Adaptive Tests for Association Analysis of Rare Variants

Wei Pan[1] and **Xiaotong Shen**[2]

[1]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455

[2]School of Statistics, University of Minnesota, Minneapolis, MN 55455

Abstract

In anticipation of the availability of next-generation sequencing data, there has been increasing interest in association analysis of rare variants (RVs). Due to the extremely low frequency of a RV, single variant based analysis and many existing tests developed for common variants (CVs) may not be suitable. Hence, it is of interest to develop powerful statistical tests to assess association between complex traits and RVs with sequence data. Recently a pooled association test based on variable thresholds (VT) was proposed and shown to be more powerful than some existing tests (Price et al 2010). Here we generalize the VT test of Price et al in several aspects. We propose a general class of adaptive tests that covers the VT test of Price et al as a special case. In particular, we show that some of our proposed adaptive tests may substantially improve the power over the pooled association tests, including the VT test of Price et al, especially so in the presence of many neural RVs and/or of causal RVs with opposite association directions, in which cases most of the existing pooled association tests suffer from significant loss of power. Our proposed tests are also general and flexible with the ability to incorporate weights on RVs and to adjust for covariates.

Keywords

complex traits; logistic regression; Neyman's adaptive test; permutation; pooled association tests; SSU test; Sum test; Variable threshold

INTRODUCTION

The current wave of genome-wide association studies (GWASs) has successfully identified hundreds of common variants (CVs) associated with complex traits. However, these identified CVs can explain only a small proportion of heritable variability (Maher 2008). One possible explanation for the "dark matter" of missing heritability is the existence of undiscovered rare variants (RVs) associated with complex traits (Pritchard and Cox 2002; Bodmer and Bonilla 2008; Gorlov et al 2008; Schork et al 2009). The recent advance of next-generation sequencing has made it both technically and financially feasible to sequence large parts of or whole genomes. The availability of sequencing data offers an exciting opportunity for the first time to conduct large scale studies to discover complex trait-RV associations. However, due to the extremely low frequencies of RVs, typically lower than 1%, there is only a limited amount of information contained within each RV, rendering the most popular analysis strategy adopted in GWASs, single variant-based analysis, powerless; see Basu and Pan (2011) for some numerical examples. Equally, many other statistical tests

Correspondence author: Wei Pan, Telephone: (612) 626-2705, Fax: (612) 626-0660, weip@biostat.umn.edu, Address: MMC 303, A460 Mayo, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455–0392, USA.

developed for CVs in GWASs may not be applicable with low power for RVs. Hence, there is a compelling need to develop novel and powerful statistical tests meeting the challenge with RVs; see two recent reviews (Asimit and Zeggini 2010; Bansal et al 2010).

A class of new statistical tests, called pooled association tests, have been developed specifically to analyze RVs (Morgenthaler and Thilly 2007; Li and Leal 2009; Madsen and Browning 2009). They all share a common feature of pooling or collapsing multiple RVs into a "super" single nucleotide variant (SNV) and then analyze the association between the trait and the super SNV. In this way, they utilize the information in each of the multiple RVs while avoiding the low power approach to analyzing each individual RV separately. However, as pointed out by several authors (Han and Pan 2010; Bansal et al 2010; Li et al 2010), a common weakness of the pooled association tests caused by the pooling or collapsing strategy is that the tests will lose power if there are opposite association directions among the pooled/collapsed RVs. More importantly, Basu and Pan (2011) discovered that the performance of pooled association tests deteriorated quickly as the number of neutral or non-functional RVs to be pooled increased. Note that it is in general inevitable to have non-functional RVs to be pooled in any candidate gene or region of the genome. Han and Pan (2010) proposed a test to adaptively determine the coding of the RVs to overcome the shortcoming of pooled association tests; a similar idea was used in the test of Li et al (2010). However, the power of their adaptive test largely depends on the difficult choice of a parameter value that controls the adaptiveness of the test, and the resulting test may have low power if an inappropriate parameter value is used.

