Efficient Strategy for Detecting Gene × Gene Joint Action and Its Application in Schizophrenia

Sungho Won,[1,2]* Min-Seok Kwon,[3] Manuel Mattheisen,[4,5] Suyeon Park,[1] Changsoon Park,[1,2] Daisuke Kihara,[6] Sven Cichon,[4,5,7] Roel Ophoff,[8] Markus M. Nöthen,[4,5] Marcella Rietschel,[9] Max Baur,[10,11] Andre G. Uitterlinden,[12,13] A. Hofmann,[14] GROUP Investigators,[15] and Christoph Lange[11,16,17,18]

[1]Department of Applied Statistics, Chung-Ang University, Seoul, Korea; [2]Research Center for Data Science, Chung-Ang University, Seoul, Korea; [3]Bioinformatics Program, Seoul National University, Seoul, Korea; [4]Institute of Human Genetics, University of Bonn, Bonn, Germany; [5]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany; [6]Department of Computer Sciences, Purdue University, West Lafayette, Indiana, United States of America; [7]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany; [8]Department of Medical Genetics, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; [9]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Mannheim, Mannheim, Germany; [10]Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany; [11]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; [12]Department of Internal Medicine, Genetics Laboratory, Eramsmus Medical Center Rotterdam, The Netherlands; [13]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; [14]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; [15]For a full list of members see Acknowledgments; [16]Institute for Genomic Mathematics, University of Bonn, Bonn, Germany; [17]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America; [18]Center for Genomic Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America

ABSTRACT: We propose a new approach to detect gene × gene joint action in genome-wide association studies (GWASs) for case-control designs. This approach offers an exhaustive search for all two-way joint action (including, as a special case, single gene action) that is computationally feasible at the genome-wide level and has reasonable statistical power under most genetic models. We found that the presence of any gene × gene joint action may imply differences in three types of genetic components: the minor allele frequencies and the amounts of Hardy-Weinberg disequilibrium may differ between cases and controls, and between the two genetic loci the degree of linkage disequilibrium may differ between cases and controls. Using Fisher's method, it is possible to combine the different sources of genetic information in an overall test for detecting gene × gene joint action. The proposed statistical analysis is efficient and its simplicity makes it applicable to GWASs. In the current study, we applied the proposed approach to a GWAS on schizophrenia and found several potential gene × gene interactions. Our application illustrates the practical advantage of the proposed method.
Genet Epidemiol 38:60–71, 2014. © 2013 Wiley Periodicals, Inc.

KEY WORDS: gene × gene joint action; minor allele frequency; Hardy-Weinberg disequilibrium; linkage disequilibrium

Introduction

The availability of high-throughput genotyping technology has enabled genome-wide association studies (GWASs), which have provided an unprecedented opportunity to identify genetic variants for many human diseases, including age-related macular degeneration [Klein et al., 2005] , body mass index [Church et al., 2010], Crohn's disease [Barrett et al., 2008], and Alzheimer's disease [Harold et al., 2009]. However, although GWASs have successfully identified many susceptibility variants, the findings explain only a fraction of the expected overall heritability and replication of significant results has often failed. It has been speculated that some of the unexplained heritability and inconsistency among replica-tion studies may be caused by epistatic effects [Culverhouse et al., 2002]. Simulations, for instance, have shown that in the presence of gene × gene interaction, differences in the allele frequency between the initial study and the replication populations can affect the power of single-locus strategies and hamper the reproducibility of true findings [Marchini et al., 2005]. At the same time, increasing empirical evidence supports that general gene × gene interactions [Zerba et al., 2000] play a key role in the variation of complex traits. Consequently, the systematic statistical analysis of gene × gene interaction is critical for the success of GWASs in genetic epidemiology.

However, in spite of the importance of gene × gene interaction, the meaning of gene × gene interaction has not been clearly understood, and statistical gene × gene interaction has often been confused with biological gene × gene interaction. In particular, the inference on the biological mechanism is

*Correspondence to: Sungho Won, Department of Applied Statistics, Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul 156-756, Korea. E-mail: swon@cau.ac.kr

complicated because of the lack of direct correspondence between statistical and biological interaction [Ueki and Cordell, 2012]. Statisticians usually define statistical interaction as a departure from additivity in a linear model using a selected measurement scale. However, as was pointed out by Wang et al. [Wang et al., 2011], if one aims to infer biological interactions, statistically modeled interactions and main effect terms should not be separately interpreted. In this context, it has been suggested that the term "biological interaction" can be detected by the statistical model that compares joint multilocus genotype frequencies [Wang et al., 2010]. In this study we focus on the biological interaction and it will be denoted as the joint action to avoid confusion.

Numerous parametric and nonparametric approaches have been proposed to detect gene × gene joint action. Parametric approaches are usually based on logistic regression [Cordell, 2002] and shrinkage approaches have been applied to decrease the multiple testing problems [Park and Hastie, 2008]. Nonparametric approaches, including the multifactor dimensionality reduction (MDR) [Ritchie et al., 2001], multivariate adaptive regression spline [Cook et al., 2004], and random forest approaches [Jiang et al., 2009], have also been adapted to detect gene × gene joint action. However, most statistical methods for detecting gene × gene joint action are affected by some intractable issues. First, gene × gene joint action analysis at the genome-wide scale is computationally intensive and computational requirements can detract from the strength of each method. MDR has been shown to be a powerful approach for complex gene × gene joint action [Ritchie et al., 2001]; however, the initial marker selection step, which alleviates the computational intensity, can reduce the statistical efficiency of this approach. Cordell concluded that the semiexhaustive search of two-locus interactions implemented in PLINK [Purcell et al., 2007] and the random forests analysis implemented in Random Jungle [Cordell, 2009] are computationally feasible approaches. Recent computational achievement based on graphic processing units [Chikkagoudar et al., 2011; Yung et al., 2011] or multithreaded parallelization [Gyenesei et al., 2012] also enables the genome-wide gene × gene joint action analysis. However, further investigation on the efficient algorithm is still necessary. Second, even for a two-locus joint action model, there are 50 different models for gene × gene joint action [Li and Reich, 2000], and an exhaustive search of all these interaction models generates a substantial statistical burden. To maximize statistical efficiency, it may be better to search for significant differences in some parameters obtained by transforming the two-locus genotype frequencies.

In genetic association analysis for multiple SNPs, three potential statistically independent quantities provide information about the genetic association at unlinked loci [Won and Elston, 2008; Won et al., 2009a]. First, the association of single or multiple SNPs with disease produces the different allele frequencies between cases and controls [Sasieni, 1997]. Second, the amount of Hardy-Weinberg disequilibrium (HWD) at single or multiple loci may differ between cases and controls [Nielsen et al., 1998]. Third, the amount of linkage disequilibrium (LD) may differ between cases and

controls [Nielsen et al., 2004], where the composite LD can be used if genotypes are not phased [Zaykin et al., 2006]. The difference of these quantities between cases and controls can be utilized to detect the gene × gene joint action. In this report, information about the presence of a gene × gene joint action that is included in these three different quantities is utilized to maximize the statistical power of the analysis, while at the same time, the numerical complexity is minimized. Thus, the proposed approach enables gene × gene joint action analysis on a genome-wide scale via exhaustive analysis of all possible two-way interactions.

Furthermore, we applied our method to a GWAS on schizophrenia (SCZ). SCZ is a common psychiatric disorder that results from interactions between genetic and environmental factors [Sullivan, 2005]. Recent GWASs [Athanasiu et al., 2010; Djurovic et al., 2010] have identified genetic risk variants for SCZ, but the findings explain only a small fraction of the estimated heritability, which is around 80% [Cardno and Gottesman, 2000]. One hypothesis is that gene × gene interactions may be the cause of the missing heritability. We applied the proposed method to a GWAS on SCZ, and discovered the occurrence of some potential gene × gene joint action. Our analysis results for SCZ illustrate the substantial advantages of our approach in terms of both computational intensity and statistical power.

