Published in final edited form as:
Genet Epidemiol. 2011 November ; 35(7): 606–619. doi:10.1002/gepi.20609.

Comparison of Statistical Tests for Disease Association with Rare Variants

Saonli Basu and **Wei Pan**
Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455

Abstract

In anticipation of the availability of next-generation sequencing data, there is increasing interest in investigating association between complex traits and rare variants (RVs). In contrast to association studies for common variants (CVs), due to the low frequencies of RVs, common wisdom suggests that existing statistical tests for CVs might not work, motivating the recent development of several new tests for analyzing RVs, most of which are based on the idea of pooling/collapsing RVs. However, there is a lack of evaluations of, and thus guidance on the use of, existing tests. Here we provide a comprehensive comparison of various statistical tests using simulated data. We consider both independent and correlated rare mutations, and representative tests for both CVs and RVs. As expected, if there are no or few non-causal (i.e. neutral or non-associated) RVs in a locus of interest while the effects of causal RVs on the trait are all (or mostly) in the same direction (i.e. either protective or deleterious, but not both), then the simple pooled association tests (without selecting RVs and their association directions) and a new test called kernel-based adaptive clustering (KBAC) perform similarly and are most powerful; KBAC is more robust than simple pooled association tests in the presence of non-causal RVs; however, as the number of non-causal CVs increases and/or in the presence of opposite association directions, the winners are two methods originally proposed for CVs and a new test called C-alpha test proposed for RVs, each of which can be regarded as testing on a variance component in a random-effects model. Interestingly, several methods based on sequential model selection (i.e. selecting causal RVs and their association directions), including two new methods proposed here, perform robustly and often have statistical power between those of the above two classes.

Keywords

C-alpha test; kernel machine regression; logistic regression; model selection; permutation; pooled association tests; random-effects models; SSU test; Sum test; statistical power

INTRODUCTION

Genome-wide association studies (GWASs) have successfully identified thousands of common genetic variants, mainly common single nucleotide variants (SNVs), associated with complex traits, including many common diseases (Hindorff et al 2010). However, these identified variants can only explain a small proportion of inheritable phenotypic variance (Maher 2008), leaving the door open for many more yet to be discovered variants. A popular hypothesis is that many more rare variants (RVs) may contribute to the missing heretability unexplained by discovered common variants (CVs) (Bodmer and Bonilla 2008; Gorlov et al

Correspondence author: Wei Pan, Telephone: (612) 626-2705, Fax: (612) 626-0660, weip@biostat.umn.edu, Address: Division of Biostatistics, MMC 303, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455–0392, U.S.A.

2008; Pritchard 2001; Pritchard and Cox 2002). At the same time, biotechnological advances have made it feasible to re-sequence parts of or whole genomes.

In anticipation of the arrival of massive amounts of next-generation sequencing data, the chance of success in detecting association between complex traits and RVs largely depends on statistical analysis strategies for RVs; see two excellent timely reviews (Asimit and Zeggini 2010; Bansal et al 2010). Since frequencies of RVs are very low, even with high penetrance, it will be difficult to detect association with any single RV. Hence, the most popular statistical test for GWAS based on testing single SNVs is not expected to perform well. In fact, in light of the significant difference in variant frequencies between RVs and CVs, common wisdom might suggest that many existing methods for CVs would not work either, motivating the development of new statistical tests specifically targeting RVs. The most striking feature of several recently proposed new tests for RVs is the idea of pooling or collapsing: rather than testing on individual SNVs one by one (as in GWASs), one would pool or collapse multiple rare SNVs together such that collectively they would have a reasonably high frequency, and then apply a test to the collapsed genotype (Morgenthaler and Thilly 2007; Li and Leal 2008; Madsen and Browning 2009; Price et al 2010). Albeit well motivated and shown to perform better than single SNV-based testing, such a pooling strategy has its own limitations. If the RVs to be pooled are associated with the trait in different directions, i.e. some are associated positively while others negatively, the strategy of pooling may weaken or diminish the signal in associated RVs. Furthermore, if many of the RVs are non-causal, i.e. they are not associated with the traits, pooling will inevitably introduce noises into the collapsed genotype and thus have reduced statistical power. Note that, the effects of RVs are not always in the same direction: they can be protective or deleterious. For example, some RVs in gene PCSK9 are associated with lower plasma levels of low-density lipoprotein cholesterol (LDL-C) while others associated with higher levels of LDL-C (Kotowski et al 2006). In recognition of these limitations, several methods based on model selection have been proposed recently (Han and Pan 2010; Hoffmann et al 2010; Bhatia et al 2010; Zhang et al 2010). The main idea is to determine whether a RV should be pooled, and if so, what is its association direction. Since these methods are based on either a marginal test or a step-up procedure on each individual RV, the power of selecting a RV and determining its association direction may be limited. Here we propose two new model-selection procedures that improve over the existing pooled association tests while maintaining low computational cost, borrowing the idea of Basu et al (2010) in linkage analysis.

Very recently several new tests, including a kernel-based adaptive clustering (KBAC) (Liu and Leal 2010), a C-alpha test (Neale et al 2011) and a replication-based test (RBT) (Ionita-Laza et al 2011), specifically designed for RVs and aiming to overcome various weaknesses of the pooled association tests, have appeared. However, no comparison was made among these new tests and model-selection approaches for RVs. More generally, in the current literature, there is no evaluation on the applicability of most existing tests to RVs. Although most existing tests have been proposed for and mainly applied to CVs, some were originally developed for high-dimensional data and thus are likely to be robust to the large number of parameters facing the analysis of RVs, and may have reasonable power for RVs. Goeman's score test (Goeman et al 2006) and kernel machine regression (KMR) (Liu et al 2008) are two such examples. Since Goeman's test is permutation-based and is equivalent to a test called the sum of squared score (SSU) test (Pan 2009), we consider the SSU test here. As to be shown, perhaps surprisingly, both the SSU test and KMR, along with the C-alpha test specifically proposed for RVs (Neale et al 2011), performed extremely well under certain situations when the pooled association tests had low power. In summary, given the compelling interest of the scientific community in detecting association between complex traits and RVs while little is known about the relative performance and merits of various existing and new tests, it is timely to have a comparative evaluation of the tests, an endeavor taken here.

METHODS

To be concrete, we restrict the attention to the case-control design with a binary trait, say disease, though many of the methods discussed are based on logistic regression and can be easily extended to generalized linear models (GLMs) for other types of traits. We do not consider adjusting for covariates, such as environmental factors, though again methods based on logistic regression can easily accommodate covariates. We assume that the analysis goal is to detect whether there is any association between the disease and a group of rare SNVs, for example, SNVs in a sliding window or in a functional unit such as gene. We denote the binary trait $Y_i = 0$ for n_0 controls, and $Y_i = 1$ for $n_1 = n - n_0$ cases. The k variants are coded by an additive genetic model: $X_{ij} = 0$, 1 or 2 for the number of the rare variant (minor allele) for SNV j, $j = 1, \ldots, k$.

Methods originally proposed for common variants

Logistic regression—Several most popular statistical tests are based on logistic regression:

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

The null hypothesis to be tested is H_0: $\beta = (\beta_1, \ldots, \beta_k)' = 0$. Maximum likelihood can be utilized to derive asymptotically equivalent score test, Wald's test and likelihood ratio test (LRT); here we focus on the score test for its computational simplicity. For model (1), the score vector and its covariance matrix are

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

where $X_i=(X_{i1},\ldots,X_{ik})'$, $\overline{Y}=\sum_{i=1}^{n} Y_i/n$ and $\overline{X}=\sum_{i=1}^{n} X_i/n$.

1 The most popular test for CVs in GWAS is the (univariate) minP (UminP) method that tests on each single SNVs one-by-one and then takes the minimum of their p-values. The corresponding UminP score test statistic is

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

where U_j is the jth element of U and V_{jj} is the (j, j)th element of V. An adjustment for multiple testing has to be made. Although the Bonferroni and permutation methods are most commonly used, a better way is to derive the null distribution of T_{UminP} and thus a p-value based on numerical integration with respect to a multivariate Gaussian density (Conneely and Boehnke 2007).

A joint test as an alternative to the UminP test is the multivariate score test:

2

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

which has an asymptotic chi-squared distribution with degrees of freedom (DF) k. If DF k is large, the test may not have high power.

Pan (2009) proposed two tests, called sum of squared score (SSU) and sum of weighted squared score (SSUw) tests:

3

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

where Diag(V) is a diagonal matrix with the diagonal elements of V. Under H_0, each of the two test statistics has an asymptotic distribution of a mixture of χ_1^2's, which can be approximated by a scaled and shifted chi-squared distributions (Pan 2009). The two tests can be regarded as modified score test by ignoring the non-diagonal elements of V, i.e. correlations among the components of U, which is known to be advantageous for high-dimensional data (Chen and Qin 2010). More importantly, as shown by Pan (2009), the SSU test is equivalent to the permutation-based version of Goeman's (2008) test, which is derived as a variance component score test for a random-effects (R-E) logistic regression model. Specifically, in model (1), if we assume β_j's as random effects drawn from a distribution with $E(\beta) = 0$ and $Cov(\beta) = \tau\Sigma$, then Goeman's score test on H_0: $\tau = 0$ is

$$S = \frac{1}{2} U' \Sigma U - \frac{1}{2} \text{tr}(\Sigma U), \quad (2)$$

← test variance component

where tr(A) is the trace of matrix A. Observing that V is invariant to permutations of Y, we know that, under permutations, using S is equivalent to using $S_P = U'\Sigma U$, which is equivalent to the SSU, SSUw and score test statistics with $\Sigma = I$, $\Sigma = \text{Diag}(V)$ and $\Sigma = V$, respectively. Note that Goeman's test was originally derived to test on a large number of parameters for high-dimensional microarray data, though its good performance for lower-dimensional SNV data have been empirically confirmed too (Chapman and Whittaker 2008; Pan 2009).