More recently, Price et al (2010) proposed an improved pooled association test based on variable thresholds (VT). They observed that the minor allele frequencies (MAFs) of causal RVs might be different from those of non-functional ones, and hence proposed to use the MAFs as multiple thresholds to form various collapsing groups of RVs, to each of which a standard pooled association test is applied, and then their results are summarized. In addition, they also proposed incorporating various weights, e.g. based on computational prediction of the likelihood of each RV's being functional, to further improve the power. However, the VT test still shares the common weakness of pooled association tests: if there are causal RVs with opposite association directions, and/or if there are many non-functional RVs to be pooled, the test will have dramatic power loss, as to be confirmed in this paper. Our main contribution is that, by recognizing the adaptiveness of the VT test, we generalize the idea in two aspects. First, we propose a class of adaptive tests, some of which can overcome the weakness of the pooled association tests; that is, some proposed new tests maintain higher power than that of the VT test of Price et al or other standard pooled association tests when there are a large number of neutral RVs, and/or when some causal RVs are associated with the trait with opposite association directions. In addition, the VT test of Price et al is a special case of this new class of adaptive tests. Second, rather than using the MAF as variable thresholds, we propose using some other criteria as variable thresholds, which can largely boost the power in some situations. Our proposed tests can also accommodate various weights for RVs as does the VT test. We use simulated data to show much improved performance of our proposed tests over existing ones.

METHODS

We focus on the case with a binary trait, say a disease indicator, such as arising in a case-control design, though all the methods discussed here are based on logistic regression and thus can be easily extended to generalized linear models (GLMs) for other types of traits. Equally the methods discussed can also take account of covariates, though we do not pursue it here. The analysis goal is to detect whether there is any association between the binary trait and a group of rare SNVs, which, for example, are 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, $j = 1, \ldots, k$, though other codings can be also used.

Logistic regression

Since we are to modify several existing tests originally proposed for CVs to fit in the case with RVs, we briefly review these tests. For the binary trait and a group of SNVs, it is natural to build various tests based on a logistic regression model:

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

The null hypothesis to be tested is H_0: $\beta = (\beta_1, \ldots, \beta_k)' = 0$, for which one of the asymptotically equivalent score test, Wald's test and likelihood ratio test (LRT) can be applied. Since the score test is computationally fastest without the need to iteratively obtain the maximum likelihood estimate (MLE) of β, we will exclusively focus on the score test. For model (1), the score vector and its covariance matrix are

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

where $=\sum_{i=1}^{n}Y_i/n$ and $\bar{X}=\sum_{i=1}^{n}X_i/n$ are the sample means, and $\mathbf{1} = (1, \ldots, 1)'$ is the k-vector of all 1's.

The multivariate score test is

$$T_{Score}=T_{Score}(U,V)=U'V^{-1}U,$$

which has an asymptotic chi-squared distribution with degrees of freedom (DF) $k = rank(V)$ under H_0. As shown by Clayton et al (2004), the score test is closely related to Hotelling's T^2 test used for CVs (Xiong et al 2002; Fan and Knapp 2003). While accounting for the correlation structure of the score vector, the score test may lose power for high dimensional data with a larger DF k.

To overcome the possible shortcoming of the score test, Pan (2009) proposed two tests, called sum of squared score (SSU) and weighted sum of squared score (SSUw) tests:

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

where $V_d = \text{Diag}(V)$ is a diagonal matrix with the same 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 distribution (Pan 2009). Each of the two tests can be regarded as a 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 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 $Cov(\beta) = \tau I$, then Goeman's permutation-based score test on $H_0: \tau = 0$ is equivalent to the SSU test. Interestingly, Goeman's test and the SSU test also performed well with high power for relatively lower-dimensional CV data, as empirically confirmed by Chapman and Whittaker (2008) and Pan (2009).

Another test with high power under certain situations for CVs is the so-called Sum test, as noted by Chapman and Whittaker (2008) and Pan (2009). The Sum test aims to utilize multiple SNVs with a minimum DF, for which it imposes a generally incorrect working assumption that the SNVs are all associated with the trait with a common association strength:

$$\text{Logit } \Pr(Y_i=1)=\beta_{c,0}+\sum_{j=1}^{k}X_{ij}\beta_c, \tag{2}$$

where β_c reflects the common odds ratio (OR) between the trait and each SNV under the working assumption. 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) shared the same spirit and thus similar performance as the Sum test. Note that in model (2) we regress Y on a new "super-SNV" that is the sum of the genotype scores of all the SNVs, hence we call the resulting test the Sum test. The Sum test based on the score vector is

$$T_{Sum}=T_{Sum}(U,V)=\frac{\mathbf{1}'U}{\sqrt{\mathbf{1}'V\mathbf{1}}}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{k}X_{ij}(Y_i-)}{\sqrt{\mathbf{1}'V\mathbf{1}}},$$

which has an asymptotic null distribution of $N(0, 1)$.