Method

We consider the case-control studies, and the causal gene × gene joint action of two loci with two alleles is assumed. The two alleles at the two loci are denoted by E and e, and F or f, respectively, and we used $p_{EF|EF}^a$ and $p_{EF|EF}^u$ to indicate the frequencies of the phase-known genotype ($EF|EF$) in cases and controls, respectively. It should be noted that a and u as superscripts are used to indicate cases and controls, respectively. We considered p_E and p_F as allele frequencies in the population, and $p_E^a (p_E^u)$ and $p_{EE}^a (p_{EE}^u)$ as allele and genotype frequencies, respectively, in cases (controls). We denote the joint frequency at two different gametes in cases and controls by $p_{E:F}^a$ and $p_{E:F}^u$, respectively. $\Delta_{EF}^a = p_{EF}^a - p_E^a \cdot p_F^a$ and $\Delta_{EF}^u = p_{EF}^u - p_E^u \cdot p_F^u$ are the amounts of LD in cases and controls. Because LD is usually not observed, we consider the composite LD [Weir, 1996; Zaykin, 2004] and the composite LDs in cases and controls are denoted as $\delta_{EF}^a = p_{EF}^a + p_{E:F}^a - 2p_E^a \cdot p_F^a$ and $\delta_{EF}^u = p_{EF}^u + p_{E:F}^u - 2p_E^u \cdot p_F^u$, respectively. We denote the amount of single-locus level and haplotype-level HWD for the cases by $D_E^a = p_{EE}^a - 2p_E^a p_E^a$ and $D_{EF|Ef}^a = p_{EF|Ef}^a - 2p_{EF}^a p_{Ef}^a$, respectively. The numbers of cases and controls are assumed to be n_a and n_u, respectively.

Analysis of gene × gene joint action typically tests the homogeneity of genotype frequencies in cases and controls. There are 9 and 10 different genotypes at two SNPs when the genotypes are unphased and phased, respectively, and the differences in the two-locus genotype frequencies between cases and controls suggest the presence of gene × gene joint action, a special case of which could be single gene action. Therefore, the presence of gene × gene joint action can be

detected by comparing the two-locus genotype frequencies. In particular, when genotypes are phased, the null hypothesis for any joint action analysis is as follows:

$$H_0 : p^a_{EF|EF} = p^u_{EF|EF}, \ldots, p^a_{ef|ef} = p^u_{ef|ef}.$$

The corresponding null hypothesis for the phase-unknown genotypes can be simply constructed. This null hypothesis is equivalent to the following null hypothesis under a transformation [Won and Elston, 2008]:

$$H_0 : p^a_E = p^u_E, p^a_F = p^u_F, \Delta^a_{EF} = \Delta^u_{EF},$$
$$D^a_{EF|Ef} = D^u_{EF|Ef}, \ldots, D^a_{eF|ef} = D^u_{eF|ef}.$$

Here, nine components that comprise the two minor allele frequencies, six components for the level of haplotype HWD, and one component for the level of LD are required to test the equivalence of the phased genotype frequencies, and they can be categorized into three types of informative components for association analysis [Won and Elston, 2008]: the allele frequency, the amount of HWD, and the level of LD. Under this transformation, general gene × gene joint action can be detected by comparing the differences in these three types of components instead of the two-locus genotype frequencies.

In the proposed null hypothesis, however, the haplotype is often not available, and estimating haplotype frequencies for the two loci increases the additional computational complexity. Consequently, we consider the amount of single-locus level HWD instead of haplotype-level HWD, and composite LD [Wang et al., 2007; Won et al., 2009a; Zaykin, 2004] for the computation of Δ^a_{AB} and Δ^u_{AB}. Furthermore, because Hardy-Weinberg equilibrium (HWE) at each marker and linkage equilibrium (LE) between two markers are usually preserved in a population and cases are often less common than controls, HWE and LE tests computed in cases often constitute a more efficient approach than testing the different amounts of HWD and LD between cases and controls [Wang and Shete, 2008; Won and Elston, 2008]. The following testing strategy therefore seems to be justified. If the results for the tests for HWE and LE in controls are not significant, we propose to use tests for HWE and LE only in cases; if they are significant, the amounts of HWD or LD in cases and controls have to be compared. Consequently, when HWE and LE are preserved in controls, we consider the following null hypothesis:

$$H_0 : p^a_E = p^u_E, \quad p^a_F = p^u_F, D^a_E = D^a_F$$
$$= \delta^a_{EF} = 0 \quad \text{vs} \quad H_1 : \text{not } H_0.$$

If two markers are linked and the HWE at each locus is not preserved in controls, the null hypothesis is equivalent to the following:

$$H_0 : p^a_E = p^u_E, p^a_F = p^u_F, D^a_E = D^u_E, D^a_F = D^u_F,$$
$$\delta^a_{EF} = \delta^u_{EF} \quad \text{vs} \quad H_1 : \text{not } H_0.$$

For statistical testing under this null hypothesis, we first calculated the statistics for each equality and then combined their P-values into a single statistic. When HWE and LE were preserved in controls, our individual statistics for each equality in the null hypothesis were as follows:

$$T^1_E = \frac{(\hat{p}^a_E - \hat{p}^u_E)^2}{\text{var}(\hat{p}^a_E - \hat{p}^u_E)}, \quad T^1_F = \frac{(\hat{p}^a_F - \hat{p}^u_F)^2}{\text{var}(\hat{p}^a_F - \hat{p}^u_F)},$$

$$T^{2|a}_E = \frac{(\hat{D}^a_E)^2}{\text{var}(\hat{D}^a_E)}, \quad T^{2|a}_F = \frac{(\hat{D}^a_F)^2}{\text{var}(\hat{D}^a_F)}, \quad T^{3|a}_{EF} = \frac{(\hat{\delta}^a_{EF})^2}{\text{var}(\hat{\delta}^a_{EF})}.$$

Here the variances in the amount of HWD and composite LD under HWE and LE [Weir, 1996] can be calculated as follows:

$$\text{var}(\hat{D}^a_E) = \frac{1}{n_a}(p^a_E)^2(1 - p^a_E)^2,$$

$$\text{var}(\hat{\delta}^a_{EF}) = \frac{1}{2n_a} p^a_E(1 - p^a_E)p^a_F(1 - p^a_F).$$

When HWE and LE are not guaranteed in controls, we should test the homogeneity of the level of the HWD and LD between cases and controls, rather than $T^{2|a}_E$, $T^{2|a}_F$, and $T^{3|a}_{EF}$:

$$T^2_E = \frac{(\hat{D}^a_E - \hat{D}^u_E)^2}{\text{var}(\hat{D}^a_E - \hat{D}^u_E)}, \quad T^2_F = \frac{(\hat{D}^a_F - \hat{D}^u_F)^2}{\text{var}(\hat{D}^a_F - \hat{D}^u_F)},$$

$$T^3_{EF} = \frac{(\hat{\delta}^a_{EF} - \hat{\delta}^u_{EF})^2}{\text{var}(\hat{\delta}^a_{EF} - \hat{\delta}^u_{EF})}.$$

After we calculated these statistics, the second step of our approach was to provide overall inference based on these derived individual statistics. Although the most powerful test can be constructed when some information is available about effect sizes [Won et al., 2009b], there does not exist the uniformly most powerful method for combining P-values of different test statistics [Birnbaum, 1954]. We decided to use Fisher's method in this study because this method has been shown to be the most efficient approach in terms of Bahadur's relative efficiency [Little and Folks, 1971]. Therefore, we propose the following overall test statistic to test for the presence of a gene × gene joint action. Under HWE and LE in the population, the test statistics, T^1_E, T^1_F, $T^{2|a}_E$, $T^{2|a}_F$, and $T^{3|a}_{EF}$ are mutually independent [Won and Elston, 2008], and Fisher's method can be used to combine these statistics. Therefore, if we denote their P-values by P^1_E, P^1_F, $P^{2|a}_E$, $P^{2|a}_F$, and $P^{3|a}_{EF}$, respectively, our overall statistic under HWE and LE in the population becomes

$$-2\big[\log P^1_E + \log P^1_F + \log P^{2|a}_E + \log P^{2|a}_F + \log P^{3|a}_{EF}\big]$$
$$\sim \chi^2(df = 10) \quad \text{under } H_0,$$

and they are denoted as FIS^1. In the absence of HWE or LE, we cannot guarantee their independence, and in this case, or when the sample sizes are small, permutation tests have to be used to maintain the correct significance level of the overall test statistic. If we consider the P-values for T^2_E, T^2_F, and T^3_{EF} as P^2_E, P^2_F, and P^3_{EF}, respectively, we can calculate $-2[\log P^1_E + \log P^1_F + \log P^2_E + \log P^2_F + \log P^3_{EF}]$ for both the original and permuted samples, and that from the original sample needs to

be compared with those from permuted samples to calculate the empirical P-value.