Another test performed well under certain situations for CVs is the so-called Sum test, as noted by Chapman and Whittaker (2008) and Pan (2009). The Sum test was motivated to strike a balance between jointly testing on multiple SNVs and its resulting DF. The Sum test is based on a key and generally incorrect working assumption that the SNVs are all associated with the trait with a common association strength:

4.

$$\text{Logit Pr}(Y_i = 1) = \beta_{c,0} + \sum_{j=1}^{k} X_{ij} \beta_c, \quad (3)$$

where β_c reflects the common odds ratio (OR) between the trait and each SNV under the working assumption. While utilizing all the SNVs, the Sum test avoids the possibly too large DF, and thus loss of power, of other multivariate tests. It only requires to test on a single parameter with H_0: $\beta_c = 0$ by a score test (or its asymptotically equivalent Wald's test or LRT). Pan (2009) pointed out that the weighted score test of Wang and Elston (2007) share the same spirit and thus similar performance as the Sum test. Note that in model (3) we regress Y on a new "super-SNV" that is the sum of the genotype values of all the SNVs, hence we call the resulting test Sum test.

Σ·X

Weighted sum

To incorporate prior biological information, one may want to weight SNVs using some suitable weights, e.g. based on their MAFs (Madsen and Browning 2009) or their predicted likelihoods of being functional (Price et al 2010). It is straightforward to do so in logistic regression: with a set of weights $w = (w_1, \ldots, w_k)'$, we can simply weight the codings for SNVs; that is, rather than using $X_i = (X_{i1}, \ldots, X_{ik})'$ for subject i, we use $X_{i,w} = (w_1X_{i1}, \ldots, w_kX_{ik})'$ in logistic regression model (1). It is easy to see that, the UminP, score and SSUw tests are invariant to such weighting, while the SSU and Sum tests do depend on such weighting. In fact, a careful examination of the SSU and Sum test statistics indicates that the above two tests treat X_{ij}'s more or less equally across j. By the expression of V, we see that those variants with larger MAFs tend to have larger variances for their components of the score vector. Hence, without weighting, the SSU and Sum tests essentially give heavier weights to the variants with larger MAFs, implying that they will be sensitive to the presence of non-causal CVs, as to be confirmed.

To overcome the above weakness of the SSU test, a simple strategy is to weight each variant j inversely by its sample standard deviation $SD(X_{1j}, \ldots, X_{nj})$, which is equivalent to standardizing each predictor j to have a sample SD=1. The resulting SSU test is essentially the same as the SSUw test. This point highlights a key difference between the SSU and SSUw tests, illustrating when one of the two is more powerful than the other. For example, if the causal variants tend to have lower MAFs than that of non-causal ones, the SSU test is expected to be less powerful; otherwise, the SSU test is more powerful. A potential problem with the above weighting scheme (and with the SSUw test) is that, since a causal variant may have a higher MAF in cases but a lower MAF in controls, and thus a higher overall MAF across both cases and controls, it will downweight this causal variant, leading to reduced power. This is a reason that Madsen and Browning (2009) proposed using the MAFs of only controls to construct weights. Specifically, if there are n_j^0 minor alleles for variant j in all the controls, then we can use weights

$$w_j=1/\sqrt{nq_j(1-q_j)}, \quad q_j=(n_j^0+1)/(2n_0+2). \tag{4}$$

→ 只用control的算 q_j

With such weights, which already use the disease labels, the asymptotic SSU test (and Sum test) would have inflated Type I error rates. Alternatively, we use a standard permutation to calculate p-values, and denote the resulting test wSSU-P.

Logistic kernel machine regression and genomic similarity-based methods— Rather than testing the effects of the SNVs parametrically (i.e. linearly in our specified model (1)), one can adopt a nonparametric model:

$$\text{Logit}\ \Pr(Y_i=1)=\beta_0+h(X_{i1},\ldots,X_{ik}), \tag{5}$$

where $h(.)$ is an unknown nonparametric function to be estimated, offering the flexibility in modeling the effects of the SNVs on the trait. In a specific approach called kernel machine regression (KMR) (Liu et al 2008), the form of $h(.)$ is determined by a user-specified positive and semi-definite (psd) kernel function $K(X_i.,X_j)$, which measures the genomic similarity between the genotypes of subjects i and j. Some commonly used kernels include linear, identity-by-descent (IBS) and quadratic kernels. By the representer theorem

IBS

(Kimeldorf and Wahba 1971), $h_i=h(X_i)=\sum_{j=1}^{n}\gamma_jK(X_i,X_j)$ with some $\gamma_1,\ldots,\gamma_n$. To test the null hypothesis of no association between the phenotype and SNVs, one can test H_0: $h = (h_1(X_1), \ldots, h_n(X_n))' = 0$. Denote K as the $n \times n$ matrix with the (i, j)th element as $K(X_i,X_j)$ and $\gamma = (\gamma_1, \ldots, \gamma_n)'$, then we have $h = K\gamma$. Treating h as subject-specific random effects with mean 0 and covariance matrix τK, testing H_0: $h = 0$ for no SNV effects is equivalent to testing H_0: $\tau = 0$. The corresponding variance component score test statistic is

$$Q=(Y-\overline{Y}1)'K(Y-\overline{Y}1),$$

whose asymptotic null distribution is a mixture of χ_1^2's, which can be approximated by a scaled chi-squared distributions (Wu et al 2010).

The above logistic KMR can be extended to include other covariates and for other traits, e.g. linear models for quantitative traits (Kwee et al 2008). Since the kernel function measures the similarity of two genotypes, KMR is expected to be related to genomic-distance based regression (GDBR) of Wessel and Schork (2006); see Schaid (2010 a, b) for a review on the topic. More specifically, as shown by Pan (2011), both KMR and GDBR are equivalent to the SSU test on H_0: $b = 0$ in a new logistic regression model:

$$\text{Logit Pr}(Y=1)=b_0+Zb, \tag{6}$$

where $K = ZZ'$. Hence, the difference between the SSU test for model (1) and logistic KMR is only in the transformation of SNV codings in model (6), while both tests are actually an SSU test applied to two different regression models. A special case is that, for a linear kernel K, we have $K = XX'$; that is, $Z = X$, under which the SSU and KMR are equivalent.

Empirically it has been found that GDBR and KMR performed very well in detecting disease association with CVs (Lin and Schaid 2009; Wu et al 2010; Han and Pan 2010b). Albeit proposed for and mainly applied to CVs, first Wessel and Schork (2006), and more recently Bansal et al (2010) commented that GDBR (and thus KMR) could be applied to sequence data to detect association with RVs.

To our knowledge, the above statistical tests originally proposed for CVs have never been applied to RVs. Intuition might argue against their application to RVs. However, as to be shown, perhaps quite surprisingly, some of them performed quite well in our numerical studies. We will offer some explanations in Discussion.

Methods for rare variants

1.

Pooled association tests—The first test specifically designed for RVs is perhaps the cohort allelic sums test (CAST) (Morgenthaler and Thilly 2007). CAST works by first collapsing the genotypes across RVs to generate a "super-variant": $X_{i,C} = 1$ 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 trait and this new $X_{i,C}$. It can be regarded as fitting a logistic regression model

$$\text{Logit Pr}(Y_i=1)=\beta_{C,0}+X_{i,C}\beta_C, \tag{7}$$

1.2

都类通

sum test.

and testing H_0: $\beta_C = 0$. The most striking feature of CAST, as the Sum test, is its testing on a single parameter, thus low DF and possibly increased power.

As pointed out by Han and Pan (2010), the CAST is closely related to the Sum test: both test on only a single parameter representing some average effect of the multiple SNVs. They differ in their coding of the “super-variant”: $X_{i,C} = \vee_{j=1}^{k} X_{ij}$ versus $X_{i,S} = \sum_{j=1}^{k} X_{ij}$, similar to the use of a dominant genetic model versus an additive genetic model for the effect of an individual variant. Note for rare variants, we have $X_{i,C} \approx X_{i,S}$. Other codings for the “super-variant” are also possible, as considered by Morris and Zeggini (2010).

2

Li and Leal (2009) proposed a new test called Combined Multivariate and Collapsing (CMC) test, which modifies the CAST to improve its performance when both rare and common variants are present. Specifically, for any rare mutations with their minor allele frequencies (MAFs) less than some threshold, say 0.05, they will be combined into a new group as in the CAST, while each common variant (e.g. with $MAF > 0.05$) forms its own group, and the generalized Hotelling’s test (Fan and Knapp 2003; Xiong et al 2002) is applied to such formed multiple groups. Note that the generalized Hotelling’s test is closely related to the score test in logistic regression (Clayton et al 2004). Hence, for only RVs, the CMC test is essentially the same as the CAST (and the Sum test).