Pooled association tests

The best-known tests specifically designed for RVs with sequence data belong to the family of pooled association tests, including the cohort allelic sums test (CAST) (Morgenthaler and Thilly 2007), Combined Multivariate and Collapsing (CMC) test (Li and Leal 2009), which modifies the CAST to improve its performance when both RVs and CVs are present, and a weighted sum test of Madsen and Browning (2009). The basic idea of the pooled association tests is to combine or pool multiple RVs together to form a "super SNV" and then test its association with the trait. The rationale is that, due to the low frequency of any RV, there is only a weak association between each RV and the trait; pooling may boost the signal while reducing the DF. In this sense, the Sum test also belongs to pooled association tests. For example, the CAST creates a super SNV that equals to 1 or 0 depending on whether there is any minor allele in any RVs; that is, the super SNV is coded as $X_{C,i}=\mathrm{V}_{j=1}^{k}I(X_{ij}>0)$, in contrast to that of the Sum test $X_{S,i}=\sum_{j=1}^{k}X_{ij}$; note that for RVs, $X_{C,i} \approx X_{S,i}$. Plugging-in the super SNV as the predictor in a logistic regression model (e.g. equation 2), one tests its association with the trait. Basu and Pan (2011) showed similar performance between the Sum test and other pooled association tests for RVs, hence we will use the Sum test as a representative for the pooled association tests.

Price et al (2010) proposed a variable threshold (VT) test based on the assumption that the MAFs of the causal RVs may be different from those of non-functional RVs, which, if true,

can be utilized to improve the power of the corresponding pooled association tests. Suppose that H is the set of all observed MAFs across all RVs. For a given MAF threshold $h \in H$, suppose $\xi_j^h = I(h > \text{MAF of RV } j)$ is the indicator of whether the MAF of RV j is no bigger than h. Then the test statistic is

$$Price = \max_{h \in H} z(h) = \max_{h \in H} \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} \xi_j^h X_{ij}(Y_i - \bar{Y})}{[\sum_{i=1}^{n} \sum_{j=1}^{k} (\xi_j^h X_{ij})^2]^{1/2}}.$$

Note that, the numerator of $z(h)$ is the same as that of the Sum test if $\xi_j^h = 1$ for all j; the denominator of $z(h)$ is presumably to estimate the standard error of the numerator since

$Var[\sum_{i=1}^{n} \sum_{j=1}^{k} \xi_j^h X_{ij}(Y_i - \bar{Y})] = \sigma^2 \sum_{i=1}^{n} (\sum_{j=1}^{k} \xi_j^h X_{ij})^2$ with $\sigma^2 = Var(Y_i - \bar{Y}|H_0) \approx \bar{Y}(1 - \bar{Y})$. The above test can easily accommodate any weighting scheme on RVs: if we assign a weight w_j to RV j, we can simply replace X_{ij} by $w_j X_{ij}$.

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 significant power loss if the association directions of the functional variants are opposite. Perhaps even more importantly and less obviously, the pooled association tests are not robust to pooling over non-functional RVs, as shown by Basu and Pan (2011) and to be confirmed later. The reason is simple in retrospect: pooling over both functional and neutral RVs will add noises into the signal and thus reduce the power. Note that, in general it is unavoidable to have neutral RVs in a group of RVs to be tested. A main contribution of this paper is to extend the idea of the VT test of Price et al (2010) to a large class of tests, so-called adaptive tests. This class of tests include an adaptive SSU (aSSU) test, which, as the SSU test, is robust to different association directions *and* a large number of neutral RVs. Furthermore, it is shown that the VT test of Price et al is closely related to an adaptive Sum (aSum) test, a member of the proposed family of adaptive tests. Importantly, we also generalize the idea of the VT test: rather than using the MAF as the VT, other criteria can be also utilized, which will further boost the power of the corresponding adaptive tests.

