



Genome-Wide Analysis of Gene-Gene and Gene-Environment Interactions Using Closed-Form Wald Tests

Zhaoxia Yu,^{1*} Michael Demetriou,^{2,3} and Daniel L. Gillen¹

¹Department of Statistics, University of California, Irvine, California, United States of America; ²Department of Neurology, University of California, Irvine, California, United States of America; ³Department of Microbiology & Molecular Genetics, University of California, Irvine, California, United States of America

Received 20 November 2014; Revised 25 February 2015; accepted revised manuscript 6 May 2015.

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/gepi.21907

ABSTRACT: Despite the successful discovery of hundreds of variants for complex human traits using genome-wide association studies, the degree to which genes and environmental risk factors jointly affect disease risk is largely unknown. One obstacle toward this goal is that the computational effort required for testing gene-gene and gene-environment interactions is enormous. As a result, numerous computationally efficient tests were recently proposed. However, the validity of these methods often relies on unrealistic assumptions such as additive main effects, main effects at only one variable, no linkage disequilibrium between the two single-nucleotide polymorphisms (SNPs) in a pair or gene-environment independence. Here, we derive closed-form and consistent estimates for interaction parameters and propose to use Wald tests for testing interactions. The Wald tests are asymptotically equivalent to the likelihood ratio tests (LRTs), largely considered to be the gold standard tests but generally too computationally demanding for genome-wide interaction analysis. Simulation studies show that the proposed Wald tests have very similar performances with the LRTs but are much more computationally efficient. Applying the proposed tests to a genome-wide study of multiple sclerosis, we identify interactions within the major histocompatibility complex region. In this application, we find that (1) focusing on pairs where both SNPs are marginally significant leads to more significant interactions when compared to focusing on pairs where at least one SNP is marginally significant; and (2) parsimonious parameterization of interaction effects might decrease, rather than increase, statistical power.

Genet Epidemiol 0:1–10, 2015. © 2015 Wiley Periodicals, Inc.

KEY WORDS: closed-form; epistasis; genome-wide

Introduction

Genome-wide association studies (GWAS) have led to the discovery of hundreds of common variants for complex human traits. However, the identified variants to date only explain a small fraction of heritability, leaving the majority of genetic determinants yet to be discovered. Emerging evidence suggests that gene-gene and gene-environment interactions, may explain part of the missing heritability [Carlborg and Haley, 2004; Eichler et al., 2010; Manolio et al., 2009]. There is little doubt that testing interactions is of fundamental importance for defining the etiology of diseases. The knowledge of interactions may also be used to further the goal of personalized medicine [Moore and Williams, 2009]. For example, if a treatment alters the function of a gene, then patients with varying genetic background are expected to respond differently to the treatment if the targeting gene interacts with other genes.

In the literature, a variety of definitions for interactions have been used by investigators in different disciplines. For

example, Bateson [1907] first defined a gene-gene interaction as a phenomenon in which one gene masks the effect of another. If we limit the definitions to those used by statisticians, there is no unique definition either, as the presence and magnitude of an interaction is model and scale dependent. Some excellent discussions about varying definitions can be found in Cordell [2002], Moore and Williams [2009], and Wang et al. [2010a].

In this article, we define an interaction as the departure from a main effects linear model on the log odds of disease. For example, we model the disease risk for a subject with genotype g at single-nucleotide polymorphism (SNP) 1 and genotype h at SNP 2 using the following logistic regression

$$\begin{aligned} \text{logit}(\pi_{gh}) = & \mu + \alpha_1 I(g = 1) + \alpha_2 I(g = 2) + \beta_1 I(h = 1) \\ & + \beta_2 I(h = 2) + \lambda_{11} I(g = 1)I(h = 1) + \lambda_{12} I(g = 1)I(h = 2) \\ & + \lambda_{21} I(g = 2)I(h = 1) + \lambda_{22} I(g = 2)I(h = 2). \end{aligned} \quad (1)$$

Here, we code genotypes at each SNP using the number 0 for the common homozygote, 1 for the heterozygote, and 2 for the rare homozygote. In model (1), the parameter μ is not related to genetic effects, α_1 and α_2 are the main effects of SNP

Supporting Information is available in the online issue at wileyonlinelibrary.com.

*Correspondence to: Dr. Zhaoxia Yu, Department of Statistics, University of California, Donald Bren Hall 2019, Irvine, CA 92697, USA. E-mail: zhaoxia@ics.uci.edu

1, β_1 and β_2 are the main effects of SNP 2, and λ_{11} , λ_{12} , λ_{21} , and λ_{22} are the interaction effects between the two SNPs.

To test the interaction between two SNPs, one often uses the likelihood ratio test (LRT) stemming from a binomial probability model [Cordell, 2002; North et al., 2005]. In the LRT, the maximized likelihood under the full model (model 1) is compared to that under the reduced model (model 2) given by

$$\begin{aligned} \text{logit}(\pi_{gh}) = & \mu + \alpha_1 I(g = 1) + \alpha_2 I(g = 2) \\ & + \beta_1 I(h = 1) + \beta_2 I(h = 2). \end{aligned} \quad (2)$$

In general, the maximum likelihood estimates (MLEs) for logistic regression parameters do not have a closed-form solution, and iterative algorithms are generally employed to obtain the MLEs for the reduced model. This implies that use of the LRT can be quite computationally demanding for a genome-wide analysis of gene-gene interactions. In the past few years, a considerable number of computationally efficient tests have been proposed [Bhattacharya et al., 2011; Dong et al., 2008; Fan et al., 2011; Hu et al., 2014; Millstein et al., 2006; Ueki and Cordell, 2012; Wu et al., 2010b; Yang et al., 2009b; Zhao et al., 2006]. These tests are often used as a screening tool to identify promising pairs and the LRT is then used to formally test the identified pairs.

The vast majority of computationally feasible methods for testing gene-gene interactions are correlation-based and have one degree of freedom (1 *df*). In each of these tests, a chosen linkage disequilibrium (LD) measure is compared between cases and controls. For example, Wu et al. [2010b] use a measure calculated from haplotype data; the “fast-epistasis” option in PLINK [Purcell et al., 2007] and the test of Wellek and Ziegler [2009] are based upon genotype data. However, as reported in Ueki and Cordell [2012], they are invalid for testing gene-gene interactions in the presence of both main effects and LD. Ueki and Cordell [2012] recently propose a joint test and illustrated that the test is valid for testing gene-gene interactions for rare diseases.

Tests with four degrees of freedom (4 *df*) have also been considered. Yang et al. [2009b] proposed to test interactions by partitioning chi-squares. Plackett [1962] has pointed out that the test statistic used in the chi-square partitioning method does not necessarily follow a chi-squared distribution in the absence of interactions. Because the MLEs of the parameters under the reduced model do not exist in closed-form, Wan et al. [2010] approximated the MLEs using marginal frequencies of low orders. However, the approximation performs poorly except in “perfect” contingency tables, which require several constraints on the marginal probabilities [Darroch, 1962; Roy and Kastenbaum, 1956]. Thus, neither of the two four-degree-freedom tests is valid for testing gene-gene interactions in general.