3

The weighted sum (w-Sum) test of Madsen and Browning (2009) is also based on the idea of collapsing RVs. It differs from the Sum test in i) using a weighted sum, instead of a simple sum, of RVs by their MAFs, and ii) comparing the ranks of the weighted sums, rather than the sums themselves, between the case and control groups. Hence, putting aside the difference in weighting, the wSum test is analogous to the Mann-Whitney-Wilcoxon rank test, while other pooled association tests are analogous to the t-test.

The main advantage of the above pooled association tests is their minimum DF at 1, hence no loss of power due to large DF or multiple testing adjustment. However, as pointed out by Han and Pan (2010), they all share a common weakness: they suffer from possibly significant power loss if the association directions of the causal variants are opposite. This can be most clearly seen from the Sum test. Generally, the common association parameter β_c in (3) can be viewed as a weighted average of the individual $\beta_1, \ldots, \beta_k$; see a closed-form expression for $\hat{\beta}_c$ for linear regression given in Pan (2009). Hence, depending on the signs of $\beta_1, \ldots, \beta_k$, $|\beta_c|$ may be very small, leading to loss of power in the Sum test. To overcome this limitation, several methods based on model selection have been proposed, as to be presented next.

Methods based on model selection—A general model has been proposed by Hoffmann et al (2010):

1 ✓

$$\text{Logit}\ \Pr(Y_i=1)=\beta_{c0}+\sum_{j=1}^{k}\gamma_j X_{ij}\beta_c, \qquad (8)$$

with $\gamma_j = w_j s_j$, where w_j is a weight assigned to SNV j, $s_j = 1$ or -1 indicating whether the effect of SNV j is positive or negative, and $s_j = 0$ indicating the exclusion of SNV j from the model (i.e. the SNV is unlikely to be associated with the trait). Madsen and Browning (2009) suggested to weight RVs with the weights depending on their MAFs. However, it is still debatable on how to appropriately weight the SNVs, and if needed, it is not difficult to incorporate a weighting scheme into most methods discussed here. Hence we do not discuss the use of weights and always assume $w_j = 1$ for any test except the wSum test.

The pooled association tests correspond to fixing $s_j = 1$ for all j’s. Several existing model-selection-based methods can be classified into one of the two classes:

1. Choosing $s_j = 1$ or -1 in a data-dependent manner. Han and Pan (2010) proposed an adaptive Sum (aSum) test, in which the value of each s_j is determined based on a

univariate test on the marginal association between the trait and SNV j for $j = 1, \ldots, k$.

2. Choosing $s_j = 1$, 0 or −1 in a data-dependent manner. A Step-up procedure (Hoffmann et al (2010) and a covering method (called RareCover) (Bhatia et al 2010) have been proposed to determine the value of s_j 's, both in a manner of forward variable selection: starting from a null model without any SNV, SNVs are selected one by one based on their statistical significance and then added into the model.

Here we propose two new methods, both of which start from the Sum test with all $s_j = 1$. The main motivation is that, since the individual effect of each RV is hard to detect while the Sum test (or any other pooled association test) has proven useful for RVs, rather than starting from a null model (as in the Step-up and the RareCover procedures) or testing on marginal association (as in the aSum test), we would like to start from the Sum test and make any necessary adjustment on the values of s_j's, which may result in higher power. In the first method, called Sequential Sum test (Seq-Sum), for each SNV j with $j = 1$ and increased to k, we determine which of the two models, the current model with $s_j = 1$ and the other model with $s_j = -1$ (while all other s_j's fixed at their current values for both models), is preferred based on which model yields a larger (maximized) likelihood; then we increase j by one and repeat the above process until we have tried $j = 1, \ldots, k$. In the second method, called Sequential Sum test with variable selection (Seq-Sum-VS), starting from SNV $j = 1$, we consider three models with $s_j = 1$, 0, and −1 respectively, and choose the model with the largest (maximized) likelihood; then we increase j by 1, and repeat the above process until having tried $j = 1, \ldots, k$ sequentially. Hence Seq-Sum considers only the coding of each SNV (i.e. its protective or harmful effect), while Seq-Sum-VS considers selecting both SNVs and their association directions. It is noted that the two methods consider only a total of k +1 and 2k +1 candidate models respectively. Due to the nature of their sequential search and dependence on the order of the SNVs, unlike the Step-up and CoverRare procedures, it is unlikely that they will select the best model (in terms of the largest maximized likelihood). Nevertheless, there are two possible benefits. One is the obviously reduced computational cost when compared to an exhaustive search for exponentially many (i.e. 2^k and 3^k) models. The second benefit is less obvious: there is also lower cost for multiple testing adjustment due to a reduced number of model comparisons. Computationally, rather than using the maximized likelihood as the criterion to select models, which requires fitting each model by an iterative algorithm to obtain the maximum likelihood estimates, we adopt a score test, which is computationally much faster. The proposed Seq-Sum test is closely related to a new adaptive Sum test of Pan and Shen (2011), which is more flexible while overcoming a weakness of the Seq-Sum method, namely, its dependence on an arbitrary ordering of the SNVs.

In general, it is difficult to analytically derive the null distribution of a test statistic after model selection. For each procedure above, we use permutations to calculate p-values.

Kernel-based adaptive clustering—Liu and Leal (2010) proposed a method called kernel-based adaptive clustering (KBAC) for RV association testing. KBAC works by
2 grouping/clustering mutation patterns across the variants, and assigning each mutation pattern a kernel-based weight adaptively determined by data. Specifically, suppose that among the cases and controls, we have $M + 1$ mutation patterns across all k variants, denoted as $G_0, G_1, \ldots, G_M$, where G_0 represents the wild-type without any mutation. We also assume that there are $n_{1,i}$ cases and $n_{0,i}$ controls with mutation pattern G_i; denote $n_{.i} = n_{1,i} + n_{0,i}$. For mutation pattern G_i, the risk of having disease is estimated as $R_i = n_{1,i}/n_{.i}$. The KBAC test statistic is

$$T_{KBAC}=\left(\sum_{i=1}^{M}(n_{1,i}/n_1-n_{0,i}/n_0)w_i\right)^2,$$

where the weight w_i is determined by a hyper-geometric kernel:

$$w_i=\int_0^{R_i}k_i^0(r)dr=\sum_{r\in\{\frac{0}{n_{.i}},\frac{1}{n_{.i}},\ldots R_i\}}\frac{C(n_{.i},n_{.i}r)C(n-n_{.i},n_1-n_{.i}r)}{C(n,n_1)}$$

with $C(a, b)$ as the combination number of choosing b out of a. The p-value is calculated by standard permutations (for small samples while a Monte Carlo approximation is used for large samples).

From the expression of T_{KBAC}, we see that its performance may deteriorate in the presence of both protective and harmful causal variants: some positive and negative components $(n_{1,i}/n_1 - n_{0,i}/n_0)w_i$ may cancel out with each other in the sum, though the use of weight w_i may alleviate the problem. A simple modification as shown below may help overcome the problem:

$$T_{mKBAC}=\sum_{i=1}^{M}(n_{1,i}/n_1-n_{0,i}/n_0)^2w_i,$$

though we do not pursue it here. In addition, KBAC includes non-causal variants in forming mutation patterns, which may dramatically increase the number of mutation patterns (M) and thus effectively reduce the group sizes $n_{.i}$'s, leading to loss of power. Nevertheless, the KBAC test is attractive in detecting possible interactions among the variants, though we do not pursue this issue here.

3rd test 3

C-alpha test—Neale et al (2011) proposed using the C-alpha test of Neyman and Scott (1966). It is based on testing for a common value (i.e. homogeneity) for a set of binomial proportions, not on logistic regression.

For SNV j, assume there are n_j subjects with the rare mutation (or minor allele); among those n_j subjects, we have m_j cases with mutation (and $n_j - m_j$ controls with mutation). We assume $m_j \sim Bin(n_j, p_j)$. Under the null hypothesis of no association between the disease and SNV j, we have $p_j = p_0$ for some common p_0 for all $j = 1, \ldots, k$. For a case-control study as considered here, we have $p_0 = 1 - n_0/n$. The C-alpha test is based on the following:

$$T_c=\sum_{j=1}^{k}T_{c,j}=\sum_{j=1}^{k}(m_j-n_jp_0)^2-n_jp_0(1-p_0),$$

$$V_c=\sum_{j=1}^{k}Var(T_{c,j})=\sum_{j=1}^{k}E[(m_j-n_jp_0)^2-n_jp_0(1-p_0)]^2,$$

where

$$Var(T_{C,j})=\sum_{u=0}^{n_j}[(u-n_jp_0)^2-n_jp_0(1-p_0)]^2f(u|n_j,p_0)$$

and $f(u|u_j, p_0)=C(n_j, u)p_0^u(1-p_0)^{n_j-u}$ is the binomial probability $Pr(U = u)$ for $U \sim Bin(n_j, p_0)$. *If all m_j's are independent*, then under the null hypothesis of no association between any SNV and the disease, the test statistic $Z=T_C/\sqrt{V_C}$ has an asymptotic distribution of $N(0, 1)$, from which a p-value can be calculated. Alternatively, one can permute the disease labels Y, calculate Z's for permuted data and thus a p-value. We denote the two versions of the tests using the asymptotic distribution and the permutation distribution respectively as C-alpha-A and C-alpha-P.