Adaptive tests

Adaptive Neyman's test—Since our proposed tests (and Price et al's VT test) are closely related to the adaptive Neyman's test, we first review the latter. Let $Z \sim N(\theta, I)$ be a k-dimensional normal random vector. To test $H_0: \theta = 0$ versus $H_1: \theta \neq 0$, the score test, LRT and the Wald test all share the same test statistic $\|Z\|^2 = \sum_{j=1}^{k} Z_j^2$, where $\|.\|$ is the L_2 norm. The three conventional tests suffer from low power for high dimensional data with a large k: the power of the tests at $\theta = \theta_0 \neq 0$ is approximately

$$1 - \Phi(z_{1-\alpha} - \|\theta_0\|^2 / \sqrt{2k})$$

if $\|\theta_0\|^2 = o(k)$. Hence the power of the tests tends to the nominal Type I error rate α even if $\|\theta_0\| \to \infty$ with $\|\theta_0\|^2 = o(k^{1/2})$. Note $\Phi(.)$ is the cumulative distribution function of the standard normal $N(0, 1)$.

As an alternative, Neyman (1937) proposed testing the first few dimensions with a test statistic $\sum_{j=1}^{m} Z_j^2$ with $m \le k$ with its power approximately equal to

$$1 - \Phi\left(z_{1-\alpha} - \sum_{j=1}^{m} \theta_{0,j}^2 / \sqrt{2m}\right).$$

To maximize the power, Fan (1996) proposed an adaptive Neyman's test with test statistic

$$T_{AN} = \max_{1 \le m \le k} \left\{ \sum_{j=1}^{m} (Z_j^2 - 1) / \sqrt{2m} \right\}.$$

Note that $\sum_{j=1}^{m} (Z_j^2 - 1) / \sqrt{2m}$ is an unbiased estimator of $\sum_{j=1}^{m} \theta_{0,j}^2 / \sqrt{2m}$. It is obvious that the adaptive Neyman's test depends on the ordering of the components of Z.

New adaptive tests—The adaptive Neyman's test is based on the assumption that we have a test statistic $Z \sim N(\theta, I)$. Although many test statistics have a normal distribution, many others do not. We extend the idea of the adaptive Neyman's test in two aspects. First, we use both normal and non-normal test statistics that are based on the score vector. Second, since it may be difficult to obtain a closed form of the power function for a test, we use its p-value as a surrogate for its power. Specifically, suppose that $U = (U_1, \ldots, U_k)'$ is the score vector, and denote $U_{(m)} = (U_1, \ldots, U_m)'$ is the score subvector containing the first m components with $m \le k$. For any test statistic $T = T(U)$, we can equally apply it to only the first few components of U, say $T(U_{(m)})$, and obtain its p-value as $P_{T(U_{(m)})}$, Then we construct an adaptive version of the test T as

$$aT = aT(U) = \min_{1 \le m \le k} P_{T(U_{(m)})}.$$

Note that the aT test depends on the order of the components of the score vector U, as the adaptive Neyman's test depends on the ordering of Z_j's. Generally, the distribution of aT is complex; we recourse to permutation to obtain its p-value. Specifically, we randomly shuffle the response Y to yield a permuted version $Y^{(b)}$, then apply the adaptive test with $Y^{(b)}$ to yield the corresponding test statistic $aT^{(b)}$. We repeat the process for $b = 1, \ldots, B$. The final p-value is $\sum_{b=1}^{B} I(aT^{(b)} < aT) / B$.

If we substitute the score, SSU, SSUw and Sum test statistics as T respectively, then we have the corresponding adaptive tests called the aScore, aSSU, aSSUw and aSum tests.

In the logistic regression model, each component of U, say U_j, corresponds to a RV j. In the default aT test, denoted as aT-Loc, the ordering of the components of U is determined by the chromosome locations of their corresponding RVs; other ordering schemes are possible, leading to other versions of the aT tests. The first is to order the components of U based on the minor allele counts of the corresponding RVs, leading to the adaptive test denoted as aT-MAF. Another is to order the components of U based on their (standardized) magnitudes, denoted as aT-Ord. For the score test, we first take a transformation: $T^s = V^{-1/2} U$. Since $T^s \sim N(\theta_0, I)$, we can order the components of U based on $|T_j^s|$ in a descending order. For the other three tests, it is not clear how to optimally order the components of the score vector U;

for the aSSU-Ord, aSSUw-Ord and aSum-Ord tests, we simply order the components of U based on the magnitudes of $|U_j|$, $|U_j|/\sqrt{v_j}$ and $U_j/\sqrt{v_j}$ respectively.