Due to the computational challenges in genome-wide interaction analysis, it is tempting to derive closed-form solutions in order to reduce computational cost. For example, Schwender et al. [2012] derived closed-form solutions to testing gene-gene and gene-environment interac-

tions for case-parent studies; Chen et al. [2012] provided closed-form estimates of gene-environment interactions under the constraints of multiplicative gene-environment interactions, gene-environment dependence, Hardy-Weinberg equilibrium, binary environment factors, and rare disease prevalence. Here, we propose to use closed-form Wald tests for genome-wide gene-gene and gene-environment interaction analysis for case-control studies. The Wald test enjoys the same asymptotic properties as the LRT, including being optimal under standard regularity conditions. On the other hand, compared to the LRT, the Wald test is much more computationally feasible for genome-wide interaction analysis. The remainder of the article is organized as follows. We derive a Wald test with 4 *df* and a modified Wald test with 1 *df* for gene-gene interactions, describe how to apply this strategy to test gene-environment interactions, demonstrate feasibility using simulations, and then apply the method to a genome-wide study of multiple sclerosis. A discussion is provided at the end of the article.

Methods

In this section, we first propose closed-form Wald tests and then describe several existing tests.

The Wald Test for Gene-Gene Interactions

The data for testing the interaction between a pair of SNPs when modeling the probability of a binary disease indicator can be summarized using a $3 \times 3 \times 2$ table, as each SNP has three levels (0, 1, and 2) and the disease status has two levels (0 for normal and 1 for diseased). Let n_{ghk} denote the number of subjects with genotype g at SNP 1, genotype h at SNP 2, and disease status k . Here, $g = 0, 1, 2$; $h = 0, 1, 2$ and $k = 0$ or 1. Similarly, let p_{ghk} denote the probability of observing a subject with genotype g at SNP 1, genotype h at SNP 2, and disease status k . For simplicity, we use $\bar{\theta} = (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})^T$ to denote the vector of all parameters in the full model and $\bar{\theta}_\lambda = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})^T$ to denote the vector of interaction parameters. In the literature of testing interactions for three-dimensional contingency tables, Plackett [1962] proposed to use closed-form estimates. These estimates correspond to the MLEs of the saturated model, i.e., model 1. It is easy to verify that the MLEs are given by

$$\hat{\lambda}_{gh} = \log \left(\frac{n_{gh1} n_{001} / n_{gh0} n_{000}}{n_{g01} n_{0h1} / n_{g00} n_{0h0}} \right), \quad g = 1, 2; h = 1, 2.$$

Let $\hat{\theta}$, $\hat{\theta}_\lambda$ and $I(\hat{\theta})$ denote the MLE of the vector of all the parameters, the MLE of the vector of interaction parameters, and the observed Fisher information, respectively. When all p_{ghk} 's are greater than 0 and the sample size is large, it is easily shown that the joint distribution of the MLEs is approximately multivariate normal, i.e., $\hat{\theta} \sim N_9(\bar{\theta}, I^{-1}(\hat{\theta}))$. We calculate the asymptotic covariance of $\hat{\theta}_\lambda$, which is denoted

by $\text{cov}(\hat{\theta}_\lambda)$ (supplementary Section A). From this, the Wald test statistic for testing $H_0 : \theta_\lambda = 0$ is

$$\text{Wald} = \hat{\theta}_\lambda^T [\text{cov}(\hat{\theta}_\lambda)]^{-1} \hat{\theta}_\lambda.$$

Under the null hypothesis, the above quadratic form then follows the chi-squared distribution with 4 *df* asymptotically.

When an SNP has a low minor allele frequency (MAF), the number of subjects with the rare homozygote is small and using it as a separate genotype category is likely to reduce power. In our study, when the rare homozygote has less than 20 subjects, we collapse it with the heterozygote. The degrees of freedom of the Wald test are then reduced accordingly.

The Modified Wald Test with 1 *df*

One concern of using the 4 *df* tests is that the power could be low due to the large number of *df* [Song and Nicolae, 2009]. As a result, several tests with 1 *df* have been proposed [Barhdadi and Dube, 2010; Chatterjee et al., 2006; Hoffmann et al., 2009; Jiao et al., 2012], including Tukey's 1 *df* test [Tukey, 1949]. These tests usually consider some parsimonious functional form when modeling interactions. One particularly interesting model is the single interaction-parameter model used in VanderWeele and Laird [2011], which incorporates an additive interaction term with unconstrained main effects:

$$\begin{aligned} \text{logit}(\pi_{gh}) = & \mu + \alpha_1 I(g = 1) + \alpha_2 I(g = 2) \\ & + \beta_1 I(h = 1) + \beta_2 I(h = 2) + \lambda gh. \end{aligned} \quad (3)$$

By assuming an additive interaction of the total number of alleles, focus lies in testing a single interaction parameter in model (3) and hence reduces the *df* to 1. Note that this model allows flexible main effects, thereby avoiding the potential bias in testing interaction that could be caused by misspecifying the main effects [Chen et al., 2012; Tchetgen and Kraft, 2011; VanderWeele and Laird, 2011; Vansteelandt et al., 2008; Yu, 2011].

To generalize the idea of VanderWeele and Laird [2011], we assume that the four interaction parameters satisfy some constraints such that we can rewrite the vector $\bar{\theta}_\lambda$ as $\bar{\theta}_\lambda = \lambda A \mathbf{1}$, where A is a 4×4 diagonal matrix, λ is a univariate interaction parameter, and $\mathbf{1} = (1, 1, 1, 1)^T$. For example, with $A = \text{diag}(1, 2, 2, 4)$, the full logistic regression model in (1) reduces to the single interaction-parameter model in (3), and with $A = \text{diag}(1, 1, 1, 1)$, the parameterization of interactions is the same as that of Jiao et al. [2012]. The conventional Wald test requires the MLE of λ , which does not have a closed-form solution. To derive a computationally feasible test, similar to Ueki and Cordell [2012], we combine the information in the MLEs of $\bar{\theta}_\lambda$. Note that the vector $A^{-1} \hat{\theta}_\lambda \xrightarrow{p} A^{-1} \bar{\theta}_\lambda$ as $n \rightarrow \infty$ and hence provides four consistent estimates of λ . We consider using a weighted average to estimate λ , where the weight w is chosen to minimize the variance of $w^T A^{-1} \hat{\theta}_\lambda$. Using the Lagrange multiplier, it is not difficult to see that

$w = cA[\text{cov}(\hat{\theta}_\lambda)]^{-1} A \mathbf{1}$, where c is a constant. Therefore, the corresponding test statistic is given by

$$\text{Wald}_1 = \frac{(w^T A^{-1} \hat{\theta}_\lambda)^2}{w^T A^{-1} \text{cov}(\hat{\theta}_\lambda) A^{-1} w} = \frac{(\mathbf{1}^T A [\text{cov}(\hat{\theta}_\lambda)]^{-1} \hat{\theta}_\lambda)^2}{\mathbf{1}^T A [\text{cov}(\hat{\theta}_\lambda)]^{-1} A \mathbf{1}}.$$

Under the null hypothesis of no interaction, i.e., $\lambda = 0$, the test statistic Wald_1 follows the chi-squared distribution with 1 *df* asymptotically. In this article, we define Wald_1 as the test that corresponds to additive interaction, i.e., $A = \text{diag}(1, 2, 2, 4)$.