The C-alpha test treats SNV-specific mutation rates p_j's as a random sample drawn from some common distribution, say G. Under H_0, the distribution reduces to a point mass at p_0. Hence, the C-alpha test can be regarded as testing on the variance component of G: the variance of p_j's is 0 under H_0. The C-alpha test is a score test for such a homogeneity problem (Zelterman and Chen 1988), bearing some similarity to the framework of the variance component testing for a R-E model, under which the SSU and KMR can be formulated. In fact, as shown in Appendix, the general homogeneity score test of Zelterman and Chen (1988) has the same form of Goeman's test.

Each component of the C-alpha test statistic, $T_{C,j}$, contrasts the sample variance for variant j with its theoretical variance under H_0. Since the 4th central moment is

$$E(m_j-n_jp_0)^4=3(n_jp_0q_0)^2+n_jp_0q_0(1-6p_0q_0)$$

with $q_0 = 1- p_0$, under H_0, we have

$$Var(T_{C,j})=2(n_jp_0q_0)^2+n_jp_0q_0(1-6p_0q_0),$$

which is an increasing function of n_j. Thus, similar to the SSU test (and KMR), since the C-alpha test statistic is a simple sum of the statistics for the variants, $T_C=\sum_{j=1}^k T_{C,j}$, it may be dominated by the variants with large $Var(T_{C,j})$, e.g. those with high MAFs; it is possible, and even productive, to weight the components suitably with a set of weights w_j's to yield a weighted version of the C-alpha test:

$$T_{C,w}=\sum_{j=1}^k w_jT_{C,j}.$$

As to be shown, similar to the SSU test, the C-alpha test does not perform well in the presence of non-causal CVs, in which case its weighted versions are more powerful. We can use weights w_j as shown in (4), or $w_j=1/\sqrt{Var(T_{C,j})}$, and calculate their-p-values using permutations; we denote the resulting tests as w1C-alpha-P and w2C-alpha-P respectively. Since $V\,ar(T_{C,j}) = 0$ and 0.25 for $n_j = 1$ and $n_j = 2$ respectively, we define $w_j = 1/0.5$ for $n_j = 1$, the same as w_j for $n_j = 2$, in the w2C-alpha-P test. As to be shown, the two weighted C-alpha tests did not perform as well as the wSSU-P test, and will be skipped in most of simulations.

4

Replication-based test—Ionita-Laza et al (2011) proposed a new test called replication-based test (RBT). The RBT is similar to a pooled association test but purposefully designed to deal with possibly different association directions. In addition, a new weighting scheme is adopted to improve power. Using the same notation as before, suppose that for variant *j* there are n_j mutations in cases and $m_j - n_j$ mutations in controls. Define a statistic to measure the enrichment of mutations in cases:

$$S_+ = \sum_{j=1}^{k} I(n_j > m_j/2) w(n_j, m_j)$$

with weight

$$w(n_j, m_j) = -\log Pr(n_j, m_j) = -\log (ppois(m_j - n_j, m_j/2)[1 - ppois(n_j - 1, m_j/2)]),$$

where *ppois*(*a*, *b*) is the cumulative distribution function of a Poisson distribution *Pois*(*b*) evaluated at *a*. Similarly we measure the enrichment of mutations in controls with

$$S_- = \sum_{j=1}^{k} I(n_j < m_j/2) w(m_j - n_j, m_j).$$

The final test statistic is $T_R = \max(S_+, S_-)$. The p-value is calculated by permutations

Although the RBT was designed to differentiate between protective and harmful variants, it treats and tests the two groups separately, hence may lose power. Furthermore, for a non-causal variant *j*, it is likely that $n_j \neq m_j/2$, under which case non-causal variant *j* will be pooled over into the test statistic, though its weight w_j may be relatively small; nonetheless, the RBT may lose power in the presence of a large number of non-causal RVs.

A summary—We compare the above tests in several aspects as shown in Table 1. We do not include CAST and CoverRare since they are similar to CMC and Step-up respectively. We note that the wSum test uses permutations to estimate the mean and variance of its asymptotic Normal distribution, and does not need a large number of permutations to reach high statistical significance, which is required by other permutation-based tests. We also note that the CMC was proposed to use the generalized Hotelling's test, which does not accommodate covariates and other types of traits as shown in Table 1. However, since Hotelling's test is equivalent to the score test in logistic regression (Clayton et al 2004), it is easy to generalize the CMC test to accommodate covariates and other types of traits if the score test in a GLM is adopted. Finally, we will call the Sum, CMC and wSum tests loosely as the pooled association tests (that do not consider selecting SNVs and their association directions).

Simulated data

We generated simulated data as in Wang and Elston (2008) and Pan (2009). Specifically, we simulated *k* SNVs with the sample size of 500 cases and 500 controls. Each RV had a mutation rate or MAF uniformly distributed between 0.001 and 0.01, while for a CV it was between 0.01 and 0.1. First, we generated a latent vector $Z = (Z_1, \ldots, Z_k)'$ from a multivariate normal distribution with a first-order auto-regressive (AR1) covariance structure: there was an correlation $Corr(Z_i, Z_j) = \rho^{|i-j|}$ between any two latent components. We used $\rho = 0$ and $\rho = 0.9$ to generate (neighboring) SNVs in linkage equilibrium and in linkage disequilibrium

(LD) respectively. Second, the latent vector was dichotomized to yield a haplotype with MAFs each randomly selected. Third, we combined two independent haplotypes and obtained genotype data $X_i = (X_{i1}, \ldots, X_{ik})'$. Fourth, the disease status Y_i of subject i was generated from the logistic regression model (1). For the null case, we used $\beta = 0$; for non-null cases, we randomly selected 8 non-zero components of β while the remaining ones were all 0. Fifth, as in any case-control design we sampled 500 cases and 500 controls in each dataset.

We considered several simulation set-ups. Throughout the simulations, we fixed the test significance level at $\alpha = 0.05$ (or $\alpha = 0.01$ in a few cases), and used 500 permutations for each permutation-based method. The results were based on 1000 independent replicates for each set-up.

We used the R code of Wu et al (2010) implementing the KMR methods. We used the linear, IBS and quadratic kernels; since the first two performed similarly across all simulations, we present results for the linear and quadratic kernels. We used the R package `thgenetics` implementing the Step-up procedure, and a C++/R implementation of KBAC. We implemented all other tests in R. For the CMC test, we used the default cut-off of MAF ≤ 0.05 for RVs, though we explored using the cut-off ≤ 0.01 in a few cases.

RESULTS

Independent RVs

We first consider that there is no linkage disequilibrium (LD) between any two RVs, mimicking the situation where mutations are all completely random and independent of each other. To investigate the possible dependence of performance on the significance level, we used both $\alpha = 0.05$ and $\alpha = 0.01$.

Table 2 shows that all the tests had satisfactory Type I error rates except that the C-alpha-A test might have some inflated Type I error rates at $\alpha = 0.01$, suggesting that perhaps a larger sample size is needed for using its asymptotic distribution with a more stringent significance level.

For power comparison, the overall conclusions are the same with either $\alpha = 0.05$ and $\alpha = 0.01$. First, for the non-null case that the eight causal RVs shared a common OR (Table 3), which is ideal for the pooled association tests (Sum, CMC, wSum), the pooled association tests and KBAC were most powerful if there were no or few non-causal RVs (i.e. RVs not associated with the trait). As the number of non-causal RVs increased, the SSU and KMR gradually became the most powerful while the C-alpha test and the model selection approaches also had much improved performance relatively. The KBAC was most powerful except for the case with 64 non-causal RVs. Note that the aSum test maintained power as high as that of the Sum test, while the single SNV-based test, UminP, most commonly used in GWAS, had consistently low power.

For the case that the association strengths of the causal RVs were not constant with possibly opposite directions (Table 4), it is confirmed that the pooled association tests performed similarly and suffered from substantial loss of power. Across all the situations, the SSU, KMR and C-alpha performed similarly and were most powerful. Although the three sequential model selection approaches (Step-up, Seq, Seq-VS), the KBAC and the aSum test performed well with no or few non-causal RVs, surprisingly, as the number of non-causal RVs increased, their performance deteriorated more than that of the SSU, KMR and C-alpha tests. Nevertheless, the above procedures did improve over the pooled association tests.

It is noted that the CMC(0.01) test (with $MAF \leq 0.01$ as the cut-off for RVs) was less powerful than the default CMC, i.e. CMC(0.05), test in Table 3 because the former unnecessarily formed a few extra groups for CVs and increased the DF of the test; in contrast, the former performed better than the latter in Table 4, presumably because some causal RVs might have an overall $MAF > 0.01$ (due to their enrichment in cases) and the CMC(0.01) test grouped these causal RVs into separate groups, avoiding pooling them over with other causal RVs with opposite association directions and thus improving power. Hence, the choice of the MAF threshold for RVs is important for CMC, but it is unclear how to do so generally in practice.

It is noted that since the true model was the main-effects model (1), KMR with a linear kernel corresponded to using the true model, thus it was more powerful than using a quadratic kernel; the small performance difference between using the two kernels demonstrated the robustness of the KMR method. It is also noted that, since all the RVs were independent, the covariance matrix V was nearly diagonal, and thus the score test and SSUw test performed similarly. Finally, since in the current simulation set-up, causal RVs were randomly chosen with various MAFs, it was not informative to weight the variants according to their MAFs, suggesting why the SSU test outperformed the wSSU-P and SSUw tests.

