

Statistical Tests of Genetic Association in the Presence of Gene-Gene and Gene-Environment Interactions

WEI PAN

*Division of Biostatistics, School of Public Health, University of Minnesota,
Minneapolis, MN 55455*

May 11, 2009; revised June 25, July 25, 2009

Correspondence author: Wei Pan

Telephone: (612) 626-2705

Fax: (612) 626-0660

Email: weip@biostat.umn.edu

Address: Division of Biostatistics, MMC 303,
School of Public Health, University of Minnesota,
Minneapolis, Minnesota 55455-0392, U.S.A.

ABSTRACT

Background: While its importance is well recognized, it remains challenging to test genetic association in the presence of gene-gene (or gene-environment) interactions. A major technical difficulty lies in the fact that a general model of gene-gene interactions calls for the use of often a large number of parameters, leading to possibly reduced statistical power. An emerging theme of some recent work is to reduce the number of such parameters through dimension reduction. Wang et al (2009) proposed such an approach based on the partial least squares (PLS) for dimension reduction. They compared their method with several others using simulated data, establishing that their PLS test performed best. Unfortunately, Wang et al did not include in their evaluations several powerful tests just recently discovered for analyzing multiple SNPs in a candidate gene or region.

Methods: In this paper, we first extend these tests to the current context to detect gene-gene interactions in the presence of nuisance parameters, then compare these tests with the PLS test using the simulated data of Wang et al (2009).

Results: It is confirmed that some other tests can be more powerful than the PLS test, though there is no uniform winner. Some interesting, albeit not new, observations are also made: some of the new tests are more robust to the large number of parameters in a model and may thus perform well; on the other hand, even for a purely epistatic genetic model, some of the tests applied to a logistic main-effects model without any interaction terms may be superior to that based on a full model that explicitly accounts for gene-gene interactions.

Conclusion: The proposed statistical tests are potentially useful in practice.

Key Words: Epistasis; Genome-wide association study; Logistic regression; Main-effects model; Score test; SNP; Sum of squared score tests.

1 Introduction

Genotyping a large number of single nucleotide polymorphisms (SNPs) has made it possible to detect genetic variants associated with human diseases and other complex traits. Even if a disease-causing genetic variant is not genotyped, due to linkage disequilibrium (LD), its nearby SNPs may still be associated the disease. Most existing statistical methods of detecting genetic association focus on analyzing SNPs one-by-one or testing multiple SNPs inside a candidate gene region (Clayton et al 2004; Roeder et al 2005) while ignoring other genes or environmental factors. However, it has been recognized that such an approach may not be optimal in the presence of gene-gene or gene-environment interactions, which are expected to be commonly present (e.g., Moore 2003; Millstein et al 2006; Zhao et al 2006; Zheng et al 2006; Zhang and Liu 2007; Lo et al 2008; Kooperberg and LeBlanc 2008, and references therein). Some of the existing studies (e.g. Kooperberg and LeBlanc 2008) aim to detect only interactions either by significance testing (e.g. Kooperberg and LeBlanc 2008) or data mining (reviewed by Musani et al 2007), which differs from *the goal here to conduct significance testing to detect either main- or epistatic-effects of a gene of primary interest in the presence of its interaction with other genes or environmental factors*. With multiple SNPs genotyped inside each gene region, there is a technical difficulty in dealing with gene-gene or gene-environment interactions: to model interactions in a general way induces often a large number of parameters. For example, suppose that there are k_g SNPs in candidate gene g for $g = 1$ or 2 , and we assume an additive main effect and additive-additive epistasis, then there are in total $k_1 k_2$ two-way interaction terms between the two genes, in contrast to only k_1 and k_2 main effects of the two genes respectively. In general, there may be high cost associated with estimating or testing on a large number of parameters: the resulting test may lose power due to a large number of degrees of freedom (DF) (or multiple testing adjustment). Hence, recent work in this area aims to increase the statistical efficiency or reduce the model complexity by exploring gene-gene or gene-environment independence (Piegorsch

et al 1994; Chatterjee and Carroll 2005; Mukherjee and Chatterjee 2008), by restricting the parameter space (Wang 2008; Song and Nicolae 2008), or more directly, by reducing the number of parameters for interaction terms while accounting for interactions to some extent. Chatterjee et al (2006) proposed a latent model to reduce the extra number of parameters for multiple interactions to only one in a logistic regression model; their method is related to Tukey’s 1-DF test. Wang et al (2009) proposed a new dimension-reduction method based on the partial least squares (PLS) to reduce the number of interaction terms to only one. Using simulated data, Wang et al showed that their PLS test was more powerful than that of Chatterjee et al (2006) and the multivariate score (or Wald or likelihood-ratio) test based on either a main-effects or a full logistic regression model with all two-way interactions. Unfortunately, they did not compare their method with some recently emerging tests shown to be powerful in LD mapping with multiple SNPs in a gene region (Pan 2009), namely an empirical Bayes test (Goeman et al 2006) and its closely related sum of squared score (SSU and SSUw) tests, and a so-called minP test that combines single-locus tests by taking the minimum of their p-values. Because the SSU and SSUw tests perform similarly to the Goeman’s test, while the former two, with their known asymptotic null distributions, do not require the use of permutations or simulations to calculate p-values (Pan 2009), we will focus on the former two tests, along with the minP test. From a technical point, since the SSU and SSUw tests have not been applied to the situation in the presence of some nuisance parameters under the null hypothesis, corresponding to the main-effects of gene two or of environments in the current context, we extend their use here. From a more practical point, we wish to investigate the robustness of such tests to larger numbers of parameters for interaction terms than that in main-effects models as previously used.

There are two main arguments for developing new methods to deal with gene-gene interactions. First, by ignoring gene-gene interactions such as in a main-effects logistic regression model, a test focuses on the marginal effects of the gene of interest (say, gene

one) and may have low power if the marginal effects do not exist or are weak. Second, if one takes account of gene-gene interactions by including many interaction terms in a regression model, because of the large number of interaction terms, a resulting test with large DF may lose power. Although the above two arguments are intuitively reasonable, as to be shown here, they may not always hold, depending on which test is to be applied. Here is a sketch of our arguments. First, we will show that a score statistic for main-effects (with uncentered genotype scores) can indirectly test on interaction effects, and similarly, a score statistic for interactions may also indirectly contribute to testing on the main-effects. Hence, as to be verified in our numerical examples, even for a purely epistatic genetic model, some tests based on a main-effects model containing no interaction terms may turn out to be more powerful than those aiming to explicitly account for the presence of gene-gene interactions, such as the PLS test. Second, although a large model containing too many nuisance parameters in general leads to reduced power, some tests, such as the SSU and SSUw tests, are more robust to the large number of parameters; in fact, the above two tests can be motivated as an empirical Bayes approach to testing in a high-dimensional parameter space (Goeman et al 2006; Pan 2009). We will show that the above two tests based on an expanded regression model with a large number of interaction terms could be more powerful than those based on dimension reduction, such as the PLS test, though there is no uniform winner.

2 Methods

2.1 Data and Models

To be concrete, we only consider the case-control study design with a binary trait and some marker genotypes in two candidate genes. The question of interest is whether a gene of primary interest, say gene 1, is associated with the trait. Although gene 2 (or an environmental factor) is of no direct interest, we may need to account for its possible

effects (as nuisance parameters).

For any subject $i = 1, \dots, m$, the trait or response is $Y_i = 0$ or 1 , and the marker SNPs in the two genes are coded as $X_{i1} = (X_{i1,1}, \dots, X_{i1,k_1})'$ and $X_{i2} = (X_{i2,1}, \dots, X_{i2,k_2})'$ respectively. In this article, we consider only the dosage coding of genotypes based on an additive inheritance mode: the genotype score $X_{i,g,j} = 0, 1$ or 2 , representing the copy number of one of the two alleles present at locus j of gene g of subject i , though other coding schemes can be equally used.