The Wald Tests for Gene-Environment Interactions

The strategies presented above were motivated by the problem of testing gene-gene interactions. They are readily applicable to testing gene-environment (G-by-E) interactions when the environmental factor is a categorical variable or can be converted to a categorical variable in a scientifically meaningful way. For example, for an environmental factor with K categories, the full model is saturated; similar to (1), the MLE of the interaction parameters exists in closed-form. Thus, we can derive a Wald test with $2(K-1)$ *df*. If we assume that the genetic factor plays an additive role in the G-by-E interactions, we can construct a test with $K-1$ *df*. The degrees of freedom may be further reduced when meaningful numerical scores can be assigned to the environmental factor.

Existing Methods for Gene-Gene Interactions

In supplementary Section B, we describe several commonly used existing tests, including the LRTs, Boost [Wan et al., 2010], PLINK's fast-epistasis [Purcell et al., 2007], and the joint test of Ueki and Cordell [2012]. We will also briefly discuss the conditions under which they are valid.

Simulations

To examine the performance of the proposed methods, we conduct simulations in six scenarios (Table 1). The null hypothesis of no gene-gene interaction holds in scenarios 1–3. In scenario 1, there is no LD between the two SNPs. In scenarios 2 and 3, the two SNPs are in LD, with the Pearson correlation coefficients varying from 0.1 to 0.9. These three scenarios are used to examine how LD and other violations of model assumptions jointly affect type I error rates. Specifically, scenario 2 assumes additive main effects at SNP1 and no main effect at SNP2; scenario 3 assumes dominant main effects at SNP1 and no main effect at SNP2. In scenarios 4 and 5, the SNPs are interacting with each other on disease risk. Scenario 4 assumes that the interaction is in an additive manner, i.e., the value of the interaction parameters depends on the number of risk alleles. Scenario 5 assumes that the four interaction parameters have the same value.

We fix the MAF of SNP1 at 0.5 and vary the MAF of SNP2 from 0.1 to 0.5 for scenarios 1, 4, and 5. For scenarios 2 and 3, the MAF at both SNPs is 0.5. In each of the scenarios,

Table 1. Parameters used in the simulation study

		Main effect of SNP1		Main effect of SNP2	Interaction effects	LD	MAF
	Scenario	μ	(α_1, α_2)	(β_1, β_2)	$(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})$	r	p_2
No interaction	1	-2.61	(0.2, 0.4)	(0.2, 0.4)	(0, 0, 0, 0)	0	0.1–0.5
	2	-2.40	(0.2, 0.4)	(0, 0)	(0, 0, 0, 0)	0.1–0.9	0.5
	3	-2.42	(0.3, 0.3)	(0, 0)	(0, 0, 0, 0)	0.1–0.9	0.5
Additive interaction	4	-3.10	(0.2, 0.4)	(0.2, 0.4)	(0.4, 0.8, 0.8, 1.6)	0	0.1–0.5
Dominant interaction	5	-3.10	(0.2, 0.4)	(0.2, 0.4)	(0.8, 0.8, 0.8, 0.8)	0	0.1–0.5

We consider five scenarios, with the first three under the null hypothesis (absence of gene-gene interaction) and the last two under the alternative hypothesis (presence of gene-gene interaction).

the parameter μ is chosen such that the disease prevalence is approximately 0.1 when the MAF of both SNPs is 0.5 and $r = 0$. For each scenario, 1,000 simulations are used; in each simulation, 1,000 cases, and 1,000 controls are sampled from a homogenous population.

To assess the performance of the proposed methods, we compared the performance of three 4 *df* tests and four 1 *df* tests. The three 4 *df* tests are the Wald test (*Wald*), the *LRT*, and the test uses MLE approximation in “perfect” tables (*Boost*). The four 1 *df* tests are the modified Wald test with 1 *df* (*Wald*₁), the *LRT* with 1 *df* (*LRT*₁), the fast-epistasis test in *PLINK*, and the joint test of Ueki and Cordell [2012] (*UC*). The proportion of simulations with *P*-values less than 0.05 is used to estimate the type I error rates (scenarios 1–3) or power (scenarios 4 and 5) of the seven tests.

A Real Study

We apply the proposed tests to a genome-wide study of multiple sclerosis. In this study, the control group includes 1,476 subjects enrolled in the 1958 British Birth Cohort Study and the case group includes 994 patients diagnosed with multiple sclerosis. Subjects in both groups were self-identified as white and of European descent [Wellcome Trust Case Control Consortium and The Australo-Anglo-American Spondylitis Consortium (TASC), 2007]. Approximately 15k nonsynonymous SNPs across the genome were genotyped. We exclude SNPs with a low MAF (<1%) or filtered out by the Wellcome Trust Case Control Consortium due to quality concerns. This leaves approximately 11k SNPs for interaction analysis.

Results

Simulation Results

The estimated type I error rates under scenario 1 (Fig. 1A) indicate that all tests have reasonable control of false positives in the absence of LD. Among the seven tests, *Wald* is slightly conservative and *Boost* is anticonservative when the MAF of SNP2 is small. These results are not surprising, as *Boost* gives an upper bound of the *LRT* statistic [Wan et al., 2010] and the Wald test is known to be conservative under some situations [Newcombe, 1998]. In the presence of LD (Fig. 1B and C), *Boost* has inflated false positives. This is because the conditions for “perfect” tables do not hold in the presence of both main effects and LD; as a result, the approximation

of MLE in *Boost* is incorrect. It is also clear that the inflation increase with LD.

PLINK has an appropriate control of false positives when the true main effect is additive (scenarios 1 and 2 and Fig. 1A and B) but not in the presence of both LD and non-additive main effects (scenario 3 and Fig. 1C). This can be explained by the fact that the fast-epistasis test of *PLINK* is an alternative to tests based upon the logistic regression that assumes both additive main effects and interactions [Ueki and Cordell, 2012]. When the true underlying main effects are nonadditive, forcing them to be additive corresponds to under fitting. As a result, when the two SNPs are in LD, we have biased estimates of gene-gene interaction parameters.