We can also accommodate any weighting scheme as in the VT test of Price et al (2010). If there is any prior, say w_j, on the likelihood of RV j to be functional, e.g., based on computational predictions, we can incorporate such weights in the above tests: we can simply replace X_{ij} by $X_{w,ij} = w_j X_{ij}$. Specifically, denote $W = \mathrm{Diag}(w_1, \ldots, w_k)$ be a diagonal matrix with the weights $w_j > 0$ as diagonal elements. It is easy to verify that the corresponding score vector and its covariance matrix are

$$U_w = \sum_{i=1}^{n} (Y_i -)X_{w,i} = WU, \quad V_w = Cov(U_w) = WVW.$$

Consequently, we have

$$
\begin{aligned}
T_{Score}(U_w, V_w) &= U_w' V_w^{-1} U_w = U'V^{-1}U = T_{Score}(U, V), \\
T_{SSU}(U_w, V_w) &= U'W^2 U, \\
T_{SSUw}(U_w, V_w) &= U_w' V_{w,d}^{-1} U_w = U_w'(WV_d W)^{-1} U_w = T_{SSUw}(U, V), \\
T_{Sum}(U_w, V_w) &= 1'U_w/\sqrt{1'V_w 1} = 1'WU/\sqrt{1'WVW1}.
\end{aligned}
$$

Hence, the score and SSUw tests are invariant to weighting while the other two tests are not. We also note that, by the correspondence between the Sum test and $z(h)$, the aSum-MAF test is similar to the VT test of Price et al (2010).

Simulated data

We conducted simulation studies to evaluate and compare the performance of various tests. The simulated data were generated as in Wang and Elston (2008). Specifically, we simulated k SNVs with the sample size of 500 cases and 500 controls. Each SNV had a mutation rate or MAF uniformly distributed in a small interval, e.g. between 0.001 and 0.01. 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 between 0.001 and 0.01, or between 0.001 and 0.005 (see below for more details). 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 two simulation set-ups. In Case I, all SNVs were independent (i.e. in linkage equilibrium) with $\rho = 0$, and we used a common association strength OR=3 for each of the eight causal RVs. To mimic the situation that the MAFs of causal RVs differ from those of non-functional ones, the MAFs of the causal RVs were randomly drawn from a uniform distribution between 0.001 and 0.005, $U(0.001, 0.005)$, while the MAFs for non-functional RVs were from $U(0.001, 0.01)$. We also generated a predicted score for the likelihood of each RV's being functional or causal. To reflect the practical performance of existing computational algorithms (Wei et al 2010), the score for a causal RV was randomly generated from $U(0.4, 0.8)$ while that for a non- functional RV was from $U(0.2, 0.6)$.

Although such scores were informative in the sense that the score for a functional RV tended to be higher than that of a neutral RV, there was a non-ignorable probability that the score of a neutral RV could be higher than that of a causal one.

The simulation set-up for Case II was similar to a set-up in Basu and Pan (2011). It differed from Case I in three aspects. First, there was no distributional difference of MAFs between causal and non-causal RVs: the MAF of any RV was randomly generated from $U(0.001, 0.01)$. Second, within each of the two groups of the 8 causal RVs and other non-functional RVs, there was LD with $\rho = 0.9$, though there was no LD between the two groups. Third, we also used varying ORs with $OR = (3, 1/3, 2, 2, 2, 1/2, 1/2, 1/2)'$ for the eight causal SNVs to reflect possibly different association directions and association strengths.

Throughout the simulations, we fixed the test significance level at $\alpha = 0.05$, and used $B = 200$ permutations for each permutation-based method. The results were based on 1000 independent replicates for each set-up.

RESULTS

Simulation Case I

We first consider the case with no LD between any two RVs, mimicking the situation where mutations are all completely random and independent of each other. We applied the tests with or without weighting with predicted functional scores. Table 1 shows that in general all the tests had satisfactory Type I error rates. Although some tests had slightly inflated Type I error rates, it was likely due to relatively larger variations caused by the small number of permutations, $B = 200$; there is no theoretical reason that permutation tests would not work here. In practice, we can always use a larger B to improve the accuracy.

For the non-null case that the eight causal RVs shared a common OR (Table 2), which is ideal for the pooled association tests, the Sum test or adaptive Sum (aSum) test was most powerful if there were no or few non-functional RVs. As the number of non-functional RVs increased, the SSU or its adaptive versions (aSSU) gradually caught up and became the most powerful.