Given the data for m independent subjects, we would like to test whether there is any association between the trait and gene 1 in the possible presence of interactions between gene 1 and gene 2. For this purpose, a full two-way interaction (Full) logistic regression model can be considered:

$$\text{Logit Pr}(Y_i = 1) = \beta_0 + \sum_{j=1}^{k_1} X_{i1,j} \beta_{1,j} + \sum_{j=1}^{k_2} X_{i2,j} \beta_{2,j} + \sum_{j=1}^{k_1} \sum_{l=1}^{k_2} X_{i1,j} X_{i2,l} \beta_{12,jl}, \quad (1)$$

and we would like to test the null hypothesis $H_{0,1}$: $\beta_1 = (\beta_{1,1}, \beta_{1,2}, \dots, \beta_{1,k_1})' = 0$ and $\beta_{12} = (\beta_{12,11}, \dots, \beta_{12,k_1 k_2})' = 0$ versus $H_{1,1}$: $\beta_1 \neq 0$ or $\beta_{12} \neq 0$. Throughout this article, we refer β_1 as *main-effects* and β_{12} as *interacting- or epistatic-effects*. Note that β_2 is a vector of nuisance parameters: although β_2 is of no direct interest, we need to estimate β_2 even under $H_{0,1}$ to draw inference on β_1 and β_{12} . The goal here is to detect any effect of the gene of primary interest, either marginal or epistatic, which differs from that of detecting only interacting effects (i.e. with a null hypothesis $\beta_{12} = 0$) as discussed by Kooperberg and LeBlanc (2008).

A possible drawback of the above Full model is the large number of the interaction parameters β_{12} that need to be estimated (or tested on). Alternatively, we can consider a simpler main-effects (Main) logistic regression model:

$$\text{Logit Pr}(Y_i = 1) = \beta_0 + \sum_{j=1}^{k_1} X_{i1,j} \beta_{1,j} + \sum_{j=1}^{k_2} X_{i2,j} \beta_{2,j}, \quad (2)$$

and then test the null hypothesis $H_{0,2}$: $\beta_1 = 0$ versus $H_{1,2}$: $\beta_1 \neq 0$. Note that our targeted null hypothesis is $H_{0,1}$. Rejecting $H_{0,2}$ would imply rejecting $H_{0,1}$, but $H_{0,2}$ may

hold even if $H_{0,1}$ is false. Hence, if the main effects are weak while the interaction effects are strong, using the main-effects model may lose power. Here the main-effects β_1 is also referred as *marginal effects* (after adjusting for gene 2).

Intermediate between the above two extremes is to first conduct dimension reduction, then replace the multiple interaction terms in (1) by a single interaction between two composite predictors Z_1 and Z_2 :

$$\text{Logit Pr}(Y_i = 1) = \beta_0 + \sum_{j=1}^{k_1} X_{i1,j} \beta_{1,j} + \sum_{j=1}^{k_2} X_{i2,j} \beta_{2,j} + \gamma Z_{i1} Z_{i2}. \quad (3)$$

Chatterjee et al (2006) proposed using Tukey's 1-DF model with $Z_{ig} = \sum_{j=1}^{k_g} X_{ig,j} \beta_{g,j}$ for $g = 1$ and 2; Wang et al (2009) proposed using Z_{i1} as the first partial least squares (PLS) component of regressing X_{i1} on (Y_i, X_{i2}) , and Z_{i2} as the fitted values of regressing Y_i on X_{i2} . As another option, one can use the first principle component (PC) of X_{i1} as Z_{i1} , and the first PC of X_{i2} as Z_{i2} . For the reduced model (3), one would like to test $H_{0,3}$: $\beta_1 = 0$ and $\gamma = 0$ versus $H_{1,3}$: $\beta_1 \neq 0$ or $\gamma \neq 0$. The results of simulation studies in Wang et al (2009) indicated that, among the three approaches, the PLS method was the winner.

2.2 Statistical Tests

2.2.1 Multivariate score test

To test $H_{0,1}$ in the Full model, a multivariate score test can be applied. First, the full score vector under $H_{0,1}$ is

$$U_f = \begin{pmatrix} U_{\beta_1} \\ U_{\beta_{12}} \\ U_{\beta_2} \end{pmatrix} = \sum_{i=1}^m (Y_i - \hat{\mu}_{0i}) X_i = \begin{pmatrix} \sum_{i=1}^m (Y_i - \hat{\mu}_{0i}) X_{i1} \\ \sum_{i=1}^m (Y_i - \hat{\mu}_{0i}) X_{i12} \\ \sum_{i=1}^m (Y_i - \hat{\mu}_{0i}) X_{i2} \end{pmatrix}, \quad (4)$$

where $X_i = (X'_{i1}, X'_{i12}, X'_{i2})'$ is the vector of covariates for subject i with $X_{i12} = (X_{i1,1} X_{i2,1}, \dots, X_{i1,k_1} X_{i2,k_2})'$ as the vector of all two-way interactions, and $\hat{\mu}_{0i}$ is the maximum likeli-

hood estimate (MLE) of $\Pr(Y_i = 1)$ under $H_{0,1}$; that is,

$$\text{Logit}(\hat{\mu}_{0i}) = \hat{\alpha}_0 + \sum_{j=1}^{k_2} X_{i,2j} \hat{\alpha}_{2j},$$

where $\hat{\alpha}_0$ and $\hat{\alpha}_{2j}$'s are the MLEs of the regression coefficients in the above logistic model. The covariance matrix of U_f can be consistently estimated by the expected Fisher information matrix

$$V_f = \sum_{i=1}^m \hat{\mu}_{0i}(1 - \hat{\mu}_{0i})(X_i - \bar{X})(X_i - \bar{X})',$$

where $\bar{X} = \sum_{i=1}^m X_i/m$.

To test $H_{0,1}$, we use the score vector $U_{\beta_1, \beta_{12}} = (U'_{\beta_1}, U'_{\beta_{12}})'$, the first $(k_1 + k_1 k_2)$ components of U_f , and construct the test statistic

$$T_{F-Sco} = U'_{\beta_1, \beta_{12}} V_{\beta_1, \beta_{12}}^{-1} U_{\beta_1, \beta_{12}},$$

where $V_{\beta_1, \beta_{12}}^{-1}$ is the upper-left submatrix of V_f^{-1} with size $(k_1 + k_1 k_2) \times (k_1 + k_1 k_2)$ (Cox and Hinkley 1974). Note that, with the presence of nuisance parameter β_2 , we cannot use a formula similar to V_f to directly estimate $Cov(U_{\beta_1, \beta_{12}})$. Under $H_{0,1}$, T_{F-Sco} has an asymptotic chi-squared distribution χ_r^2 with degrees of freedom $r = \text{rank}(V_{\beta_1, \beta_{12}})$, typically $r = k_1 + k_1 k_2$. A potential problems with the test is that, for a large $k_1 k_2$, the test can be low-powered because of the cost of the large DF.

Similarly, to test $H_{0,2}$, we use the score vector U_{β_1} , the first k_1 components of U_f ; the covariance matrix of U_{β_1} is V_{β_1} , the upper-left submatrix of $V_{\beta_1, \beta_{12}}$ with size $k_1 \times k_1$. The test statistic is

$$T_{M-Sco} = U'_{\beta_1} V_{\beta_1}^{-1} U_{\beta_1}.$$

Under $H_{0,2}$, T_{M-Sco} has an asymptotic chi-squared distribution χ_r^2 with degrees of freedom $r = \text{rank}(V_{\beta_1})$, typically k_1 .

