genotyping rare variants (e.g. by next-generation resequencing) multiply, there is an urgent demand to develop powerful association tests that are able to detect disease associations with both common and rare variants. In this article we present such a test. It is based on data-adaptive modifications to a so-called Sum test originally proposed for common variants, which aims to strike a balance between utilizing information on multiple markers in linkage disequilibrium and reducing the cost of large degrees of freedom or of multiple testing adjustment. When applied to multiple common or rare variants in a candidate region, the proposed test is easy to use with 1 degree of freedom and without the need for multiple testing adjustment. We show that the proposed test has high power across a wide range of scenarios with either common or rare variants, or both. In particular, in some situations the proposed test performs better than several commonly used methods.

Copyright © 2010 S. Karger AG, Basel

## Introduction

Genome-wide association studies (GWASs) have been successful in identifying many DNA variants, namely single nucleotide polymorphisms (SNPs), associated with complex and common diseases. The basic mechanism of GWASs is indirect mapping of disease loci through linkage disequilibrium (LD) among physically close DNA variants, facilitating the use of tag SNPs to survey the whole genome. Their effectiveness also depends on the validity of the common disease-common variant (CDCV) hypothesis. In spite of the success of GWASs, for most common diseases the proportion of the overall phenotypic variance explained by the discovered disease-susceptibility loci remains very small [Maher, 2008]. Among several possible explanations, we consider two. First, because the effect sizes of most discovered risk loci are quite small, and they account for only remarkably small amounts of the phenotypic variances, it is likely that many more loci with much smaller effects remain to be detected. As Flint and Mackay [2009] showed in their review, the distribution of effect sizes at 140 loci discovered so far for 20 human disease phenotypes ranged from odds ratio (OR) 1 to 2.4, with the overwhelming majority between 1 and 1.4 and with a mode of only 1.2. However, as observed in a recent multi-stage study [Willer et al., 2009] that detected six previously unknown common variants influencing body mass index (BMI), despite the huge sample sizes ( $n = 32,387$  in stage 1 and  $n = 45,018$  in

stage 2), the eight known loci explained only 0.84% of the variance, although family and twin studies have shown that genetic factors account for 40–70% of the population variation in BMI. Second, an alternative to the CDCV hypothesis is the common disease-rare variant (CDRV) hypothesis, which states that for complex diseases there is extreme genetic heterogeneity and that the disease etiology is caused collectively by multiple rare variants with moderate to high penetrance, as supported by some recent studies [Azzopardi et al., 2008; Cohen et al., 2004; Ji et al., 2008; Walsh et al., 2008]. More theoretical and practical arguments supporting the CDRV hypothesis have appeared in the literature [e.g. Bodmer and Bonilla, 2008; Gorlov et al., 2008; Pritchard, 2001; Pritchard and Cox, 2002]. Although still highly debated, both hypotheses have merits, and it is likely that, for most complex diseases, the allelic architecture of susceptibility variants has a wide range of allele frequencies and effect sizes [McCarthy, 2009; Schork et al., 2009; Wen et al., 2004].

Therefore, from a methodological point of view, it would be desirable to develop statistical methods that can address both of the issues raised above. First, dealing with small effect sizes of common susceptibility variants requires the use of a statistical test with the highest possible statistical power. Due to the existence of LD, it seems more effective to use multiple variants or SNPs in an LD block or a candidate gene region. Although in theory there is no single most powerful test for multiple SNPs across all scenarios [Cox and Hinkley, 1974; Pan, 2009], some tests tend to have higher power in many common situations [Chapman and Whittaker, 2008; Pan, 2009] than others. In particular, as advocated by Wang and Elston [2007], one needs to strike a balance between using multiple SNPs and the thus incurred high cost of large degrees of freedom (DF) or of multiple test adjustment. A simple test called the Sum test, similar to a test proposed by Wang and Elston [2007], summarizes information across multiple SNPs in LD with only DF = 1 and needs no multiple test adjustment. It was shown to have high power in some situations, but, unfortunately, not in others [Chapman and Whittaker, 2008; Pan, 2009]. It is the first goal of this paper to modify the Sum test in a data-adaptive way so that it can maintain high power across most scenarios.

Importantly, to address the second issue raised above, the Sum test or its modifications should be applicable to detect disease association with rare variants. Note that most of the existing tests were proposed under the CDCV hypothesis, and may not be applicable to a situation with rare variants. To our knowledge, there are only three tests

specifically designed under the CDRV hypothesis [Morgenthaler and Thilly, 2007; Li and Leal, 2008; Madsen and Browning, 2009]. A key idea shared by the three tests for rare variants is to group or collapse multiple variants, a feature closely related to the Sum test. It is the second goal of this paper to show the effectiveness of the modified Sum test for rare variants, and more generally, to extend it to detect disease association with both common and rare variants. In particular, we show that in some situations the modified Sum test significantly outperforms any of the three existing tests, while for others it performs similarly to the three tests.

## Methods

### *Logistic Regression for Common Variants*

Given  $m$  independent observations  $(Y_i, X_i)$  with  $Y_i = 0$  or 1 as disease status and  $X_i = (X_{i1}, \dots, X_{ik})$  as genotypes at  $k$  SNPs in LD for subject  $i$ , we would like to test for any possible association between the disease and genotypes. As usual, we use the dosage coding for  $X_{ij}$  under an additive genetic model:  $X_{ij} = 0, 1$  or 2, representing the copy number of one of the two alleles present in SNP  $j$  of subject  $i$ , though other genetic models can be adopted. Many multilocus association tests are based on fitting a logistic regression model

$$\text{Logit Pr}(Y_i = 1) = \beta_0 + \sum_{j=1}^k X_{ij}\beta_j. \quad (1)$$

A global test of any possible association between the disease and SNPs can be formulated as jointly testing on the multiple  $\beta_j$  parameters with the null hypothesis  $H_0: \beta = (\beta_1, \dots, \beta_k)' = 0$  by one of the three asymptotically equivalent tests: the likelihood ratio test, the Wald test and the score test. Under  $H_0$ , any of the three test statistics has an asymptotic  $\chi^2$  distribution with DF =  $k$ . The generalized Hotelling's  $T^2$  test [Fan and Knapp, 2003; Xiong et al., 2002] is closely related to the score test [Clayton et al., 2004]. A potential problem with the above tests is the power loss due to large DF for a large  $k$ .

In contrast to a global test, another extreme is to conduct a single-locus test for each SNP [Roeder et al., 2004]. Rather than including all  $k$  SNPs, we can include each SNP sequentially in a series of univariate or marginal regression models:

$$\text{Logit Pr}(Y_i = 1) = \beta_{M,0j} + X_{ij}\beta_{M,j}, \quad (2)$$

where we explicitly distinguish  $\beta_M = (\beta_{M,1}, \dots, \beta_{M,k})'$  in marginal models (2) from  $\beta$  in the joint model (1). One would proceed by testing  $H_0: \beta_{M,j} = 0$  for each  $j = 1, \dots, k$  sequentially. Each test can be done with only one DF, but a multiple test adjustment has to be made based on numerical solutions [Conneely and Boehnke, 2007], permutations or other methods. The method is equivalent to choosing the univariate test for  $\beta_{M,j}$  with the minimum p value, and is hence also called UminP method.