Figure 1 also indicates that the inflation of the joint test (*UC*) is ignorable in most situations (scenarios 1–3 and Fig. 1A–C) except when both the disease prevalence and main effects are quite large (data not shown). Ueki and Cordell [2012] illustrated that under the assumption of rare diseases, the joint test is valid for testing gene-gene interactions. Here, we found that it has acceptable control of type I error rates even when the disease prevalence is 10% (scenarios 1–3 and Fig. 1A–C). Compared to *Boost* and *PLINK*, *UC* is only moderately inflated, suggesting that the type I error rate of the joint test probably is not a concern for most situations.

The estimated power under additive and dominant interactions is summarized in Figures 2 and 3, respectively. In general, *Wald* agrees well with *LRT*, and *Wald*₁ agrees well with *LRT*₁. This demonstrates that we can obtain accurate results while avoiding iterative algorithms. When the true interactions are additive (scenario 4 and Fig. 2), the 1 *df* tests, namely *Wald*₁, *LRT*₁, *PLINK*, and *UC*, are more powerful than *Wald* or *LRT*. On the other hand, when the true underlying interaction model is dominant (scenario 5 and Fig. 3), the 1 *df* tests are less powerful, among which *PLINK* has the lowest power, then followed by *UC*.

Depending on association parameters, the power of *Boost* can be higher or lower than the other tests. However, since *Boost* results in inflated type I error rates in the presence of LD (Fig. 2), we do not recommend it as a test for interactions.

The WTCCC Study of Multiple Sclerosis

We tested the 6×10^7 pairs of two-way interactions from the 11k SNPs using all the seven methods. To quantify the speeds of different methods, we implemented all methods using C++ (source code available upon request). Using the

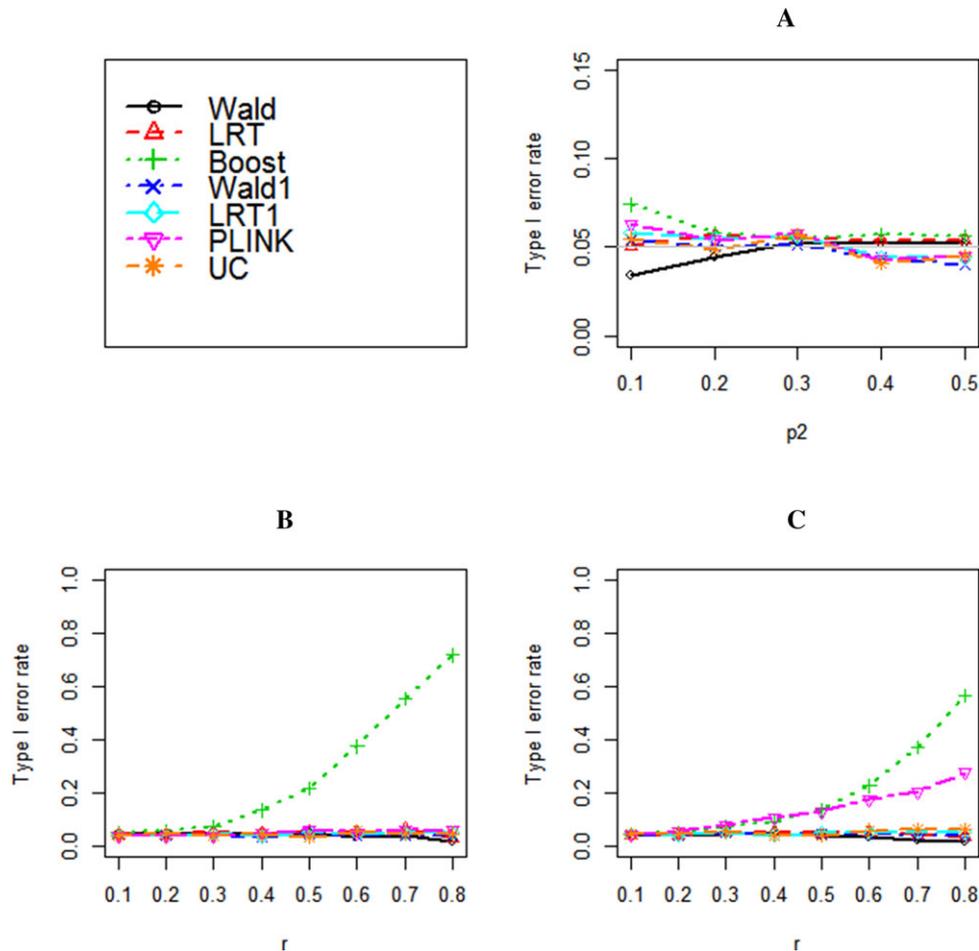


Figure 1. Estimated type 1 error rates. (A) Scenario 1, i.e., additive main effects at both SNPs and no LD. The x-axis shows the MAF of SNP2. (B) Scenario 2, i.e., additive main effects at SNP1 and LD between the two SNPs. (C) Scenario 3, i.e., dominant main effects at SNP1 and LD between the two SNPs. For (B) and (C), the x-axis shows r , i.e., the Pearson correlation coefficient based on haplotypes.

slowest method, namely LRT_1 , as the reference, the speeds of the other methods are 11.5 ($Wald$), 11.5 ($Wald_1$), 2.1 (LRT), 12.6 ($PLINK$), 10.5 ($Boost$), and 5.5 (UC) times faster, respectively. The details are provided in supplementary Table S1.

We first discuss the results based on the 4 *df* test $Wald$. The Bonferroni criterion was used for bounding the family-wise type I error rate at 0.05. In this case, pairs with nominal P -values less than $0.05/(6 \times 10^7)$ are considered significant. According to this standard, three SNP pairs are significant, as shown in the first three rows of supplementary Table S2. For a complex trait like multiple sclerosis, a large number of interactions are likely present. However, applying the conservative Bonferroni correction, only very strong interactions survive the stringent P -value cutoff.