When detection of only a statistical gene × gene interaction that is orthogonal to the main effects is required, inference with the proposed overall statistic is not valid because a difference in minor allele frequencies by itself can generate significant results. Therefore, it is better to consider only T_{EF}^3 or $T_{EF}^{3|a}$. Notably, $\mathrm{var}(\hat{\delta}_{EF}^a)$ and $\mathrm{var}(\hat{\delta}_{EF}^u)$ for T_{EF}^3 and $T_{EF}^{3|a}$ in the presence of marginal effects are affected by the disease mode of inheritance. For instance, even when HWE and LE are present in the population, the presence of dominant or recessive disease effects generates HWD in cases and controls, and then the variances of $\hat{\Delta}_{EF}^a$ and $\hat{\Delta}_{EF}^u$ in the absence of gene × gene interaction are respectively shown by Weir [Weir, 1996] as follows:

$$\mathrm{var}\left(\hat{\delta}_{EF}^a\right) = \frac{1}{2n_a}\left[p_E^a\left(1-p_E^a\right)+D_E^a\right]\left[p_F^a\left(1-p_F^a\right)+D_F^a\right],$$

$$\mathrm{var}\left(\hat{\delta}_{EF}^u\right) = \frac{1}{2n_u}\left[p_E^u\left(1-p_E^u\right)+D_E^u\right]\left[p_F^u\left(1-p_F^u\right)+D_F^u\right].$$

Results

Simulation Studies for Joint Action Analysis

All possible two-locus models can be explained with 50 different models, and in our simulations we considered the 8 typical two-locus models for gene × gene joint action [Dong et al., 2008], M1–M8 (see Fig. 1): the joint recessive-recessive model, joint recessive-dominant model, single-gene recessive model, threshold model, modifying-effect model, joint dominant-dominant model, exclusive OR model, and the diagonal model. The diagonal model is the disease model where the diagonal elements in the contingency table of two-locus genotypes have higher risk than the off-diagonal elements (see Fig. 1). In Figure 1, λ_1 and λ_2 denote the disease genotype relative risk in the case of low and high risk, respectively, and λ_1 is 1.

We calculated the noncentrality parameters of the statistics for each equality in the alternative hypothesis for gene × gene joint action analysis. The disease prevalence and the disease allele frequency were assumed to be 0.2. We assumed HWE and LE between two SNPs in the population and, under these

M1 (RR)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_1
Ee	λ_1	λ_1	λ_1
ee	λ_1	λ_1	λ_2

M2 (RD)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_1
Ee	λ_1	λ_1	λ_1
ee	λ_1	λ_2	λ_2

M3 (1L:R)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_1
Ee	λ_1	λ_1	λ_1
ee	λ_2	λ_2	λ_2

M4 (T)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_1
Ee	λ_1	λ_1	λ_2
ee	λ_1	λ_2	λ_2

M5 (Mod)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_1
Ee	λ_1	λ_1	λ_2
ee	λ_2	λ_2	λ_2

M6 (DD)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_1
Ee	λ_1	λ_2	λ_2
ee	λ_1	λ_2	λ_2

M7 (XOR)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_2
Ee	λ_1	λ_1	λ_2
ee	λ_2	λ_2	λ_1

M8 (Diagonal)			
SNP 1	SNP 2		
	FF	Ff	ff
EE	λ_1	λ_1	λ_2
Ee	λ_1	λ_2	λ_1
ee	λ_2	λ_1	λ_1

Figure 1. Eight typical two-locus models. λ_1 denotes low-risk genotype combinations and λ_2 denotes high-risk genotype combinations. M1, joint recessive-recessive model (RR); M2, joint recessive-dominant model (RD); M3, single-gene recessive model (1L: R); M4, threshold model (T); M5, modifying-effect model (Mod); M6, joint dominant-dominant model (DD); M7, exclusive OR model (XOR); M8, diagonal model (Diagonal).
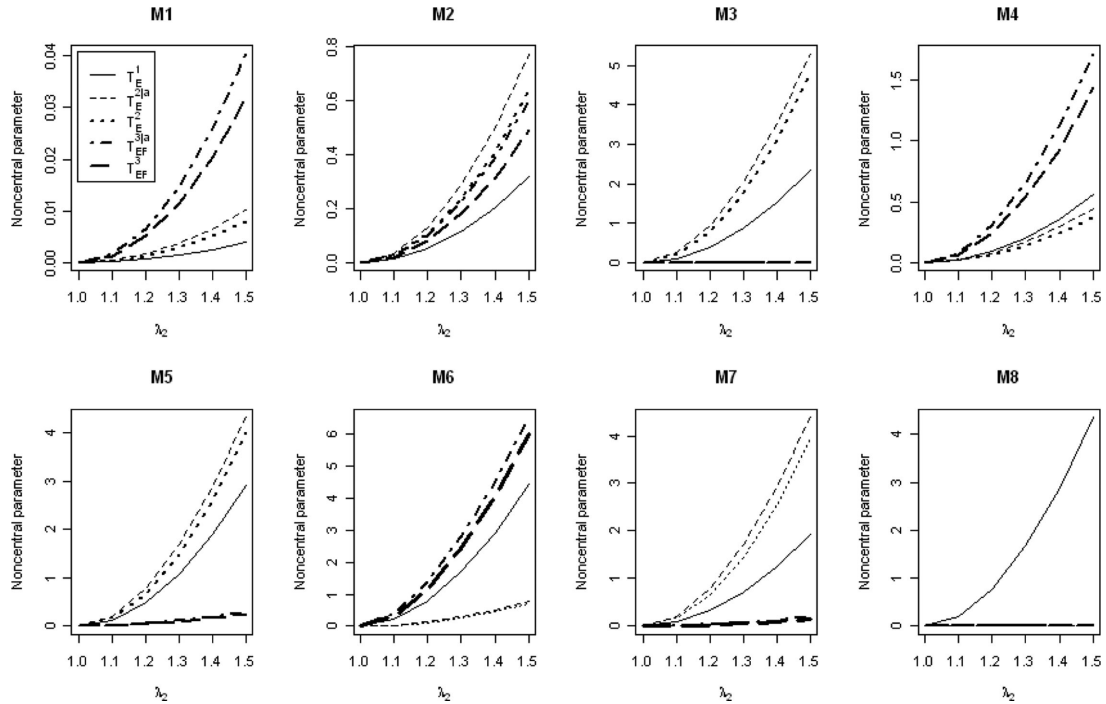
Figure 2. Noncentrality parameters for T_E^1, T_E^2, $T_E^{2|a}$, T_{EF}^3, and $T_{EF}^{3|a}$ under M1, M2, ..., M8. Noncentrality parameters were calculated for the three types of components under the four gene × gene interaction models (see Fig. 1). The disease allele frequencies for both SNPs were assumed to be 0.2. The disease prevalence was assumed to be 0.2, and the penetrance for each disease genotype was calculated under the assumed disease prevalence.

circumstances, it should be noted that the statistics for the three components are asymptotically mutually independent [Won and Elston, 2008]. The penetrance for each disease genotype was calculated under the assumed disease model. Figure 2 shows the noncentrality parameters for M1–M8. Our results show that the three types of components are usually informative for each joint action model except M8. Furthermore, testing differences in the HWD and LD between cases and controls is a less powerful approach than testing HWE and LE in cases alone.

Furthermore, the empirical estimates of type-1 error and power were calculated with the simulated data. The disease prevalence was assumed to be 0.2. We assumed HWE and LE in the population. The disease genotypes were generated from the multinomial distribution and their phenotypes were then sampled with the penetrances of their genotypes. Sampling was repeated until the given numbers of cases and controls were obtained. In our simulation, FIS^1 combines P_E^1, P_F^1, $P_E^{2|a}$, $P_F^{2|a}$, and $P_{EF}^{3|a}$ with Fisher's method. FIS^2 combines P_E^1, P_F^1, P_E^2, P_F^2, and P_{EF}^3. Table 1 shows the empirical type-1 error estimates from 5,000 replicates at the 0.1 significance level. In Table 1, we assumed the disease genotype relative risks $\lambda_1 = \lambda_2 = 1$ for any combination of two SNPs. The disease allele frequencies for the two disease loci were randomly generated from the uniform distribution $U(0.1, 0.9)$. Our results show that there is no evidence of inflation for the proposed methods for detecting joint action.