For the simulated data, since the chromosome locations of the causal RVs were random, using the location information in the adaptive tests (e.g. the aSSU-Loc test) did not lead to improved power as compared to the original (unadaptive) tests (e.g. the SSU test). On the other hand, since the causal RVs had slightly lower MAFs, an adaptive test with the MAF as VT improved over the original test as the number of non-functional RVs increased. It is confirmed that the aSum-MAF and Price's tests gave almost equal power.

It is noted that ordering the components of the score vector by their standardized magnitudes might improve the power of the adaptive tests, more so as the number of non-functional RVs increased. For example, without weighting and with 64 neutral RVs, the power of the aSSU-Ord was 0.330, compared to 0.265 of the SSU test. Most strikingly, the power of the aSum-Ord test was dramatically boosted, which however depended on the assumption that the causal RVs were all associated with the trait in the same direction.

It is interesting to see that, with partially informative weights, weighting could dramatically improve the power of the SSU and Sum tests and their adaptive versions, while the Score and SSUw tests and their adaptive versions were invariant to weighting as shown theoretically.

Simulation Case II

Now we consider the case where causal RVs were correlated, non-functional RVs were also correlated, but there was no LD between causal and non-functional RVs; the simulation set-up is similar to that of Basu and Pan (2011). For the null case (Table 3), again all the tests had satisfactory Type I error rates.

For the power comparison, due to the presence of the causal RVs with opposite association directions, even with no or few non-functional RVs, the pooled association tests, including the Sum, aSum and Price's test, did not perform well with low power. In particular, differing from that in Case I, here even the aSum-Ord test was not the winner, though it had much improved power over the Price et al's test, the Sum test and other versions of the aSum test; similarly, even though the chromosome locations of causal RVs were random, the aSum-Loc test was more powerful than the Sum test since the former adaptively determined which RVs to be pooled to maximize the use of signals. In general, the SSU test or an aSSU test was the most powerful, more so with larger numbers of non-functional RVs. Since for the simulated data there was no difference of MAFs between causal and non-causal RVs, the adaptive tests based on MAF did not improve over the original tests. Impressively, the power of the aSum-Ord was much closer to that of the winners, and in particular, much higher than the Sum test and the VT test of Price et al.

Similar results (not shown) were obtained for two other simulation set-ups as used in Basu and Pan (2011).

DISCUSSION

We have proposed and studied a class of adaptive tests, which generalize the idea of the VT test of Price et al (2010). In particular, our proposed aSSU test, as the SSU test, is robust to the large number of neutral RVs and to causal RVs with opposite association directions, while the aSum-Ord test maintains high power if there are a large number of neutral RVs, and it substantially improves over other pooled association tests when there are causal RVs with opposite association directions, thus largely overcoming the common weakness of existing pooled association tests, including the VT test of Price et al (2010). In addition, rather than using the MAF as the variable thresholds as in the VT test of Price et al, we propose using some other criteria as variable thresholds. In particular, a version of our proposed adaptive Sum test using a standardized magnitude of the components of the score vector as variable thresholds, the aSum-Ord test, can largely boost the power to be much higher than that of other pooled association tests, including the VT test, in the case with multiple neutral RVs and/or with causal RVs with opposite association directions. We have also demonstrated the improved performance of the proposed tests with imperfect weighting schemes. Due to their superior performance in differing situations, we recommend the use of the SSU, aSSU and aSum tests.

It is noted that our proposed methods can incorporate the use of computational predictions for functional RVs (Adzhubei et al 2010; Tavtigian et al 2008). As illustrated in Price et al (2010), some predicted scores representing the likelihoods of RVs' being functional, either deleterious or protective, can be used as the weights in a pooled or adaptive pooled association test. Since current algorithms cannot give perfect predictions (Wei et al 2010), it is both informative and robust to use such scores as weights. In our simulations, we have demonstrated substantial power gains of our proposed adaptive tests when such informative but not-perfect weights are used in the presence of neutral RVs. As discussed by Madsen and Browning (2009) for other pooled association tests, our proposed adaptive tests can be applied to both candidate regions or genome scans; in the latter case, the analysis unit can be based on genes, other functional units or simply some sliding windows. The performance of

any test may depend on the choice of weights and analysis units; practical guidelines are needed and may be gained through future applications to real sequence data.