2.2.2 Combining single-locus-based tests

The multivariate score test involves the use of a possibly large covariance matrix, which may cause problems. A single-locus-based analysis would first test individual loci one-by-one, e.g., by univariate score tests, then combine the individual tests (e.g. Roeder et al 2005). The most popular is the so-called minP test that combines the individual tests by taking the minimum of their p-values. Specifically, to test $H_{0,1}$, we can combine individual score tests by

$$T_{F-minP} = \max (U_{\beta_1, \beta_{12}}^2 \text{Diag}(V_{\beta_1, \beta_{12}})^{-1}) = \max_{1 \leq j \leq k_1 + k_1 k_2} U_j^2 / v_j,$$

where $\text{Diag}(M)$ denotes a diagonal matrix with the diagonal elements of M , U_j is the j th component of $U_{\beta_1, \beta_{12}}$ and $v_j = \text{Var}(U_j)$ is the j th diagonal element of $V_{\beta_1, \beta_{12}}$ for $j = 1, \dots, k_1 + k_1 k_2$. That is, T_{F-minP} is the maximum of the individual standardized score statistics U_j^2 / v_j . Equivalently, we could first obtain a p-value, say p_j , by referring U_j^2 / v_j to χ_1^2 , then take $p = \min_j \{p_j\}$; however, p is no longer a p-value, and some multiple test adjustment has to be made.

Similarly, to test $H_{0,2}$, we use

$$T_{M-minP} = \max (U_{\beta_1}^2 \text{Diag}(V_{\beta_1})^{-1}).$$

A multiple test adjustment based on either permutation or the Bonferroni method is commonly used. Because the Bonferroni adjustment is known to be conservative, it is more common to use a permutation method by permuting Y , which however is computationally demanding for its requirement of fitting models (or re-calculating test statistics) repeatedly. Here we propose using simulations to estimate the null distribution (Seaman and Muller-Myhsok 2005). First, we note that each of the two minP tests is based on a component of or the whole score vector $U_{\beta_1, \beta_{12}}$. Second, because of the asymptotic null distribution of $U_{\beta_1, \beta_{12}}$ is known to be $U_{\beta_1, \beta_{12}} \sim N(0, V_{\beta_1, \beta_{12}})$, we can simulate B iid copies of $U^{(b)}$'s from $N(0, V_{\beta_1, \beta_{12}})$ with $b = 1, 2, \dots, B$. Based on each $U^{(b)}$,

we can calculate $T_{F-minP}^{(b)}$ and $T_{M-minP}^{(b)}$. Third, the p-values for the two tests are simply $\sum_{b=1}^B I[T_{F-minP} < T_{F,minP}^{(b)}]/B$ and $\sum_{b=1}^B I[T_{M-minP} < T_{M,minP}^{(b)}]/B$, respectively. We used $B = 1000$ throughout.

2.2.3 Sum of squared score tests: SSU and SSUw

To test $H_{0,1}$ with the Full model, in contrast to the usual multivariate score test with statistic $T_{F-Sco} = U'_{\beta_1, \beta_{12}} V_{\beta_1, \beta_{12}}^{-1} U_{\beta_1, \beta_{12}}$, an alternative is to simply use the sum of the squared scores as the test statistic

$$T_{F-SSU} = U'_{\beta_1, \beta_{12}} U_{\beta_1, \beta_{12}} = \sum_{j=1}^{k_1+k_1k_2} U_j^2,$$

ignoring the covariance matrix of the score vector $U_{\beta_1, \beta_{12}}$. This test is related to the test of Goeman et al (2006); in fact, the above SSU test is equivalent to the permutation-based version of the Goeman test. Although the Goeman test was derived under an empirical Bayes framework to test on a large number of parameters, as arising in microarray gene expression data, it has been shown empirically that the Goeman test works impressively well across a wide range of scenarios.

A weighted form of the above test is the SSUw

$$T_{F-SSUw} = U'_{\beta_1, \beta_{12}} \text{Diag}(V_{\beta_1, \beta_{12}})^{-1} U_{\beta_1, \beta_{12}} = \sum_{j=1}^{k_1+k_1k_2} U_j^2/v_j.$$

Hence, T_{F-SSUw} is simply the sum of the locus-specific univariate score statistics, in contrast to T_{F-minP} as the maximum of the univariate score statistics. The SSUw test can be interpreted as an *estimated* most powerful test (Pan 2009), which also partially explains the good performance of SSU. Often the SSU and SSUw perform similarly, but not always.

Asymptotically, each of the two test statistics is a quadratic form of Normal variates, $Q = U'_{\beta_1, \beta_{12}} W^{-1} U_{\beta_1, \beta_{12}}$, with $W = I$ or $W = \text{Diag}(V_{\beta_1, \beta_{12}})$ respectively. It is well known (e.g. Johnson and Kotz, 1970, p.150) that the distribution of Q is a weighted sum

of k independent chi-squared variates with $DF=1$, $\sum_{j=1}^k c_j \chi_1^2$, where c_j 's are the eigen values of $V_{\beta_1, \beta_{12}} W^{-1}$. Furthermore, by the results of Zhang (2005), $\sum_{j=1}^k c_j \chi_1^2$ can be well approximated by $a\chi_d^2 + b$ with

$$a = \frac{\sum_{j=1}^k c_j^3}{\sum_{j=1}^k c_j^2}, \quad b = \sum_{j=1}^k c_j - \frac{\left(\sum_{j=1}^k c_j^2\right)^2}{\sum_{j=1}^k c_j^3}, \quad d = \frac{\left(\sum_{j=1}^k c_j^2\right)^3}{\left(\sum_{j=1}^k c_j^3\right)^2}.$$

To calculate a p-value, for example for the SSU test,

$$Pr(T_{F-SSU} > s | H_0) \approx Pr(\chi_d^2 > (s - b)/a).$$

Similarly, we can construct the SSU and SSUw tests for $H_{0,2}$ for the main-effects model (2):

$$T_{M-SSU} = U'_{\beta_1} U_{\beta_1}, \quad T_{M-SSUw} = U'_{\beta_1} \text{Diag}(V_{\beta_1})^{-1} U_{\beta_1}.$$

Their null distributions can be approximated as scaled and shifted chi-squared distribution as shown above.

2.2.4 The PLS test

To test $H_{0,3}$, Wang et al (2009) proposed a likelihood ratio test (LRT) by comparing model (2) to model (3), which we call the PLS test throughout.

2.2.5 Effects of linear transformation of SNP codings

It is known that the multivariate score test is invariant to linear transformation of predictors; in fact, it is the uniformly most powerful invariant test. In contrast, the SSU, SSUw and minP tests do not possess this invariance property, which implies that the power of the tests may increase or decrease depending on what linear transformation is taken on the SNP codings. Here we illustrate this point with a simple transformation: we center each coded SNP at sample mean 0; that is, we replace $X_{ig,j}$ by its centered

version $X_{ig,j}^* = X_{ig,j} - \bar{X}_{g,j} = X_{ig,j} - \sum_{i=1}^m X_{ig,j}/m$ in Full model (1), leading to

$$\text{Logit Pr}(Y_i = 1) = \beta_0^* + \sum_{j=1}^{k_1} X_{i1,j}^* \beta_{1,j}^* + \sum_{j=1}^{k_2} X_{i2,j}^* \beta_{2,j}^* + \sum_{j=1}^{k_1} \sum_{l=1}^{k_2} X_{i1,j}^* X_{i2,l}^* \beta_{12,jl}^*, \quad (5)$$

and we would like to test the null hypothesis $H_{0,1}^*: \beta_1^* = (\beta_{1,1}^*, \beta_{1,2}^*, \dots, \beta_{1,k_1}^*)' = 0$ and $\beta_{12}^* = (\beta_{12,11}^*, \dots, \beta_{12,k_1 k_2}^*)' = 0$ versus $H_{1,1}^*: \beta_1^* \neq 0$ or $\beta_{12}^* \neq 0$. The resulting new score statistic is obtained by replacing X_i by the corresponding X_i^* in (4):

$$U_f^* = \begin{pmatrix} U_{\beta_1}^* \\ U_{\beta_{12}}^* \\ U_{\beta_2}^* \end{pmatrix} = \sum_{i=1}^m (Y_i - \hat{\mu}_{0i}) X_i^* = \begin{pmatrix} U_{\beta_1} \\ U_{\beta_{12}} - \bar{X}_2 U_{\beta_1} - \bar{X}_1 U_{\beta_2} \\ U_{\beta_2} \end{pmatrix},$$

where $\bar{X}_g = \sum_{i=1}^m X_{ig}/m$ for $g = 1$ and 2 . It is clear that, due to the fact that $U_{\beta_{12}}^* \neq U_{\beta_{12}}$ while $U_{\beta_1}^* = U_{\beta_1}$, any of T_{F-SSU} , T_{F-SSUw} and T_{F-minP} will change accordingly, while their counterparts for the main-effects model will not. In addition, because U_f^* is just a linear transformation of U_f , it is easy to see that the multivariate score test statistic will not change.