To address the burden of multiple testing, several multistage approaches have been considered and discussed [Dai et al., 2012; Daly and Altshuler, 2005; Emily et al., 2009; Evans et al., 2006; Hsu et al., 2012; Ionita and Man, 2006; Kooperberg and LeBlanc, 2008; Lewinger et al., 2013; Liu et al., 2011; Ma et al., 2012; MacGregor and Khan, 2006; Marchini et al.,

2005; Millstein et al., 2006; Murcray et al., 2009; Musani et al., 2007; Wang et al., 2010b; Wu et al., 2010a; Yang et al., 2009a]. In most of these approaches, a subset of pairs was focused on in order to reduce the number of tests. Here, we utilize two of these approaches, namely the conditional approach [Daly and Altshuler, 2005] and the simultaneous approach [Kooperberg and LeBlanc, 2008; Marchini et al., 2005]. In the conditional approach, at least one of the two SNPs in the pair is marginally significant; whereas in the simultaneous approach both SNPs in the pair are significant. To compare the two strategies, we provide QQ plots with varying nominal P -value cutoffs (supplementary Fig. S1). The P -values for marginal effects are calculated using the Armitage trend test [Sasieni, 1997]. The numbers of significant interactions are also given in supplementary Figure S1. It is obvious that the proportion of significant interactions is enriched with the simultaneous approach, but not with the conditional approach. As a result, the simultaneous approach identifies a greater number of significant interactions when compared to the conditional approach. In particular, using a nominal P -value cutoff of 0.001, we identified 67 interactions. It is worth

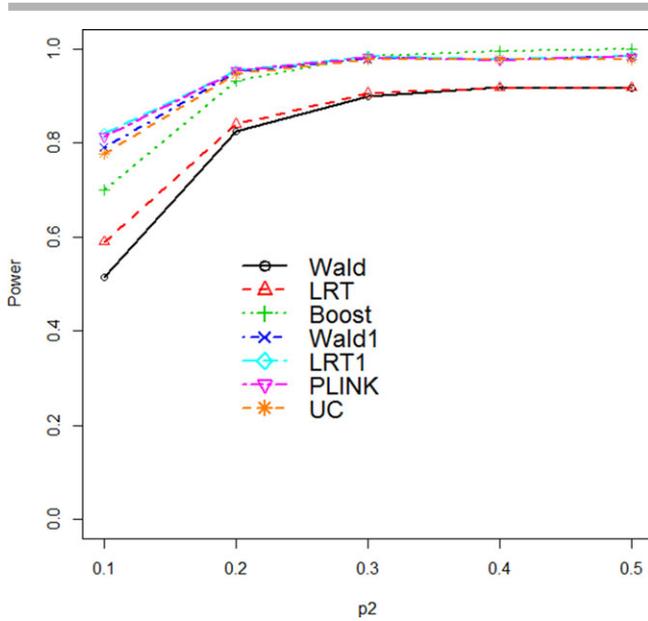


Figure 2. Estimated power under additive interactions (scenario 4). The x-axis is the MAF of SNP2.

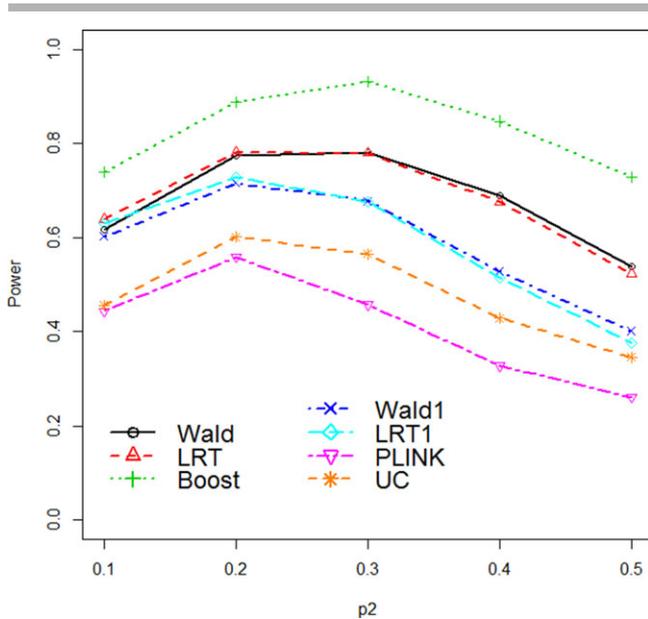


Figure 3. Estimated power under dominant interactions (scenario 5). The x-axis is the MAF of SNP2.

pointing out that we also examined the two two-stage strategies using the false discovery rate criterion [Storey, 2003]. Again, the simultaneous approach identifies more interactions than does the conditional approach (data not shown).

We next examined whether different tests agree with each other for the 67 pairs detected by the simultaneous approach using the 4 *df* test *Wald*. Detailed information about the involved SNPs, interactions, and estimated parameters are

given in supplementary Table S2. Most of the SNPs are highly significant marginally. The MAFs are relatively large, ranging from 0.19 to 0.50. The Pearson correlation ranges from near zero to about 0.30, indicating most of the identified pairs are not in strong LD. Comparing the *P*-values resulting from the considered tests echoes what was observed in the simulation study. First, *Wald* gives similar results to *LRT* and *Wald*₁ gives similar results to *LRT*₁ (supplementary Table S2 and the top panel of Fig. 4). Second, the comparison between *Boost* and *LRT* (lower left panel of Fig. 4) shows that *Boost* is anticonservative, especially when SNPs are in LD. In the two remaining tests, *UC* is more similar to *LRT*₁ than *PLINK*. Note that *PLINK* is an alternative to the logistic regression that assumes additive main effects. The difference between *PLINK* and *LRT*₁ indicates that the main effects of the SNPs in the top 67 pairs are unlikely to be additive.

Both supplementary Table S2 and the lower panel of Figure 4 indicate that *Wald*₁ and *LRT*₁ are lower in magnitude than *Wald* and *LRT* in most situations. This difference suggests that the additive interaction model in (3) does not fit well for many of the 67 pairs. Indeed, the estimated interaction parameters (supplementary Table S2) for most SNP pairs do not follow the additive interaction model. To assess the degree to which the additive interaction assumption is violated, we conduct a test. Specifically, define the following matrix

$$B = \begin{pmatrix} 1 & -0.5 & 0 & 0 \\ 1 & 0 & -0.5 & 0 \\ 1 & 0 & 0 & -0.25 \end{pmatrix}.$$

The test statistic for additive interaction is

$$(B\hat{\theta}_\lambda)^T [B\hat{\text{cov}}(\hat{\theta}_\lambda)]^{-1} B\hat{\theta}_\lambda.$$

For large sample size and under the null hypothesis of additive interaction, the test statistic is approximately distributed as a chi-squared random variable with 3 *df*. Note that when one homozygote group of an SNP is rare, we collapse it with the heterozygote; as a result, the *B* matrix and the *df* will be modified accordingly. The resulting *P*-values (the last column of supplementary Table S2) reveal that the additive assumption does not hold for many of the SNP pairs.