Recently some new tests have been proposed that overcome the weakness of pooled association tests, including a test of Neale et al (2010) and variable selection-based approaches (Hoffmann et al 2010; Basu and Pan 2011). In a numerical comparison, Basu and Pan (2011) found that Neale's test (and a kernel method) performed similarly to the SSU test while variable selection approaches tended to perform robustly but less well. Since our proposed adaptive tests could improve over the original SSU test as shown here, we expect our proposed adaptive tests to be competitive as compared to the other tests or their extensions with variable thresholds, though more studies are needed. It is noted that the adaptive Neyman's test and thus our proposed adaptive tests can be regarded as a restrictive form of sequential variable selection to determine which of the first few components/RVs to be included to construct a test; in particular, the aSum test is in the same spirit of a sequential variable selection method (called Seq-aSum) of Basu and Pan (2011), but differs with different selection criteria. Due to the low MAF of any RV, there is only quite limited information contained in each RV, thus it may not be productive to conduct full variable selection of RVs. The restrictive form of variable selection in our proposed adaptive tests differs from other more direct variable selection or weighting schemes implemented in other methods (Bhatia et al 2010; Han and Pan 2010; Hoffmann et al 2010; Yi and Zhi 2011). In addition, we emphasize that our proposed adaptive tests are for the case with sequence data, though it seems straightforward to extend them to the case with GWAS data to infer association with relatively more frequent RVs in the MAF range of 0.1%–5% (Zhu et al 2010; Li et al 2010). Finally, we point out that the proposed adaptive tests can be easily extended to the case with covariates since the tests are based on logistic regression for binary traits, or generalized linear models for other types of traits.

R code will be posted on our web site at http://www.biostat.umn.edu/~weip/prog.html.

Acknowledgments

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7 248249.

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

Basu, S.; Pan, W. Division of Biostatistics. University of Minnesota; 2011. Comparison of Statistical Tests for Association with Rare Variants. Research Report 2011-010. Available online at http://www.biostat.umn.edu/~weip/paper/RV1.pdf

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]

Fan J. Test of significance based on wavelet thresholding and Neyman's truncation. JASA. 1996; 91:674–688.

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

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]

Han F, Pan W. A data-adaptive sum test for disease association with multiple common or rare variants. Hum Hered. 2010; 70:42–54. [PubMed: 20413981]

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

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]

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]

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. 2010 Manuscript.

Neyman J. Smooth test for goodness of fit. Skandinavisk Aktuarietidskrift. 1937; 20:149–199.

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

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, Cox NJ. The allelic architecture of human disease genes: common disease-common variant… or not? Hum Mol Genet. 2002; 11:2417–2423. [PubMed: 12351577]

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]

Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Human Mutation. 2008; 29:1342–1354. [PubMed: 18951461]

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]

Wei P, Liu X, Fu YX. A comparative study of incorporating predicting functions of nonsynonymous variants into gene-based analysis of exome sequencing data. 2010 Manuscript.

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]

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

Type I error rates for simulation Case I with 8 "causal" RVs and a number of other non-functional RVs. There is no LD among the RVs.

| Test | No weighting | | | | | | With weighting | | | | | |
| | # of non-functional RVs | | | | | | # of non-functional RVs | | | | | |
	0	8	16	32	64	0	8	16	32	64
Score	.036	.028	.040	.034	.031	.036	.028	.040	.034	.031
SSU	.047	.051	.049	.040	.047	.043	.044	.050	.040	.053
SSUw	.037	.036	.037	.036	.028	.037	.036	.037	.036	.028
Sum	.046	.036	.052	.058	.036	.047	.037	.048	.062	.037
aScore-Loc	.066	.050	.050	.051	.052	.070	.051	.048	.052	.053
aSSU-Loc	.066	.049	.059	.045	.043	.062	.053	.065	.054	.042
aSSUw-Loc	.074	.055	.058	.049	.046	.067	.063	.055	.051	.052
aSum-Loc	.054	.048	.062	.054	.041	.054	.045	.055	.049	.042
aScore-MAF	.073	.046	.059	.050	.050	.062	.053	.057	.048	.052
aSSU-MAF	.069	.049	.063	.045	.042	.067	.056	.062	.036	.037
aSSUw-MAF	.074	.051	.058	.043	.045	.064	.052	.054	.049	.054
aSum-MAF	.057	.052	.057	.062	.039	.053	.046	.048	.062	.045
Price-VT	.054	.055	.056	.061	.038	.055	.049	.055	.059	.047
aScore-Ord	.054	.048	.059	.036	.044	.055	.047	.063	.039	.050
aSSU-Ord	.064	.056	.062	.043	.053	.060	.046	.065	.039	.051
aSSUw-Ord	.056	.057	.050	.045	.041	.057	.046	.057	.050	.047
aSum-Ord	.052	.057	.054	.042	.059	.060	.055	.057	.048	.054

Table 2

Empirical power for simulation Case I with 8 causal RVs with a common association strength *OR* = 3 and a number of non-functional RVs. There is no LD among the RVs.