The empirical power estimates at the 0.01 significance level have been shown in Figure 3. In our simulations, we considered nine different joint action models, that is, M1–M9. For M1–M8, the joint action models are shown in Figure 1, and λ_2 was assumed to be 1.5. For M9, the disease genotype relative risks for each two-locus genotype were randomly sampled from the uniform distribution $U(1, 1.5)$. In particular, the detectable relative risks for 80% power by using FIS^1 for M1–M8, for the case illustrated in Figure 3, are 7.9, 2.4, 1.7, 1.9, 1.6, 1.4, 1.5, and 1.6, respectively, which indicates that M1 contrary to the other disease models have low power because only one genotype has high relative risk. For disease allele frequencies, we assumed that $p_E = p_F = 0.2$, but we found that the results for other disease allele frequencies were similar. For comparison with the traditional methods, the generalized

Table 1. Empirical type-1 error at the 0.1 significance level

N	$T_{EF}^{3|a}$	T_{EF}^3	FIS^1	FIS^2
1,000	0.103 ± 0.009	0.116 ± 0.009	0.098 ± 0.008	0.114 ± 0.009
2,000	0.099 ± 0.008	0.099 ± 0.008	0.108 ± 0.009	0.096 ± 0.008
3,000	0.096 ± 0.008	0.09 ± 0.008	0.099 ± 0.008	0.097 ± 0.008

The empirical type-1 errors at the 0.1 significance level and 95% confidence intervals were calculated from 5,000 replicates. Cases (n_a = 1,500) and controls (n_u = 1,500) were generated, and the disease allele frequencies for the two disease loci were randomly generated from the uniform distribution $U(0.1, 0.9)$. FIS^1 combines P_E^1, P_F^1, $P_E^{2|a}$, $P_F^{2|a}$, and $P_{EF}^{3|a}$ with Fisher's method and FIS^2 combines P_E^1, P_F^1, P_E^2, P_F^2, and P_{EF}^3.
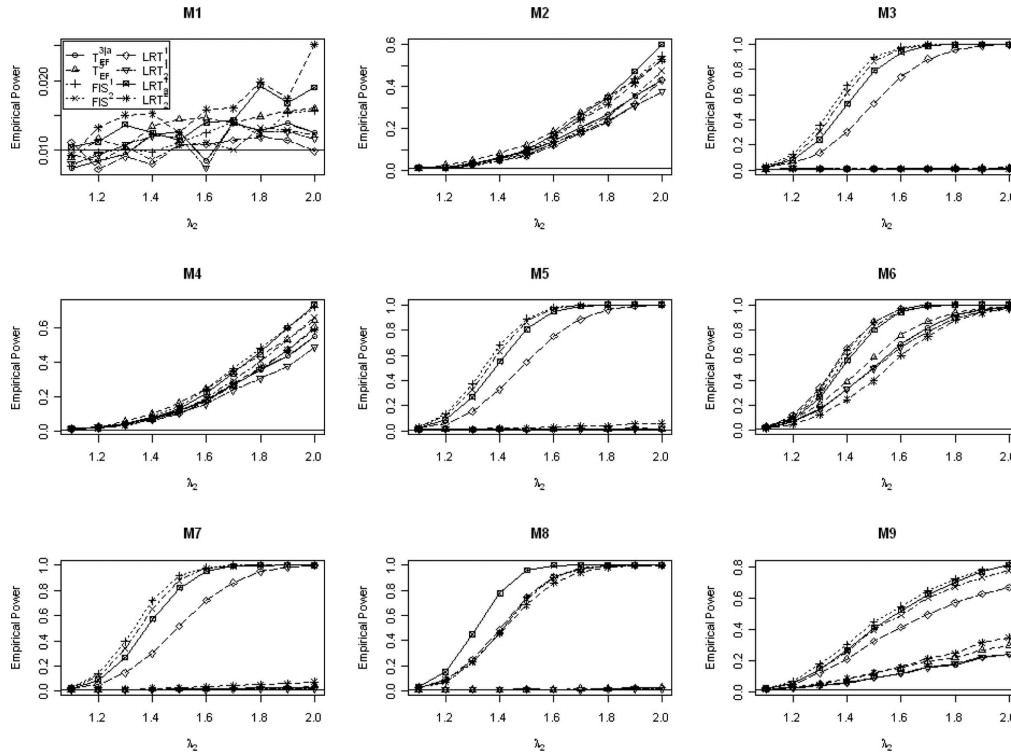
Figure 3. Empirical power estimates at the 0.01 significance level for gene × gene joint action analysis. Empirical power for various values of λ_2 was estimated with 5,000 replicates at the 0.01 significance level. The disease allele frequencies for two disease SNPs were assumed to be 0.3 and 0.1, respectively. The disease prevalence was assumed to be 0.2. FIS^1 combines P_E^1, P_F^1, $P_E^{2|a}$, $P_F^{2|a}$, and $P_{EF}^{3|a}$ with Fisher's method, and FIS^2 combines P_E^1, P_F^1, P_E^2, P_F^2, and P_{EF}^3. LRT_1^1 (LRT_2^1) indicates the likelihood ratio test that compares the linear interaction model in logistic regression with the intercept model (the linear model). For LRT_1^2 and LRT_2^2, dummy variables are utilized, and both indicate the likelihood ratio tests that compare the linear interaction model in logistic regression with the intercept model and the linear model, respectively.

linear model for a binary trait with a logit link was considered. LRT_1^1 indicates the likelihood ratio test that compares the linear interaction model in logistic regression with the intercept model, and in LRT_2^1 the linear interaction model in logistic regression has been compared with the linear model without the interaction terms. Therefore, under the null hypothesis, LRT_1^1 follows the chi-square distribution with three degrees of freedom, and LRT_2^1 follows the chi-square distribution with one degree of freedom. For LRT_1^2 and LRT_2^2, dummy variables are utilized for each two-locus genotype score instead of additive genotype scores. The linear model with nine dummy variables, one for each genotype, in logistic regression is compared with the intercept model for the former, and with the linear model without interaction terms for the latter. Therefore, LRT_1^2 and LRT_2^2 indicate likelihood ratio tests with eight and four degrees of freedom, respectively.

FIS^1 seems to generally be the most powerful approach (Fig. 3), although the most powerful approach varies depending on gene × gene joint action model. In particular, FIS^1 usually performs better than FIS^2, and this indicates that testing HWE and LE in cases is more efficient than testing the differences in HWE and LE between cases and controls. Comparison of FIS^1 with $T^{3|a}$ showed the power improvement of the overall statistics compared to the individual statistic,

$T^{3|a}$. The results for M9 indicate that the proposed method is more informative than the other methods when the two-locus genotype effects are randomly assigned. LRT_1^2 (LRT_2^2) is usually more efficient than LRT_1^1 (LRT_2^1), and the gene × gene joint action model using dummy variables for each two-locus genotype score seems to be better than additive genotype scoring. These results suggest that for many gene × gene joint action models, the overall differences between cases and controls are captured by the proposed statistic and that the individual statistics for the marginal effects are also informative for detecting gene × gene joint action. The logistic regressions simply compare the differences between the joint disease genotype frequencies, but their performance can be affected by the genotype scoring.

Simulation Studies for Statistical Gene × Gene Interaction

The empirical type-1 error and power estimates for statistical gene × gene interaction analysis were also evaluated with the simulated data. The disease prevalence was set to 0.2, and HWE and LE in the population were assumed at the disease locus. We assumed that both p_E and p_F were 0.2. The disease genotypes were generated from a multinomial distribution, and their phenotypes were then sampled with the penetrance

Table 2. Empirical type-1 error at the 0.1 significance level

Disease mode of inheritance	λ_2	$T_{EF}^{3|a}$		T_{EF}^{3}		$LRT_2{}^1$	$LRT_2{}^2$	$WALD_1{}^2$
		HWE	HWD	HWE	HWD			
Recessive	2	0.1222 ± 0.0092	0.0980 ± 0.0084	0.1178 ± 0.0091	0.1094 ± 0.0088	0.1026 ± 0.0086	0.1150 ± 0.0090	0.1018 ± 0.0086
	3	0.1364 ± 0.0097	0.0972 ± 0.0084	0.1150 ± 0.0090	0.0988 ± 0.0084	0.0890 ± 0.0081	0.1008 ± 0.0085	0.0880 ± 0.0080
	4	0.1482 ± 0.0100	0.1000 ± 0.0085	0.1244 ± 0.0093	0.1064 ± 0.0087	0.0766 ± 0.0075	0.1022 ± 0.0086	0.0758 ± 0.0075
Dominant	2	0.0794 ± 0.0076	0.1052 ± 0.0087	0.0946 ± 0.0083	0.1066 ± 0.0087	0.1142 ± 0.0090	0.1146 ± 0.0090	0.1138 ± 0.0090
	3	0.0654 ± 0.0070	0.1100 ± 0.0088	0.0806 ± 0.0077	0.1024 ± 0.0086	0.1328 ± 0.0096	0.1124 ± 0.0089	0.1320 ± 0.0096
	4	0.0486 ± 0.0061	0.1064 ± 0.0087	0.0706 ± 0.0072	0.1044 ± 0.0086	0.1360 ± 0.0097	0.1156 ± 0.0090	0.1348 ± 0.0097

