



# A Powerful Gene-Based Test Accommodating Common and Low-Frequency Variants to Detect Both Main Effects and Gene-Gene Interaction Effects in Case-Control Studies

Ren-Hua Chung\* and Chen-Yu Kang

Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan

## OPEN ACCESS

### Edited by:

Cellia M. T. Greenwood,  
Lady Davis Institute (LDI), Canada

### Reviewed by:

Yu-Ru Su,  
Fred Hutchinson Cancer Research  
Center, United States  
Pingzhao Hu,  
University of Manitoba, Canada

### \*Correspondence:

Ren-Hua Chung  
rchung@nhri.org.tw

### Specialty section:

This article was submitted to  
Statistical Genetics and Methodology,  
a section of the journal  
Frontiers in Genetics

**Received:** 30 August 2017

**Accepted:** 18 December 2017

**Published:** 08 January 2018

### Citation:

Chung R-H and Kang C-Y (2018) A  
Powerful Gene-Based Test  
Accommodating Common and  
Low-Frequency Variants to Detect  
Both Main Effects and Gene-Gene  
Interaction Effects in Case-Control  
Studies. *Front. Genet.* 8:228.  
doi: 10.3389/fgene.2017.00228

Next-generation sequencing (NGS) has been widely used in genetic association studies to identify both common and rare variants associated with complex diseases. Various statistical association tests have been developed to analyze NGS data; however, most focus on identifying the marginal effects of a set of genetic variants on the disease. Only a few association tests for NGS data analysis have considered the interaction effects between genes. We developed three powerful gene-based gene-gene interaction tests for testing both the main effects and the interaction effects of common, low-frequency, and common with low-frequency variant pairs between two genes (the IGOF tests) in case-control studies using NGS data. We performed a comprehensive simulation study to verify that the proposed tests had appropriate type I error rates and significantly higher power than did other interaction tests for analyzing NGS data. The tests were applied to a whole-exome sequencing dataset for autism spectrum disorder (ASD) and the significant results were evaluated in another independent ASD cohort. The IGOF tests were implemented in C++ and are available at <http://igof.sourceforge.net>.

**Keywords:** gene-gene interaction, next-generation sequencing, case-control study, rare variant association, simulations, autism spectrum disorders, association test

## INTRODUCTION

Next-generation sequencing (NGS) has become a popular technology used in genetic studies to identify rare variants (i.e., genetic variants with minor allele frequencies <5%) as well as common variants associated with complex diseases. Many statistical association tests have been developed for analyzing NGS data (Li and Leal, 2008; Madsen and Browning, 2009; Ionita-Laza et al., 2011; Wu et al., 2011). Since testing individual rare variants can result in low statistical power, these methods mainly test a set of rare variants in an exon or gene by aggregating association signals from the set of variants. For example, the weighted-sum test aggregates association signals from individual variants in a genomic region, while more weights are assigned to variants with lower minor allele frequencies (MAFs) in the test statistic (Madsen and Browning, 2009). Furthermore, some association tests that consider both common and rare variants are also available for analyzing

NGS data (Ionita-Laza et al., 2013; Chung et al., 2014). However, most of the aforementioned tests are restricted to identifying marginal effects of genetic variants on the disease. Gene-gene interactions, defined as the departure of the additive effects of two or more genetic variants in a statistical model (Cordell, 2009), are also expected to be responsible for a portion of missing heritability for complex traits (Manolio et al., 2009). Therefore, developing a powerful association test that considers both main effects and gene-gene interactions for analyzing NGS data has become important in genetic association studies.

Only a few tests that specifically model gene-gene interactions in genetic studies using NGS data are available. The Sequence Kernel Association Test (SKAT), a variance component test via different kernels, can be used to test for gene-gene interactions using a 2wayIX kernel, which considers two-way interactions between variants in a region (Wu et al., 2010). Recently, a gene-gene interaction test was developed based on a functional logistic regression model (FGLM) for analyzing NGS data (Zhao et al., 2016). In FGLM, the interaction signals for all possible pairs of variants between two genes were transformed to a few functional principal components in the model. Then, the test statistic is constructed based on the principal components. On the other hand, the kernel-based adaptive cluster (KBAC) method (Liu and Leal, 2010), which adaptively assigned more weights to multi-variant genotypes enriched in cases using a kernel function, was developed to identify main effects and gene-gene interaction effects for rare variants. Compared to the weighted-sum test that sums the association signals from individual variants, gene-gene interactions were implicitly modeled in KBAC, as it considered the joint effects of multi-variant genotypes. Therefore, simulation studies suggest that KBAC can be more powerful than methods focusing on main effects such as the weighted-sum test in the presence of gene-gene interactions (Liu and Leal, 2010). However, the power of KBAC, SKAT via the 2wayIX kernel, and FGLM has not been compared.

Gene-based interaction tests considering all possible pairwise interactions of variants between two genes, such as SKAT or FGLM, can result in a large number of variant pairs. For example, there were approximately 1,800,000 exonic variants reported in the Exome Sequencing Project (<http://evs.gs.washington.edu/EVS>) in approximately 18,000 genes, resulting in an average of approximately 100 exonic variants per gene. Therefore, there would be 10,000 pairwise interactions when testing two genes with 100 variants, but there may be only a small portion of variant pairs with interaction effects on the disease. When considering combining signals from multiple test statistics, the threshold algorithm, which selects a subset of promising signals for further testing, can be a powerful approach to aggregating signals. For example, the truncated product method, which combines  $p$ -values less than a certain threshold, is more powerful than are methods combining all  $p$ -values, such as Fisher's test (Zaykin et al., 2002). A gene-based association test that combines single-variant statistics with  $p$ -values less than a threshold has also been shown to be more powerful than tests combining all association signals within a gene (Wang et al., 2015). Moreover, incorporating the truncated product method in a gene-based interaction test results in a significant

increase in power compared with other tests for common variants (Ma et al., 2013). However, SKAT and FGLM consider signals from all variant pairs, and their power can be compromised if there is only a small portion of variant pairs with interaction effects. Incorporating the threshold algorithm in a gene-based interaction test for NGS data thus becomes attractive as it maintains the power of the test.

We developed gene-based tests accommodating variants with all allele frequency spectrums to identify main effects and gene-gene interaction effects for case-control studies based on threshold algorithms. The test statistic for a variant pair was based on Pearson's goodness-of-fit (GOF) statistic, which compares the difference between the observed and expected genotype distributions at two variants in cases and controls. The GOF statistic follows a chi-square distribution when all the expected values in the contingency table are greater than 5 (Yates, 1934). Therefore, variant pairs between two genes were classified as common or low-frequency (LF) variant pairs in the test based on their expected genotype counts in the sample. That is, a variant pair was a common variant pair if the expected genotype counts for the two variants in either cases or controls under the null hypothesis of no association were all  $\geq 5$ ; otherwise, the variant pair was an LF variant pair. Two different threshold algorithms were applied to the common and LF variant pairs, respectively. The statistic for common variant pairs was constructed based on the GOF statistics for common variant pairs with  $p$ -values calculated based on the chi-square distribution less than a pre-determined threshold, and the test is referred to as the  $IGOF_{\text{common}}$  test. For LF variant pairs, assuming that minor alleles have risk effects, statistics for variant pairs with observed minor allele counts in cases that were greater than those in controls were combined as the LF variant pair statistic. The test is referred to as the  $IGOF_{\text{LF}}$  test. A  $p$ -value for a combined test (referred to as the  $IGOF_{\text{combined}}$  test) was then calculated based on the  $IGOF_{\text{common}}$  and  $IGOF_{\text{LF}}$  test  $p$ -values. We performed a comprehensive simulation study to evaluate the type I error rates for the proposed tests and to compare the power of the  $IGOF_{\text{combined}}$  test with that of SKAT, FGLM, and KBAC. Finally, the tests were applied to the whole-exome sequencing dataset of an autism spectrum disorder (ASD) association study, and the ten most significant gene pairs were evaluated in another replication cohort.

## MATERIALS AND METHODS

### The Pearson's GOF Statistic

We first review Pearson's GOF statistic, which is the basis of the  $IGOF$  test statistics. Assume we have  $N$  cases and  $M$  controls. For two variants where one variant has alleles  $A$  and  $a$  and the other has alleles  $B$  and  $b$ , there are 9 categories of genotypes (i.e.,  $AABB$ ,  $AABb$ ,  $AAbb$ ,  $AaBB$ ,  $AaBb$ ,  $Aabb$ ,  $aaBB$ ,  $aaBb$ , and  $aabb$ ). Assume that  $n_i$  and  $m_i$  are the numbers of cases and controls in category  $i$ . Pearson's GOF statistic is calculated as:

$$C^2 = \sum_{i=1}^9 \left[ (n_i - E(n_i))^2 / E(n_i) + (m_i - E(m_i))^2 / E(m_i) \right],$$

where

$$E(n_i) = N \times f_i, E(m_i) = M \times f_i, \text{ and } f_i = (n_i + m_i)/(N + M). \tag{1}$$

With the assumption that affection status is independent of the genotype categories,  $C^2$  asymptotically follows a chi-square distribution with 8 degrees of freedom. The null hypothesis of the test is that the two variants are not associated with the disease, and a variant pair with main effects and/or interaction effects could result in the rejection of the null. Note that the calculation of  $f_i$  does not assume linkage equilibrium between the two variants, and therefore the statistic could be used to test for interaction of variants in linkage disequilibrium (LD).