Most of the existing methods fall into one of the two extreme categories mentioned above. The limitations of the two extremes have been well recognized and induced the development of new

methods. A general strategy is to capture full information in all  $\beta_j$ 's while reducing the cost of large DF or adjustment for multiple testing, but the key is how to reconcile the two possibly conflicting goals. A weighted score test [Wang and Elston, 2007] and a Sum test [Chapman and Whittaker, 2008; Pan, 2009] are two such attempts.

#### The Sum Test

The Sum test can be used to strike a balance between jointly modeling with multiple SNPs and its resulting large DF: while using all the SNPs, it adopts a key and generally incorrect working assumption that all SNPs are associated with the disease with a common OR:

$$\text{Logit Pr}(Y_i = 1) = \beta_{c0} + \sum_{j=1}^k X_{ij}\beta_c, \quad (3)$$

where  $\beta_c$  reflects the common association strength between the disease and each SNP under the working assumption. To address the question of whether there is any association between the disease and any SNP, one only needs to test on one parameter with  $H_0: \beta_c = 0$ , which can easily be done by the likelihood ratio, Wald or score test in fitting model (3). Note that fitting model (3) is equivalent to regressing  $Y$  on a new covariate that is the sum of the genotypes of the multiple SNPs, and hence we call the resulting test the Sum test.

The main advantage of the Sum test is that, because it tests on only one parameter  $\beta_c$ , there will be no power loss due to large DF or multiple testing adjustment. Generally, the common association parameter  $\beta_c$  in (3) is an 'average' (or more precisely, a function) of the individual  $\beta_1, \dots, \beta_k$ . In linear regression, Pan [2009] showed that

$$\hat{\beta}_c = \frac{\sum_{j=1}^k \sum_{i=1}^m X_{ij}^2 \hat{\beta}_{M,j}}{\sum_{i=1}^m \left( \sum_{j=1}^k X_{ij} \right)^2}, \quad (4)$$

where  $\hat{\beta}_c$  is the normal-based maximum likelihood estimate (or least-squares estimate) of  $\beta_c$  in a linear model analogous to (3), and  $\hat{\beta}_M$  is the maximum likelihood estimate of  $\beta_M$  in the marginal linear models analogous to (2). Now it becomes clear that the main problem with the Sum test is its dependence on the signs of  $\hat{\beta}_{M,j}$ , or on the codings of each SNP (i.e. which allele is chosen as the reference category). If the signs of  $\beta_{M,j}$  are quite different, it may result in a small  $\hat{\beta}_c$ , and thus in power loss. Hence, before applying the Sum test, based on some ad hoc heuristic, one needs to choose the codings of the SNPs to maximize the number of their positive pairwise correlations [Pan, 2009]: (1) calculate pairwise Pearson's correlation coefficients between any two SNPs; (2) see which SNP  $s$  has the largest number of negative correlation coefficients, say  $n_s$ , with other SNPs; (3) if  $n_s > k/2$ , then flip the coding of SNP  $s$ , and repeat the above process; otherwise, stop. Unfortunately, such an endeavor does not guarantee any good performance of the Sum test. A similar issue exists with the weighted score test of Wang and Elston [2007]. Below we propose a data-adaptive Sum test that overcomes the above-mentioned limitation of the original Sum test.

#### A Data-Adaptive Sum Test

The problem with the Sum test is that, as shown in equation (4), because the estimated common effect  $\hat{\beta}_c$  is a linear combina-

tion of the components of  $\hat{\beta}_M$  with always positive coefficients, the test may have reduced power with a small  $\hat{\beta}_c$  when the components of  $\hat{\beta}_M$  have different signs. Hence, a natural approach is to choose the coding of each SNP  $j$  based on the sign of  $\hat{\beta}_{M,j}$ , which is data-adaptive and may lead to inflated type I error rates if no adjustment is made with the null distribution. Our proposed data-adaptive Sum (aSum) test is based on such an idea. However, critically, we need to address the following two issues. First, how to modify the null distribution of the resulting test statistic such that the type I error rate will be well controlled, and second, how to increase the power of the aSum test. Our proposed implementation below addresses both issues.

#### Step 1

With the original data  $\{(Y_i, X_{ij})\}$ , we fit the marginal regression model (2) to each SNP  $j$ , obtaining  $\hat{\beta}_{M,j}$  and p value  $p_{M,j}$ , for  $j = 1, \dots, k$ .

#### Step 2

With a chosen cutoff  $\alpha_0$  for each  $j = 1, \dots, k$ , if  $\hat{\beta}_{M,j} < 0$  and  $p_{M,j} \leq \alpha_0$ , we change the coding of SNP  $j$  from  $X_j$  to  $X_j^* = 2 - X_j$ ; otherwise, no change is done with  $X_j^* = X_j$ . Note the resulting data  $\{(Y_i, X_i^*)\}$ .

#### Step 3

With the new data  $\{(Y_i, X_i^*)\}$ , we fit the common-effect model (3), obtaining the usual score statistic  $U$  under  $H_0: \beta_c = 0$ , its variance  $V$  and a p value  $p$  from the test statistic  $U/\sqrt{V}$  with the usual null distribution  $N(0, 1)$ .

#### Step 4

By repeatedly permuting the disease indicators of the original data, we obtained  $B$  sets of permuted data  $\{(Y_i^{(b)}, X_i)\}$  for  $b = 1, \dots, B$ . For each permuted data  $\{(Y_i^{(b)}, X_i)\}$ , we repeat the above steps (1) to (3), obtaining the score statistic  $U^{(b)}$  for  $H_0: \beta_c = 0$  and its usual normal-based p value  $p^{(b)}$ .

A proper permutation-based p value for the aSum test, called aSum-P, is calculated as  $\sum_{b=1}^B I(p^{(b)} < p)/B$ .

#### Step 5

Calculate the sample mean and sample variance (or covariance matrix, in general) of  $U^{(1)}, \dots, U^{(B)}$  as  $U_0$  and  $V_0$ . The aSum test statistic is

$$aSum = (U - U_0)' V_0^{-1} (U - U_0),$$

which is compared to a null distribution  $a\chi_d^2 + b$  to obtain a p value, where  $d = \dim(U)$ ,  $a$  and  $b$  are obtained to match the first two moments of  $a\chi_d^2 + b$  to that of the empirical distribution of  $U^{(1)}, \dots, U^{(B)}$ .

#### Remark 1

In step 2, the choice of  $0 \leq \alpha_0 \leq 1$  is related to the power of the aSum test. Trivially, if  $\alpha_0 = 0$ , the aSum test reduces to the usual Sum test. If  $\alpha_0 = 1$ , we choose the coding of SNP  $j$  based on the sign of  $\hat{\beta}_{M,j}$  which is perhaps the most intuitive way to overcome the problem with the Sum test; however, as will be shown later, the corresponding aSum test has low power because the resulting null distribution has heavy tails. Hence, we would like to choose a relatively small  $\alpha_0$  such that (i) under  $H_0$ , we do not tend to change the coding of any SNP, and (ii) when  $H_0$  does not hold,

we would like to change the coding of an SNP with negative  $\hat{\beta}_{M,j}$ . As a compromise, by default, we used  $\alpha_0 = 0.1$  throughout the study. As will be shown, we found that  $\alpha_0 = 0.1$  performed well across all situations considered here. Nevertheless, the optimal choice of  $\alpha_0$  should depend on the data at hands, and hence some data-driven methods of choosing  $\alpha_0$  will be useful. One possibility is to choose  $\alpha_0$  to maximize the (estimated) power of the resulting test, as discussed in Pan et al. [2010]; more research is needed.