RVs in LD

We next consider the case where all the RVs, both causal and non-causal ones, were possibly correlated. In this case, if a RV was associated with the disease, so were the other RVs since they were in LD. For the null case (Table 5), all the tests except C-alpha-A had their Type I error rates well controlled. Since the asymptotic distribution of the C-alpha test is derived under the assumption that all the RVs are independent, which was violated here, one has to use its permutational distribution, which appears to work well.

For the non-null case with varying association strengths (Table 5), again all the pooled tests suffered from significant power loss, while the SSU, KMR and C-alpha-P tests were most powerful. The three sequential model selection approaches and KBAC performed similarly and better than the aSum test, and all improved over the pooled association tests.

Due to the LD among the RVs, the score test and SSUw test performed differently: When there was no non-causal RVs, the score test was more powerful; however, as the number of non-causal RVs increased, the SSUw test became much more powerful than the score test.

No LD between causal RVs and non-causal RVs

Now we consider the case where causal RVs were correlated, non-causal RVs were also correlated, but there was no LD between causal and non-causal RVs. For the null case (Table 6), again all the tests except the C-alpha-A had satisfactory Type I error rates. The C-alpha-A did not work because its independence assumption on the RVs was violated.

For power comparison, again due to the presence of opposite association directions, the pooled association tests performed similarly and had the lowest power. With no or few non-causal RVs, Seq-aSum and Seq-aSum-VS performed best, closely followed by the C-alpha-P, KMR and SSU tests; for a larger number of non-causal RVs, the C-alpha-P, KMR and SSU tests were the winners.

Independent RVs and CVs

Finally we considered the case with independent RVs and four non-causal CVs (with MAFs randomly between 0.01 and 0.1). Although the aSum test was proposed by Han and Pan

(2010) to group CVs separately from RVs, as done in CMC, for simplicity here, we did not do such groupings. All the tests had satisfactory Type I error rates (not shown). For power comparison (Table 7), it is most notable that the SSU, KMR and C-alpha tests were all low-powered, due to the undue influence of the CVs, as analyzed before. The performance of the pooled association tests and KBAC also degraded. With CVs, variable selection worked well as evidenced by the good performance of the Step-up procedure, and by that the Seq-aSum-VS performed much better than Seq-aSum without variable selection. We also note that the weighted C-alpha tests were much more powerful than the original unweighted C-alpha test; between the two weighted C-alpha tests, the first one with weights inversely proportional to the MAFs performed much better than the second one in the presence of a large number of non-causal RVs, but not so otherwise; neither was more powerful than the wSSU-P test. Overall, the weighted SSU test (wSSU-P) performed best, closely followed by the Step-up procedure, then by the SSUw and score tests.

DISCUSSION

The three pooled association tests (i.e., Sum, CMC and wSum) performed similarly for RVs. They were most powerful when there were no opposite association directions and when there were no or only few non-causal RVs; otherwise, they suffered from a substantial loss of power.

Perhaps the most surprising and interesting finding is that, overall, in the presence of opposite association directions and non-causal RVs, the SSU, KMR and C-alpha-P performed similarly and best. Although the three methods appear quite different, they share a common feature: all can be regarded as testing on a variance component in a random-effects model, thus are robust to a large number of parameters induced by a large group of RVs. This is related to the success story of the class of gene set tests, including both the SSU-equivalent Goeman's test (Goeman et al 2004) and KMR (2008), applied to high-dimensional microarray data. Furthermore, the SSU test and KMR are themselves closely related to each other (Pan 2011), and share some advantages: they can be applied to other GLMs for other types of traits, such as quantitative or survival traits, and to adjust for other covariates, such as environmental factors, which are important as argued by Bansal et al (2010).

The approaches based on model selection (aSum, Step-up, Seq-aSum, Seq-aSum-VS) improve over the pooled association tests in the presence of opposite association directions. However, in spite of their strong motivation for model selection, their performance might not be as impressive as expected, especially in the presence of a large number of non-causal RVs. A possible explanation lies in the trade-off between the gain and the cost of model selection: in spite of possibly a strong association with the trait, due to its low frequency of any single RV, often there is only minimal power to detect its own association with the trait, rendering it difficult to distinguish whether the RV is or is not associated with the trait, and if so, whether its effect is protective or harmful. As shown by the close performance between Seq-aSum and Seq-aSum-VS, there is only minimal gain or loss in selecting causal RVs. On the other hand, as shown here, when there were both protective and deleterious causal RVs and few non-causal RVs, our newly proposed Seq-aSum and Seq-aSum-VS were or nearly were the most powerful, suggesting their applicability not only to RVs, but also to CVs: in analyzing multiple common SNVs in an LD region, if the untyped causal SNV is in LD with the multiple typed SNVs, the two methods could be powerful. In addition, leveraging on the idea of pooling and thus reduced degrees of freedom, they can be also applied to detect epistasis, as done in He et al (2010).

Several approaches are not considered here, including penalized regression (e.g. Malo et al 2008 for CVs, Zhou et al 2010 for both CVs and RVs) and some non-parametric regression techniques, such as logic regression (Kooperberg et al 2001) for CVs, and a Bayesian GLM (Yi and Zhi 2011), largely due to their difficulty in controlling Type I error rates (which is required to make a formal and fair comparison with other statistical tests) and associated high computing cost in permutation tests, especially if one aims to take account of the uncertainty in choosing optimal penalization or tuning parameters. Penalized regression and logic regression belong to the class of model selection-based approaches. Compared to the four selection methods compared here, penalized regression and logic regression are believed to have some advantages. However, existing penalized regression and logic regression methods do not incorporate the strategy of collapsing RVs or of random-effects models, two key elements for the success of the compared methods for RVs; further studies are needed to evaluate their performance. Another approach is based on haplotype inference (Zhu et al 2010; Li et al 2010), which is appealing for its applicability to GWAS for association analysis of more frequent RVs in the MAF range of 0.1%–5%.

We note that Price et al (2010) proposed using multiple thresholds and (possibly predicted) biological functional annotations to group RVs and empirically showed its advantage over using only one group. For simplicity, we have only considered the use of a single group. However, our conclusion should be useful for the case with multiple groups for RVs: a test with high power for a single group is likely to be even more powerful for multiple groups that are appropriately constructed, as shown by Pan and Shen (2011). As shown by our simulation studies, mixing non-causal CVs (or RVs with relatively higher MAFs) with RVs may degrade the performance of several tests, especially the SSU, KMR, C-alpha and KBAC tests. Hence it is a critical question in practice how to define RVs, to which a test is applied. There are two possible ways. The first way is to use the multiple cut-offs of MAF to define RVs and then combine the test results, as implemented in the multiple threshold test of Price et al (2010) and in the adaptive tests of Pan and Shen (2011). Second, as shown here and by other authors (Madsen and Browning 2009), weighting the variants in a test with suitably chosen weights (e.g. inversely proportional to their MAFs) may improve the performance of the test. We have not investigated these issue extensively and more studies are needed in the future. Finally, the simulation set-ups considered here are similar to Li and Leal (2008), but may still be over-simplified. Although there is no compelling statistical argument for the strong dependence of our conclusions on the simulation set-ups, it would be helpful to consider more practical set-ups, such as using real sequencing data; we did not pursue it here due to lack of publicly available large samples of sequencing data. With a sample size of currently only several hundreds with multiple racial/ethnic groups provided by the 1000 Genome Project, it is not clear how to best construct simulated data to mimic real data while maintaining the low MAFs of RVs. Although we acknowledge the limitation of our current simulation set-ups, they do illustrate some useful properties of various tests, such as how they perform in the presence of opposite association directions, of non-causal RVs and/or CVs, and of correlated SNVs.

In summary, since there is a large power difference between the pooled association tests and the random-effects model-based approaches (SSU, KMR and C-alpha-P) at either of the two extremes (i.e. whether there are opposite association directions), we recommend the use of a test from each class if it is unclear which extreme is likely to hold. We also recommend the use of the KBAC test and a variable selection-based approach, e.g. our newly proposed Seq-Sum-VS test; the former may be able to explore some complex interactions among RVs, while the latter may shed light on which SNVs are associated with the trait and if so, their association directions. Among the pooled association tests, they all perform similarly, while for the other class, the SSU and KMR have certain advantages: their known asymptotic distributions avoid the use of computationally demanding permutations, they can be implemented in any GLMs, which implies their applicability to binary, quantitative and other types of traits, and their ability to adjust for other covariates such as environmental factors and to detect environment-gene interactions, their applicability to CVs and/or RVs no matter whether they are in LD or not. We note that the SSU test can be applied to more complex regression models, e.g. with both main effects and some interaction terms; equally, in KMR we can use a kernel that can capture some complex interactions among the SNVs. Of course, as shown earlier, with any given kernel and its decomposition, we can have an SSU test equivalent to KMR. It would be of interest to compare the performance of the SSU/KMR and KBAC in the presence of interactions among RVs.

A potentially useful resource resulting from this work is freely available software: we have implemented most of the compared methods in R; R code will be posted on our web site at http://www.biostat.umn.edu/~weip/prog.html.

Acknowledgments

WP thanks Drs Roeder and Witte for sharing their manuscripts, Drs Wu, Ionita-Laza and Leal and Gao Wang for sharing their computer programs. The authors are grateful to the reviewers for helpful and constructive comments. This research was partially supported by NIH grant R21DK089351; WP was also supported by R01HL65462 and R01HL105397.