### The IGOF Statistic for Common Variant Pairs

A common variant pair is defined as a variant pair with the minimum of  $E(n_i)$  and  $E(m_i)$  for  $i = 1, 2, \dots, 9$  greater than or equal to 5. A modification of the truncated  $p$ -value product method is applied to the IGOF<sub>common</sub> test. For a pair of genes, assume there are  $n$  common variant pairs. We calculate the  $C^2$  statistics and their  $p$ -values for the  $n$  pairs. Let  $\Omega$  be a set of common variant pairs with  $p$ -values less than  $\tau$ . The  $p$ -value threshold  $\tau$  was set as 0.05 in both Zaykin et al. (2002) and Ma et al. (2013). Therefore,  $\tau$  was also set to be 0.05 as a default value in the IGOF<sub>common</sub> test. We later used simulations to evaluate the power of the IGOF<sub>common</sub> test with different values of  $\tau$ . The IGOF<sub>common</sub> test statistic is defined as:

$$I_{common} = \sum_{j \in \Omega} \Phi^{-1}(p_j) \tag{2}$$

where  $p_j$  is the  $p$ -value for the GOF statistic  $C_j^2$  for variant pair  $j$  and  $\Phi$  is the cumulative distribution function (cdf) of a standard normal random variable. Note that  $I_{common}$  is not the product of the truncated  $p$ -values, as calculated by Zaykin et al. (2002) and Ma et al. (2013). Instead, it is the sum of the inverse normal statistics of the  $p$ -values. As explained in the section on "Performance Improvement", the transformation of  $p$ -values to the normal statistics is used to calculate an approximated  $p$ -value for  $I_{common}$ .

### The IGOF Statistic for LF Variant Pairs

Variant pairs that are not common variant pairs are referred to as LF variant pairs. That is, an LF variant pair has at least one of  $E(n_i)$  or  $E(m_i)$  for  $i = 1, 2, \dots, 9$  less than 5. A threshold algorithm is also applied to the calculation of the statistic for LF variant pairs. However, the selection of LF variant pairs for testing is not based on the  $p$ -value of the  $C^2$  statistic because the chi-square property may not hold for  $C^2$  calculated based on LF variant pairs. Instead, the selection criteria are based on including variant pairs with possible risk interaction effects. To be more specific, assume there are  $N$  cases. For an LF variant pair where one variant has major and minor alleles of  $A$  and  $a$ , respectively, and the other variant has major and minor alleles of  $B$  and  $b$ ,

respectively, a  $3 \times 3$  table as shown in **Table 1** for genotype counts in the  $N$  cases can be constructed, where  $\sum_{i=1}^9 n_i = N$ . There are  $4N$  alleles in the variant pair, and a  $2 \times 2$  table in **Table 2** is constructed based on **Table 1** for allele counts in the variant pair. For example,  $N_{AB}$  is the number of alleles  $A$  and  $B$  in individuals carrying both alleles  $A$  and  $B$ . The construction procedures for **Tables 1, 2** are the same as the procedure for constructing the fast epistasis statistic in PLINK (Purcell et al., 2007). Similarly, assume that there are  $M$  controls. Then,  $M_{AB}, M_{Ab}, M_{aB},$  and  $M_{ab}$  are calculated for the allele counts of the variant pair in controls. A variant pair is included in the IGOF test if  $N_{ab}$ , the minor allele count of the two variants (in **Table 2**) calculated in cases, is larger than  $M_{ab}$  in controls. This is based on the assumption that if the minor alleles of the two variants have risk interaction effects, more minor alleles would be observed in cases than in controls. Assume  $\Psi$  is a set of  $l$  LF variant pairs selected based on this threshold between the two genes. The IGOF<sub>LF</sub> test statistic is calculated as

$$I_{LF} = \sum_{j \in \Psi} C_j^2 \tag{3}$$

where  $C_j^2$  is the  $C^2$  statistic for variant pair  $j$  in  $\Psi$ .

### The IGOF<sub>common</sub>, IGOF<sub>LF</sub>, and IGOF<sub>combined</sub> Tests

$I_{common}$  is the sum of the  $k$  largest normal random variables from the  $n$  correlated normal random variables, where  $k$  is the size of  $\Omega$ .  $I_{LF}$  is the sum of  $l$  correlated chi-square random variables, where the random variables were pre-selected based on the affection status. The distributions for  $I_{common}$  and  $I_{LF}$  are both unknown. Therefore, permutations in which the case and control status are randomly shuffled are used to calculate the  $p$ -values for  $I_{common}$  and  $I_{LF}$ . The permutations are performed based on the adaptive permutation algorithm (Che et al., 2014). Briefly, given the number of tests, the maximum number of permutations  $b$  and the cutoff value  $r$  can be estimated such that the standard error of the  $p$ -value estimation is less than 20% of the significance level  $\alpha$ . Permutations are terminated when the number of permuted statistics  $>$  the observed statistic is larger

**TABLE 1** | The genotype counts between two variants.

	BB	Bb	bb
AA	$n_1$	$n_2$	$n_3$
Aa	$n_4$	$n_5$	$n_6$
aa	$n_7$	$n_8$	$n_9$

**TABLE 2** | The allele counts between two variants.

	B	b
A	$N_{AB} = 4n_1 + 2n_2 + 2n_4 + n_5$	$N_{Ab} = 4n_3 + 2n_2 + 2n_6 + n_5$
a	$N_{aB} = 4n_7 + 2n_4 + 2n_8 + n_5$	$N_{ab} = 4n_9 + 2n_6 + 2n_8 + n_5$

than  $r$  or when the number of permutations is larger than  $b$ . If the number of permuted statistics  $>$  the observed statistic is larger than  $r$ , the  $p$ -value is calculated as  $r/B$ , where  $B$  is the number of permutations to achieve  $r$ . Otherwise, the  $p$ -value is calculated as  $(R+1)/(b+1)$ , where  $R$  is the number of permuted statistics  $>$  the observed statistic in the  $b$  permutations. Assume that the  $p$ -values for  $I_{common}$  and  $I_{LF}$  are  $p_{common}$  and  $p_{LF}$ , respectively. Note that if  $\Omega$  (or  $\Psi$ ) is an empty set, no permutations will be performed and the  $IGOF_{common}$  (or  $IGOF_{LF}$ )  $p$ -value will be 1. Similar to the extended Simes' test for combining  $p$ -values into a global test  $p$ -value (Li et al., 2011), a  $p$ -value for the IGOF test accommodating common and LF variant pairs (i.e., the  $IGOF_{combined}$  test) is calculated as:

$$P_{combined} = \text{Min}(2\text{Min}(p_{common}, p_{LF}), \text{Max}(p_{common}, p_{LF})) \quad (4)$$

The null hypothesis for each of the IGOF tests is that none of the variant pairs between the two tested genes are associated with the disease. Any variants with main effects or variant pairs with interaction effects could result in the rejection of the null. Hence, the IGOF tests can detect both main effects and interaction effects between two genes.

## Performance Improvement

Although the calculation of  $C^2$  is fast for a pair of variants, calculating  $I_{common}$  and  $I_{LF}$  and the permuted statistics can still be very time consuming. We propose a two-stage approach to improve the performance of the IGOF tests. In the first stage, an approximated  $p$ -value based on a theoretical distribution is rapidly calculated. Adaptive permutations to calculate the  $p$ -value will be performed only if the approximated  $p$ -value is less than an elevated significance level  $m\alpha$ , where  $m$  is a pre-specified constant and  $\alpha$  is the significance level. The  $IGOF_{common}$  or  $IGOF_{LF}$   $p$ -value will be 1 if no permutation is performed.

It has been shown that the sum of the  $k$  largest from  $n$  independent normal random variables is asymptotically normal (Wiens et al., 2006). For  $n$  independent standard normal random variables, the distribution of the sum of the  $k$  largest random variables,  $Y$ , can be simplified as.

$$\sqrt{n} \left( \frac{(Y/n) - \mu_{\beta}(\bar{F}_n)}{\sigma_{\beta}(\bar{F}_n)} \right) \sim N(0, 1) \quad (5)$$

where  $\bar{F}_n$  is a function of a standard normal random variable, and  $\mu_{\beta}$  and  $\sigma_{\beta}$  are the mean and standard deviation functions of  $\bar{F}_n$ , respectively, to estimate the mean and standard deviation of  $Y$ . More details on these functions can be found in Wiens et al. (2006). The  $p$ -value calculated based on the distribution of  $Y$  was referred to as  $p_N$ . Due to correlations between common variant pairs,  $I_{common}$  would not have the same distribution as  $Y$ . Therefore,  $p_N$  for  $I_{common}$  is only an approximation of the true  $p$ -value.