#### Remark 2

In step 3, we implement the aSum test with the score statistic, although the Wald test statistic can be equally applied.

#### Remark 3

The permuted data in step 4 are generated to estimate the effects of the data-adaptive nature in step 2, and thus accordingly to adjust its null distribution. Note that the normal-based p value  $p$  is no longer a genuine p value because the usual null distribution  $N(0, 1)$  is not valid and a necessary adjustment has to be made; as shall be shown later, the direct use of  $p$  (called naive aSum, or na-Sum test) will lead to dramatically inflated type I error rates.

#### Remark 4

Step 5 is optional in calculating the p value of the aSum test by a theoretical null distribution. Although it depends on the use of permutation, in contrast to the aSum-P, it may not require a large number of permutations,  $B$ , to reach a small p value; by default we used  $B = 100$ . However, this theoretical null distribution is only an approximation, which may or may not work well.

#### Remark 5

By the Central Limit Theorem with a large sample size,  $U$  should have an approximately normal distribution, and thus the test statistic aSum should have an approximately  $\chi_d^2$  as its null distribution. However, as is to be shown later, we found empirically that the null distribution of the standardized  $U$  tends to be slightly right-skewed (due to its data-adaptive nature), and hence we used a scaled-shifted  $\chi_d^2$  as a null distribution to correct for its skewness.  $a$  and  $b$  are obtained by the well-known Satterthwaite [1946] approximation.

#### Tests for Rare Variants

To our knowledge, there are only three existing methods to test the association for rare variants for case-control data: the cohort allelic sums test (CAST) [Cohen et al., 2004; Morgenthaler and Thilly, 2007], the Combined Multivariate and Collapsing (CMC) method [Li and Leal, 2008], and a groupwise weighted Sum (w-Sum) test [Madsen and Browning, 2009]. For rare variants, neither joint modeling nor univariate regression on individual SNPs works, and some grouping is needed. Since most disease-associated rare variants discovered so far are nonsynonymous sequence variants in some candidate genes [Cohen et al., 2004], it is natural to group nonsynonymous variants in a gene. However, a recent study has identified some susceptible rare non-coding variants [Haller et al., 2009]. Madsen and Browning [2009] suggest grouping variants according to a functional element, such as gene, pathway and ultra-conserved area. In particular, the CAST works by first collapsing the genotypes across variants into a new coding:  $X_{i,C} = 1$  if any  $X_{ij} > 0$  (i.e. any rare variant is present), and  $X_{i,C} = 0$  otherwise. It then tests the association between

the disease and this new  $X_{i,C}$ . It can be regarded as fitting a logistic regression model

$$\text{Logit Pr}(Y_i = 1) = \beta_{C0} + X_{i,C}\beta_C, \quad (5)$$

and testing  $H_0: \beta_C = 0$ . Comparing models (3) and (4), it is clear that the CAST is closely related to the Sum test: both test on only one parameter representing some common effect of the variants; they differ in their use of the new genotypic coding  $X_{i,C} = \bigvee_{j=1}^k X_{ij}$  versus  $X_{i,S} = \sum_{j=1}^k X_{ij}$ , similar to the use of a dominant genetic mode versus an additive genetic mode for the effects of the individual variants. Note that for rare variants, we have  $X_{i,C} \approx X_{i,S}$ . Hence, it seems natural to apply the Sum test to rare variants. More importantly, if the effects of the rare variants are in opposite directions (i.e. some are protective while others deleterious), the CAST will encounter the same problem as the Sum test discussed in the context of common variants. In that case, it would be better to use the aSum test. Note that recent studies have confirmed that nonsynonymous sequence variants in certain genes are preferentially enriched at either or both of the extremes of a phenotype's population distribution, suggesting that the effects of nonsynonymous variants could be either harmful or beneficial, although most are harmful [Cohen et al., 2004]. In fact, some nonsynonymous variants in gene PCSK9 have been discovered to be associated with lower plasma levels of low-density lipoprotein cholesterol (LDL-C) (e.g. only present in low-LDL subjects), while others are associated with higher levels of LDL-C [Kotowski et al., 2006].

Li and Leal [2009] modified the CAST to improve its performance with both rare and common variants. Specifically, for any rare mutations with minor allele frequencies (MAFs) of less than 0.01, they will be combined into a new group as in the CAST, while each common variant (with  $\text{MAF} > 0.01$ ) forms its own group, and the generalized Hotelling's test is applied to the thus formed multiple groups. We can modify our aSum test in a similar way: we combine the rare variants into one group and combine the common variants into another group by summing over their genotypic codings (which are data-adaptively determined as before), then test on the two corresponding regression coefficients in a logistic regression model. We call the resulting test aSumC test.

There are two potential advantages of the aSumC test over the CMC test, as will be shown. First, the aSumC test overcomes the problem with different association directions of the functional variants, from which both the CMC test and the w-Sum test [Madsen and Browning, 2009] suffer with possibly significant power loss. Second, with only two groups, the aSumC may have a much smaller number of DF and thus higher power than the CMC test.

#### Other Tests to Be Compared

We will numerically compare the proposed aSum test with several other tests: the multivariate score test, the sum of squared score (SSU) test and its weighted version (SSUw) on  $H_0: \beta = 0$ , all based on the joint model (1), and the UminP test for the marginal models (2), as considered in Pan [2009].

Based on model (1), we can derive the score vector and its covariance matrix as

$$U = \sum_{i=1}^m (Y_i - \bar{Y}) X_i, \\ V = \bar{Y}(1 - \bar{Y}) \sum_{i=1}^m (X_i - \bar{X})(X_i - \bar{X})',$$



where  $\bar{Y} = \sum_{i=1}^m Y_i / m$  and  $\bar{X} = \sum_{i=1}^m X_i / m$ . The four test statistics are

$$T_{\text{Score}} = U'V^{-1}U,$$

$$T_{\text{SSU}} = U'U,$$

$$T_{\text{SSUw}} = U'V_d^{-1}U \quad \text{with} \quad V_d = \text{Diag}(V),$$

$$T_{\text{UminP}} = \max_{j=1}^k U_j^2 / v_j,$$

where  $U_j$  is the  $j$ -th element of  $U$  and  $v_j$  is the  $(j, j)$ -th diagonal element of  $V$ . Under  $H_0$ , the first three statistics have either asymptotic or approximate  $\chi^2$  distributions [Pan, 2009], while the distribution of  $T_{\text{UminP}}$  can be numerically obtained [Conneely and Boehnke, 2007].