It is easy to verify that $\beta_{1,j}^* = \beta_{1,j} + \sum_{l=1}^{k_2} \bar{X}_{2,l} \beta_{12,jl}$ and $\beta_{12,jl}^* = \beta_{12,jl}$. If the SNPs in the two regions are unlinked, then the predictors in (5) are uncorrelated, and thus, as for the linear model (e.g. Tang and Siegmund 2002), the maximum likelihood estimates (MLEs) for the main-effects and interactions, $\hat{\beta}_1^*$ and $\hat{\beta}_{12}^*$, are asymptotically independent; that is, based on the multivariate score test (or its asymptotically equivalent Wald test or LRT), testing the main-effects β_1^* has nothing to do with the interaction effects β_{12}^* , and vice versa. A special case is that, even if $\beta_1 \neq 0$ and $\beta_{12} \neq 0$, it is possible, though not very likely, that β_1^* is 0 or quite small, leading to vanishing power in testing on the marginal effects β_1^* . On the other hand, because of the relation between $\beta_{1,j}^*$ and $\beta_{12,jl}$, the power of testing $\beta_{1,j}$ in (4) with the score vector U_{β_1} does depend on β_{12} , as to be verified in our numerical examples. Similarly, the power of testing β_{12} with $U_{\beta_{12}}$ depends on the value of β_1 . In summary, in the original scale of uncentered genotype scores, testing the main-effects and testing the epistatic effects are related.

3 Simulations

3.1 Simulation Set-ups

We performed a simulation study following the set-ups given in Wang et al (2009) with $k_1 = k_2 = 6$ marker SNPs in each gene. Only one disease-causing SNP u_g was assumed for each gene; u_g was assumed to be at the first position, followed by $k = 6$ marker SNPs in gene g . First, independently for each gene, we generated a latent vector from a multivariate normal distribution with a so-called AR-1-flip correlation structure; that is, the correlation between the two latent variables at loci i and j was $|\rho_{ij}| = \rho_0^{|i-j|}$ with a sign randomly determined to be either positive or negative with an equal probability; if the resulting covariance matrix was positive definite, we kept and used it; otherwise it was discarded. Second, the latent vector was dichotomized element-wise to yield a haplotype with minor allele frequencies (MAFs) randomly between 0.1 and 0.4, while the MAFs for the two disease-causing SNPs, say MAF1 and MAF2, were fixed at 0.1 or 0.2. Specifically, suppose $L = (L_0, L_1, \dots, L_6)'$ was such a simulated continuous latent vector for gene 1 with L_j for locus j ; for any $0 \leq j \leq 6$, if the MAF at locus j is f_j , then the allele at locus j is $a_j = I(L_j < \Phi^{-1}(f_j))$, where $I(\cdot)$ is the indicator function, and Φ is the cumulative distribution function (CDF) for a standard Normal variate. Third, we combined two independently generated haplotypes to obtain a vector of marker genotypes X_i as the number of the minor alleles for subject i . Fourth, the disease status Y_i of subject i was generated from a logistic regression model:

$$\text{Logit Pr}(Y_i = 1) = \theta_0 + \theta_1 I_{i,u_1} + \theta_2 I_{i,u_2} + \theta_{12} I_{i,u_1} I_{i,u_2}, \quad (6)$$

where I_{i,u_1} and I_{i,u_2} were binary indicators of the presence of the minor alleles in the two disease-causing SNPs for subject i . Note that the dominant mode of inheritance (MOI) for each causal SNP was assumed throughout in the above model (6), differing from the additive MOI for other marker SNPs used to infer disease association as implied by the dosage coding in the working models (1)-(3). We used $\theta_0 = -4$, and various values of

other parameters as in Wang et al (2009). Finally, following the case-control design, we sampled n cases (with $Y_i = 1$) and n controls (with $Y_i = 0$). We excluded the two disease-causing SNPs, supplying only $\{(Y_i, X_i) : i = 1, 2, \dots, 2n\}$ as a dataset to various statistical tests; each dataset contained $n = 400$ cases and $n = 400$ controls. For each set-up, we simulated **10000 datasets for null models**, or 1000 datasets for non-null models, to obtain an empirical size or power for each test, which was defined as the proportion of incorrectly or correctly rejecting the null hypothesis. In particular, **the Monte Carlo standard error of a Type I error estimate around 0.05 is $\sqrt{.05(1 - .05)/10000} = 0.002$** , while that of an empirical power \hat{p} is $\sqrt{\hat{p}(1 - \hat{p})/1000} \leq 0.016$.

The main reason to use the latent variables for haplotypes is to (indirectly) generate some complex LD patterns of genotypes. By varying ρ_0 for the continuous latent variables, we obtained varying LD strengths from strong to weak as used in Wang et al (2009). Specially, we considered four simulation cases: Case I with $\rho_0 = 0.95$, MAF1=0.2 and MAF2=0.1; Case II with $\rho_0 = 0.95$, MAF1=0.1 and MAF2=0.1; Case III with $\rho_0 = 0.8$, MAF1=0.2 and MAF2=0.1; and Case IV with $\rho_0 = 0.6$, MAF1=0.2 and MAF2=0.1. The LD patterns for the seven SNPs (with the first one as disease-causing) in Case I were measured by r from 10000 sets of simulated genotype scores as:

$$\begin{pmatrix} 1.000 & -0.199 & 0.168 & -0.498 & -0.552 & 0.538 & -0.395 \\ & 1.000 & -0.733 & 0.386 & 0.307 & -0.240 & 0.406 \\ & & 1.000 & -0.332 & -0.265 & 0.206 & -0.370 \\ & & & 1.000 & 0.744 & -0.583 & 0.646 \\ & & & & 1.000 & -0.728 & 0.633 \\ & & & & & 1.000 & -0.534 \\ & & & & & & 1.000 \end{pmatrix}.$$

For each simulation case, we considered several models, each with several sets of parameter values, and we averaged the powers of each test over these sets. For the null model, we used $\theta_1 = \theta_{12} = 0$ but $\theta_2 = 0$ or 2. We considered three types of non-null

Table 1: Marginal and conditional disease probabilities for Cases I, III and IV. With MAF1=0.2 and MAF2=0.1, the probabilities of $(I_{u1}, I_{u2}) = (0, 0)$, $(1, 0)$, $(0, 1)$ and $(1, 1)$ are respectively 0.5184, 0.2916, 0.1216 and 0.0684 respectively.