For  $I_{LF}$ , two approximated  $p$ -values are calculated. The first approximated  $p$ -value, referred to as  $p_D$ , is calculated based on the assumption that  $I_{LF}$  is distributed as a chi-square distribution with  $\sum_{i=1}^l d_i$  degrees of freedom, where  $d_i$  is the number of non-zero genotype categories for variant pair  $i$ . Although  $p_D$  can

be rapidly calculated, we found that  $p_D$  was seriously inflated using simulation studies because of the following reasons: (1)  $C_i$  in  $\Psi$  can be correlated; (2) each  $C_i$  in  $\Psi$  is not necessarily distributed as a chi-square with  $d_i$  degrees of freedom; and (3)  $C_i$  is pre-selected based on the assumption of risk effects. Therefore, if  $p_D$  is less than  $m\alpha$ , another method based on moment-matching, (i.e., Satterthwaite-Welch approximation of the distribution of the weighted sum of chi-square random variables; Welch, 1937; Satterthwaite, 1941), was used to calculate the second approximated  $p$ -value, referred to as  $p_{SW}$ . Briefly, as noted in Box (1954), assume that the first and second moments of the random variable  $Q$ , which is the weighted sum of chi-square random variables with different degrees of freedoms, are  $k_1$  and  $k_2$ , respectively. Then  $Q$  is approximately distributed as  $g\chi_h^2$ , where  $g = \frac{k_2}{2k_1}$  and  $h = \frac{2k_1^2}{k_2}$ . As the first and second moments are unknown for  $I_{LF}$ , we perform 1,000 permutations and estimate the values based on the 1,000 permuted statistics of  $I_{LF}$  under the null. Note also that, due to the correlations among LF variant pairs,  $C^2$  for LF variant pairs are not necessarily distributed as chi-square, and  $C_i$  is pre-selected,  $p_{SW}$  calculated based on the distribution of  $Q$  is also an approximated  $p$ -value for  $I_{LF}$ .

## Simulations

We used simulations to evaluate type I error rates for the  $IGOF_{common}$ ,  $IGOF_{LF}$ , and  $IGOF_{combined}$  tests and to compare the power of the  $IGOF_{combined}$  test with SKAT, FGLM, and KBAC. We first used the coalescent-based simulator (COSI) (Schaffner et al., 2005) to generate a population of 10,000 sequences (i.e., haplotypes) in two independent 30 kb regions based on a European ancestry. Each region consisted of approximately 600 variants. SeqSIMLA2 (Chung et al., 2015) was then used to simulate case-control samples based on the 10,000 sequences. For type I error simulations, the penetrance function for SeqSIMLA2 was specified as:

$$\text{logit}(P(\text{affected})) = \alpha_0 \quad (6)$$

where  $\alpha_0$  determined the baseline disease prevalence, which was assumed to be 5%. We considered four sets of sample sizes (i.e., 500 cases and 500 controls; 1,000 cases and 1,000 controls; 1,500 cases and 1,500 controls; 2,000 cases and 2,000 controls) and four sets of gene pairs (i.e., 10, 30, 50, and 100 variants in each gene). The variants were randomly selected from the 600 variants in the two genes, and they consisted of both common and LF variants. When testing the interactions between two genes that were located close together on the same chromosome, there could be LD at variants between the two genes. To evaluate the effects of LD on the type I error rates for the IGOF tests, we tested two regions with strong and weak LD. To generate two regions with strong LD, from one 30 kb region, we selected the first 50 variants as the first region and the subsequent 50 variants as the second region. Similarly, to generate two regions with weak LD, from one 30 kb region, we selected the first 50 variants as the first region and the last 50 variants in the same 30 kb region as the second region. Furthermore, we evaluated the effects of population stratification on the type I error rates. COSI was used to generate another set of 10,000 sequences

based on an African ancestry. Simulated samples based on the European and African ancestries were combined for the IGOF tests. Type I error rates were calculated based on 5,000 and 10,000 simulated replicates of samples for significance levels of 1 and 0.1%, respectively. That is, the IGOF tests were performed for each replicate and the  $p$ -values for the IGOF tests from the 5,000 (10,000) replicates were used to calculate the type I error rates at the 1% (0.1%) significance level. To ensure that the standard error of the  $p$ -value estimation in the adaptive permutation algorithm was less than 20% of the significance level, the maximum number of permutations  $b$  and the cutoff value  $r$  were set at 2,500 and 36, respectively, for the significance level of 1%, and  $b$  and  $r$  were set at 25,000 and 36, respectively, for the significance level of 0.1%. The values  $b$  and  $r$  for the simulation study and real data analysis for ASD were estimated based on the R script available online (<http://motsingerreiflab.jigsy.com/software>) from the author of the adaptive algorithm (Che et al., 2014). The parameter  $m$  was always set to 10 in the two-stage approach for performance improvement in our simulations unless otherwise specified.

For power simulations, we first followed the simulation scenario in Zhao et al. (2016), where both strong main effects and interaction effects were simulated. The penetrance function for SeqSIMLA2 was specified as:

$$\begin{aligned} \text{logit}(P(\textit{affected})) = & \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j \\ & + \beta_{12} x_1 x_2 + \beta_{34} x_3 x_4 + \dots + \beta_{ij} x_i x_j \end{aligned} \quad (7)$$

where  $\alpha_0$  determined the baseline disease prevalence;  $\beta_1, \beta_2, \dots, \beta_j$  are the log of odds ratios of the main effects for variants 1, 2, ...,  $j$ ;  $\beta_{12}, \beta_{34}, \dots, \beta_{ij}$  are the log of odds ratios of interaction effects for the variant pairs; and  $x_1, x_2, \dots, x_j$  are the minor allele counts of the causal variants assuming an additive model. Each of the two genes had 250 variants randomly selected from the 600 variants in each of the genes, and 25 pairs of variants between the two genes were randomly selected as the disease variants. We also simulated a model where the 25 pairs of disease variants all had MAFs < 5%. Similar to the setting in Zhao et al. (2016), all of the log of odds ratios for the main effects and interaction effects were set to  $\log(2)$ ,  $\log(1.5)$ , or  $\log(1.2)$ . A total of 2,000 cases and 2,000 controls were simulated. This simulation setting is referred to as Scenario 1.

We also simulated a model with weaker main effects for the power simulations (referred to as Scenario 2). The penetrance function for SeqSIMLA2 was specified as:

$$\text{logit}(P(\textit{affected})) = \alpha_0 + \beta_{12} x_{12} + \beta_{34} x_{34} + \dots + \beta_{ij} x_{ij} \quad (8)$$

where the notations of  $\alpha_0, \beta_{12}, \beta_{34}, \dots$ , and  $\beta_{ij}$  are the same as those in Equation (7), and  $x_{12}, x_{34}, \dots, x_{ij}$  are the genotype coding of the causal variants based on a given interaction model. Following Wan et al. (2010), we considered four types of interaction models with main effects, including additive, exclusive OR (XOR), classical epistasis, and a model that was used to describe the genetics of handedness and the color of swine (referred to as the color model). For a pair of causal variants  $i$  (with alleles  $A$  and  $a$ ,  $a$  as the minor allele) and  $j$  (with alleles  $B$  and  $b$ ,

$b$  as the minor allele), under the additive model,  $x_{ij}$  was the multiplication of minor allele counts in the two genotypes. Under the XOR model,  $x_{ij}$  was coded as 0 for genotypes  $AABB, AAbb, AaBb, aaBB$ , and  $aabb$ , while  $x_{ij}$  was coded as 1 for the other genotypes. Under the classical epistasis model,  $x_{ij}$  was coded as 1 for genotypes  $AAbb, AaBb$ , and  $aaBB$ , while  $x_{ij}$  was coded as 0 for the other genotypes. Furthermore, under the color model,  $x_{ij}$  was coded as 1 for genotypes  $AAbb, AAbb, AaBB$ , and  $aaBB$ , and was coded as 0 for the other genotypes. We further considered pure epistasis models without main effects (referred to as Scenario 3). GAMETES (Urbanowicz et al., 2012), which generates random, pure and strict epistasis models given disease heritability and prevalence, was used to generate penetrance functions assuming different disease heritability, and the penetrance values calculated by GAMETES were used directly in SeqSIMLA2 instead of using Equation (8). The sample size was fixed at 1,000 cases and 1,000 controls in the power simulations for Scenarios 2 and 3.