Next we review the w-Sum test of Madsen and Browning [2009] for rare variants. It is still based on the basic idea of grouping rare variants, but differs from the Sum test in (i) using a weighted sum, instead of a simple sum, of the rare variant scores for each subject, and (ii) comparing the ranks of the weighted sums, rather than the sums, between the case and control groups. Specifically, if  $A = \{i : Y_i = 1\}$  and  $\bar{A} = \{i : Y_i = 0\}$  denote the sets of cases and controls, respectively, the w-Sum test is constructed as follows.

#### Step 1

Calculate weight  $w_j = \sqrt{m_j q_j (1 - q_j)}$ , where  $q_j = (\sum_{i \in \bar{A}} X_{ij} + 1) / (2n_j^U + 2)$ ,  $m_j$  is the total number of subjects genotyped for variant  $j$ , and  $n_j^U$  is the total number of controls genotyped for variant  $j$ .

#### Step 2

Calculate  $\gamma_i = \sum_{j=1}^k X_{ij} / w_j$  and  $R = \sum_{i \in A} \text{rank}(\gamma_i)$ .

#### Step 3

Generate a permuted dataset  $b$  (by randomly shuffling the disease labels), and calculate the corresponding  $R^{(b)}$ . Repeat for  $B$  times.

#### Step 4

The test statistic is  $T_{\text{w-Sum}} = (R - \bar{R}) / \sqrt{\text{Var}(R)}$ , where  $\bar{R}$  and  $\text{Var}(R)$  are the sample mean and variance of  $\{R^{(1)}, \dots, R^{(B)}\}$ . A  $p$  value can be obtained by referring  $T_{\text{w-Sum}}$  to a standard normal distribution. Alternatively, a permutation  $p$  value can be obtained by comparing  $R$  with  $\{R^{(1)}, \dots, R^{(B)}\}$ .

## Results

### Common Variants

#### Simulated Data with Compound Symmetry Correlation Structure: Ideal for the Sum Test

We first consider an ideal situation for the Sum test: there is a common association strength between any SNP and the disease, as considered by Wang and Elston [2007] and Pan [2009]. Specifically, we simulated  $k = 10$  marker SNPs with sample size  $n = 500$  cases or controls. The disease-causing SNP was assumed to be in the center of the

marker SNPs, but was removed from the data. First, we generated a latent vector from a multivariate normal distribution with a compound symmetry covariance structure: there was an equal correlation  $\rho = 0.4$  between any two SNPs. Second, the latent vector was dichotomized to yield a haplotype with the allele frequencies of the marker SNPs randomly between 0.2 and 0.8, while the MAF for the disease-causing SNP was fixed at 0.2. Third, we combined two haplotypes and obtained marker genotype data  $X_i = (X_{i1}, \dots, X_{ik})'$  (after removing the disease-causing SNP  $X_{0i}$ ) for subject  $i$ . Fourth, the disease status  $Y_i$  of subject  $i$  was generated from a logistic regression model:

$$\text{Logit Pr}(Y_i = 1) = \beta_0 + \log(\text{OR})X_{i0}, \quad (6)$$

where we chose  $\beta_0 = -\log 4$  to give a background (i.e. not caused by the causal SNP) disease probability of 0.2, and the OR ranged from 1 (i.e. no association) to 2. Finally, following the case-control design, we sampled  $n$  cases (with  $Y_i = 1$ ) and  $n$  controls (with  $Y_i = 0$ ). We excluded the disease-causing SNP, supplying  $\{(Y_i, X_i) : i = 1, 2, \dots, 2n\}$  as a dataset to various statistical tests. For each setup, we simulated 1,000 datasets from which we obtained an empirical size or power for each test as its proportion of correctly or incorrectly rejecting its  $H_0$ .

The results for the sample sizes  $n = 500$  (i.e. 500 cases and 500 controls) and  $n = 1,000$  based on 1,000 replicates are shown in table 1. It is clear that the aSum test and its permutation-based aSum-P performed best with the properly controlled type I error rates and highest power. In particular, we note that the naSum test (which does not make any adjustment to its null distribution) yielded much-inflated type I error rates  $> 0.50$ .

Figure 1 depicts the empirical distributions of the various statistics from 1,000 replicates under the null hypothesis (i.e. OR = 1) with  $n = 500$ . It is noteworthy that the usual score test statistics  $U/\sqrt{V}$  were centered at 0.3, not 0 as dictated by the usual null distribution  $N(0, 1)$  in the conventional situation, due to the data-adaptive choice of the SNP codings, explaining why the use of the  $N(0, 1)$  as the null distribution would lead to inflated type I error rates for the naive naSum test. By an adjustment with permuted data, the distribution of  $(U - U_0)/\sqrt{V_0}$  was close to  $N(0, 1)$ , though the former seemed to be slightly right-skewed (fig. 1c, d), which motivated the use of  $a\chi_1^2 + b$  as a reference distribution for the aSum statistic (fig. 1e, f). If  $\chi_1^2$  was used as the null distribution for the aSum test, we found that the resulting test was conservative with some power loss (results not shown). The empirical distribution of  $a$  was characterized by its three quartiles 0.927, 0.997 and 1.090, and by its mean 1.019,

**Table 1.** Type I error and power based on 1,000 replicates for simulated data with a CS covariance structure

<i>n</i>	OR	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	naSum
500	1.0	0.045	0.043	0.044	0.046	0.051	0.052	0.053	0.552
	1.4	0.085	0.193	0.199	0.149	0.235	0.278	0.262	–
	1.6	0.139	0.356	0.358	0.237	0.393	0.447	0.440	–
	1.8	0.245	0.519	0.518	0.365	0.577	0.622	0.608	–
	2.0	0.346	0.662	0.661	0.489	0.711	0.747	0.740	–
	2.2	0.448	0.771	0.767	0.599	0.808	0.832	0.827	–
1,000	1.0	0.033	0.033	0.032	0.029	0.040	0.047	0.050	0.548
	1.2	0.053	0.119	0.116	0.086	0.139	0.177	0.167	–
	1.4	0.145	0.349	0.344	0.232	0.386	0.426	0.411	–
	1.6	0.310	0.641	0.641	0.449	0.673	0.717	0.718	–
	1.8	0.499	0.836	0.835	0.658	0.872	0.892	0.890	–
	2.0	0.691	0.935	0.933	0.816	0.965	0.968	0.966	–

$k = 10$  variants,  $a_0 = 0.1$ ,  $n$  cases and  $n$  controls. The significance level is at 0.05.