Model	Set-ups	$Pr(Y = 1)$	$Pr(Y = 1 (I_{u1}, I_{u2}))$			
			$(I_{u1}, I_{u2}) = (0, 0)$	$(1, 0)$	$(0, 1)$	$(1, 1)$
Null $(\theta_1 = \theta_{12} = 0)$	$\theta_2 = 0$	0.018	0.018	0.018	0.018	0.018
	$\theta_2 = 2$	0.037	0.018	0.018	0.119	0.119
Multiplicative $(\theta_{12} = 0)$	$\theta_1 = \theta_2 = 0.3$	0.022	0.018	0.024	0.024	0.032
	$\theta_1 = \theta_2 = 0.9$	0.034	0.018	0.043	0.043	0.100
Epistatic $(\theta_1 = \theta_2 = 0)$	$\theta_{12} = 1.2$	0.021	0.018	0.018	0.018	0.057
	$\theta_{12} = 2.4$	0.028	0.018	0.018	0.018	0.168
Cross-over $(\theta_1 = -0.5, \theta_2 = 0)$	$\theta_{12} = 0.5$	0.016	0.018	0.011	0.018	0.018
	$\theta_{12} = 3.2$	0.029	0.018	0.011	0.018	0.214

models: i) multiplicative (relative risk) models as discussed by Chatterjee et al 2006) with various $\theta_1 = \theta_2 > 0$ and $\theta_{12} = 0$; ii) epistatic models with $\theta_1 = \theta_2 = 0$ and various $\theta_{12} > 0$; iii) cross-over models with $\theta_1 = -0.5$, $\theta_2 = 0$ and various θ_{12} . Table 1 shows marginal and conditional disease probabilities for Case I with two sets of parameters (the smallest and the largest parameters) used for each model. **Note that the relatively small marginal probabilities of disease were due to the use of the small $\theta_0 = -4$ in the disease model (6), leading to a small background disease probability of 0.018 for subjects without any disease-causing mutations. If we use $\theta_0 = -2$, the background disease probability will be 0.119 and the marginal probabilities of disease will range from about 0.1 to 0.2 for the various models in Table 1. As expected, the change of the intercept term θ_0 would have small effects on the power of any test on other parameters in model (6); that is,**

the conclusions drawn from our simulated data should be applicable to the case with a higher disease probability.

Note that the causal SNPs were used only to generate the disease status, but were not available from the observed data for hypothesis testing.

3.2 Simulation Results

3.2.1 Type I error rate

For the null models, in which no genetic variants in gene 1 were associated with the disease while those in gene 2 might or might not, all the tests except the PLS test seemed to maintain a correct test size: their Type I error rates were close to the nominal significance level at 0.05 (Table 2). Although the test size of the PLS test was close to 0.05, it tended to be larger than the nominal level with its 95% confidence interval sometimes not covering 0.05. The slightly inflated Type I error rate of the PLS test could be due to the double use of the traits Y in both constructing and then testing the PLS component.

3.2.2 Power

We first consider Case I with strong correlations among the SNPs within the same gene (Fig 1a). With a multiplicative genetic model, the M-SSU (SSU test with the main-effects model (2), and similarly for others), M-SSUw and M-minP tests performed similarly and were winners, closely followed by the F-SSU (SSU test with the Full model (1), and similarly for others) and F-SSUw tests with only slightly lower power. Next came in descending order the F-minP, M-score and PLS tests; the F-score was the obvious loser with much lower power. We note the minimal power difference between the M-SSU (or M-SSUw) and F-SSU (or F-SSUw) tests in spite of the much larger numbers of parameters tested in the Full model than that in the main-effects model (i.e. 42 vs 6), highlighting the robustness of the SSU and SSUw tests to the large number of parameters. In contrast,

Table 2: Empirical sizes of the various tests with the nominal significance level 0.05 for simulated data.

Case	θ_2	Full model					Main-effects			
		PLS	SSU	SSUw	Sco	minP	SSU	SSUw	Sco	minP
I	0	.0551	.0482	.0493	.0505	.0518	.0485	.0475	.0532	.0509
	2	.0597	.0524	.0516	.0492	.0511	.0521	.0531	.0527	.0514
II	0	.0494	.0467	.0493	.0478	.0497	.0513	.0496	.0446	.0495
	2	.0536	.0463	.0471	.0519	.0480	.0487	.0497	.0502	.0503
III	0	.0545	.0461	.0482	.0490	.0503	.0457	.0469	.0516	.0498
	2	.0538	.0465	.0457	.0493	.0492	.0461	.0465	.0504	.0487
IV	0	.0534	.0488	.0482	.0515	.0503	.0505	.0495	.0481	.0512
	2	.0531	.0481	.0485	.0501	.0528	.0480	.0477	.0508	.0490

there was an absolute power difference around 5% between the M-minP and F-minP tests, while the difference was dramatic for the score test with the power of F-score being only a half of that of M-score. As noted by Wang et al (2009), the power difference between the PLS and M-score tests was minor, presumably due to only one extra DF in the former. However, if we compare the PLS test with the M-SSU, M-SSUw and M-minP tests, the power loss of the PLS was substantial.

When the underlying genetic model is purely epistatic, the F-minP was the winner, closely followed by F-SSUw, F-SSU and PLS, then by F-SSU, M-SSU and M-SSUw. Next came M-minP, then M-score, and finally F-score. In this situation, it is clear that either score test was low powered, but surprisingly, using the main-effects model with the M-SSU, M-SSUw and M-minP tests did not lose much power; in fact, they performed similarly to the PLS test. A possible reason is that, although there were no main effects of the causal SNPs with $\theta_1 = \theta_2 = 0$, due to the correlations between the genotypes X_1

in gene 1 and the two-way interactions X_{12} , the score statistic U_{β_1} measuring correlations between Y and X_1 , might still carry information about association between Y and X_{12} ; see section 2.2.5 for more discussions. The above argument also offers an alternative explanation to the competitiveness of the M-score test as compared to the Tukey’s 1-DF method and the PC method that explicitly account for interactions, as noted by Wang et al (2009).

When the underlying genetic model is a cross-over model, the F-minP test was the winner, closely followed by the PLS test. The F-SSUw test came up next, followed by F-SSU, then by M-SSU, M-SSUw and M-minP tests. The least powerful were clearly the F-score and M-score tests, and between the two, the F-score test was better than M-score.

If we reduced the MAF of the causal SNP in gene 1 from 0.2 to 0.1 while keeping other parameters the same, the relative performance of the methods hardly changed (Fig 1b). In particular, for both the epistatic and cross-over models, the F-minP test was a clear winner over the PLS and the other tests.

Next we consider Case III with weaker correlations among the SNPs with $\rho_0 = 0.8$, and the causal SNPs’ MAF1=0.2 and MAF2=0.1 (Fig 1c). In overall, the M-SSU and M-SSUw were clear winners across all the underlying genetic models. Specifically, first, when the underlying genetic model is multiplicative, again the M-SSU and M-SSUw tests performed similarly and were best, then closely followed by F-SSU, F-SSUw and M-minP, and then by F-minP, PLS and M-score. The F-score test was again the worst with much reduced power as compared to others. Second, when the underlying genetic model is a purely epistatic model, differing from that in Fig 1a, the M-SSU and M-SSUw tests (based on the main-effects model) were two surprising winners, closely followed by F-SSU, F-SSUw and M-minP, then in descending order by F-minP, PLS, and M-score tests. Again the F-score performed much worse. Third, when the underlying genetic model is a cross-over model, the M-SSU, M-SSUw, F-SSU, F-SSUw, M-minP and PLS

tests performed similarly with much higher power than the M-score and F-score tests, and the M-score test was much better than the F-score test.

When the correlation strengths among the SNPs within the same gene were further reduced to $\rho_0 = 0.6$ (Fig 1d), again the M-SSUw and M-score tests were winners, but the PLS test could work equally or almost equally well, closely followed by M-minP, M-SSU, F-SSU and F-SSUw, then by F-minP. The F-score test consistently had lowest power.

3.2.3 Effects of centering genotypic codings

It might be surprising that the F-SSU, F-SSUw and F-minP tests are not invariant to a linear transformation of the predictors (i.e. SNP codings). For example, compared with using the original uncentered SNP codings, the power of the F-SSU, F-SSUw and F-minP tests with centered SNPs may decrease or increase, whereas the other tests' power remains the same; the power differences for Case I are shown in Fig 2. When the underlying genetic model was a multiplicative model, the power of F-SSU and F-minP was much reduced, while that of F-SSUw was the lowest, about the same as that of F-score. The reason is that, after centering the genotype scores, the interaction terms were no longer correlated with the main-effects terms; that is, the score statistic for the interactions did not carry any information about the main effects. For the case that the underlying genetic model was purely epistatic, the power of F-SSU slightly increased, while that of F-SSUw and F-minP substantially decreased. For the case with a cross-over genetic model, centering the genotypic codings greatly improved the power of F-SSU, slightly increased the power of F-SSUw, but decreased the power of F-minP. For the last two cases, as a result, the F-SSU test replaced F-minP as the most powerful test.