All the significant interactions are from SNPs in the major histocompatibility complex (MHC) region on Chromosome 6. This is not surprising, as it is already known that the MHC region plays a major role in the function of the immune system and numerous variants in the region have been associated with autoimmune diseases [Fernando et al., 2008]. In fact, gene-gene interactions in this region have been identified in other autoimmune diseases, such as type 1 diabetes and rheumatoid arthritis [Lippert et al., 2013; Wan et al., 2010; Wu et al., 2010a]. Figure 5 visualizes the positions of the SNPs of the 67 identified interactions. All except two interactions are within or between class II and class III. Lincoln et al. [2009] also reported interactions within the MHC class II region in multiple sclerosis. Thus, the interactions identified here are likely to be true. Interpreting the interactions requires biological experiments, as a significant *P*-value could be due to other reasons such as haplotype

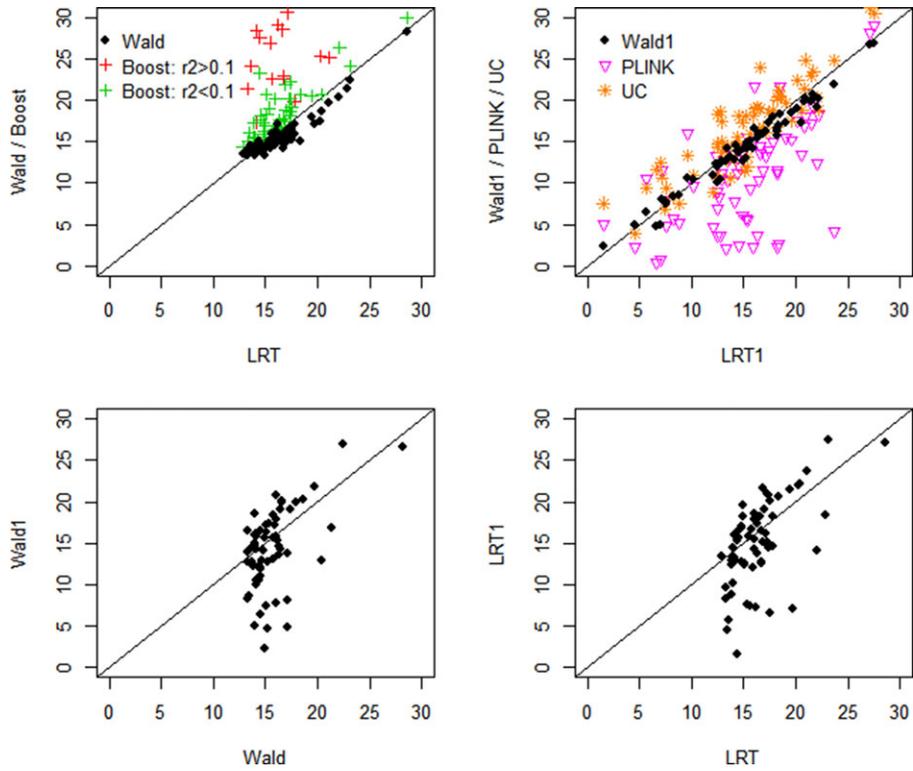


Figure 4. The $-\log_{10} P$ -values for the 67 pairs selected using the simultaneous approach. Upper left: *Wald* and *Boost* vs. *LRT*; upper right: *Wald*₁, *PLINK*, and *UC* vs. *LRT*₁; lower left: *Wald*₁ vs. *Wald*; lower right: *LRT*₁ vs. *LRT*.

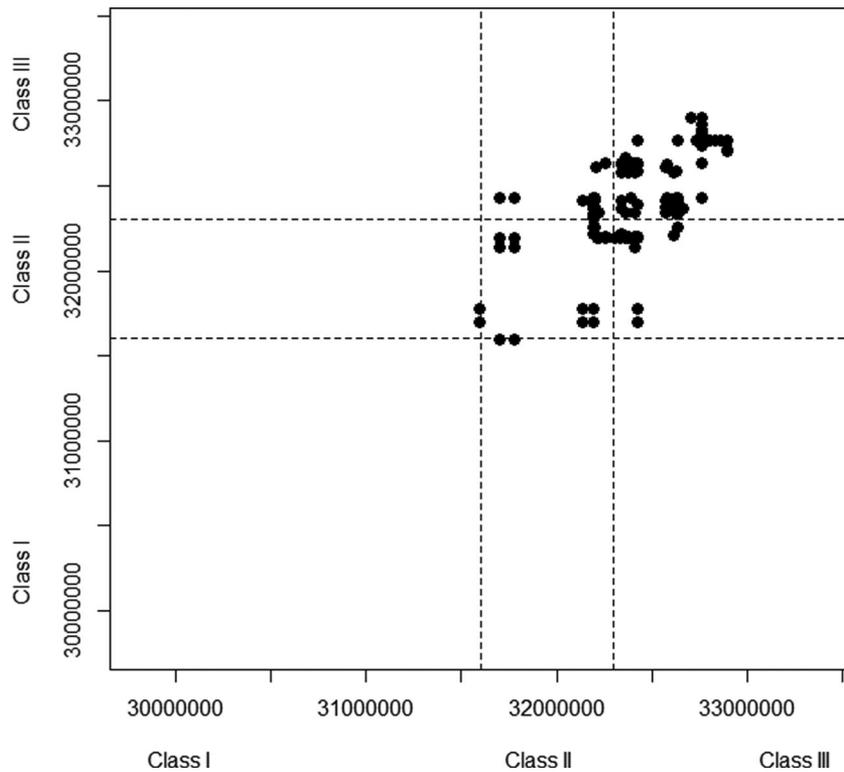


Figure 5. The 67 interactions identified by the simultaneous approach.

effects or an SNP pair that is flanking an unobserved functional variant. Recently, a functional assay in humanized mice suggested functional epistasis between two MHC class II alleles in multiple sclerosis [Gregersen et al., 2006]. Indeed, six of the identified interactions are within the MHC class II region (Fig. 5).

Discussion

In this article, we proposed closed-form Wald tests for genome-wide analysis of gene-gene and gene-environment interactions. Here, we did not consider the score test using model 2 or the 1-*df* Wald test using model 3. Conducting the score test based on model 2 or the 1-*df* Wald test based on model 3 requires obtaining the MLEs for the parameters in model 2 or model 3, in which a closed-form solution does not exist. Our simulations also show that the 1-*df* Wald test based on model 3 has a similar performance with the proposed 1-*df* Wald test (data not shown). For these reasons, we did not show the performance of these two tests in our article. Note that our closed-form Wald tests place no constraints on the main effects. Here, we provide three reasons of not considering the following constrained gene-gene interaction model

$$\text{logit}(\pi_{gh}) = \mu + \alpha g + \beta h + \lambda gh. \quad (4)$$

First, when a constraint on the main effects disagrees with the underlying true model, the main effects are misspecified. Similar to the effect of under fitting in regression models, misspecifying main effects leads to biased estimates of the interaction parameters [Chen et al., 2008, 2012; Tchetgen and Kraft, 2011; VanderWeele and Laird, 2011; Vansteelandt et al., 2008]. By allowing full flexibility in modeling main effects, our tests are still valid in the presence of LD between genes or dependence between genetic and environmental factors. Second, the MLEs for the parameters in model 4 do not have a closed-form solution. Last, when the true model is model 4, our 1-*df* Wald test has minimal power loss, as it is based on model 3, which only uses two more nuisance parameters than model 4.