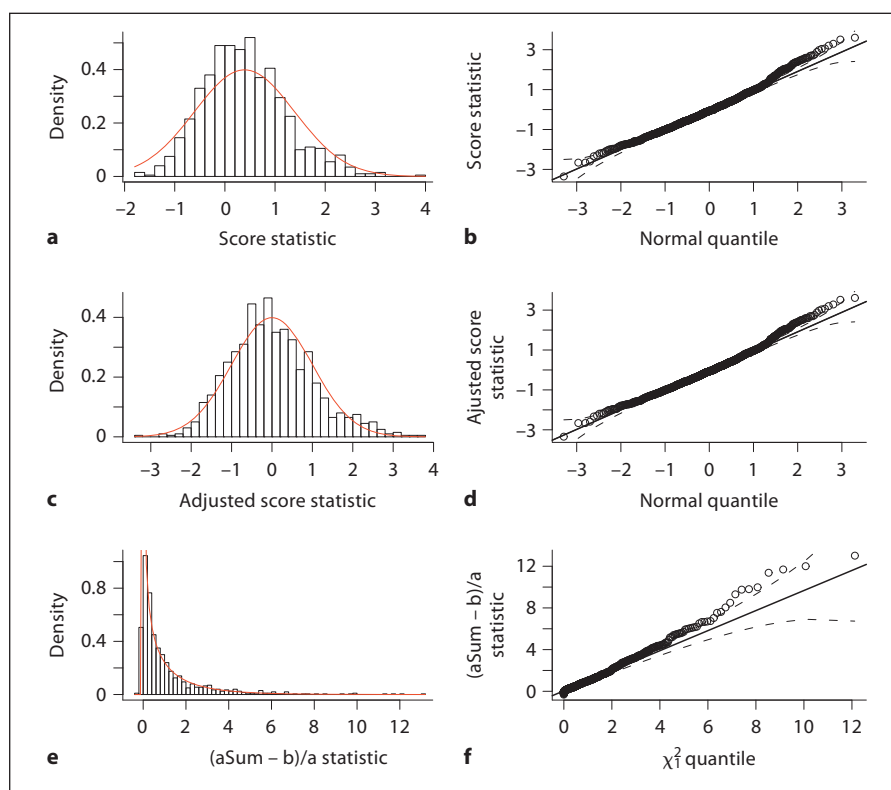
**Table 2.** Effects of  $a_0$  on type I error and power for the aSum test based on 1,000 replicates for simulated data in table 1

OR	$a_0 = 0.05$	$a_0 = 0.1$	$a_0 = 0.2$	$a_0 = 0.5$	$a_0 = 1.0$
1.0	0.053	0.053	0.055	0.052	0.052
1.4	0.272	0.262	0.255	0.208	0.165
1.6	0.455	0.440	0.429	0.369	0.351
1.8	0.629	0.608	0.587	0.523	0.496
2.0	0.756	0.740	0.725	0.650	0.626
2.2	0.837	0.827	0.812	0.753	0.721

**Table 3.** Type I error and power based on 1,000 replicates for gene CHI3L2

<i>n</i>	OR	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	naSum
500	1.0	0.050	0.050	0.049	0.050	0.050	0.059	0.049	0.211
	1.1	0.068	0.125	0.123	0.107	0.129	0.170	0.175	–
	1.2	0.144	0.363	0.358	0.288	0.346	0.399	0.412	–
	1.3	0.293	0.671	0.672	0.589	0.633	0.676	0.680	–
	1.4	0.470	0.891	0.894	0.852	0.839	0.884	0.910	–
	1.5	0.718	0.974	0.972	0.950	0.952	0.968	0.971	–
1,000	1.0	0.049	0.042	0.045	0.052	0.045	0.045	0.047	0.201
	1.1	0.079	0.204	0.199	0.170	0.190	0.247	0.239	–
	1.2	0.271	0.644	0.636	0.569	0.608	0.648	0.664	–
	1.3	0.571	0.928	0.930	0.896	0.898	0.926	0.933	–

$k = 17$  variants,  $a_0 = 0.1$ ,  $n$  cases and  $n$  controls. The significance level is at 0.05.



**Fig. 1.** Distributions of the naive score test statistics  $U/\sqrt{V}$  (a), the adjusted score statistics  $(U - U_0)/\sqrt{V_0}$  (c) and  $(a\text{Sum} - b)/a$  (e), and their observed values against the quantiles from  $N(0.3, 1)$  (b),  $N(0, 1)$  (d) and  $\chi^2_1$  (f), respectively. All are based on  $n = 500$  and 1,000 replicates of the simulated data with a compound symmetry covariance structure.

while that of  $b$  had three quartiles  $-0.100$ ,  $-0.007$  and  $0.064$ , and mean  $-0.029$ . Note that 95% pointwise confidence envelopes (implemented in R package *car*) were drawn on the Q-Q plots (fig. 1b, d and f): there were 125, 109 and 86 points outside the confidence bounds in figure 1b, d, f, respectively. When the Kolmogorov-Smirnov goodness-of-fit test was applied to cases b, d and f, we obtained the p values of  $2 \times 10^{-7}$ , 0.068 and 0.163, respectively. Hence, in summary, the aSum test statistics had the best fit to the theoretical null distribution  $a\chi^2_1 + b$ .

Table 2 lists the results for the aSum test with various values of the cutoff  $\alpha_0$  for  $n = 500$ . It seems that the results were not too sensitive to  $\alpha_0$  for  $\alpha_0$  around 0.1. However, for  $\alpha_0 = 0.5$  or 1, there was a substantial loss of power.

#### HapMap Data for Gene CHI3L2

Wang and Elston [2007] and Pan [2009] conducted a simulation study based on real LD patterns within the CHI3L2 gene as observed in the HapMap data, for which the Sum test performed well. The SNPs of the CHI3L2 gene for the 90 CEU (Utah residents with ancestry from northern and western Europe) individuals from the HapMap site were downloaded in June 2008. For data processing, the same procedure as the previous authors used

was followed: first, we excluded SNPs with  $\text{MAF} \leq 0.2$ , leaving 23 SNPs. Second, we did a single imputation for each of the missing genotypes by randomly drawing an observed genotype of the same SNP. Third, we used the dosage coding for the SNPs and applied the algorithm described earlier to minimize the number of negative pairwise SNP-SNP correlations. Fourth, we deleted redundant SNPs that were perfectly correlated with other SNPs. There was substantial LD among the remaining 17 SNPs as indicated by the distribution of their pairwise Pearson's correlation coefficients, which ranged from  $-0.388$  to  $0.989$ . Fifth, we repeatedly sampled (with replacement) subjects from the 90 CEU individuals. Finally, we chose SNP rs2182114 as disease-causing and generated disease indicators from the logistic model (6) with the same  $\beta_0$  and other possible values of OR.

The simulation results for the sample sizes  $n = 500$  and  $n = 1,000$  with 1,000 replicates are shown in table 3. Here the aSum test had a correct type I error rate. Furthermore, it had the highest power with a slight edge over the SSU and SSUw tests. In particular, it outperformed the Sum test, although the latter also worked quite well. We also noted that the naive naSum test without adjustment again had highly inflated type I error rates.

**Table 4.** Type I error and power based on 1,000 replicates for gene IL21R

<i>n</i>	OR	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	naSum
500	1.0	0.039	0.042	0.044	0.055	0.050	0.047	0.048	0.354
	1.2	0.123	0.202	0.208	0.182	0.164	0.225	0.212	–
	1.3	0.205	0.402	0.402	0.417	0.304	0.422	0.415	–
	1.4	0.369	0.594	0.589	0.652	0.432	0.644	0.626	–
	1.5	0.526	0.737	0.740	0.829	0.527	0.784	0.768	–
	1.6	0.691	0.828	0.836	0.909	0.607	0.851	0.877	–
1,000	1.0	0.042	0.044	0.043	0.047	0.034	0.052	0.057	0.343
	1.2	0.214	0.390	0.391	0.408	0.286	0.381	0.367	–
	1.3	0.441	0.679	0.670	0.730	0.499	0.714	0.699	–
	1.4	0.726	0.826	0.826	0.918	0.604	0.839	0.869	–
	1.5	0.905	0.912	0.920	0.975	0.699	0.944	0.941	–

$k = 27$  variants,  $a_0 = 0.1$ ,  $n$  cases and  $n$  controls. The significance level is at 0.05.