Test	No weighting					With weighting				
	# of non-functional RVs					# of non-functional RVs				
	0	8	16	32	64	0	8	16	32	64
Score	.823	.693	.570	**.397**	.269	.823	.693	.570	.397	.269
SSU	.851	.632	.515	.330	.265	.822	.744	.695	.570	**.473**
SSUw	.816	.689	**.569**	.396	**.297**	.816	.689	.569	.396	.297
Sum	**.983**	**.773**	.568	.369	.206	**.982**	**.888**	**.761**	**.577**	.378
aScore-Loc	.780	.576	.434	.285	.213	.785	.590	.434	.292	.208
aSSU-Loc	.807	.580	.452	.293	.232	.773	.683	.636	**.476**	**.393**
aSSUw-Loc	.805	.634	**.510**	**.374**	**.279**	.799	.642	.519	.372	.282
aSum-Loc	**.955**	**.713**	.492	.325	.163	**.949**	**.827**	**.665**	.473	.302
aScore-MAF	.796	.685	.574	.437	.319	.800	.688	.579	.442	.319
aSSU-MAF	.807	.680	.582	**.451**	**.353**	.777	.741	.713	**.632**	**.564**
aSSUw-MAF	.793	.687	.574	.439	.339	.796	.689	.571	.447	.335
aSum-MAF	**.961**	**.818**	**.612**	.411	.237	**.953**	**.889**	**.761**	.588	.413
Price-VT	.958	.813	.609	.407	.236	.952	.888	.757	.591	.416
aScore-Ord	.747	.623	.548	.413	.345	.751	.622	.563	.403	.353
aSSU-Ord	.765	.639	.551	.401	.330	.737	.658	.601	.498	.437
aSSUw-Ord	.788	.680	.613	.459	.391	.799	.673	.604	.450	.391
aSum-Ord	**.962**	**.916**	**.819**	**.698**	**.537**	**.957**	**.922**	**.852**	**.753**	**.623**

Table 3

Type I error rates and power for simulation Case II with 8 "causal" RVs with varying ORs and a number of other non-functional RVs. There is LD among the 8 "causal" RVs and among other non-functional RVs.

Test	Type I error					Power				
	# of non-functional RVs					# of non-functional RVs				
	0	8	16	32	64	0	8	16	32	64
Score	.029	.029	.028	.019	.021	.631	.480	.373	.241	.160
SSU	.049	.051	.035	.034	.034	**.642**	**.553**	**.475**	**.444**	**.334**
SSUw	.045	.040	.027	.015	.036	.562	.450	.352	.272	.187
Sum	.046	.059	.046	.046	.046	.345	.229	.159	.110	.079
aScore-Loc	.045	.060	.054	.044	.037	.544	.399	.308	.203	.143
aSSU-Loc	.063	.064	.049	.050	.041	**.611**	**.502**	**.460**	**.420**	**.319**
aSSUw-Loc	.061	.062	.049	.052	.042	.582	.451	.358	.305	.215
aSum-Loc	.055	.066	.047	.058	.052	.433	.310	.237	.171	.130
aScore-MAF	.065	.049	.039	.044	.059	**.595**	.455	.359	.238	.167
aSSU-MAF	.061	.042	.044	.035	.057	.585	**.470**	**.415**	**.356**	**.282**
aSSUw-MAF	.063	.045	.046	.034	.062	.528	.400	.324	.246	.182
aSum-MAF	.055	.062	.053	.044	.067	.298	.156	.130	.078	.068
Price-VT	.058	.065	.045	.051	.047	.334	.210	.140	.100	.078
aScore-Ord	.056	.053	.051	.049	.039	.553	.459	.416	.335	.262
aSSU-Ord	.066	.047	.045	.045	.046	.598	.511	**.464**	**.424**	**.337**
aSSUw-Ord	.069	.045	.043	.043	.048	**.604**	**.519**	.457	.408	.320
aSum-Ord	.047	.058	.043	.053	.045	.597	.503	.418	.374	.273