Existing computationally efficient tests for gene-gene interactions are invalid when some of their model assumptions, such as additive main effects, main effects at no more than one SNPs, and no LD, are violated. Among them, the joint test of Ueki and Cordell [2012] has the best control of false positives under various violations of assumptions. Although their test was motivated by the haplotype-based correlation of Wu et al. [2010b], it is insensitive to nonadditive main effects or LD. Instead of calculating correlations from reconstructed haplotypes, collapsed tables, or numerical genotypes, they calculated four correlations from sub two-by-two tables. This explains why their test is insensitive to LD and nonadditive main effects. Ueki and Cordell [2012] showed that the joint test is valid for testing gene-gene interactions for rare diseases and we proved that the test is valid as long as there is at least an SNP with no main effects. Our simulations indicate that its inflation in the type I error rates is ignorable even the disease

prevalence is as high as 10%. Compared to PLINK and Boost, the inflation of the joint test is still moderate when the prevalence is over 45%. Therefore, the inflation in false positives of the joint test is unlikely to be a concern in practice. Compared to the methods in Ueki and Cordell [2012], ours are derived from a formal statistical framework, are valid under all circumstances, and are more general in several perspectives. First, we derived Wald-type tests from MLEs, which ensure asymptotic efficiency of the resulted tests. Second, our tests cover both constrained and unconstrained parameterizations of interaction effects, making them useful under multiple modeling assumptions.

From the interaction study of multiple sclerosis, we concluded that (1) the simultaneous approach, where both SNPs in an SNP pair show marginal effects, identified a greater number of significant interactions when compared to the conditional approach, where either of the SNPs has a marginal effect; and (2) additive interactions are unlikely to be common. One needs to be cautious when generalizing these conclusions to other situations. It is well understood that the MHC region plays an important role in autoimmune diseases, including multiple sclerosis. Genes in this region often have very large effect sizes. It is unclear whether the two conclusions hold for other complex diseases. Additive interactions might be more common in interactions with small to moderate effect sizes than those with large effect sizes. For example, we have observed additive interactions between *MGAT1* and other genes in both multiple sclerosis and type 1 diabetes [Mkhikian et al., 2011; Yu et al., 2014]. Similarly, when there are a large number of interactions with small to moderate sizes, the conditional approach might also be useful. It is of great future interest to utilize and evaluate these strategies in more GWAS for complex diseases.

Last but not least, our work has several limitations. First, similar to other computationally efficient methods aforementioned, our closed-form solutions do not allow adjusting covariates. In practice, one can use our tests to identify a small number of gene-gene interactions to follow-up and then model disease risks using both the identified interactions and other covariates. Second, we proposed to improve computational efficiency from a statistical point-of-view. One can combine our closed-form tests with other strategies, such as parallel computing and GPU-accelerated computing, to further improve computation speed.

Acknowledgments

The research was supported in part by grant R01AI082266 from the National Institute of Allergy and Infectious Diseases to M.D., and by grant P30CA062203 from the National Cancer Institute to D.L.G. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113.

Web Resources

The C++ source code that implements the proposed tests is available from <http://www.ics.uci.edu/~zhaoxia/>.