**Table 5.** Power based on 1,000 replicates of the phased-haplotype data for gene IL21R

<i>n</i>	OR	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum
500	1.2	0.121	0.140	0.134	0.139	0.076	0.143	0.123
	1.4	0.333	0.345	0.359	0.425	0.162	0.391	0.344
	1.5	0.447	0.448	0.473	0.524	0.228	0.488	0.430
1,000	1.2	0.202	0.231	0.219	0.247	0.137	0.220	0.198
	1.3	0.391	0.398	0.412	0.491	0.200	0.435	0.394
	1.4	0.567	0.562	0.576	0.632	0.284	0.578	0.530

$k = 19$  variants,  $a_0 = 0.1$ ,  $n$  cases and  $n$  controls. The significance level is at 0.05.

### HapMap Data for Gene IL21R

As shown by Chapman and Whittaker [2008] and Pan [2009], the Sum test did not perform well in the region of gene IL21R. We followed exactly the same steps as for gene CHI3L2 except that the disease-causing SNP was selected randomly and then excluded from the data in each simulation run. At the end, we had 28 SNPs.

The simulation results based on  $n = 500$ ,  $n = 1,000$  and 1,000 replicates are shown in table 4. As expected, the Sum test had low power, while the UminP test was most powerful. Nevertheless, the power of the aSum test (and its permutation-based version, aSum-P) was only slightly lower than that of the UminP test. Note that the naive naSum test had highly inflated type I error rates.

As suggested by a reviewer, we also downloaded phased haplotype data of the 60 HapMap CEU subjects. We followed a similar procedure as before (except that

there was no need to impute because there was no missing value), leading to 20 SNPs. The results for the two sample sizes based on 1,000 replicates are shown in table 5. The same conclusions can be drawn: the UminP test had the highest power, closely followed by the aSum-P and aSum tests; in particular, the aSum-P and aSum tests were much more powerful than the Sum test.

### Rare Variants

#### Simulation Setups

We independently generated  $k$  rare mutations, each with a mutation rate or MAF uniformly distributed between 0 and 0.05 with the constraint  $\sum_{j=1}^k \text{MAF}_j = 0.05$  or 0.01. The disease outcome was generated from a joint logistic regression model (1) with sample size  $n = 500$ . Several cases were considered. (i) Ideal case: with all  $\beta_1 = \dots = \beta_k = \log(\text{OR})$ ; (ii) nonideal:  $\beta_1 = \beta_2 = \beta_3 =$



**Table 6.** Type I error and power based on 1,000 replicates for an ideal case for rare variants with a common association strength OR between any of nine causal SNPs and the disease

OR	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	aSumC-P	aSumC	CMC	w-Sum
1.0	0.023	0.034	0.026	0.016	0.046	0.062	0.058	0.052	0.055	0.036	0.046
1.2	0.047	0.062	0.050	0.023	0.144	0.183	0.138	0.131	0.143	0.090	0.161
1.4	0.130	0.176	0.133	0.074	0.416	0.419	0.365	0.324	0.317	0.263	0.407
1.6	0.318	0.388	0.317	0.169	0.675	0.661	0.615	0.578	0.558	0.504	0.663
1.8	0.502	0.591	0.503	0.278	0.863	0.841	0.816	0.771	0.759	0.715	0.854
2.0	0.707	0.775	0.704	0.421	0.970	0.963	0.926	0.911	0.903	0.868	0.960

The sum of the MAFs is set to be 0.05,  $a_0 = 0.1$ ,  $n = 500$ . The significance level is at 0.05.

**Table 7.** Power based on 1,000 replicates for an ideal case for nine causal rare variants with a common association strength OR = 1.8 and a number of nonfunctional common variants (NF)

NF	MAF	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	aSumC-P	aSumC	CMC	w-Sum
0	–	0.502	0.591	0.503	0.278	0.863	0.841	0.816	0.771	0.759	0.715	0.854
1	0.02	0.511	0.493	0.506	0.308	0.807	0.769	0.739	0.712	0.705	0.671	0.795
1	0.05	0.510	0.286	0.510	0.309	0.668	0.629	0.592	0.640	0.619	0.670	0.647
2	0.02	0.464	0.411	0.457	0.293	0.678	0.629	0.577	0.633	0.590	0.618	0.669
2	0.05	0.465	0.191	0.467	0.292	0.499	0.507	0.466	0.537	0.508	0.616	0.507
4	0.02	0.444	0.355	0.445	0.310	0.536	0.540	0.499	0.584	0.542	0.553	0.520
4	0.05	0.444	0.140	0.453	0.310	0.348	0.386	0.341	0.463	0.467	0.548	0.338

The sum of the MAFs of the nine causal variants is set to be 0.05,  $a_0 = 0.1$ ,  $n = 500$ , and the significance level is at 0.05.

$\log(OR_p)$ ,  $\beta_4 = \beta_5 = \beta_6 = \log(0.5)$  and  $\beta_7 = \beta_8 = \beta_9 = \log(2)$ . In addition, for each case, we also considered adding a few independent and nonfunctional (i.e. noncausal) common variants or rare variants, and investigated the corresponding power properties of the various tests.

#### Simulation Results

For the ideal situation with all causal associations in the same direction, the simulation results based on  $n = 500$  and 1,000 replicates are shown in table 6. It is clear that the Sum test and the w-Sum test of Madsen and Browning [2009] performed best, closely followed by the aSum-P and aSum tests. Next came the aSumC-P and aSumC tests, which had higher power than the CMC test. Note that the naive naSum test had a hugely inflated type I error rate at 0.617.

When more nonfunctional (i.e. noncausal) common SNPs were added (table 7), the CMC test became the overall winner, closely followed by the aSumC-P and aSumC

tests when the MAF of the nonfunctional SNPs was low. On the other hand, when multiple nonfunctional rare variants were added (table 8), the Sum and aSum-P tests appeared to be the winners, closely followed by the aSum, w-Sum, aSumC-P and aSumC tests. Next came the CMC test.

For the nonideal situation with both positive and negative associations between the causal rare variants and the disease, the simulation results based on  $n = 500$  and 1,000 replicates are shown in table 9. Evidently our proposed aSum test and its permutation-based version aSum-P had higher power than the CMC, and even more significantly than the w-Sum test (and the Sum test). It is interesting to note that the SSU, SSUw and score tests also performed very well here.