We selected various variant pairs to evaluate the power (e.g., four LF variant pairs, three common variant pairs, a mixture of two common and two LF variant pairs) for Scenario 2. A variant pair with MAFs of  $p$  and  $q$  was defined as an LF variant pair if  $1000 \times p^2 q^2$ , which is the expected genotype count of homozygous minor alleles in either cases or controls, was less than 5. Otherwise, the variant pair was defined as a common variant pair. For an LF variant pair  $i$  and  $j$ ,  $\beta$  in Equation (8) was determined by  $0.3 \lfloor \log_{10}(MAF_i \times MAF_j) \rfloor$ , similar to the function used in Wu et al. (2011). For a common variant pair,  $\beta$  was fixed at  $\log(1.25)$  unless otherwise specified. We also considered interaction pairs with different directions of effects by changing the signs of half of the  $\beta$ s in Equation (8). For the pure and strict epistasis model, it was difficult in GAMETES to generate multiple LF variant pairs with pure and strict interaction effects. Therefore, only one LF variant pair with such effects was simulated. Furthermore, six common variants with MAFs > 30% were specified in GAMETES to generate a complicated 6-locus pure and strict epistasis model. A total of 24 models were simulated for the power comparison and the setting for each model can be found in **Table S1** in the Supplementary Material.

We compared the power of the IGOF<sub>combined</sub> test with that of SKAT and FGLM. KBAC was included in the power comparisons only when LF variant pairs had interaction effects, as KBAC aimed to identify associations in rare variants. The power for SKAT was evaluated based on the 2wayIX kernel, which considers 2-way interactions between variants. For FGLM, the parameter  $\lambda$ , which is a smooth parameter for the Fourier expansion, was specified as  $0.01/(\text{number of variants in the genes} \times \text{sample size})$  and the number of basis functions was specified as 23, as suggested in the user manual of FGLM. Power was calculated based on 1,000 batches at the significance levels of 1 and 0.1%.

The threshold parameter  $\tau$  was specified as 0.05 as a default value in the aforementioned simulation settings. We then evaluated the power of the IGOF<sub>combined</sub> test when different values of  $\tau$  (i.e.,  $\tau = 0.05, 0.01$ , and  $0.001$ ) were specified. Furthermore, the parameter  $m$  was specified as 10 in the aforementioned simulation settings. The type I error rates for the IGOF tests may become conservative with smaller  $m$  due to a more stringent threshold for performing the adaptive

permutations. We therefore evaluated the type I error rates and power for the IGOF tests under different values of  $m$  (i.e.,  $m = 10, 1, \text{ and } 0.1$ ).

### ASD Analysis

We first applied the IGOF test to a whole-exome sequencing dataset from the ARRA autism sequencing consortium study (Liu et al., 2013). The study samples were all of European ancestry, as determined based on genetic analysis (i.e., eigen-vector analysis), and European origin. The dataset consisted of two batches of samples sequenced by two centers (Baylor College of Medicine and Broad Institute), and different sequencing platforms and genotype calling pipelines were used at the two centers. We followed the same quality control (QC) procedures as Liu et al. (2013) to merge the two batches of samples except for the selection of variants. Only variants with  $MAF < 1\%$  were selected for the analysis of Liu et al. (2013), as they aimed to identify rare variants associated with ASD. Because the  $IGOF_{\text{combined}}$  test accommodates both common and LF variants, variants with all allele frequency spectrums were used in our analysis. We then evaluated the 10 most significant interaction pairs identified by the  $IGOF_{\text{combined}}$  test in another genome-wide association study (GWAS) dataset for autism obtained from the Autism Genome Project (AGP) (Autism Genome Project et al., 2007; Anney et al., 2010; Pinto et al., 2010). The dataset consisted of approximately 2,700 nuclear families that contained two unaffected parents and one affected child. The samples were genotyped on Illumina Infinium 1 M-single or 1 M-duo SNP arrays. We applied the same QC procedures as described by Anney et al. (2010) to the dataset, including eigen-vector analysis to remove samples not of European ancestry. We then randomly sampled 1,300 unrelated affected children as cases and 1,358 unrelated parents as controls. To increase the density of the SNP set, these samples were imputed against a large reference panel consisting of more than 64,000 haplotypes from the Haplotype Reference Consortium (McCarthy et al., 2016) on the Michigan Imputation Server (Das et al., 2016). Exonic variants were extracted for the analysis. The analysis in the present study was approved by the Institutional Review Board (IRB) of the National Health Research Institutes in Taiwan (IRB protocol # EC1020503-E).

Several studies have suggested that testing gene-gene interactions within protein-protein interaction (PPI) networks can significantly reduce the number of gene pairs to be tested, which alleviated the multiple testing correction burden and hence increased testing power (Baranzini et al., 2009; Emily et al., 2009; Lin et al., 2016). Therefore, we downloaded PPI pairs from the STRING PPI database (von Mering et al., 2005), consisting of 542,895 PPI pairs with combined scores  $> 800$ . As a PPI pair with a combined score  $> 700$  was considered a high confidence pair (von Mering et al., 2005), the selection of a more stringent threshold in our analysis ensured a high quality set of PPI pairs. A total of 207,412 gene pairs, each of which had variants in both genes in the ARRA dataset and proteins encoded by both genes in the STRING PPI pairs, were evaluated by the  $IGOF_{\text{combined}}$  test. The significance threshold  $\alpha$  was set at  $2 \times 10^{-7}$ , which was the conservative Bonferroni correction threshold for the 207,412 tests. The maximum number of permutations  $b$  and the cutoff value  $r$  were set as  $1.25 \times 10^8$  and 36, respectively, in the adaptive permutation algorithm so that the standard error of the  $p$ -value estimation was less than 20% of the significance level. In the analysis,  $m\alpha$  was set at 0.1, so  $IGOF_{\text{combined}}$   $p$ -values less than 0.01 were correctly calculated.

### RESULTS

**Table 3** shows the type I error rates for the  $IGOF_{\text{common}}$ ,  $IGOF_{\text{LF}}$ , and  $IGOF_{\text{combined}}$  tests for different sample sizes and different numbers of variant pairs between the two genes at the 1% significance level. The type I error rates at the 0.1% significance level are shown in **Table S2**. All 95% confidence intervals for the type I error rate estimates contained the expected levels. **Table 3** also shows the proportions of variant pairs passing the selection thresholds (i.e.,  $p < 0.05$  for common variant pairs and  $N_{ab}$  in cases  $> N_{ab}$  in controls for LF variant pairs) in the threshold algorithms. As expected, approximately 5% of common variant pairs were selected for calculating the IGOF statistic, while approximately 5–9% of LF variant pairs were selected. **Table 4** shows the type I error rates for the IGOF tests in the presence of LD or population stratification. The IGOF tests maintained appropriate type I error rates in the presence of either strong

**TABLE 3 |** Type I error rates for the IGOF tests at the 1% significance level.

No. of variants <sup>a</sup>	Sample size <sup>b</sup>	IGOF <sub>common</sub>		IGOF <sub>LF</sub>		IGOF <sub>combined</sub>
		Type I error	Prop <sup>c</sup> (%)	Type I error	Prop <sup>d</sup> (%)	Type I error
10,10 (8,6)	2,000	0.0088	4.4	0.0094	9.0	0.0080
30,30 (26,25)	2,000	0.0082	5.1	0.0102	5.3	0.0094
50,50 (42,41)	1,000	0.0124	4.6	0.0110	4.9	0.0092
	2,000	0.0100	5.5	0.0092	6.0	0.0080
	3,000	0.0096	4.6	0.0098	6.9	0.0098
	4,000	0.0078	4.7	0.0106	7.6	0.0084
100,100 (77,83)	2,000	0.0106	5.0	0.0076	7.2	0.0086

<sup>a</sup>Number of variants in the two genes and the numbers in parentheses show the numbers of variants with  $MAFs < 1\%$ .

<sup>b</sup>Number of cases and controls where the numbers of cases and controls are equal.

<sup>c</sup>Proportion of common variant pairs selected by the threshold algorithm.

<sup>d</sup>Proportion of LF variant pairs selected by the threshold algorithm.

or weak LD between variants in the two genes for testing 1,000 cases and 1,000 controls. However, in the presence of strong population stratification, the IGOF tests were conservative. This was not surprising as population stratification was not modeled in the IGOF tests, and the presence of population stratification may result in incorrect calculations of the expected values in the  $C^2$  statistic.