The empirical type-1 error at the 0.1 significance level was calculated from 5,000 replicates. 1,500 cases and 1,500 controls were generated and the disease allele frequencies for the two disease loci were randomly generated from the uniform distribution $U(0.1, 0.9)$. $LRT_2{}^1$ indicates the likelihood ratio test that compares the linear interaction model in logistic regression with the linear model. For $LRT_2{}^2$, dummy variables are utilized and it indicate the likelihood ratio tests that compare the linear interaction model in logistic regression with the linear model. $WALD_1{}^2$ indicates the Wald test for the coefficient of statistical gene × gene interaction in an additive genotype scoring.
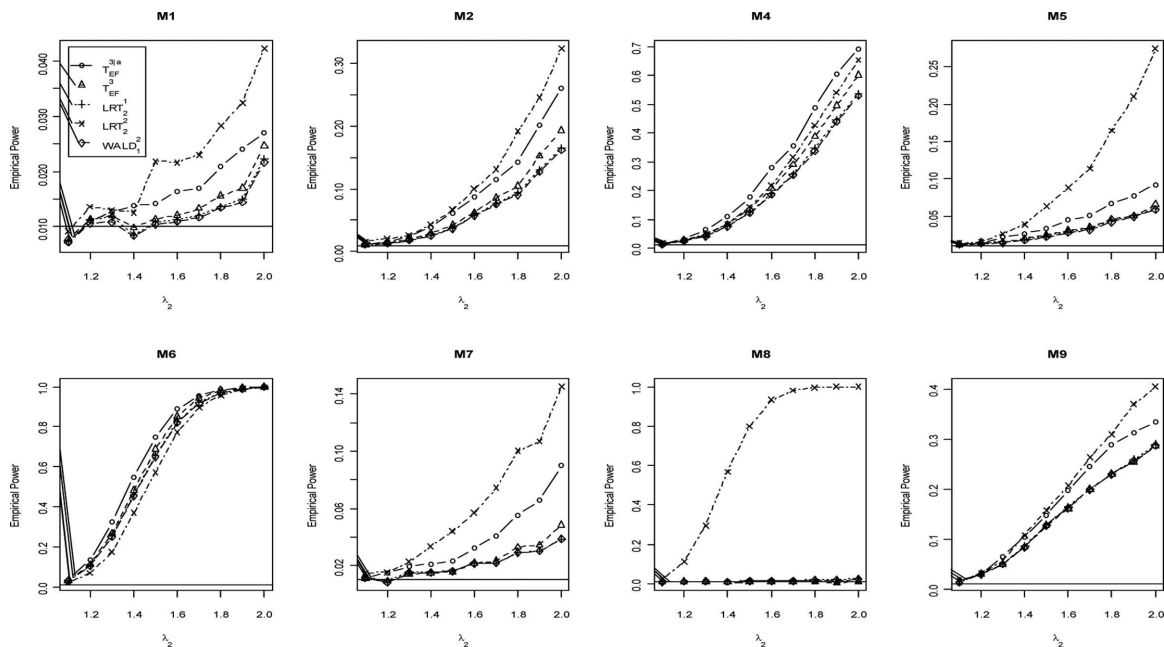
Figure 4. Empirical power estimates at the 0.01 significance level for gene × gene statistical interaction analysis. Empirical powers for various λ_2 were estimated with 5,000 replicates at the 0.01 significance level. The disease allele frequencies for two disease SNPs were assumed to be 0.2 and 0.2, respectively. The disease prevalence was assumed to be 0.2. M3 was excluded because it assumes that there is a single gene effect.

for each genotype. Sampling was repeated until 1,500 cases and 1,500 controls were obtained. We considered $P_{EF}^{3|a}$, P_{EF}^{3}, $LRT_2{}^1$, and $LRT_2{}^2$ for gene × gene interactions, and $\mathrm{var}(\hat{\delta}_{EF}^{a})$ and $\mathrm{var}(\hat{\delta}_{EF}^{a})$ in $P_{EF}^{3|a}$ and P_{EF}^{3} were calculated under both the presence and absence of HWE. In addition, we considered the Wald test for the coefficient of the statistical gene × gene interaction term in the logistic regression, and it was denoted by $WALD_1^2$.

For statistical validity, we assumed the presence of the marginal genetic effects in the absence of statistical gene × gene interactions; inference regarding the statistical gene × gene interaction term should not be affected by the presence of marginal genetic effects. Even when HWE is preserved in the population, the HWE in cases is not preserved at a disease locus with a dominant or recessive marginal genetic effect [Won and Elston, 2008], and both disease models were

considered in our simulation studies. We assumed that E was a disease allele, and the high disease-genotype risk, λ_2, for the marginal genetic effect was assumed to be 1.5. We calculated the empirical type-1 error estimates from 5,000 replicates at the 0.1 significance level (Table 2). The results show that $P_{EF}^{3|a}$ and P_{EF}^{3} always preserve the significance level if $\mathrm{var}(\hat{\delta}_{EF}^{a})$ and $\mathrm{var}(\hat{\delta}_{EF}^{u})$ are calculated under HWD but that the empirical type-1 errors are inflated otherwise. The empirical type-1 error estimates have been slightly inflated for $LRT_2{}^1$, $LRT_2{}^2$, and for $WALD_1{}^2$ under the dominant marginal effect model, and these statistics appear to be affected by the presence of marginal effects. We also calculated the empirical power estimates (Fig. 4). M1–M9, excluding M3, were considered, because M3 has no gene × gene interaction effect. Our results show that $LRT_2{}^2$ with four degrees of freedom is the most efficient and it is followed by $P_{EF}^{3|a}$. However, the

parameterization for $LRT_2{}^2$ includes the marginal genetic effect because the empirical type-1 error rate estimates have not preserved the nominal significance level and the computation of the log-likelihood ratio test for logistic regression is known to be intensive. Therefore $P_{EF}^{3|a}$ can be a reasonable alternative for gene × gene interaction analysis at the genome-wide scale.

Data Analysis

We applied the proposed method to the German-Dutch GWAS on SCZ. The details of the study are provided in the Appendix. Because heterogeneity between the Dutch and German subjects was observed in the sample, the individual statistics for each equality in our null hypothesis were calculated separately for each group, i.e., Dutch samples vs. German samples. The tests of HWE in both cases and controls were used for quality-control purposes. Therefore, even though the tests of HWE in cases are informative to detect the joint action, these were not informative in our situation, and we did not include them in the overall statistics, because all SNPs with small P-values for the HWE test were excluded from the analysis. In addition, we focused on unlinked markers. Pairs of markers were considered as being linked if $T_{EF}^{3|u}$

for the pairs of SNPs are significant at the 0.05 significance level, and they were excluded from the analysis. For unlinked markers, we used $T_{EF}^{3|a}$ to detect the gene × gene interaction and it should be noted that $T_{EF}^{3|u}$ and $T_{EF}^{3|a}$ are independent. If we let the subscripts Dutch and German distinguish the P-values for Dutch samples and German samples, respectively, our overall statistic, FIS^1, is

$$-2\big[\log P_{E,Dutch}^1 + \log P_{E,German}^1 + \log P_{F,Dutch}^1$$
$$+ \log P_{F,German}^1 + \log P_{EF,Dutch}^{3|a} + \log P_{EF,German}^{3|a}\big],$$

which follows the chi-square distribution with 12 degrees of freedom under H_0. We also confirmed the presence of population admixture by estimating the amount of variance inflation for T_E^1. Our qq plots in Figure 5 show evidence for conservativeness of the statistics, and genomic control was applied to T_E^1. $T_{EF}^{3|a}$ has been shown to be robust against the presence of population stratification [Zhao et al., 2006].