When two common nonfunctional variants were added (table 10), the aSumC-P and aSumC tests outperformed the aSum-P and aSum tests but CMC had even higher power while both the Sum and w-Sum tests per-

**Table 8.** Power based on 1,000 replicates for an ideal case for nine causal rare variants with a common association strength  $OR = 1.8$  and a number of nonfunctional rare variants (NF), each with the same MAF as that of a corresponding causal variant

NF	sMAF	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	aSumC-P	aSumC	CMC	w-Sum
0	0.05	0.502	0.591	0.503	0.278	0.863	0.841	0.816	0.771	0.759	0.715	0.854
0	0.01	0.037	0.112	0.038	0.011	0.261	0.349	0.318	0.349	0.318	0.260	0.260
5	0.05	0.432	0.549	0.401	0.295	0.738	0.725	0.697	0.665	0.636	0.614	0.714
5	0.01	0.036	0.101	0.012	0.011	0.192	0.279	0.217	0.279	0.217	0.189	0.201
9	0.05	0.382	0.516	0.346	0.296	0.628	0.642	0.599	0.583	0.540	0.521	0.583
9	0.01	0.029	0.101	0.005	0.017	0.161	0.217	0.175	0.217	0.175	0.159	0.172

The sum of the MAFs of the nine causal variants is sMAF,  $a_0 = 0.1$ ,  $n = 500$ , and the significance level is at 0.05.

**Table 9.** Power based on 1,000 replicates for a nonideal case for rare variants: the effect sizes of the nine causal SNPs are (ORp, ORp, ORp, 0.5, 0.5, 0.5, 1.5, 1.5, 1.5)

ORp	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	aSumC-P	aSumC	CMC	w-Sum
1.0	0.232	0.278	0.231	0.156	0.099	0.217	0.193	0.235	0.241	0.210	0.104
1.4	0.292	0.359	0.293	0.183	0.123	0.332	0.309	0.333	0.304	0.248	0.123
1.8	0.423	0.506	0.428	0.247	0.212	0.489	0.424	0.462	0.447	0.356	0.213
2.2	0.585	0.642	0.585	0.365	0.326	0.609	0.578	0.585	0.585	0.520	0.317
2.6	0.713	0.754	0.710	0.486	0.438	0.717	0.684	0.710	0.675	0.625	0.426
3.0	0.806	0.826	0.806	0.586	0.522	0.787	0.759	0.752	0.757	0.722	0.513

The sum of the MAFs is set to be 0.05,  $a_0 = 0.1$ ,  $n = 500$ , and the significance level is at 0.05.

formed poorly. In these situations, it was somewhat surprising that the score and SSUw tests were the winners; note that because the SNPs were not in LD, the score and SSUw tests were expected to be close. When more nonfunctional rare variants were added (table 11), the various versions of the aSum tests performed similarly, better than the CMC test, and far better than the w-Sum test, although the SSU test was an overall winner. Interestingly, for a higher sum of the MAFs for nonfunctional rare variants, the score and SSUw tests performed well; in contrast, if the sum of the MAFs was low, the score and SSUw tests had minimal power.

*Example*

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease leading to paralysis and death, typically within 3–5 years from onset. Despite evidence for a role for genetics, little is known about common genetic variants linking to sporadic ALS. Schymick et al. [2007] conducted a GWAS to identify genetic variants predisposing to developing ALS in a cohort of 276

American sporadic ALS cases and 268 neurologically normal controls. The original study assayed 555,352 unique SNPs for each subject. By testing each individual SNP separately, Schymick et al. [2007] identified a list of 34 most significant SNPs, although none of them was statistically significant after a genome-wide multiple test adjustment.

We randomly picked nine SNPs from the list of 34 significant common SNPs of Schymick et al. [2007], including the most significant one, rs4363506, which had a  $p$  value of  $6.8 \times 10^{-7}$  by univariate 2-DF  $\chi^2$  test. For each of the nine SNPs, we extracted ten neighboring SNPs upstream and another ten downstream, then applied the default LD blocking algorithm implemented in Haploview (v4.1) [Barrett et al., 2005] to each 21-SNP region for the control group. Three SNPs were included in the LD block for SNP rs4363506, while four were included for SNP rs332389. As for any multimarker analysis, we had to choose an LD block [Gabriel et al., 2002] to analyze in a region of interest. The goal was to strike a balance between including as many informative SNPs as possible

**Table 10.** Power based on 1,000 replicates for a nonideal case for nine causal rare variants and a number of nonfunctional common variants (NF)

NF	MAF	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	aSumC-P	aSumC	CMC	w-Sum
0	–	0.806	0.826	0.806	0.586	0.522	0.787	0.759	0.752	0.757	0.722	0.513
1	0.02	0.751	0.713	0.755	0.574	0.424	0.655	0.611	0.624	0.595	0.583	0.403
1	0.05	0.749	0.489	0.760	0.577	0.343	0.552	0.510	0.545	0.521	0.584	0.335
2	0.02	0.712	0.659	0.707	0.570	0.371	0.599	0.543	0.601	0.573	0.566	0.362
2	0.05	0.718	0.376	0.719	0.568	0.269	0.458	0.405	0.494	0.454	0.567	0.255

The effect sizes of the nine causal SNPs are ( $OR_p$ ,  $OR_p$ ,  $OR_p$ , 0.5, 0.5, 0.5, 1.5, 1.5, 1.5) with  $OR_p = 3$ . The sum of the MAFs of the nine causal variants is set to be 0.05,  $a_0 = 0.1$ ,  $n = 500$ , and the significance level is at 0.05.

**Table 11.** Power based on 1,000 replicates for a nonideal case for nine causal rare variants and a number of nonfunctional rare variants (NF), each with the same MAF as that of a corresponding causal variants

NF	sMAF	Score	SSU	SSUw	UminP	Sum	aSum-P	aSum	aSumC-P	aSumC	CMC	w-Sum
0	0.05	0.806	0.826	0.806	0.586	0.522	0.787	0.759	0.752	0.757	0.722	0.513
0	0.01	0.075	0.211	0.076	0.030	0.140	0.220	0.181	0.220	0.181	0.139	0.141
5	0.05	0.714	0.787	0.693	0.608	0.428	0.668	0.626	0.645	0.606	0.570	0.420
5	0.01	0.067	0.178	0.031	0.029	0.131	0.184	0.152	0.184	0.152	0.127	0.137
9	0.05	0.654	0.746	0.620	0.616	0.348	0.574	0.525	0.593	0.548	0.541	0.343
9	0.01	0.059	0.167	0.010	0.044	0.121	0.172	0.141	0.172	0.141	0.117	0.125

The effect sizes of the nine causal SNPs are ( $OR_p$ ,  $OR_p$ ,  $OR_p$ , 0.5, 0.5, 0.5, 1.5, 1.5, 1.5) with  $OR_p = 3$ . The sum of the MAFs of the nine causal variants is sMAF,  $a_0 = 0.1$ ,  $n = 500$ , and the significance level is at 0.05.