Table 3: P-values of the tests with the nominal significance level 0.05 for the ALS data.

SNPs	Full model					Main-effects			
	PLS	SSU	SSUw	Sco	minP	SSU	SSUw	Sco	minP
LD blocks	1.87e-6	1.13e-7	3.88e-7	1.98e-2	3.52e-5	2.60e-7	6.18e-7	4.25e-6	2.07e-6
Single SNPs	4.63e-6					1.42e-6			

4 Example

Amyotrophic lateral Sclerosis (ALS) is a fatal neurodegenerative disease leading to paralysis and death, typically within 3-5 years from onset. Despite evidence for a role for genetics, no common genetic variants are known to link to sporadic ALS. Schymick et al (2007) conducted a genome-wide association study to identify genetic variants predisposing to developing ALS in a cohort of 276 American sporadic cases and 268 neurologically normal controls. The original study assayed 555352 unique SNPs for each of subject. By testing on each individual SNP separately, Schymick et al (2007) identified a list of 34 most significant SNPs, though none of them was statistically significant after a genome-wide multiple test adjustment.

Here, for illustration, we considered the top two of the 34 SNPs, rs4363506 and rs16984239 in intergenic region on chromosomes 10q26.13 and 2p24 respectively . We extracted 10 neighboring SNPs upstream and another 10 downstream either rs4363506 or rs16984239, then applied the default LD blocking algorithm implemented in Haploview (v4.1) (Barrett et al 2005) to each of the two 21-SNP sets for the control group. Three SNPs were included in the LD block for SNP rs4363506, while 16 were included for SNP rs16984239; see Figure 3 for their LD plots. The MAFs of the 3 SNPs in the first LD block were 0.328, 0.425 and 0.103, whereas the MAFs for the 16 SNPs in the second LD block ranged from 0.009 to 0.482 with the three quartiles as 0.136, 0.257 and 0.381

respectively. As for any multilocus analysis, we had to choose an SNP block to analyze in a region of interest. The goal was to strike a balance between including as many informative SNPs as possible and controlling for inflated DFs. Here we chose an LD block as often used in haplotype analysis, though it is not clear how to do so optimally; other strategies, such as using sliding-windows with various window sizes (e.g. Guo et al 2009 and references therein), may be adopted, though issues in high computational demand and multiple-test adjustments have to be addressed.

We wish to test whether SNP rs4363506 (or one of its nearby SNPs, either genotyped or ungenotyped) is associated with ALS, possibly considering its interactions with rs16984239 or its neighboring SNPs. We applied the tests to the two LD blocks and their p-values are shown in Table 3. For either minP test, the empirical p-value from 1000 permutation was 0.000, and thus we supplied with that based on more conservative Bonferroni adjustment. The general conclusion is that SNP rs4363506 seems to be associated with ALS based on either a main-effects or a full logistic regression model, though none would be statistically significant after a genome-wide multiple-testing adjustment. It is noted that the p-values from the SSU and SSUw tests based on the Full model were smaller than their counterparts from the main-effects model, and than that of the PLS test, suggesting possible power gain as demonstrated in simulations, but the differences were not large. Again the score test based on the full model had a much larger p-value than that from the main-effects model, confirming its degraded performance with large models. For comparison, we also gave the results based on the two SNPs (i.e. without their neighboring SNPs in the blocks), which were less significant than that from the multilocus analysis, suggesting possible power loss by ignoring multiple SNPs in LD with the SNPs of interest.

5 Discussion

We offer a summary of our main results. First, if a marker SNP is strongly correlated with a causal variant (e.g. with $\rho_0 = 0.95$ in simulations), even if the causal variant is not observed, the minP test is often most powerful. On the other hand, if the correlations between the marker SNPs and the causal variant is mild (e.g. with $\rho_0 = 0.8$), the SSU and SSUw tests based on a main-effects model are the likely winners. Furthermore, if the correlations are weak (e.g. with $\rho_0 = 0.6$), the SSUw and multivariate score tests based on a main-effects model may perform best. The above observations might be surprising, suggesting two major points: 1) The minP test, albeit sometimes called univariate (Pan 2009), is really multivariate because of its use of the max operator on multiple univariate test statistics to form a combined test statistic, whose null distribution takes account of the multivariate nature of the data. In this respect, the SSU and SSUw tests are similar. 2) Even if there are no or only weak main effects in an underlying genetic model, some tests constructed under a main-effects model, such as M-SSUw, may still be powerful; in some situations, they are even more powerful than those explicitly taking account of gene-gene interactions. Second, the power difference between SSU (or SSUw) based on a main-effects model and that based on a full model with interactions is often small, suggesting the robustness of the SSU and SSUw tests to the large number of parameters, which is in agreement to their performance in analyzing multiple SNPs in a candidate gene or region with a main-effects logistic model (Pan 2009). In contrast, the minP test is more sensitive to the number of parameters, though much better than the multivariate score test: there is always a dramatic power difference between the score test based on a main-effects model and that based on a full model, and the latter usually has the lowest power. Third, between the SSU and SSUw tests, they often perform similarly for intermediate to weak correlations among SNPs; however, the SSUw test may have higher power than the SSU test in the full model for highly correlated SNPs (e.g. with $\rho_0 = 0.95$ in Fig 1ab). Fourth, while the score test (and its asymptotically equivalent

Wald test and LRT) are well known to be invariant to a linear transformation of predictors in a regression model, the minP, SSU and SSUw tests do not maintain this invariance property. This seeming disadvantage of the latter tests may turn out to be a positive: for example, centering the genotype codings might improve the power of the tests, as shown in our simulations. A remaining challenge is to characterize the relationship between the test power and a linear transformation.

Wang et al (2009) empirically demonstrated that the PLS test was more powerful than a Tukey 1-DF test (Chatterjee et al 2006) and a PCA-based test. We used simulated data similar to that of Wang et al (2009) to show that the SSU, SSUw or minP test could perform much better than the PLS test. Another important conclusion we draw is that, even for a purely epistatic genetic model, the SSU, SSUw or minP test based on a main-effects model could yield higher power than those based on a full model including interaction terms. This lends some support for the current practice of first detecting marginal effects, then interactions for genome-wide association studies (Marchini et al 2005; Evan et al 2006; Kooperberg and LeBlanc 2008), though of course it may not work in some situations. A theoretical reason is that, in addition to a much smaller number of parameters (or DF) in a main-effects model as compared to a full model with many interaction terms, the score statistic for main-effects can indirectly carry information about the presence of epistatic effects. We also note that the proposed methods can be applied to genome-wide association studies with or without imputed SNPs (Marchini et al 2007; Servin and Stephens 2007). Finally, although we have focused on detecting gene-gene interactions, the proposed methods and conclusions can be similarly applied to detecting gene-environment interactions. In summary, due to the lack of uniformly most powerful test (Cox and Hinkley 1974), and in light of the generally good performance of the SSU, SSUw and minP tests, as for analyzing multiple SNPs in a candidate region, we recommend their routine use, along with the PLS and multivariate score tests, in regression models with or without interaction terms, as for detecting genetic association

in the presence of gene-gene and gene-environment interactions for SNP data.

R code will be posted on <http://www.biostat.umn.edu/~weip>, and available upon request.

Acknowledgement

The author thanks Dr Tao Wang for generously sharing his R code, and Fang Han for help with the ALS data. The author is also grateful to two reviewers for many helpful and constructive comments. This research was partially supported by NIH grants GM081535 and HL65462.