The results of our analysis are shown in Table 3. In our analysis, the number of statistics being tested was around 3.56×10^{10}, and our genome-wide significance level was 1×10^{-12}. The gene × gene joint action analysis by logistic regression at the genome-wide scale is computationally very intensive, and therefore, it was applied to only the top 15 pairs of SNPs
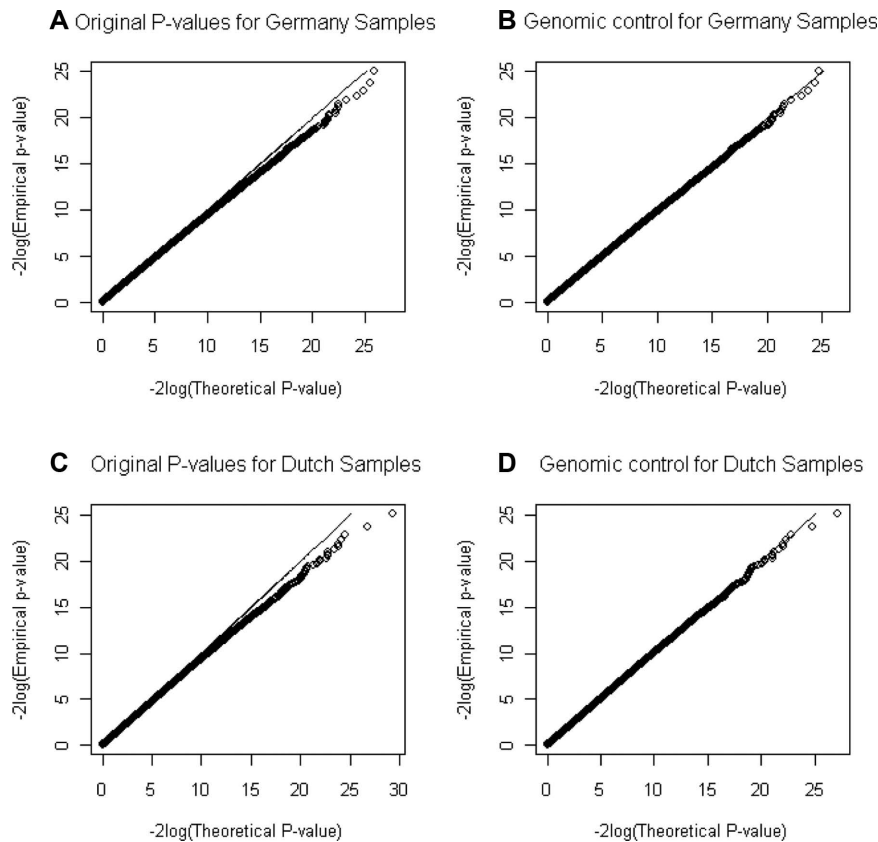
Figure 5. qq plot of T_E^1. The original P-values from T_E^1 for Germany samples are used for the qq plot (A) and genomic control is applied for plot (B). The original P-values from T_E^1 for Dutch samples are used for qq plot (C) and genomic control is applied for plot (D).

Table 3. Application to SCZ

SNP1	chromosome	position (kb)	Gene	SNP2	chromosome	position (kb)	Gene	$p_{EF}^{3|a}$	FIS^1	$LRT_1{}^1$	$LRT_2{}^1$	$LRT_1{}^2$	$LRT_2{}^2$
rs5765930	22	43,494	*PPR*	rs10812774	9	28,284	*LINGO2*	1.70·10⁻⁶	1.25·10⁻¹⁰	4.32·10⁻⁵	1.69·10⁻³	7.60·10⁻²	3.61·10⁻⁴
rs1776161	1	240,136	*EXO1*	rs4643	1	63,898	*PGM1*	6.12·10⁻⁶	2.33·10⁻¹⁰	1.08·10⁻⁴	1.77·10⁻⁵	6.24·10⁻⁵	6.86·10⁻⁴
rs2487504	10	44,575	*EIEF2AP4*	rs1381462	3	160,002	*MFSD1*	7.41·10⁻⁷	2.46·10⁻¹⁰	1.15·10⁻²	5.83·10⁻¹	9.89·10⁻¹	1.96·10⁻¹
rs2160906	22	35,823	*TMPRSS6*	rs9951150	18	50,972	*TCF4*	2.76·10⁻⁵	2.77·10⁻¹⁰	5.84·10⁻⁶	9.05·10⁻²	4.92·10⁻¹	5.39·10⁻⁴
rs1476644	7	3,894	*SDK1*	rs3755930	4	1,227	*CTBP1*	1.44·10⁻⁸	2.85·10⁻¹⁰	3.06·10⁻²	3.07·10⁻¹	7.79·10⁻¹	1.95·10⁻¹
rs228904	22	35,825	*TMPRSS6*	rs9951150	18	50,972	*TCF4*	2.76·10⁻⁵	3.03·10⁻¹⁰	4.74·10⁻⁶	1.00·10⁻¹	4.94·10⁻¹	4.33·10⁻⁴
rs1614778	11	31,911	*PAX6*	rs932012	4	140,916	*MAML3*	5.01·10⁻⁴	3.43·10⁻¹⁰	6.61·10⁻³	1.38·10⁻²	1.84·10⁻¹	1.07·10⁻¹
rs9354442	6	67,323	*Intergenic*	rs450092	3	21,433	*VENTXP7*	2.26·10⁻⁷	3.91·10⁻¹⁰	7.56·10⁻⁸	5.49·10⁻⁵	2.61·10⁻³	1.68·10⁻⁵
rs2160906	22	35,823	*TMPRSS6*	rs2051293	18	50,969	*TCF4*	2.50·10⁻⁶	4.13·10⁻¹⁰	1.98·10⁻⁴	1.99·10⁻¹	8.02·10⁻¹	1.13·10⁻²
rs4717104	7	72,784	*STX1A*	rs9870965	3	62,048	*PTPRG*	1.94·10⁻⁴	4.26·10⁻¹⁰	8.83·10⁻¹¹	7.93·10⁻¹⁰	1.68·10⁻²	1.14·10⁻⁸
rs228904	22	35,825	*TMPRSS6*	rs2051293	18	50,969	*TCF4*	2.50·10⁻⁶	4.52·10⁻¹⁰	1.63·10⁻⁴	2.31·10⁻¹	8.34·10⁻¹	9.78·10⁻³
rs10099863	8	98,944	*RPS23P1*	rs6790564	3	32,117	*CMTM8*	3.46·10⁻⁷	4.78·10⁻¹⁰	3.87·10⁻⁴	5.95·10⁻⁵	3.34·10⁻³	1.85·10⁻²
rs11779840	8	103,606	*POU5F1*	rs10110783	8	10,3147	*NCALD*	7.13·10⁻¹²	4.88·10⁻¹⁰	9.56·10⁻⁶	1.62·10⁻⁶	1.64·10⁻⁴	9.86·10⁻⁴
rs4717104	7	72,784	*STX1A*	rs2526421	3	62,046	*PTPRO*	2.01·10⁻⁴	5.28·10⁻¹⁰	1.17·10⁻¹⁰	7.94·10⁻⁴	1.66·10⁻²	1.45·10⁻⁹
rs9513475	13	98,189	*SLC15A1*	rs3751379	13	38,486	*PROSER1*	1.10·10⁻⁷	5.67·10⁻¹⁰	6.78·10⁻²	8.59·10⁻¹	3.66·10⁻¹	1.65·10⁻¹

The proposed method was applied to SCZ, and the top 15 pairs of SNPs are shown. FIS^1 combines P_E^1, P_F^1, $P_E^{2|a}$, $P_F^{2|a}$, and $P_{EF}^{3|a}$ with Fisher's method, and FIS^2 combines P_E^1, P_F^1, P_E^2, P_F^2, and P_{EF}^3. $LRT_1{}^1$ ($LRT_2{}^1$) indicates the likelihood ratio test that compares the linear interaction model in logistic regression with the intercept model (the linear model). For $LRT_1{}^2$ and $LRT_2{}^2$, dummy variables were utilized, and both indicate the likelihood ratio tests that compare the linear interaction model in logistic regression with the intercept model and the linear model, respectively. The position for each SNP is provided by NCBI reference build 36.3.

selected by FIS^1. These selected pairs of SNPs were analyzed by $T_{EF}^{3|a}$, FIS^1, $LRT_1{}^1$, $LRT_2{}^1$, $LRT_1{}^2$, and $LRT_2{}^2$. None of them was significant at the genome-wide level (Table 3). The most significant results for FIS^1 were from the pair rs5765930 and rs10812774 [Liu et al., 2012]. rs5765930 is located in the region of the genes PRR5 and LOC553158, and it is in chromosome 22q13.31, a region in which evidence for linkage to SCZ has been reported [Liu et al., 2012]. rs10812774 is located in the vicinity of LIGO2 in chromosome 9, and rs10812774 has been reported to be associated with other neurologic diseases such as essential tremor and Parkinson's disease [Wu et al., 2011]. Among SNPs in the most significant 15 SNP pairs, EIEF2AP4, PAX6, STX1A, POU5F1P2, SLC15A1, TCF4, and CTBP1 have been reported to be related with SCZ [Chiang et al., 2011; Chu et al., 2009; Thwaites and Anderson, 2007; Tsuang et al., 2005; Zhu et al., 2013]. PTPRG, NCALD, CMTM8, and PGM have been reported to be related with neurologic diseases or psychopathy [Baldi et al., 1993; Csutora et al., 2006; LeBlanc et al., 2012; Weber et al., 2011]. In particular, $P_{EF}^{3|a}$ is 7.13×10^{-12} for rs11779840 and rs10110783 and the statistical gene × gene interaction is suspected for this pair. The $LRT_1{}^1$ for rs4717104 and rs9870965 is 8.83×10^{-11} but this significance can be invalid because of the population admixture and some modification is necessary. These pairs of SNPs will be further investigated in our follow-up studies.