References

- Barhdadi A, Dube MP. 2010. Testing for gene-gene interaction with AMMI models. *Stat Appl Genet. Mol. Biol.* 9(1):1544–6115.
- Bateson W. 1907. Facts limiting the theory of heredity. *Science* 26:649–660.
- Bhattacharya K, McCarthy MI, Morris AP. 2011. Rapid testing of gene-gene interactions in genome-wide association studies of binary and quantitative phenotypes. *Genet Epidemiol* 35(8):800–808.
- Carlborg O, Haley CS. 2004. Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* 5(8):618–U4.
- Chatterjee N, Kalaylioglu 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(6):1002–1016.
- Chen JB, Kang GL, VanderWeele T, Zhang CL, Mukherjee B. 2012. Efficient designs of gene-environment interaction studies: implications of Hardy-Weinberg equilibrium and gene-environment independence. *Stat Med* 31(22):2516–2530.
- Chen YH, Chatterjee N, Carroll RJ. 2008. Retrospective analysis of haplotype-based case-control studies under a flexible model for gene-environment association. *Biostatistics* 9(1):81–99.
- Cordell HJ. 2002. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11(20):2463–2468.
- Dai JY, Kooperberg C, Leblanc M, Prentice RL. 2012. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika* 99(4):929–944.
- Daly MJ, Altshuler D. 2005. Partners in crime. *Nat Genet* 37(4):337–338.
- Darroch JN. 1962. Interactions in multi-factor contingency-tables. *J R Stat Soc Series B Stat Methodol* 24(1):251–263.
- Dong CZ, Chu X, Wang Y, Wang Y, Jin L, Shi TL, Huang W, Li YX. 2008. Exploration of gene-gene interaction effects using entropy-based methods. *Eur J Hum Genet* 16(2):229–235.
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. VIEWPOINT missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11(6):446–450.
- Emily M, Mailund T, Hein J, Schausler L, Schierup MH. 2009. Using biological networks to search for interacting loci in genome-wide association studies. *Eur J Hum Genet* 17(10):1231–1240.
- Evans DM, Marchini J, Morris AP, Cardon LR. 2006. Two-stage two-locus models in genome-wide association. *PLoS Genet* 2(9):1424–1432.
- Fan R, Zhong M, Wang S, Zhang Y, Andrew A, Karagas M, Chen H, Amos CI, Xiong M, Moore JH. 2011. Entropy-based information gain approaches to detect and to characterize gene-gene and gene-environment interactions/correlations of complex diseases. *Genet Epidemiol* 35(7):706–721.
- Fernando MMA, Stevens CR, Walsh EC, DeJager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD. 2008. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 4(4):e1000024.
- Gregersen JW, Kranc KR, Ke XY, Svendsen P, Madsen LS, Thomsen AR, Cardon LR, Bell JI, Fugger L. 2006. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* 443(7111):574–577.
- Hoffmann TJ, Lange C, Vansteelandt S, Laird NM. 2009. Gene-environment interaction tests for dichotomous traits in trios and sibships. *Genet Epidemiol* 33(8):691–699.
- Hsu L, Jiao S, Dai JY, Hutter C, Peters U, Kooperberg C. 2012. Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genet Epidemiol* 36(3):183–194.
- Hu JK, Wang XL, Wang P. 2014. Testing gene-gene interactions in genome wide association studies. *Genet Epidemiol* 38(2):123–134.
- Ionita I, Man M. 2006. Optimal two-stage strategy for detecting interacting genes in complex diseases. *BMC Genet* 7:39.
- Jiao S, Hsu L, Berndt S, Bezieau S, Brenner H, Buchanan D, Caan BJ, Campbell PT, Carlson CS, Casey G and others. 2012. Genome-wide search for gene-gene interactions in colorectal cancer. *PLoS One* 7(12):e52535.
- Kooperberg C, LeBlanc M. 2008. Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol* 32(3):255–263.
- Lewinger JP, Morrison JL, Thomas DC, Murcray CE, Conti DV, Li DL, Gauderman WJ. 2013. Efficient two-step testing of gene-gene interactions in genome-wide association studies. *Genet Epidemiol* 37(5):440–451.
- Lincoln MR, Ramagopalan SV, Chao MJ, Herrera BM, DeLuca GC, Orton SM, Dyment DA, Sadovnick AD, Ebers GC. 2009. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc Natl Acad Sci USA* 106(18):7542–7547.
- Lippert C, Listgarten J, Davidson RI, Baxter J, Poon H, Kadie CM, Heckerman D. 2013. An exhaustive epistatic SNP association analysis on expanded wellcome trust data. *Scientific Reports* 3:1099.
- Liu Y, Xu HM, Chen SC, Chen XF, Zhang ZG, Zhu ZH, Qin XY, Hu LD, Zhu J, Zhao GP and others. 2011. Genome-wide interaction-based association analysis identified multiple new susceptibility loci for common diseases. *PLoS Genet* 7(3):e1001338.
- Ma L, Brautbar A, Boerwinkle E, Sing CF, Clark AG, Keinan A. 2012. Knowledge-driven analysis identifies a gene-gene interaction affecting high-density lipoprotein cholesterol levels in multi-ethnic populations. *PLoS Genet* 8(5):e1002714.
- MacGregor S, Khan IA. 2006. GAIA: an easy-to-use web-based application for interaction analysis of case-control data. *BMC Med Genet* 7:34.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A and others. 2009. Finding the missing heritability of complex diseases. *Nature* 461(7265):747–753.
- Marchini J, Donnelly P, Cardon LR. 2005. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37(4):413–417.
- 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(1):15–27.
- Mkhikian H, Grigorian A, Li CF, Chen HL, Newton B, Zhou RW, Beeton C, Torossian S, Tatarian GG, Lee SU and others. 2011. Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. *Nat Commun* 2:334.
- Moore JH, Williams SM. 2009. Epistasis and its implications for personal genetics. *Am J Hum Genet* 85(3):309–320.
- Murcray CE, Lewinger JP, Gauderman WJ. 2009. Gene-Environment Interaction in Genome-Wide Association Studies. *Am J Epidemiol* 169(2):219–226.
- Musani SK, Shriner D, Liu NJ, Feng R, Coffey CS, Yi NJ, Tiwari HK, Allison DB. 2007. Detection of gene x gene interactions in genome-wide association studies of human population data. *Hum Hered* 63(2):67–84.
- Newcombe RG. 1998. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 17(8):857–872.
- North BV, Curtis D, Sham PC. 2005. Application of logistic regression to case-control association studies involving two causative loci. *Hum Hered* 59(2):79–87.
- Plackett RL. 1962. A note on interactions in contingency-tables. *J R Stat Soc Series B Stat Methodol* 24(1):162–166.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, deBakker PIW, Daly MJ and others. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
- Roy SN, Kastenbaum MA. 1956. On the hypothesis of no interaction in a multi-way contingency table. *Ann Math Stat* 27(3):749–757.
- Sasieni PD. 1997. From genotypes to genes: doubling the sample size. *Biometrics* 53(4):1253–1261.
- Schwender H, Taub MA, Beaty TH, Marazita ML, Ruczinski I. 2012. Rapid testing of SNPs and gene-environment interactions in case-parent trio data based on exact analytic parameter estimation. *Biometrics* 68(3):766–773.
- Song MS, Nicolae DL. 2009. Restricted parameter space models for testing gene-gene interaction. *Genet Epidemiol* 33(5):386–393.
- Storey JD. 2003. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann Stat* 31(6):2013–2035.
- Tchetgen EJT, Kraft P. 2011. On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. *Epidemiology* 22(2):257–261.
- Tukey JK. 1949. One degree of freedom for nonadditivity. *Biometrics* 5:232–242.
- Ueki M, Cordell HJ. 2012. Improved statistics for genome-wide interaction analysis. *PLoS Genet* 8(4):141–159.
- VanderWeele TJ, Laird NM. 2011. Tests for compositional epistasis under single interaction-parameter models. *Ann Hum Genet* 75:146–156.
- Vansteelandt S, Vanderweele TJ, Tchetgen EJ, Robins JM. 2008. Multiply robust inference for statistical interactions. *J Am Stat Assoc* 103(484):1693–1704.
- Wan XA, Yang C, Yang QA, Xue H, Fan XD, Tang NLS, Yu WC. 2010. BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet* 87(3):325–340.
- Wang XF, Elston RC, Zhu XF. 2010a. The meaning of interaction. *Hum Hered* 70(4):269–277.
- Wang Y, Liu X, Robbins K, Rekaya R. 2010b. AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm. *BMC Res Notes* 3:117.
- Wellcome Trust Case Control Consortium, The Australo-Anglo-American Spondylitis Consortium (TASC). 2007. Association scan of 14500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 39(11):1329–1337.
- Wellek S, Ziegler A. 2009. A genotype-based approach to assessing the association between single nucleotide polymorphisms. *Hum Hered* 67(2):128–139.

- Wu J, Devlin B, Ringquist S, Trucco M, Roeder K. 2010a. Screen and clean: a tool for identifying interactions in genome-wide association studies. *Genet Epidemiol* 34(3):275–285.
- Wu XS, Dong H, Luo L, Zhu Y, Peng G, Reville JD, Xiong MM. 2010b. A novel statistic for genome-wide interaction analysis. *PLoS Genet* 6(9):e1001131.
- Yang C, He ZY, Wan X, Yang Q, Xue H, Yu WC. 2009a. SNPHarvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies. *Bioinformatics* 25(4):504–511.
- Yang Y, He C, Ott J. 2009b. Testing association with interactions by partitioning chi-squares. *Ann Hum Genet* 73:109–117.
- Yu Z, Li CF, Mkhikian H, Zhou RW, Newton BL, Demetriou M. 2014. Family studies of type 1 diabetes reveal additive and epistatic effects between MGAT1 and three other polymorphisms. *Genes Immun* 15(4):218–223.
- Yu ZX. 2011. Testing gene-gene interactions in the case-parents design. *Hum Hered* 71(3):171–179.
- Zhao JY, Jin L, Xiong MM. 2006. Test for interaction between two unlinked loci. *Am J Hum Genet* 79(5):831–845.