APPENDIX: RELATIONSHIPS BETWEEN GOEMAN'S TEST AND ZELTERMAN AND CHEN'S HOMOGENEITY TEST

We first review Zelterman and Chen's homogeneity test. Suppose that $y_1,\ldots,y_n$ are independent random variables with respective pdf's $f_i(y_i|\lambda_i)$, conditional on a k-dimensional parameter λ_i. Under H_0, all λ_i's are equal to a fixed vector λ_0. It is assumed that λ_i's are random: $\lambda_i = \lambda_0 + az$, where z is a k-dimensional random variable with $E(z) = 0$ and $Cov(z) = \Sigma = (\sigma_{st})$. Under this formulation, testing H_0 is equivalent to testing H_0': $a = 0$. Zelterman and Chen (1988) showed that the score test statistic for H_0' is

$$T_z=\frac{1}{2}\sum_{i=1}^{n}\sum_{s=1}^{k}\sigma_{ss}\frac{\partial^2 f_i(y_i|\lambda_0)}{\partial\lambda_{0s}^2}\frac{1}{f_i(y_i|\lambda_0)}+\sum_{i=1}^{n}\sum_{s<t}\sigma_{st}\frac{\partial^2 f_i(y_i|\lambda_0)}{\partial\lambda_{0s}\partial\lambda_{0t}}\frac{1}{f_i(y_i|\lambda_0)},$$

where $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0k})'$.

For observation Y_i, the score vector is $U_{(i)}=(f'_{i,1}/f_i,\ldots,f'_{i,k}/f_i)'$, and

$$U'_{(i)}\Sigma U'_{(i)}=\sum_{s,t}\sigma_{st}f'_{i,s}f'_{i,t}.$$

For simplicity we use notation $f'_{i,s}=\partial f_i/\partial\lambda 0s$. On the other hand, we have the (s, t)th element of V as

$$V_{st}=-\sum_{I=1}^{n}\frac{\partial U_{(i),s}}{\partial\lambda_{0t}}=-\sum_{I=1}^{n}\frac{f''_{i,st}f_i-f'_{i,s}f'_{i,t}}{f_i^2}.$$

Since $E\left(U'_{(i)}\sum U_{(j)}\right)=E\left(U'_{(i)}\right)\sum E\left(U_{(j)}\right)=0$, by ignoring cross-product terms $U_{(i)}\Sigma U_{(j)}$, we have Goeman's test statistic

$$S=\frac{1}{2}U'\Sigma U - tr(\Sigma V)=\frac{1}{2}\sum_{i=1}^{n}\sum_{s,t}\sigma_{st}f''_{i,st}/f_i=T_Z.$$

Hence, Goeman's test is equivalent to Zelterman and Chen's homogeneity test, which covers the C-alpha test as a special case (with m_i as y_i and f_i as $Bin(n_i, p_i)$). By the equivalence among permutation-based Goeman's test, SSU test and KMR test with a linear kernel, we know that the SSU test, KMR test with a linear kernel, and permutation-based C-alpha test are all equivalent (if a common random variable of interest is modeled). However, in the current context, since the disease status of subject i is treated as random variable of interest y_i in the SSU test and KMR, while the mutation status of variant i is treated as y_i in the C-alpha test, the three tests are closely related but not exactly equivalent.

REFERENCES

Asimit J, Zeggini E. Rare variant association analysis methods for complex traits. Ann Rev Genet. 2010; 44:293–308. [PubMed: 21047260]

Azzopardi D, Dallosso AR, Eliason K, Hendrickson BC, Jones N, et al. Multiple rare nonsynonymous variants in the adenomatous polyposis coli gene predispose to colorectal adenomas. Cancer Res. 2008; 68:358–363. [PubMed: 18199528]

Basu S, Stephens M, Pankow JS, Thompson EA. A Likelihood-Based Trait-Model-Free Approach for Linkage Detection of Binary Trait. Biometrics. 2010; 66:205–213. [PubMed: 19459835]

Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. Nature Review Genetics. 2010; 11:773–785.

Bhatia G, Bansal V, Harismendy O, Schork NJ, Topol EJ, Frazer K. A covering method for detecting genetic associations between rare variants and common phenotypes. PLoS Computational Biology. 2010; 6 e1000954.

Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008; 40:695–701. [PubMed: 18509313]

Chapman JM, Whittaker J. Analysis of multiple SNPs in a candidate gene or region. Genetic Epidemiology. 2008; 32:560–566. [PubMed: 18428428]

Chen SX, Qin Y-L. A two-sample test for high-dimensional data with applications to gene-set testing. Ann Statist. 2010; 38:808–835.

Clayton D, Chapman J, Cooper J. Use of unphased multilocus genotype data in indirect association studies. Genet Epidemiol. 2004; 27:415–428. [PubMed: 15481099]

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. [PubMed: 17966093]

Fan R, Knapp M. Genome association studies of complex diseases by case-control designs. Am J Hum Genet. 2003; 72:850–868. [PubMed: 12647259]

Feng T, Zhu X. Genome-wide searching of rare genetic variants in WTCCC data. Human Genetics. 2010; 128:269–280. [PubMed: 20549515]

Goeman JJ, van de Geer S, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics. 2004; 20:93–99. [PubMed: 14693814]

Goeman JJ, van de Geer S, van Houwelingen HC. Testing against a high dimensional alternative. J R Stat Soc B. 2006; 68:477–493.

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. [PubMed: 18179889]

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. [PubMed: 19910025]

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]

Han, F.; Pan, W. Powerful Multi-marker Association Tests: Unifying Genomic Distance-Based Regression and Logistic Regression. To appear in Genetic Epidemiology. 2010b. Available at http://www.sph.umn.edu/biostatistics/research.asp

He H, Oetting WS, Brott MJ, Basu S. Pair-wise multifactor dimensionality reduction method to detect gene-gene interactions in a case-control study. Hum Hered. 2010; 69:60–70. [PubMed: 19797910]

Hindorff, LA.; Junkins, HA.; Hall, PN.; Mehta, JP.; Manolio, TA. [Accessed October 31, 2010] A Catalog of Published Genome-Wide Association Studies. 2010. Available at www.genome.gov/gwastudies

Hoffmann TJ, Marini NJ, Witte JS. Comprehensive approach to analyzing rare genetic variants. PLoS One. 2010; 5(11) e13584.

Ionita-Laza I, Buxbaum JD, Laird NM, Lange C. A new testing strategy to identify rare variants with either risk or protective effect on disease. PLoS Genetics. 2011; 7(2) e1001289.

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. [PubMed: 18391953]

Kimeldorf GS, Wahba G. Some results on Tchebycheffian spline function. J Math Anal Appl. 1971; 33:82–95.

Kooperberg C, Ruczinski I, LeBlanc ML, Hsu L. Sequence analysis using logic regression. Genet Epi. 2001; 21 Suppl 1:S626–S631.

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. American Journal of Human Genetics. 2006; 78:410–422. [PubMed: 16465619]

Kwee LC, Liu D, Lin X, Ghosh D, Epstein MP. A powerful and flexible multilocus association test for quantitative traits. Am. J. Hum. Genet. 2008; 82:386–397. [PubMed: 18252219]

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]

Li Y, Byrnes AE, Li M. To identify associations with rare variants, Just WHaIT: weighted haplotype and imputation-based tests. Am J Hum Genet. 2010; 87:728–735. [PubMed: 21055717]

Lin WY, Schaid DJ. Power comparisons between similarity-based multilocus association methods, logistic regression, and score tests for haplotypes. Genet Epidemiol. 2009; 33:183–197. [PubMed: 18814307]

Liu D, Ghosh D, Lin X. Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. BMC Bioinformatics. 2008; 9:292. [PubMed: 18577223]

Liu DJ, Leal SM. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. PLoS Genet. 2010; 6(10) e1001156.

Madsen BE, Browning SR. A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. PLoS Genet. 2009; 5(2) e1000384.

Maher B. Personal genomes: the case of the missing heritability. Nature. 2008; 456:18–21. [PubMed: 18987709]

Malo N, Libiger O, Schork NJ. Accommodating linkage disequilibrium in genetic-association analyses via ridge regression. Am J Hum Genet. 2008; 82:375–385. [PubMed: 18252218]

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. [PubMed: 17101154]

Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genetic Epidemiology. 2010; 34:188–193. [PubMed: 19810025]

Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Ogho-Melander M, Katherisan S, Purcell SM, Roeder K, Daly MJ. Testing for an unusual distribution of rare variants. PLoS Genetics. 2011; 7(3) e1001322.

Neyman J, Scott E. On the use of c-alpha optimal tests of composite hypothesis. Bulletin of the International Statistical Institute. 1966; 41:477–497.

Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. Genetic Epidemiology. 2009; 33:497–507. [PubMed: 19170135]

Pan, W. Relationship between Genomic Distance-Based Regression and Kernel Machine Regression for Multi-marker Association Testing. To appear in Genetic Epidemiology. 2011. Available at http://onlinelibrary.wiley.com/doi/10.1002/gepi.20567/pdf

Pan, W.; Shen, X. Adaptive Tests for Association Analysis of Rare Variants. To appear in Genetic Epidemiology. 2011. Available as Research report 2011-11, Division of Biostatistics, University of Minnesota. http://www.sph.umn.edu/biostatistics/research.asp

Price AL, Kryukov GV, de Bakker PIW, Purcell SM, Staples J, Wei L-J, Sunyaev SR. Pooled association tests for rare variants in exon-resequenced studies. Am J Hum Genet. 2010; 86:832–838. [PubMed: 20471002]

Pritchard JK. Are Rare Variants Responsible for Susceptibility to Complex Diseases? Am J Hum Genet. 2001; 69:124–137. [PubMed: 11404818]

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]

Schaid DJ. Genomic similarity and kernel methods I: advancements by building on mathematical and statistical foundations. Hum Hered. 2010a; 70:109–131. [PubMed: 20610906]

Schaid DJ. Genomic similarity and kernel methods I: methods for genomic information. Hum Hered. 2010b; 70:132–140. [PubMed: 20606458]

Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Current Opinion in Genetics & Development. 2009; 19:212–219. [PubMed: 19481926]

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. [PubMed: 17236140]

Wessel J, Schork NJ. Generalized genomic distance-based regression methodology for multilocus association analysis. Am J Hum Genet. 2006; 79:792–806. [PubMed: 17033957]

Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X. Powerful SNP-Set Analysis for Case-Control Genome-wide Association Studies. Am J Hum Genet. 2010; 86:929–942. [PubMed: 20560208]

Xiong M, Zhao J, Boerwinkle E. Generalized T^2 test for genome association studies. Am J Hum Genet. 2002; 70:1257–1268. [PubMed: 11923914]

Yi N, Zhi D. Bayesian analysis of rare variants in genetic association studies. Genetic Epidemiology. 2011; 35:57–69. [PubMed: 21181897]

Zelterman D, Chen C-F. Homogeneity tests against central-mixture alternative. JASA. 1988; 83:179–182.

Zhang L, Pei YF, Li J, Papasian CJ, Deng HW. Efficient utilization of rare variants for detection of disease-related genomic regions. PLoS One. 2010; 5(12) e14288.

Zhou H, Sehl ME, Sinsheimer JS, Lange K. Association screening of common and rare genetic variants by penalized regression. Bioinformatics. 2010; 26:2375–2382. [PubMed: 20693321]

Zhu X, Feng T, Li Y, Lu Q, Elston RC. Detecting rare variants for complex traits using family and unrelated data. Genetic Epidemiology. 2010; 34:171–187. [PubMed: 19847924]

Table 1

A summary of the properties of the tests to be compared: originally proposed to target CVs or RVs (or both), whether pooling over variants, whether sensitive to association directions (+/−), to a large number of non-causal RVs (nRVs) and to a few non-causal CVs (nCVs), requiring permutations for p-value calculations, capability to adjust for other covariates (Cov), applicability to other non-binary traits, whether can be formulated as testing on a variance component in a random-effects (R-E) model, and references for more details.

Test	Original target	Pool	Sens to +/−	Sens to nRVs	Sens to nCVs	Permut	Cov	Other traits	R-E	Refs
UminP	CV	No	No	No	No	No	Yes	Yes	No	3
Score	CV	No	No	No	No	No	Yes	Yes	Yes	1
SSU	CV	No	No	No	Yes	No	Yes	Yes	Yes	2
wSSU-P	Both	No	No	No	No	Yes	Yes	Yes	Yes	here
SSUw	CV	No	No	No	No	No	Yes	Yes	Yes	2
Sum	CV	No	Yes	Yes	Yes	No	Yes	Yes	No	2
KMR	CV	No	No	No	Yes	No	Yes	Yes	Yes	4, 5
CMC	RV	Yes	Yes	Yes	No	No	No	No	No	6
wSum	RV	Yes	Yes	Yes	Some	Some	No	No	No	7
aSum-P	Both	Yes	Some	Yes	Some	Yes	Yes	Yes	No	8
Step-up	RV	Yes	Some	Some	No	Yes	Yes	Yes	No	10
Seq-aSum	Both	Yes	Some	Some	Yes	Yes	Yes	Yes	No	here
Seq-aSum-VS	Both	Yes	Some	Some	No	Yes	Yes	Yes	No	here
KBAC	RV	No	Some	Some	Some	Yes	Some	No	No	11
C-alpha-A	RV	No	No	No	Yes	No	No	No	Yes	9
C-alpha-P	RV	No	No	No	Yes	Yes	No	No	Yes	9
RBT	RV	Yes	Some	Yes	No	Yes	No	No	No	12

Refs: 1. Clayton et al (2004); 2. Pan (2009); 3. Conneely&Boehnke (2007); 4. Kwee et al (2008); 5. Wu et al (2010); 6. Li&Leal (2008); 7. Madsen&Browning (2009); 8. Han&Pan (2010); 9. Neale et al (2011); 10. Hoffmann et al (2010); 11. Liu and Leal (2010); 12. Ionita-Laza et al (2011).

Table 2

Type I error rates at nominal level α based on 1000 replicates for 8 RVs plus a number of non-causal RVs. There is no LD among the RVs.

Test	**α = 0.05**					**α = 0.01**				
	# of neutral RVs					**# of neutral RVs**				
	0	**4**	**8**	**16**	**32**	**0**	**4**	**8**	**16**	**32**
UminP	.027	.027	.016	.011	.019	.003	.001	.004	.001	.002
Score	.043	.049	.040	.040	.040	.006	.009	.005	.005	.007
SSU	.044	.055	.045	.037	.043	.004	.013	.009	.005	.011
wSSU-P	.052	.051	.048	.048	.046	.008	.008	.014	.010	.008
SSUw	.041	.049	.039	.034	.040	.006	.011	.005	.005	.007
Sum	.047	.055	.041	.054	.038	.012	.007	.010	.010	.007
KMR(Linear)	.046	.056	.046	.042	.047	.007	.016	.011	.007	.012
KMR(Quad)	.046	.056	.047	.039	.046	.007	.016	.010	.006	.011
CMC(0.01)	.035	.053	.044	.055	.039	.008	.014	.010	.011	.009
CMC	.048	.053	.043	.056	.051	.010	.009	.011	.011	.007
wSum	.050	.057	.038	.059	.056	.010	.012	.011	.009	.006
aSum-P	.058	.064	.052	.063	.047	.012	.011	.010	.010	.011
Step-up	.046	.059	.056	.051	.051	.012	.011	.009	.009	.010
Seq-aSum	.044	.066	.056	.055	.059	.008	.013	.008	.008	.013
Seq-aSum-VS	.050	.058	.056	.051	.058	.011	.018	.011	.009	.013
KBAC	.058	.044	.053	.054	.046	.013	.007	.009	.012	.009
C-alpha-A	.045	.051	.042	.036	.043	.016	**.030**	**.022**	.010	.014
C-alpha-P	.050	.065	.058	.051	.055	.005	.016	.013	.006	.012
RBT	.045	.045	.050	.062	.044	.011	.010	.011	.011	.005

Table 3

Empirical power for tests at nominal level α based on 1000 replicates for an ideal case for 8 causal RVs with a common association strength *OR* = 2 and a number of non-causal RVs. There is no LD among the RVs.

Test	**α = 0.05**						**α = 0.01**					
	# of neutral RVs						**# of neutral RVs**					
	0	**4**	**8**	**16**	**32**	**64**	**0**	**4**	**8**	**16**	**32**	**64**
UminP	.441	.336	.296	.222	.175	.117	.142	.089	.094	.050	.043	.029
Score	.746	.632	.595	.471	.332	.245	.496	.391	.314	.221	.143	.073
SSU	.756	.702	.694	.626	.499	**.423**	.525	.479	.448	.379	.283	.205
wSSU-P	.821	.732	.714	.644	.514	.390	.573	.471	.407	.332	.222	.161
SSUw	.743	.638	.593	.477	.339	.268	.502	.389	.316	.218	.153	.082
Sum	**.951**	**.875**	**.808**	**.673**	.484	.313	**.859**	**.709**	**.605**	**.438**	.248	.116
KMR(Linear)	.762	.711	.699	.631	**.509**	**.438**	.548	.500	.473	.405	**.308**	**.234**
KMR(Quad)	.755	.707	.699	.629	.501	**.410**	.545	.497	.466	.403	.299	.215
CMC(0.01)	.853	.761	.702	.628	.484	.396	.672	.524	.452	.384	.268	**.218**
CMC	.938	.853	.777	.616	.399	.211	.831	.679	.570	.383	.196	.086
wSum	.940	.846	.782	.618	.424	.267	**.838**	**.687**	.568	.394	.216	.114
aSum-P	.933	**.858**	.780	.669	.499	.313	.781	.611	.534	.381	.257	.125
Step-up	.859	.801	.769	**.679**	**.521**	.335	.712	.608	.552	**.431**	.301	.135
Seq-aSum	.810	.705	.663	.547	.407	.312	.596	.470	.415	.320	.190	.128
Seq-aSum-VS	.798	.722	.692	.590	.420	.344	.598	.506	.452	.345	.216	.141
KBAC	**.960**	**.911**	**.867**	**.779**	**.600**	388	**.858**	**.749**	**.680**	**.529**	**.317**	.160
C-alpha-A	.741	.687	.664	.597	.460	.364	.637	.580	.538	.446	.320	.234
C-alpha-P	.771	.712	.688	.627	.484	.378	.542	.492	.459	.402	**.305**	**.219**
RBT	**.941**	.849	**.784**	.664	.463	.321	.813	.667	**.587**	.424	.238	.121

Table 4

Empirical power for tests at nominal level α based on 1000 replicates for a non-ideal case for 8 causal RVs with various association strengths *OR* = (3, 3, 2, 2, 2, 1/2, 1/2, 1/2) and a number of non-causal RVs. There is no LD among the RVs.