**Table 12.** p values from various tests on two LD blocks with 3 and 4 SNPs, respectively

SNP	Chrom	Score	SSU	SSUw	UminP	Sum	aSum-P
rs4363506	10q26.13	$4.0 \times 10^{-5}$	$9.0 \times 10^{-6}$	$1.6 \times 10^{-5}$	$1.3 \times 10^{-6}$	0.0122	$1.8 \times 10^{-6}$
rs332389	3p14.1	0.0011	$1.7 \times 10^{-6}$	$6.4 \times 10^{-6}$	$1.0 \times 10^{-5}$	0.0108	$1.0 \times 10^{-6}$

and controlling for increasing DF. Here we chose an LD block often used in haplotype analysis, although it is not clear how to do so optimally; other strategies, such as using sliding windows with various window sizes [e.g. Guo et al., 2009 and references therein], may be adopted, although issues of high computational demand and multiple testing adjustments have to be addressed.

We applied various multimarker genotype-based tests to each block for each of the nine candidate SNPs. We used the general 2-DF coding for each SNP. For two candidate SNPs, rs4363506 and rs332389, the aSum test based on  $B = 100$  gave much more significant p values

than the other tests, while the differences for the other seven SNPs were not big (not shown). However, by checking the Q-Q plots, we had some doubts about the goodness-of-fit of the far right tail of the proposed theoretical null distribution. Hence, we used  $B = 5 \times 10^6$  for the aSum-P test (table 12). It is worth pointing out the dramatic p value differences between the Sum and aSum-P tests. Albeit not conclusive, the p values of the aSum-P test, which were more significant than those of the Sum test for the two SNPs, were in agreement with those of the UminP and other tests, suggesting possible follow-up studies on the two loci.

## Discussion

We have proposed and studied a data-adaptive aSum test for genetic association with multiple common and/or rare variants. As the Sum test, it aims to minimize the cost of high DF and eliminate the need for multiple testing adjustment while utilizing information on multiple variants; these multiple variants may be SNPs in LD in a sliding window or LD block, or may be unlinked rare variants. Its major advantages include its easy use, simplicity, and applicability to both common and rare variants. We used permutations to correct for the effects of the data-adaptive aSum test on the null distribution of the resulting score (or Wald) test statistic. We also discussed the use of an approximate null distribution to derive p values based on a small number of permutations to save computing time; however, the theoretical approximation may not work well, and further studies are needed. Therefore, whenever feasible, e.g. in candidate gene association studies or with parallel computing resources, we recommend the use of the permutation-based aSum-P test. We

conducted extensive simulations to evaluate its performance for common variants, establishing its superior performance over a wide range of scenarios. We also empirically compared its statistical power with other methods for rare variants, showing its potential use and improved performance in some situations. In anticipation of the future use of next-generation resequencing and other technologies to survey genetic architectures of both common and rare variants, further studies are needed to evaluate its performance for rare variants. In particular, to maximize power in practical situations, it would be desirable to incorporate automatic variant selection before applying the test.

## Acknowledgments

This research was partially supported by NIH grants GM081535 and HL65462. The authors thank the reviewers and Dr Weihua Guan for helpful comments.

## References

- Azzopardi D, Dallosso AR, Eliason K, Hendrickson BC, Jones N, et al: Multiple rare nonsynonymous variants in the adenomatous polypsis coli gene predispose to colorectal adenomas. *Cancer Res* 2008;68:358–363.
- Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS, The Type 1 Diabetes Genetics Consortium: Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*, Epub 2009.
- Bodmer W, Bonilla C: Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008;40:695–701.
- Chapman JM, Whittaker J: Analysis of multiple SNPs in a candidate gene or region. *Genetic Epidemiology* 2008;32:560–566.
- Clayton D, Chapman J, Cooper J: Use of unphased multilocus genotype data in indirect association studies. *Genet Epidemiol* 2004;27:415–428.
- Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, et al: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004;305:869–872.
- Conneely KN, Boehnke M: So many correlated tests, so little time! Rapid adjustment of p values for multiple correlated tests. *Am J Hum Genet* 2007;81:1158–1168.
- Cox DR, Hinkley DV: *Theoretical Statistics*. Chapman and Hall, London, 1974.
- Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES: High-resolution haplotype structure in the human genome. *Nat Genet* 2001;29:229–232.
- Fan R, Knapp M: Genome association studies of complex diseases by case-control designs. *Am J Hum Genet* 2003;72:850–868.
- Flint J, Mackay TFC: Genetic architecture of quantitative traits in mice, flies and humans. *Genome Research* 2009;19:723–733.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al: The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229.
- Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI: Shifting paradigm of association studies: Value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 2008;82:100–112.
- Guo Y, Li J, Bonham A, Wang Y, Deng H: Gain in power for exhaustive analyses of haplotypes using variable-sized sliding window strategy: a comparison of association-mapping strategies. *European Journal of Human Genetics* 2009;17:785–792.
- Haller G, Torgerson DG, Ober C, Thompson EE: Sequencing the IL4 locus in African Americans implicates rare noncoding variants in asthma susceptibility. *J Allergy Clin Immunol* 2009;124:1204–1209.e9.
- Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, et al: Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 2008;40:592–599.
- Kotowski I, Pertsemlidis A, Luke A, Cooper R, Vega G, Cohen J, Hobbs H: A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet* 2006;78:410–422.
- 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.
- Madsen BE, Browning SR: A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 2009;5:e1000384.
- Maher B: Personal genomes: the case of the missing heritability. *Nature* 2008;456:18–21.
- McCarthy MI: Exploring the unknown: assumptions about allelic architecture and strategies for susceptibility variant discovery. *Genome Med* 2009;1:66.
- Morgenthaler S, Thilly WG: A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). *Mutation Research* 2007;615:28–56.



- Pan W: Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol* 2009;33:497–507.
- Pan W, Han F, Shen X: Test selection with application to detecting disease association with multiple SNPs. *Hum Hered* 2010;69:120–130.
- Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001;69:124–137.
- Pritchard JK, Cox NJ: The allelic architecture of human disease genes: common disease-common variant... or not? *Hum Mol Genet* 2002;11:2417–2423.
- Roeder K, Bacanu SA, Sonpar V, Zhang X, Devlin B: Analysis of single-locus tests to detect gene/disease associations. *Genet Epidemiol* 2005;28:207–219.
- Satterthwaite FE: An approximate distribution of estimates of variance components. *Biometrics Bulletin* 1946;2:110–114.
- Schork NJ, Murray SS, Frazer KA, Topol EJ: Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 2009;19:212–219.
- Schymick JC, Scholz SW, Fung HC, et al: Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* 2007;6:322–328.
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, et al: Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008;320:539–543.
- Wang T, Elston RC: Improved power by use of a weighted score test for linkage disequilibrium mapping. *Am J Hum Genet* 2007;80:353–360.
- Wei Z, Li M, Rebbeck T, Li H: U-statistics-based tests for multiple genes in genetic association studies. *Ann Hum Genet* 2008;72:821–833.
- Wen G, Mahata S, Cadman P, Mahata M, Ghosh S, Mahapatra N, Rao F, Stridsberg M, Smith D, Mahboubi P, Schork NJ, O'Connor DT, Hamilton BA: Both rare and common polymorphisms contribute functional variation at CHGA, a regulator of catecholamine physiology. *Am J Hum Genet* 2004;74:197–207.
- Willer CJ, Speliotes EK, Loos RJF, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C, et al: Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009;41:25–34.
- Xiong M, Zhao J, Boerwinkle E: Generalized T2 test for genome association studies. *Am J Hum Genet* 2002;70:1257–1268.