**Figure 1** shows the power comparison at the 0.1% significance level for Scenario 1. For Model 1 where strong main effects and interaction effects were simulated, all four tests maintained high power. When the effect sizes decreased to  $\log(1.2)$  (i.e., Model 3), IGOF<sub>combined</sub> and SKAT still maintained high power, but the power of FGLM and KBAC decreased significantly. For Model 4 where all the causal variants have MAFs < 5%,

IGOF<sub>combined</sub> still had high power followed by KBAC and SKAT, while FGLM resulted in negligible power. **Figure 2** shows the power comparison when common variant pairs with interaction effects were simulated for Scenario 2 (Models 5–8). IGOF<sub>combined</sub> had significantly higher power than did the other tests under Models 6–8, whereas under Model 5 (the additive model), SKAT had slightly more power than IGOF<sub>combined</sub>. **Figure 3** shows the power comparison when LF variant pairs with interaction effects were simulated (Models 9–12), as well as when LF variant pairs with different directions of interaction effects were simulated (Models 13–16). IGOF<sub>combined</sub> consistently had the highest power across all models, except that for Model 13, FGLM had the highest power. Typically, under the XOR model (Models 10 and 14), IGOF<sub>combined</sub> had significantly higher power than the other tests. Interestingly, although IGOF<sub>combined</sub> assumed risk effects when selecting LF variant pairs, it still maintained power comparable to the other tests for Models 13–16, where different directions of effects were simulated. IGOF<sub>combined</sub> also had the highest power when a mixture of common and LF variant pairs with interaction effects were simulated (Models 17–20), as shown in **Figure 4**. Finally, **Figure 5** shows the power comparison for Scenario 3 when pure epistasis was simulated. SKAT had the highest power in Model 21 (with heritability of 0.1), followed by FGLM and IGOF. However, when heritability was increased to 0.2 (Model 22), IGOF showed significantly higher power than SKAT and FGLM. More interestingly, IGOF also had significantly higher power than the other tests did under the complicated 6-locus pure and strict epistasis models (Models 23 and 24). **Figures S1–S5** show the power comparisons for Models 1–24 at the 1% significance level and the power patterns were similar to those at the 0.1% significance level.

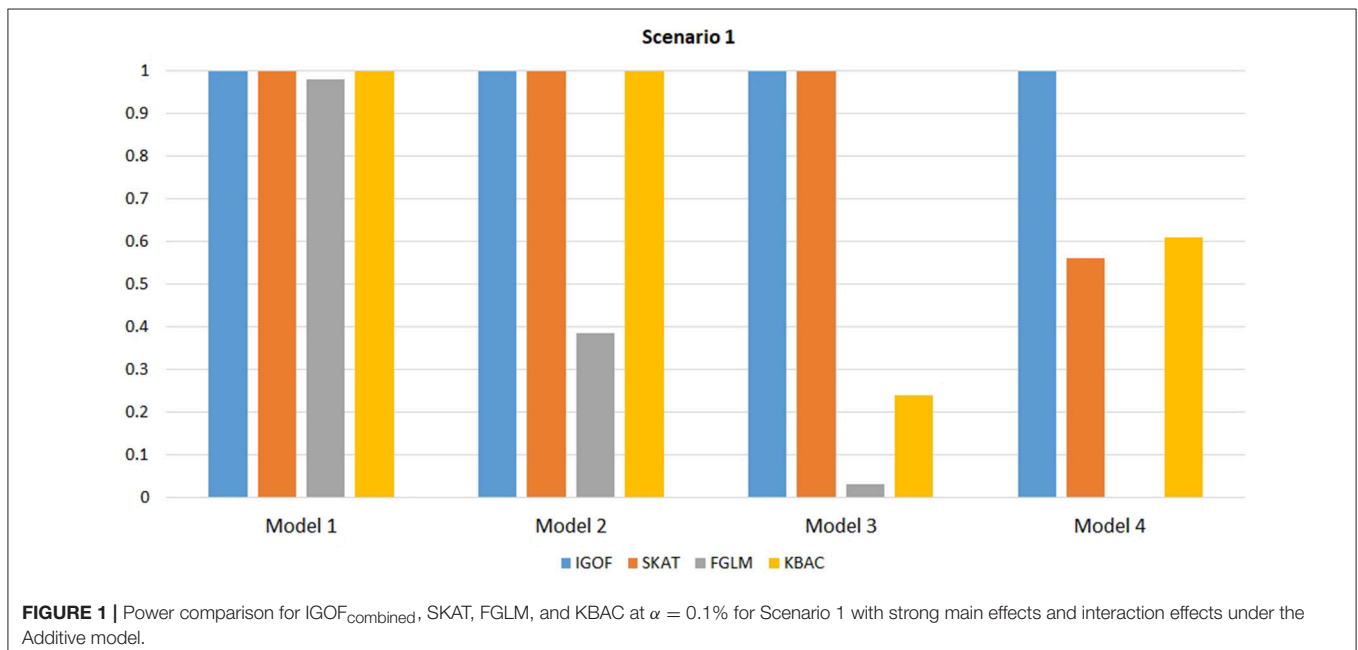
**Figure 6** shows the power changes for IGOF<sub>combined</sub> under Models 5, 18, and 23 when different values of  $\tau$ , the threshold for selecting common variant pairs in the test statistic, were specified

**TABLE 4 |** Type I error rates for the IGOF tests in the presence of LD or population stratification at the 1% significance level.

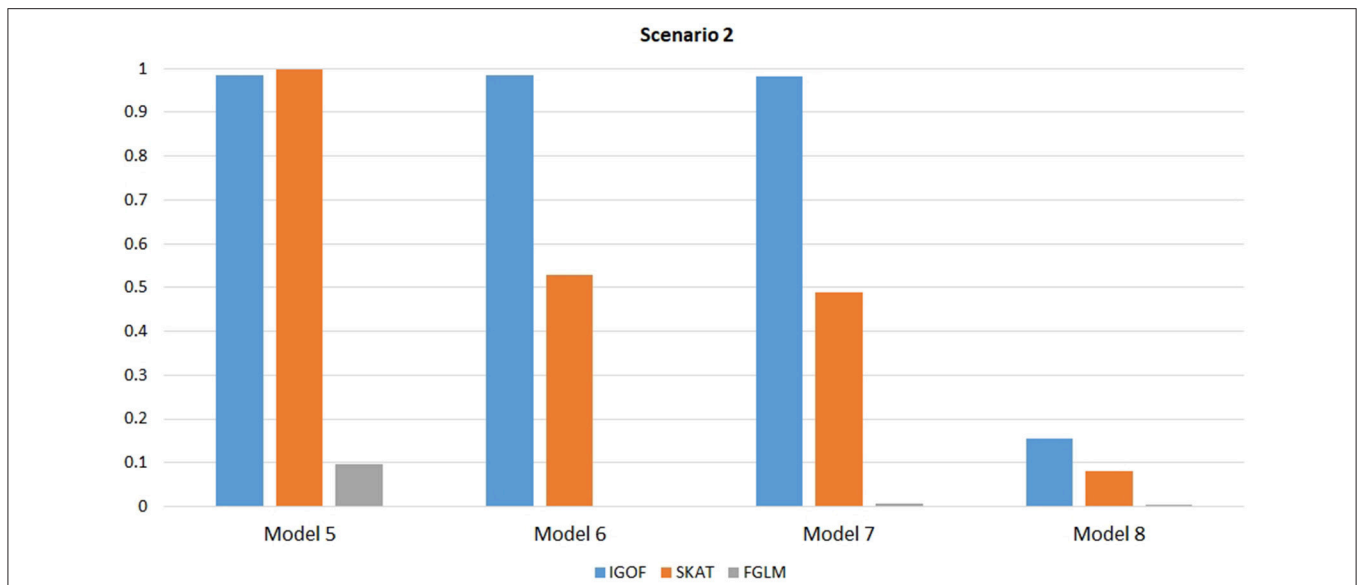
	No. of variants <sup>a</sup>	IGOF <sub>common</sub>	IGOF <sub>LF</sub>	IGOF <sub>combined</sub>
<b>LD</b>				
Strong	50,50 (38,47)	0.0100	0.0124	0.011
Weak	50,50 (36,41)	0.0086	0.0094	0.0076
<b>STRATIFICATION</b>				
AFR:800 <sup>b</sup>	50,50 (43,45)	0.0058	0	0.0026
EUR:3200				
AFR:2000	50,50 (43,45)	0.0048	0	0.0022
EUR:2000				
AFR:3200	50,50 (43,45)	0.0046	0	0.0010
EUR:800				

<sup>a</sup>Number of variants in the two genes and the numbers in parentheses show the numbers of variants with MAFs < 1%.