REFERENCES

- Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265.
- Chatterjee N, Carroll RJ (2005) Semiparametric maximum-likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* 92:399-418.
- Chatterjee N, Kalalioglu Z, Moslehi R, Peters U, Wacholder S (2006). Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *Am J Hum Genet* 79:1002-1016.
- Clayton D, Chapman J, Cooper J (2004) Use of unphased multilocus genotype data in indirect association studies. *Genet Epidemiol* 27:415-428.
- Cox DR, Hinkley DV (1974) *Theoretical Statistics*, Chapman and Hall, London.
- Evans DM, Marchini J, Morris AP, Cardon LR (2006) Two-Stage Two-Locus Models in Genome-Wide Association. *PLoS Genet* 2(9):e157.
- Goeman JJ, van de Geer S, van Houwelingen HC (2006). Testing against a high dimensional alternative. *J R Stat Soc B* 68:477-493.

- Guo Y, Li J, Bonham A, Wang Y, Deng H (2009). Gain in power for exhaustive analyses of haplotypes using variable-sized sliding window strategy: a comparison of association-mapping strategies. To appear *European Journal of Human Genetics*. Published on-line.
- Johnson NL, Kotz S (1970) *Distributions in Statistics, Continuous Univariate Distributions*, vol 2. Boston: Houghton-Mifflin.
- Kooperberg C, LeBlanc M (2008) Increasing the power of identifying gene \times gene interactions in genome-wide association studies. *Genet Epidemiol* 32: 255-263.
- Lou X, Chen G, Yan L, Ma JZ, Zhu J, Elston RC, Li MD (2007) A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet* 80:1125-1137.
- Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37:413-417.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* 39:906-913.
- Millstein J, Conti DV, Gilliland FD, Gauderman WJ (2006). A testing framework for identifying susceptibility genes in the presence of epistasis. *Am J Hum Genet* 78:15-27.
- Moore JH (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered* 56:73-82.
- Mukherjee B, Chatterjee N (2008) Exploiting gene-environment independence for analysis of case-control studies: An empirical-Bayes type shrinkage estimator to trade off between bias and efficiency. *Biometrics* 64:685-694.
- Musani SK, Shriner D, Liu N, Feng R, Coffey CS, Yi N, Tiwari HK, Allison DB (2007) Detection of gene \times gene interactions in genome-wide association studies of human population data. *Hum Hered* 63:67-84.

- Pan, W. (2009) Asymptotic tests of association with multiple SNPs in linkage disequilibrium. To appear in *Genetic Epidemiology*. Published Online on Jan 23 2009. DOI: 10.1002/gepi.20402.
Available at www.biostat.umn.edu/rrs.php as Research Report 2008-018, Division of Biostatistics, University of Minnesota.
- Piegorsch WW, Weinberg CR, Taylor JA (1994) Non-hierarchical logistic models and case-only designs for assessing susceptibility in population based case-control studies. *Statistics in Medicine* 13:153-162.
- Roeder K, Bacanu SA, Sonpar V, Zhang X, Devlin B (2005). Analysis of single-locus tests to detect gene/disease associations. *Genet Epidemiol* 28: 207-219.
- Schymick JC, Scholz SW, Fung HC, et al. (2007). Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol*, **6**, 322-328.
- Seaman SR, Muller-Myhsok B (2005) Rapid simulation of p values for product methods and multiple testing adjustment in association studies. *Am J Hum Genet* 76:399-408.
- Servin B, Stephens M (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genetics* 3:e114.
- Song M, Nicolae DL (2008). Restricted parameter space models for testing gene-gene interaction. To appear *Genet Epidemiol*.
- Tang H-K, Siegmund D (2002) Mapping multiple genes for quantitative or complex traits. *Genet Epidemiol* 22:313-327.
- Wang K (2008) Genetic association tests in the presence of epistasis or gene-environment interaction. *Genetic Epidemiology* 32:606-614.
- Wang T, Ho G, Ye K, Strickler H, Elston RC (2009) A partial least-square approach for modeling gene-gene and gene-environment interactions when multiple markers are genotyped. *Genetic Epidemiology* 33:6-15.

- Zhang J-T (2005) Approximate and asymptotic distributions of Chi-squared-type mixtures with applications. *Journal of the American Statistical Association* 100:273-285.
- Zhang Y, Liu JS (2007) Bayesian inference of epistatic interactions in case-control studies. *Nature Genetics* 39:1167-1173.
- Zhao JY, Jin L, Xiong MM (2006). Test for interaction between two unlinked loci. *Am J Hum Genet* 79:831-845.
- Zheng T, Wang H, Lo SH (2006) Backward genotype-trait Association (BGTA) - based dissection of complex traits in case-control design. *Human Heredity* 62:196-212.

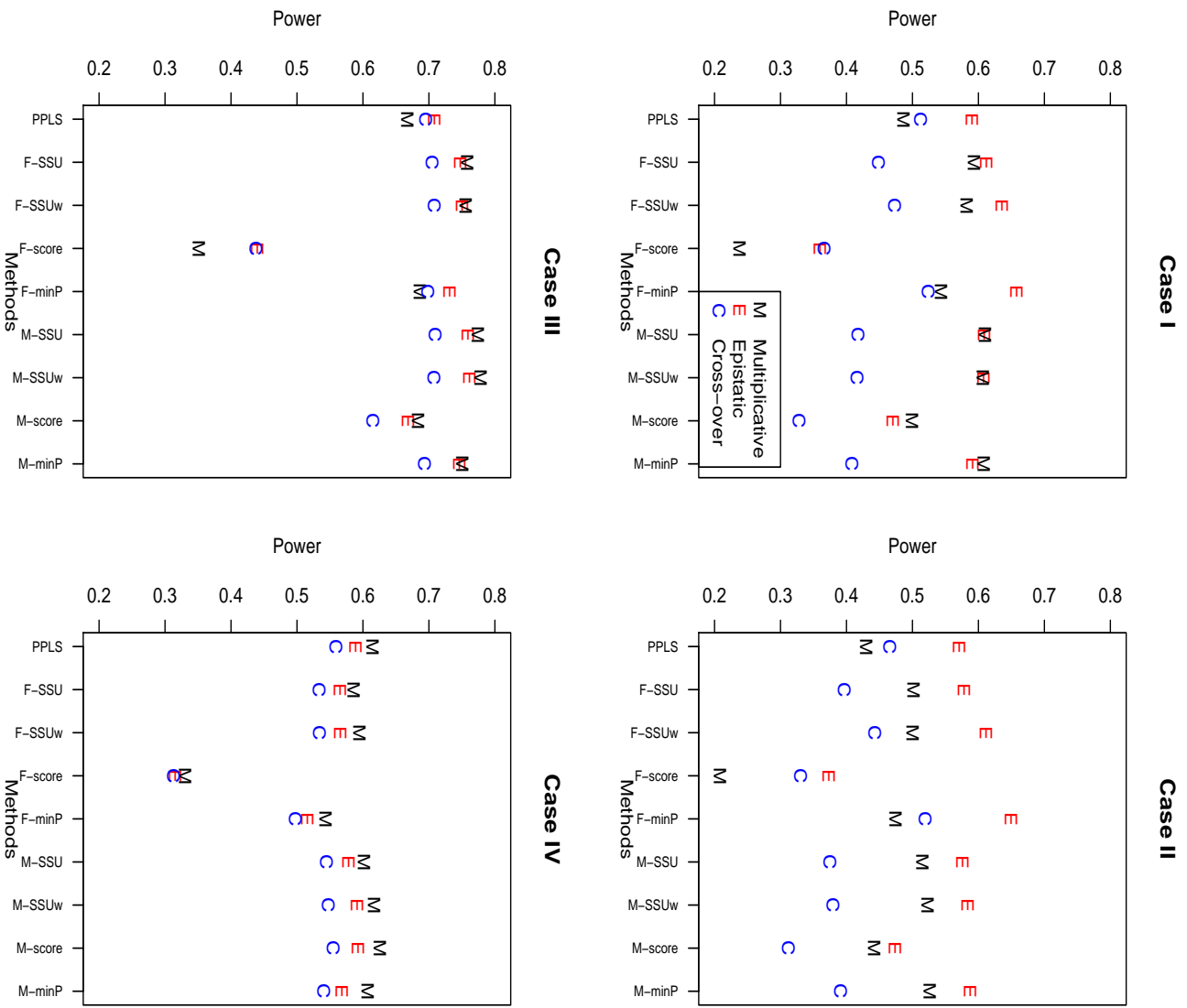


Figure 1: Average power in simulation Case I-IV for each method.