Discussion

In this study, we proposed a new statistical method to detect gene × gene joint action based on three types of components. There is no uniformly most powerful method for the analysis of gene × gene joint action; however, our results show that the proposed method performs better than logistic regression under the considered gene × gene joint action model. When the relative risk for each genotype is randomly generated, it was shown that the proposed method performs well, which indicates that the proposed method may be an efficient choice for gene × gene joint action when no prior knowledge is available about the interaction model. In particular, we assumed no haplotype effects in our simulations. If there are haplotype effects, the performance of the proposed method may improve because our LD test focuses on the co-occurrence of two alleles in the same haplotype.

The proposed method is computationally fast. For instance, our analysis of the SCZ study was completed in a week for a cluster with six nodes, which involved genome-wide examination of all pairwise joint actions. For 1,000 cases and 1,000 controls with 1,000 pairs of SNPs, the analysis of the proposed method in the statistical software, R, with Xeon E5649 @2.66 GHz, was completed in 1.7 sec, whereas a logistic regression with R required more than 34.9 sec. For 5,000 and 10,000 pairs of SNPs, the proposed methods were calculated in 8.5 sec and 16.2 sec, respectively, and the logistic regressions were in 176.5 sec and 333.5 sec. Because the logistic regression in R used the embedded C code, the difference may be larger under the same condition. This shows that our approach can be used for genome-wide analysis of gene × gene joint action. To date, many methods for detecting joint action at a genome-wide level have been investigated. However, most of them are limited by their computational burden, which often forces their analysis to be restricted to "select SNPs." Although the impact of a screening step for gene × gene interaction has not been carefully investigated, we expect that two-stage analysis approaches, for instance, for MDR, can lead to a substantial drop in efficiency. The computational simplicity of our approach is a key advantage in the GWAS context, and the software implemented in C for the proposed method can be freely downloaded at http://cau.ac.kr/∼swon.

Nevertheless, there exist some limitations in our simulation studies. First, we compared the power of the proposed method only with the logistic regression and the power comparison results of the proposed method may not readily generalize. Recently many statistical methods such as MDR [Ritchie et al., 2001] and Bayesian epistasis association mapping [Zhang and Liu, 2007] have been proposed. Even though

they were not included in the simulation studies, the power comparison of these methods with logistic regression repeatedly showed that logistic regression was not efficient [Chen et al., 2011; Wu et al., 2010; Zhang and Liu, 2007]. However, even though the logistic regression for a statistical gene × gene joint action may not be the most efficient choice of joint action analysis, it should be noted that the statistical efficiency for the logistic regression depends on the parameterization of gene × gene interaction model. It was empirically shown that the logistic regression performed the most efficient when genotypes are coded to reveal the true disease model [Ueki and Cordell, 2012], and the logistic regression with nine dummy variables has been shown to be reasonably good [Chen et al., 2011]. Our results also show that the logistic regression with nine dummy variables for each genotype performed better than the logistic regression with a linear gene × gene interaction model. Therefore, even though our comparison is limited to the logistic regression, the various choices of parameterization that we selected make our results reasonable and an exhaustive power comparison with various approaches will be conducted as our further work. Second, the proposed method has been applied only to the biallelic two-way gene × gene joint action model even though depending on the disease model, for instance, a three-way (or more) joint action model can be informative. The (composite) LD can be defined for multiple loci with multiple alleles [Yu and Wang, 2011; Zaykin et al., 2006, 2008], and our method can be extended to such a situation. For instance, for three-way joint action model, we can have three tests for the difference of minor allele frequencies, three tests for the presence of HWE, three tests for two-locus LE in cases, and one test for the presence of three-locus LE. However, the efficient way to construct the statistic for joint action analysis with the three types of components is not clear, and it will be investigated as our further work. Third, if LE and HWE are not preserved in controls, the independence of individual statistics is not guaranteed and the permutation was alternatively suggested. However for the genome-wide interaction analysis, the significance level is usually very small and thus permutation-based inference is limited at the genome-wide scale. These limitations will be investigated as part of our ongoing research.

Although GWASs have successfully identified many genetic variants for diseases, missing heritability reveals that the previous finding does not fully explain the genetic background for many diseases, and it may be partially attributable to the computational burden in the analysis of the gene × gene joint action. Our method improves the computational complexity for the analysis of gene × gene joint action at the genome-wide scale while preserving the statistically optimal efficiency, and therefore it may bridge the gap between statistical and computational issues for genome-wide studies.

References

Athanasiu L, Mattingsdal M, Kahler AK, Brown A, Gustafsson O, Agartz I, Giegling I, Muglia P, Cichon S, Rietschel M and others. 2010. Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J Psychiatr Res* 44(12):748–753.

Baldi E, Burra P, Plebani M, Salvagnini M. 1993. Serum malondialdehyde and mitochondrial aspartate aminotransferase activity as markers of chronic alcohol intake and alcoholic liver disease. *Ital J Gastroenterol* 25(8):429–432.

Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM and others. 2008. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 40(8):955–962.

Birnbaum A. 1954. Combining independent tests of significance. *J Am Stat Assoc* 49(267):559–574.

Cardno AG, Gottesman, II. 2000. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* 97(1):12–17.

Chen L, Yu G, Langefeld CD, Miller DJ, Guy RT, Raghuram J, Yuan X, Herrington DM, Wang Y. 2011. Comparative analysis of methods for detecting interacting loci. *BMC Genomics* 12:344.

Chiang CH, Su Y, Wen Z, Yoritomo N, Ross CA, Margolis RL, Song H, Ming GL. 2011. Integration-free induced pluripotent stem cells derived from schizophrenia patients with a DISC1 mutation. *Mol Psychiatry* 16(4):358–360.

Chikkagoudar S, Wang K, Li M. 2011. GENIE: a software package for gene-gene interaction analysis in genetic association studies using multiple GPU or CPU cores. *BMC Res Notes* 4:158.

Chu TT, Liu Y, Kemether E. 2009. Thalamic transcriptome screening in three psychiatric states. *J Hum Genet* 54(11):665–675.

Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, Wells S, Bruning JC, Nolan PM, Ashcroft FM and others. 2010. Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet* 42(12):1086–1092.

Cook NR, Zee RY, Ridker PM. 2004. Tree and spline based association analysis of gene-gene interaction models for ischemic stroke. *Stat Med* 23(9):1439–1453.

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.

Cordell HJ. 2009. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 10(6):392–404.

Csutora P, Karsai A, Nagy T, Vas B, L Kovács G, Rideg O, Bogner P, Miseta A. 2006. Lithium induces phosphoglucomutase activity in various tissues of rats and in bipolar patients. *Int J Neuropsychopharmacol* 9(5):613–619.

Culverhouse R, Suarez BK, Lin J, Reich T. 2002. A perspective on epistasis: limits of models displaying no main effect. *Am J Hum Genet* 70(2):461–471.

Djurovic S, Gustafsson O, Mattingsdal M, Athanasiu L, Bjella T, Tesli M, Agartz I, Lorentzen S, Melle I, Morken G and others. 2010. A genome-wide association study of bipolar disorder in Norwegian individuals, followed by replication in Icelandic sample. *J Affect Disord* 126(1–2):312–316.

Dong C, Chu X, Wang Y, Jin L, Shi T, Huang W, Li Y. 2008. Exploration of gene-gene interaction effects using entropy-based methods. *Eur J Hum Genet* 16(2):229–235.

First MB, Spitzer RL, Gibbon M, Williams JBW. 1994. *Structured Clinical Interview for DSM-IV Axis I Disorders*. New York: New York State Psychiatric Institute.

Gyenesei A, Moody J, Laiho A, Semple CA, Haley CS, Wei WH. 2012. BiForce toolbox: powerful high-throughput computational analysis of gene-gene interactions in genome-wide association studies. *Nucleic Acids Res* 40(1):W628–W632.

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A and others. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41(10):1088–1093.

Jiang R, Tang W, Wu X, Fu W. 2009. A random forest approach to the detection of epistatic interactions in case-control studies. *BMC Bioinformatics* 10(Suppl 1):S65.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST and others. 2005. Complement factor H polymorphism in age-related macular degeneration. *Science* 308(5720):385–389.

Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, El Mokhtari NE, Schreiber S. 2006. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 9(1):55–61.

LeBlanc M, Kulle B, Sundet K, Agartz I, Melle I, Djurovic S, Frigessi A, Andreassen OA. 2012. Genome-wide study identifies PTPRO and WDR72 and FOXQ1-SUMO1P1 interaction associated with neurocognitive function. *J Psychiatr Res* 46(2):271–278.

Li W, Reich J. 2000. A complete enumeration and classification of two-locus disease models. *Hum Hered* 50(6):334–349.

Little RC, Folks L. 1971. Asymptotic optimality of Fisher's method of combining independent tests. *J Am Stat Assoc* 66(336):986–998.

Liu J, Ulloa A, Perrone-Bizzozero N, Yeo R, Chen J, Calhoun VD. 2012. A pilot study on collective effects of 22q13.31 deletions on gray matter concentration in schizophrenia. *PLoS One* 7(12):e52865.

Marchini J, Donnelly P, Cardon LR. 2005. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37(4):413–417.

McGuffin P, Farmer A, Harvey I. 1991. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry* 48(8):764–770.

Nielsen DM, Ehm MG, Weir BS. 1998. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet* 63(5):1531–1540.

Nielsen DM, Ehm MG, Zaykin DV, Weir BS. 2004. Effect of two- and three-locus linkage disequilibrium on the power to detect marker/phenotype associations. *Genetics* 168(2):1029–1040.

Park MY, Hastie T. 2008. Penalized logistic regression for detecting gene interactions. *Biostatistics* 9(1):30–50.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, 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.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. 2001. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 69(1):138–147.

Sasieni PD. 1997. From genotypes to genes: doubling the sample size. *Biometrics* 53(4):1253–1261.

Schmermund A, Mohlenkamp S, Stang A, Gronemeyer D, Seibel R, Hirche H, Mann K, Siffert W, Lauterbach K, Siegrist J and others. 2002. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am Heart J* 144(2):212–218.

Spitzer RL, Endicott J, Robins E. 1978. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry* 35(6):773–782.

Sullivan PF. 2005. The genetics of schizophrenia. *PLoS Med* 2(7):e212.

Thwaites DT, Anderson CM. 2007. H+-coupled nutrient, micronutrient and drug transporters in the mammalian small intestine. *Exp Physiol* 92(4):603–619.

Tsuang MT, Nossova N, Yager T, Tsuang MM, Guo SC, Shyu KG, Glatt SJ, Liew CC. 2005. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 133B(1):1–5.

Ueki M, Cordell HJ. 2012. Improved statistics for genome-wide interaction analysis. *PLoS Genet* 8(4):e1002625.

Wang J, Shete S. 2008. A test for genetic association that incorporates information about deviation from Hardy-Weinberg proportions in cases. *Am J Hum Genet* 83(1):53–63.

Wang T, Zhu X, Elston RC. 2007. Improving power in contrasting linkage-disequilibrium patterns between cases and controls. *Am J Hum Genet* 80(5):911–920.

Wang X, Elston RC, Zhu X. 2010. The meaning of interaction. *Hum Hered* 70(4):269–277.

Wang X, Elston RC, Zhu X. 2011. Statistical interaction in human genetics: how should we model it if we are looking for biological interaction? *Nat Rev Genet* 12(1):74.

Weber H, Kittel-Schneider S, Gessner A, Domschke K, Neuner M, Jacob CP, Buttenschon HN, Boreatti-Hummer A, Volkert J, Herterich S and others. 2011. Cross-disorder analysis of bipolar risk genes: further evidence of DGKH as a risk gene for bipolar disorder, but also unipolar depression and adult ADHD. *Neuropsychopharmacology* 36(10):2076–2085.

Weir BS. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data.* Sunderland, MA: Sinauer Associates.

Wichmann HE, Gieger C, Illig T. 2005. KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 67(Suppl 1):S26–S30.

Won S, Elston RC. 2008. The power of independent types of genetic information to detect association in a case-control study design. *Genet Epidemiol* 32(8):731–756.

Won S, Kim S, Elston RC. 2009a. Phase uncertainty in case-control association studies. *Genet Epidemiol* 33(6):463–478.

Won S, Morris N, Lu Q, Elston RC. 2009b. Choosing an optimal method to combine P-values. *Stat Med* 28(11):1537–1553.

Wu X, Dong H, Luo L, Zhu Y, Peng G, Reveille JD, Xiong M. 2010. A novel statistic for genome-wide interaction analysis. *PLoS Genet* 6(9):e1001131.

Wu YW, Prakash KM, Rong TY, Li HH, Xiao Q, Tan LC, Au WL, Ding JQ, Chen SD, Tan EK. 2011. Lingo2 variants associated with essential tremor and Parkinson's disease. *Hum Genet* 129(6):611–615.

Yu Z, Wang S. 2011. Contrasting linkage disequilibrium as a multilocus family-based association test. *Genet Epidemiol* 35(6):487–498.

Yung LS, Yang C, Wan X, Yu W. 2011. GBOOST: a GPU-based tool for detecting gene-gene interactions in genome-wide case control studies. *Bioinformatics* 27(9):1309–1310.

Zaykin DV. 2004. Bounds and normalization of the composite linkage disequilibrium coefficient. *Genet Epidemiol* 27(3):252–257.

Zaykin DV, Meng Z, Ehm MG. 2006. Contrasting linkage-disequilibrium patterns between cases and controls as a novel association-mapping method. *Am J Hum Genet* 78(5):737–746.

Zaykin DV, Pudovkin A, Weir BS. 2008. Correlation-based inference for linkage disequilibrium with multiple alleles. *Genetics* 180(1):533–545.

Zerba KE, Ferrell RE, Sing CF. 2000. Complex adaptive systems and human health: the influence of common genotypes of the apolipoprotein E (ApoE) gene polymorphism and age on the relational order within a field of lipid metabolism traits. *Hum Genet* 107(5):466–475.

Zhang Y, Liu JS. 2007. Bayesian inference of epistatic interactions in case-control studies. *Nat Genet* 39(9):1167–1173.

Zhao J, Jin L, Xiong M. 2006. Test for interaction between two unlinked loci. *Am J Hum Genet* 79(5):831–845.

Zhu X, Gu H, Liu Z, Xu Z, Chen X, Sun X, Zhai J, Zhang Q, Chen M, Wang K and others 2013. Associations between TCF4 gene polymorphism and cognitive functions in schizophrenia patients and healthy controls. *Neuropsychopharmacology* 38(4):683–689.

Appendix

Data Description: Schizophrenia

Samples were obtained from either German or Dutch patients who provided written informed consent. German SCZ patients ($n_a = 487$), who were all of German descent, were recruited from consecutive hospital admissions. Lifetime best estimate diagnoses were assigned according to DSM-IV criteria based on structured interviews [First et al., 1994; McGuffin et al., 1991; Spitzer et al., 1978], medical records, and family history. Best estimate diagnoses were assigned by at least two experienced psychiatrists/psychologists. The controls were drawn from three population-based epidemiological studies: (i) PopGen [Krawczak et al., 2006] ($n_u = 490$), (ii) KORA [Wichmann et al., 2005] ($n_u = 488$), and (iii) the Heinz Nixdorf Recall (HNR, $n_u = 383$) study [Schmermund et al., 2002]. Inpatients and outpatients ($n_a = 804$) in the Netherlands were recruited from various psychiatric hospitals and institutions throughout Utrecht and Rotterdam. All the patients were diagnosed with the DSM-IV definition of SCZ. The controls from Utrecht ($n_u = 704$) were volunteers with no history of psychiatric disorders, and the Rotterdam control individuals ($n_u = 2,302$) were drawn from a large population-based project on the genetics of complex traits and diseases.

The GWAS dataset was provided by analysis at seven genotyping facilities by using HumanHap550v3 BeadArrays

with the Infinium II assay (Illumina, San Diego, CA, USA). We developed a protocol of filters for the stringent quality control of whole-genome and sub-whole-genome datasets that accounted for issues such as call rates, heterozygosity, cross-contamination, population stratification, relatedness, nonrandom missingness, HWE, and minor allele frequency. The quality-control protocol was applied to a total of 1,291 patients and 4,367 controls for the GWAS step. Consequently, 375 individuals, including 122 patients and 253 controls as well as 86,037 SNPs, were excluded prior to the association analysis. Therefore, the final GWAS dataset for gene × gene interaction analysis comprised 464,030 autosomal SNPs (including PAR1+2), genotyped in 1,169 SCZ patients and 3,714 controls.