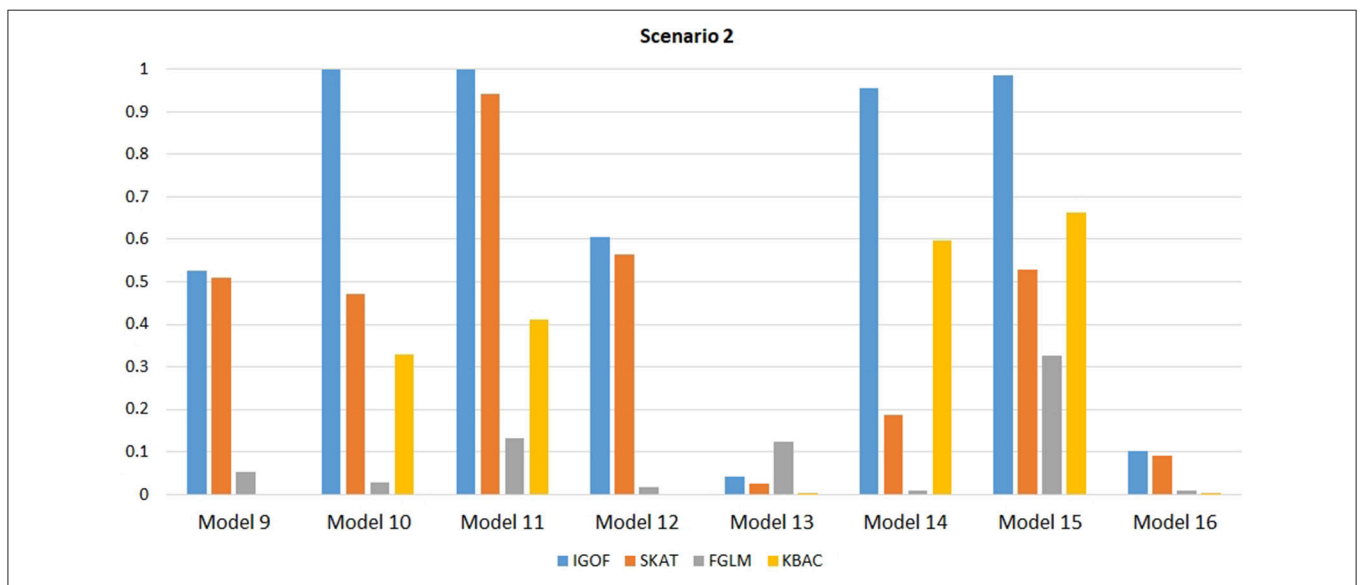
<sup>b</sup>Sample sizes in each population. The numbers of cases and controls are equal.



**FIGURE 1 |** Power comparison for IGOF<sub>combined</sub>, SKAT, FGLM, and KBAC at  $\alpha = 0.1\%$  for Scenario 1 with strong main effects and interaction effects under the Additive model.



**FIGURE 2 |** Power comparison for IGOF<sub>combined</sub>, SKAT, and FGLM at  $\alpha = 0.1\%$  for Scenario 2, where only common variant pairs had interaction effects under the Additive, XOR, Color, and Classical models.

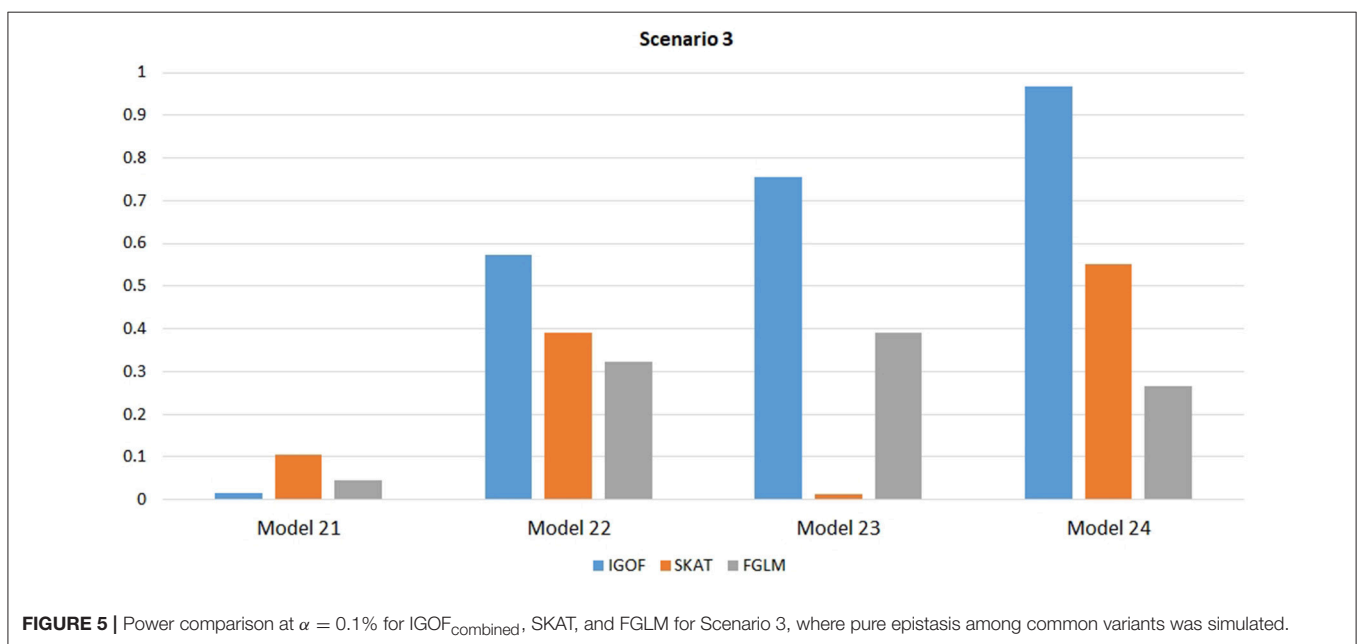
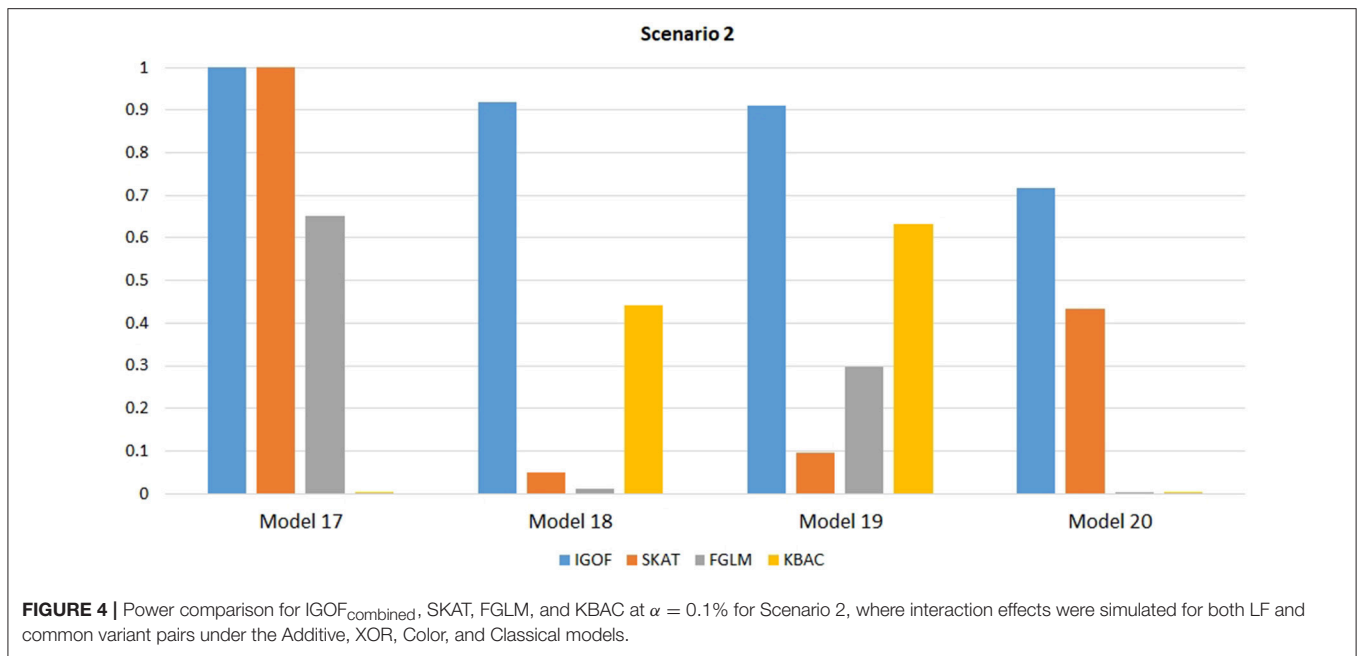


**FIGURE 3 |** Power comparison for IGOF<sub>combined</sub>, SKAT, FGLM, and KBAC at  $\alpha = 0.1\%$  for Scenario 2, where only LF variant pairs had interaction effects under the Additive, XOR, Color, and Classical models.

at the 0.1% significance level. Models 5, 18, and 23 were selected for the power evaluation because common variants were among the disease variants in these models. Furthermore, the power for IGOF<sub>combined</sub> was not close to 1 when  $\tau$  was 0.05 in these models, which would make it easier to observe the power changes when different values of  $\tau$  were specified. As seen in the Figure, the power under each model was similar with different values of  $\tau$ , suggesting that the default value of 0.05 for  $\tau$  was appropriate for IGOF<sub>combined</sub> to achieve optimal power in our simulations. **Table 5** shows the type I error rates for the IGOF tests when  $m = 1$

and  $m = 0.1$ . The type I error rates were appropriate when  $m = 1$ , but they can be conservative for IGOF<sub>LF</sub> and IGOF<sub>combined</sub> when  $m = 0.1$ . The conservative type I error rates were not surprising because when  $m = 0.1$ , permutations were not performed in a larger proportion of simulated replicates compared to  $m = 1$  and  $m = 10$ , and the replicates without permutations had  $p$ -values of 1. **Figure 7** shows the power changes for IGOF<sub>combined</sub> under Models 5, 18, and 23 when different values of  $m$  were specified at the 0.1% significance level. Although the type I error rate for IGOF<sub>combined</sub> was conservative when  $m = 0.1$ , we