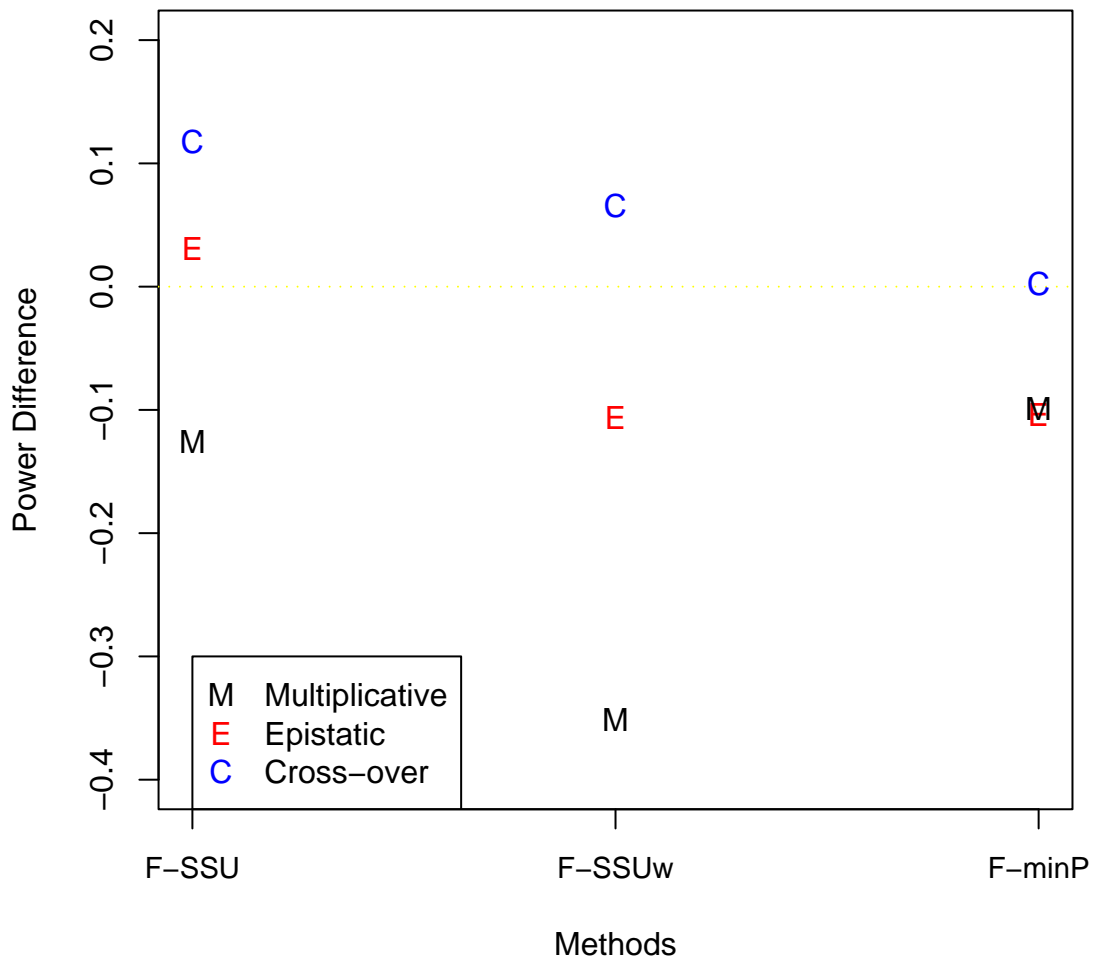


Figure 2: Average power change after centering genotype scores in simulation Case I for the F-SSU, F-SSUw, and F-minP tests.

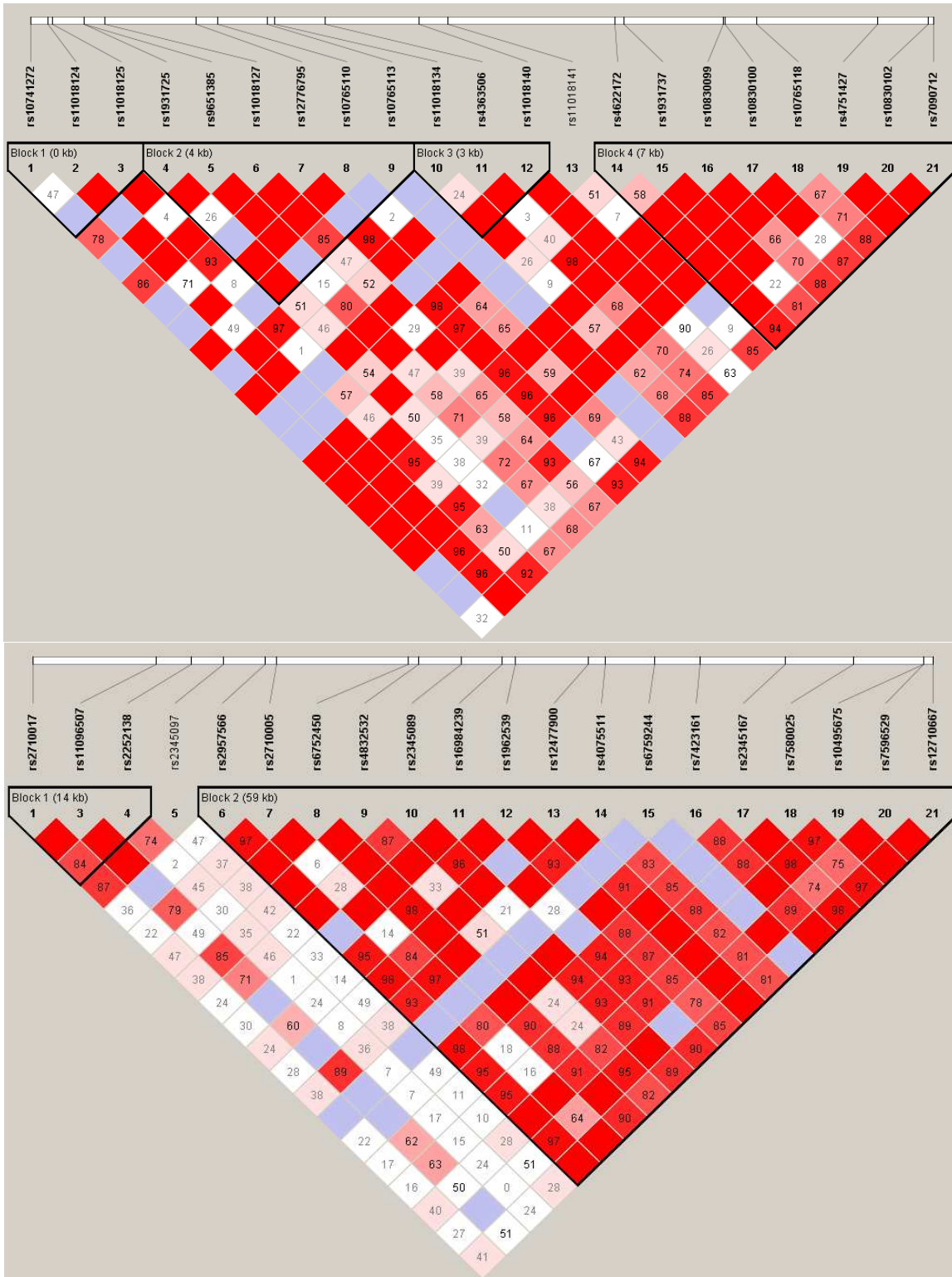


Figure 3: LD blocks around SNPs rs4363506 (top) and rs16984239 (bottom).