	α = 0.05					**α = 0.01**				
Test	**# of neutral RVs**					**# of neutral RVs**				
	0	**4**	**8**	**16**	**32**	**0**	**4**	**8**	**16**	**32**
UminP	.607	.532	.481	.417	.346	.318	.259	.227	.204	.142
Score	.869	.772	.721	.632	.483	.660	.532	.480	.356	.233
SSU	.895	**.835**	**.815**	**.774**	**.696**	.723	.662	.645	.583	.472
wSSU-P	.861	.776	.735	.685	.550	.606	.510	.460	.401	.258
SSUw	.867	.773	.732	.633	.501	.661	.550	.481	.355	.238
Sum	.682	.566	.465	.365	.258	.471	.348	.257	.172	.101
KMR(Linear)	**.897**	**.842**	**.824**	**.783**	**.707**	**.740**	**.678**	**.667**	**.619**	**.495**
KMR(Quad)	.893	.835	.815	.781	.698	.734	**.680**	**.663**	.608	**.484**
CMC(0.01)	.703	.669	.670	.670	.590	.511	.457	.470	.470	.383
CMC	.661	.544	.456	.336	.204	.461	.337	.235	.157	.086
wSum	.659	.548	.459	.335	.228	.460	.336	.236	.158	.093
aSum-P	.854	.745	.684	.574	.430	.670	.538	.430	.315	.207
Step-up	.839	.767	.724	.640	.527	.652	.564	.518	.413	.285
Seq-aSum	.892	.811	.757	.671	.528	**.752**	.620	.532	.438	.273
Seq-aSum-VS	.885	.807	.768	.686	.545	.729	.623	.567	.448	.293
KBAC	**.907**	.813	.763	.642	.436	**.737**	.607	.536	.399	.199
C-alpha-A	.892	.826	.802	.757	.655	.824	.732	.720	.653	.512
C-alpha-P	**.906**	**.844**	**.823**	**.775**	**.674**	.735	**.673**	**.661**	**.612**	**.496**
RBT	.810	.659	.603	.482	.301	.590	.429	.356	.250	.125

Table 5

Type I error (with *OR* = 1) and power (with eight causal RVs with *OR* = (3, 1/3, 2, 2, 2, 1/2, 1/2, 1/2)) for tests at nominal level α = 0.05 based on 1000 replicates for 8 RVs and a number of other non-causal RVs. There is LD among the RVs.

Test	***OR* = 1**					***OR* = (3, 1/3, 2, 2, 2, 1/2, 1/2, 1/2)**				
	# of neutral RVs					**# of neutral RVs**				
	0	**4**	**8**	**16**	**32**	**0**	**4**	**8**	**16**	**32**
UminP	.033	.027	.026	.016	.013	.489	.479	.452	.365	.318
Score	.034	.022	.025	.019	.023	.599	.538	.491	.380	.276
SSU	.040	.041	.052	.044	.036	.603	**.624**	**.635**	**.581**	**.574**
wSSU-P	.057	.043	.047	.062	.053	.566	.586	.609	.585	.491
SSUw	.035	.042	.049	.033	.034	.532	.561	.574	.506	.493
Sum	.049	.047	.059	.033	.049	.342	.312	.315	.258	.239
KMR(Linear)	.042	.045	.057	.046	.043	.611	**.630**	**.644**	**.597**	**.590**
KMR(Quad)	.038	.033	.041	.030	.025	.545	.563	.565	.493	.474
CMC	.045	.053	.056	.036	.060	.296	.283	.189	.182	.365
wSum	.045	.054	.056	.040	.063	.369	.297	.287	.191	.200
aSum-P	.050	.046	.061	.038	.053	.350	.323	.325	.258	.243
Step-up	.047	.060	.059	.042	.050	.524	.516	.532	.429	.409
Seq-aSum	.045	.062	.054	.056	.055	**.658**	.617	.596	.484	.416
Seq-aSum-VS	.043	.056	.058	.054	.049	**.658**	.606	.577	.472	.414
KBAC	.050	.054	.050	.053	.049	.497	.439	.426	.371	.275
C-alpha-A	.065	**.076**	**.092**	**.097**	**.110**	-	-	-	-	-
C-alpha-P	.050	.049	.062	.057	.048	**.629**	**.650**	**.668**	**.607**	**.598**
RBT	.047	.039	.036	.060	.056	.374	.343	.386	.357	.279

Table 6

Type I error (with *OR* = 1) and power (with eight causal RVs with *OR* = (3, 1/3, 2, 2, 2, 1/2, 1/2, 1/2)) for tests at nominal level α = 0.05 based on 1000 replicates for 8 RVs and a number of other non-causal RVs. There is LD among the 8 RVs and among other non-causal RVs, but no LD between the 8 RVs and non-causal RVs.

Test	***OR* = 1**					***OR* = (3, 1/3, 2, 2, 2, 1/2, 1/2, 1/2)**				
	# of neutral RVs					**# of neutral RVs**				
	0	**8**	**16**	**32**	**64**	**0**	**8**	**16**	**32**	**64**
UminP	.032	.018	.021	.014	.007	.506	.380	.324	.288	.208
Score	.029	.029	.028	.019	.021	.631	.480	.373	.241	.160
SSU	.049	.051	.035	.034	.034	.642	.553	**.475**	**.444**	**.334**
wSSU-P	.045	.060	.042	.050	.052	.606	.494	.424	.362	.269
SSUw	.045	.040	.027	.015	.036	.562	.450	.352	.272	.187
Sum	.046	.059	.046	.046	.046	.345	.229	.159	.110	.079
KMR(Linear)	.051	.056	.039	.040	.037	.649	**.568**	**.490**	**.459**	**.356**
KMR(Quad)	.046	.049	.022	.021	.017	.572	.487	.392	.331	.205
CMC	.046	.053	.040	.050	.047	.339	.235	.193	.124	.111
wSum	.048	.052	.041	.053	.048	.342	.237	.199	.133	.114
aSum-P	.052	.061	.049	.046	.052	.364	.239	.170	.113	.081
Step-up	.057	.055	.047	.048	.051	.554	.449	.378	.304	.213
Seq-aSum	.051	.053	.041	.046	.052	**.703**	**.584**	.453	.353	.249
Seq-aSum-VS	.053	.053	.048	.041	.054	**.701**	.572	.447	.351	.258
KBAC	.048	.058	.036	.053	.047	.527	.388	.321	.262	.180
C-alpha-A	**.076**	**.093**	**.084**	**.092**	**.118**	-	-	-	-	-
C-alpha-P	.055	.065	.043	.050	.047	**.669**	**.585**	**.504**	**.472**	**.340**
RBT	.057	.059	.049	.042	.054	.376	.285	.188	.141	.097

Table 7

Empirical power for the tests at nominal level α = 0.05 based on 1000 replicates with eight causal RVs, four neutral CVs and a number of other neutral RVs. There is no LD among the CV/RVs.

Test	***OR* = (2, 2, 2, 2, 2, 2, 2, 2)**					***OR* = (3, 3, 2, 2, 2, 1/2, 1/2, 1/2)**				
	# of neutral RVs					**# of neutral RVs**				
	0	**4**	**8**	**16**	**32**	**0**	**4**	**8**	**16**	**32**
UminP	.355	.283	.269	.213	.156	.518	.482	.441	.412	.331
Score	.628	.580	.498	.424	.348	**.766**	**.706**	.629	.584	.466
SSU	.148	.128	.134	.131	.135	.225	.201	.206	.203	.215
wSSU-P	**.777**	**.729**	**.700**	**.589**	**.518**	**.810**	**.764**	**.724**	**.655**	**.582**
SSUw	.634	.592	.515	.429	.332	**.765**	.704	**.631**	**.599**	**.489**
Sum	.455	.438	.396	.348	.299	.231	.225	.195	.199	.152
KMR(Linear)	.158	.138	.151	.145	.153	.237	.216	.222	.223	.234
KMR(Quad)	.153	.124	.136	.137	.141	.219	.198	.204	.201	.219
CMC	.575	.512	.429	.309	.212	.296	.254	.209	.155	.124
wSum	.533	.508	.469	.408	.346	.291	.285	.249	.230	.181
aSum-P	.467	.457	.414	.355	.310	.239	.245	.206	.202	.158
Step-up	**.776**	**.750**	**.715**	**.610**	**.522**	.727	**.712**	**.658**	**.605**	**.499**
Seq-aSum	.368	.314	.323	.300	.266	.453	.410	.392	.395	.342
Seq-aSum-VS	.550	.518	.502	.450	.379	.610	.617	.567	.541	.471
KBAC	.554	.537	.478	.446	.370	.415	.402	.358	.335	.270
C-alpha-A	.106	.083	.089	.088	.082	.165	.154	.146	.149	.160
C-alpha-P	.165	.150	.145	.139	.139	.245	.233	.228	.220	.225
w1C-alpha-P	.542	.527	.527	.496	.474	.670	.642	.632	.636	.593
w2C-alpha-P	.628	.568	.476	.388	.298	.773	.698	.606	.563	.422
RBT	**.826**	**.770**	**.688**	**.592**	**.453**	.630	.581	.487	.410	.321