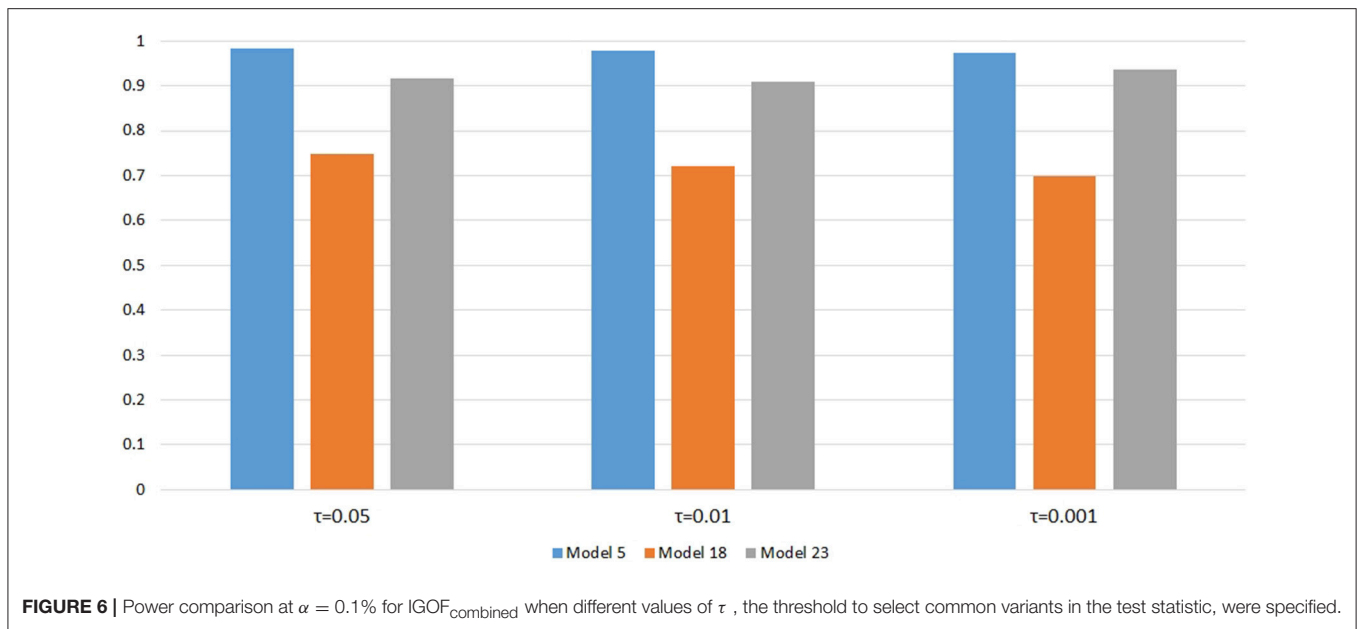




observed similar power under different values of  $m$  for these Models.

In the ASD analysis, there were 897 cases and 844 controls and 222,211 exonic variants after QC in the ARRA dataset. In the AGP dataset, there were 814,767 SNPs in 1,300 cases and 1,358 controls after QC. After imputation, there were 203,270 exonic variants, where 168,961 variants had MAFs < 5%. The tests were distributed evenly across 5 computers, while 10 parallel threads were executed on each computer. The entire analysis was completed in 1 week. The 10 most significant gene pairs identified by IGOF<sub>combined</sub> using the ARRA dataset are

shown in **Table 6**. The most significant result was the interaction between the NADH:ubiquinone oxidoreductase subunit AB1 (NDUFAB1) and NADH:ubiquinone oxidoreductase core subunit V2 (NDUFV2) genes with a  $p$ -value of  $3.30 \times 10^{-5}$ , which did not pass the stringent Bonferroni threshold of  $2 \times 10^{-7}$ . Interestingly, all 10 gene pairs had IGOF<sub>common</sub>  $p$ -values of 1, suggesting that no common interaction pairs had individual IGOF  $p < 0.05$ . Hence, the replication of the 10 gene pairs was evaluated by IGOF<sub>LF</sub> in the AGP dataset. One gene pair—ubiquinol-cytochrome *c* reductase core protein II (UQCRC2) and NDUFV2 had a significant IGOF<sub>LF</sub>  $p$ -value



**FIGURE 6** | Power comparison at  $\alpha = 0.1\%$  for  $IGOF_{combined}$  when different values of  $\tau$ , the threshold to select common variants in the test statistic, were specified.

**TABLE 5** | Type I error rates for the IGOF tests with  $m = 1$  and  $m = 0.1$ .

No. of variants	IGOF <sub>common</sub>		IGOF <sub>LF</sub>		IGOF <sub>combined</sub>	
	$m = 1$	$m = 0.1$	$m = 1$	$m = 0.1$	$m = 1$	$m = 0.1$
10,10	0.0078	0.0078	0.0086	<b>0.0016<sup>a</sup></b>	0.0080	<b>0.0054</b>
30,30	0.0090	0.0090	0.0098	<b>0.0040</b>	0.0088	0.0078
50,50	0.0102	0.0100	0.0092	<b>0.0046</b>	0.0088	0.0082
100,100	0.0108	0.0108	0.0080	<b>0.0038</b>	0.0080	0.0078

<sup>a</sup>Values in bold are outside the 95% confidence intervals for the type I error estimates.

of 0.0492. **Table 6** also shows the marginal  $p$ -values for testing the main effects of individual genes using SKAT. The kernel “linear.weighted,” which assigned more weights to rarer variants, was used. No marginal significance was observed in the gene pair of (UQCRC2, NDUFV2), suggesting that the significance of the IGOF tests resulted from the interaction effect between the two genes on ASD. **Table S3** shows the genetic variants and their MAFs in genes shown in **Table 6**. Most of the variants had MAFs < 5%, suggesting that IGOF was powerful for identifying LF variant pairs with interaction effects. Furthermore, the genes shown in **Table 6** are mostly on different chromosomes, and we did not observe LD between genes on the same chromosome.

## DISCUSSION

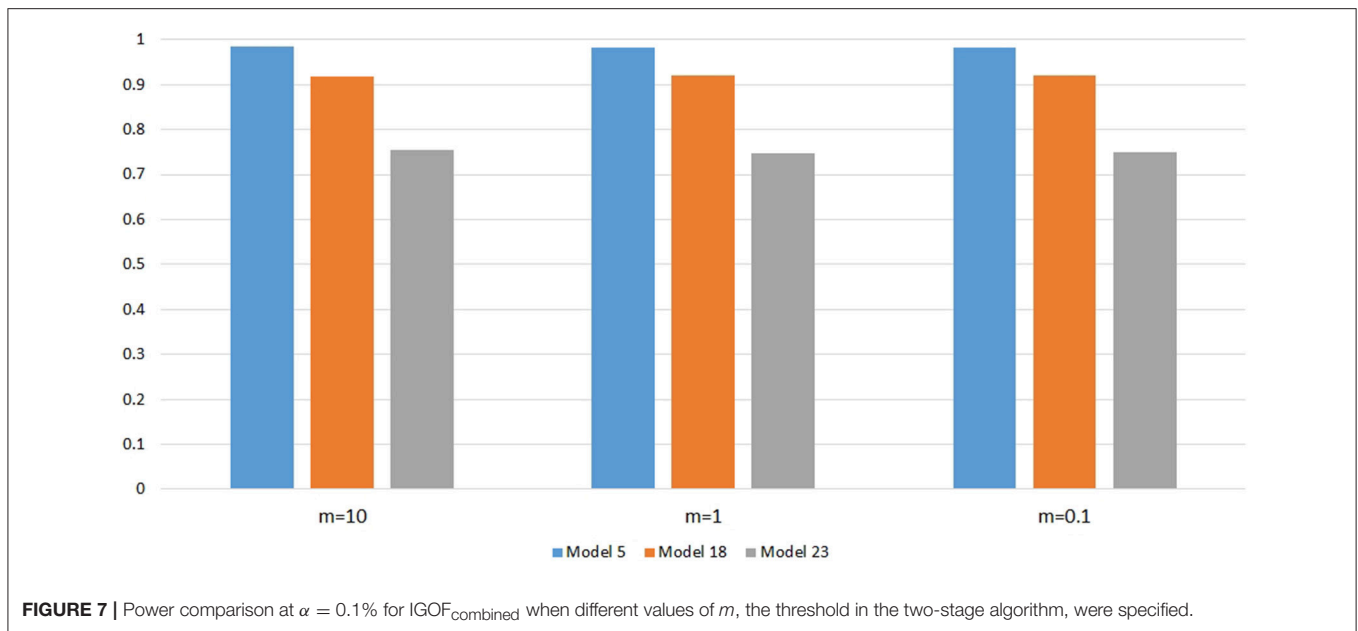
In this study, we developed the  $IGOF_{common}$ ,  $IGOF_{LF}$ , and  $IGOF_{combined}$  tests, which are powerful gene-based interaction tests for common, LF, and common combined with LF variant pairs, respectively. Based on threshold algorithms, only a subset of promising interaction pairs contributed to the final IGOF statistics. Moreover, while an adaptive permutation procedure

was required to obtain  $p$ -values for the IGOF tests, a two-stage strategy was used to improve the efficiency of the tests in terms of run time. Our simulation study results suggest that the IGOF tests are valid tests and can have significantly higher power than SKAT and FGLM.

Several rare-variant association tests considering main effects assign more weights to rarer variants in the statistics, assuming that rarer variants have larger effect sizes (Madsen and Browning, 2009; Wu et al., 2011; Ionita-Laza et al., 2013). These simulation studies suggest that weighting can improve power. It is straightforward to incorporate weights in the  $IGOF_{LF}$  statistic by calculating  $I_{LF} = \sum_{j \in \Psi} w_j C_j^2$ , where  $w_j$  is the weight for variant

pair  $j$ . Similar to the rare-variant association tests considering main effects,  $w_j$  can be determined by a function of the MAFs of the variant pair, such as the Beta distribution density function used in SKAT via the “linear.weighted” kernel. However, when we used  $w_j = Beta(pq; 1, 25)$ , where  $p$  and  $q$  were the MAFs of the variants, and 1 and 25 were the default parameters of the Beta distribution used in SKAT, there was little impact on the power of the  $IGOF_{LF}$  test. Hence, weighting was not used in this study for the  $IGOF_{LF}$  test.

The  $p$ -value for the  $IGOF_{combined}$  test is calculated using the two  $p$ -values from the  $IGOF_{common}$  and  $IGOF_{LF}$  tests based on the extended Simes’ test. The calculation assumed that the two  $p$ -values are independent. The assumption may not hold when there is LD between variants in the common and LF variant pairs. However, as seen in the results from our simulations where the LD structures were simulated among variants, the  $IGOF_{combined}$  test maintained appropriate type I error rates across different scenarios. Li et al. (2012) suggests that when combining correlated  $p$ -values from individual variants using Fisher’s method, the degree of freedom for the test statistic should be adjusted based on the correlation of  $p$ -values for each pairwise combination of the variants. The correlation of the  $IGOF_{common}$



**FIGURE 7** | Power comparison at  $\alpha = 0.1\%$  for  $IGOF_{combined}$  when different values of  $m$ , the threshold in the two-stage algorithm, were specified.

**TABLE 6** | The 10 most significant gene pairs from the ASD analysis.

Gene 1 <sup>a</sup>	Gene 2 <sup>a</sup>	$IGOF_{combined}$ (ARRA)	$IGOF_{LF}$ (AGP)
NDUFAB1 (8.33E-02)	NDUFV2 (2.98E-01)	3.30E-05	1.17E-01
ZNF217 (1.23E-04)	KDM1A (9.68E-01)	4.01E-05	7.75E-01
ATP6V1E2 (6.47E-02)	ATP6VOB (4.86E-04)	4.76E-05	7.51E-01
ATP6VOA1 (2.50E-02)	ATP6VOB (4.86E-04)	5.30E-05	7.93E-01
INSR (1.33E-01)	ATP6VOB (4.86E-04)	7.24E-05	8.76E-01
NDUFV2 (2.98E-01)	NDUFV3 (7.37E-01)	7.45E-05	1.50E-01
CDC34 (6.66E-01)	CACUL1 (9.33E-03)	8.21E-05	NA <sup>b</sup>
ATP5B (1.30E-01)	ATP6VOB (4.86E-04)	8.91E-05	6.01E-01
UQCRC2 (3.77E-01)	NDUFV2 (2.98E-01)	9.49E-05	4.92E-02
NDUFV2 (2.98E-01)	NDUFA13 (5.04E-01)	1.01E-04	2.53E-01

<sup>a</sup>Gene and its marginal  $p$ -value tested by SKAT with the "linear.weighted" kernel.

<sup>b</sup>No variants were observed in this gene so that no tests were performed.

and  $IGOF_{LF}$  test  $p$ -values may be estimated by a set of permuted  $p$ -values calculated from the permuted statistics. It is our future work to consider the correlation of the  $IGOF_{common}$  and  $IGOF_{LF}$  test  $p$ -values when calculating the  $IGOF_{combined}$  test  $p$ -value.

The IGOF test statistic is constructed based on the statistic for testing pairwise interaction. However, as shown in our simulation results, the IGOF test was also powerful for identifying high-dimensional interactions among 6 variants with MAFs > 30%. It is possible to extend the GOF statistic in Equation (1) to consider  $n$ -dimensional interactions by enumerating all possible genotype categories from the  $n$  variants, and similar IGOF statistics can be constructed. However, this would significantly increase the computational burden as substantially more combinations of variants between two genes would be tested.

The IGOF tests use a two-stage approach to improve computational efficiency. Although a large portion of tests with nonsignificant  $p$ -values will not undergo the adaptive permutation procedure, a significant amount of time is still required in the adaptive permutations for tests with very small  $p$ -values (e.g.,  $10^{-8}$ ). Therefore, the program may not be suitable for genome-wide analyses testing gene-gene interactions between all possible pairs of genes across the genome. As demonstrated in our ASD analysis, restricting the analysis to the PPI network significantly reduced the number of gene pairs tested, so the analysis can be finished in a reasonable time frame.

One major limitation of the IGOF test is that covariates such as age, sex, and population stratification are not considered. Our simulation study results suggest that in the presence of population stratification, the IGOF test can be conservative, which may result in low statistical power. In contrast to the IGOF test, covariates can be modeled in the regression-based tests SKAT and FGLM. Hence, if there is population stratification in the sample, regression-based methods such as SKAT and FGLM should be used. Another limitation of the IGOF test is that for data with imputed genotypes, it cannot use genotype dosages to account for imputation uncertainties. One strategy for analyzing imputed genotypes in the IGOF test is to use a "best guess" genotype for an individual by selecting the imputed genotype with the maximum posterior probability. This strategy was adopted in our ASD analysis for the imputed AGP data. However, as shown in the simulation studies by Zheng et al. (2011), power can be compromised by using "best guess" genotypes relative to genotype dosages.

Our application of the IGOF test to the ARRA dataset did not identify significant gene pairs passing the stringent Bonferroni correction threshold. However, the threshold may be too conservative as the IGOF tests can be highly correlated

for testing the interactions of the same gene with other genes. Furthermore, the (UQCRC2, NDUFV2) gene pair from the 10 most significant gene pairs showed a significant  $IGOF_{LF}$   $p$ -value ( $p$ -value < 0.05) in the replication AGP dataset. Rare copy number variations in the UQCRC2 gene have been found to be associated with ASD (Matsunami et al., 2013). Furthermore, the NDUFV2 gene is associated with psychiatric disorders such as schizophrenia and bipolar disorder. Further studies are warranted to investigate possible interaction effects of the two genes on ASD.

In conclusion, we developed three gene-based gene-gene interaction tests for analyzing NGS data in case-control studies. These tests will be very useful for identifying interaction effects between two genes on complex diseases. The software implementing the three tests is available at <http://igof.sourceforge.net>.

## AUTHOR CONTRIBUTIONS

R-HC developed the method. R-HC and C-YK designed and performed the simulation study, analyzed the data, and wrote the manuscript.

## ACKNOWLEDGMENTS

We are grateful to the National Center for High-performance Computing in Taiwan for computer time and facilities. The ARRA datasets used for the analysis described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number phs000298.v1.p1. The data set(s) were deposited by the ARRA Autism Sequencing Collaborative, an ARRA funded research initiative. Support for the Autism Sequencing Collaborative was provided by grants: R01-MH089208 awarded to Dr. Mark Daly, R01-MH089175 awarded to Dr. Richard Gibbs, R01-MH089025 awarded to Joseph Buxbaum, R01-MH089004 awarded to Gerard

Schellenberg, and R01-MH089482 awarded to James Sutcliffe. The AGP datasets used for the analysis described in this manuscript were also obtained from dbGaP through dbGaP accession number, phs000267.v4.p2. Submission of the data, phs000267.v4.p2, to dbGaP was provided by Dr. Bernie Devlin on behalf of the Autism Genome Project (AGP). Collection and submission of the data to dbGaP were supported by a grant from the Medical Research Council (G0601030) and the Wellcome Trust (075491/Z/04), Anthony P. Monaco, P.I., University of Oxford. This work was funded by grants from the National Health Research Institutes (PH-106-PP-10) and Ministry of Science and Technology (MOST 104-2221-E-400-004-MY2).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2017.00228/full#supplementary-material>

**Figure S1** | Power comparison for  $IGOF_{combined}$ , SKAT, FGLM, and KBAC at  $\alpha = 1\%$  for Scenario 1 with strong main effects and interaction effects under the Additive model.

**Figure S2** | Power comparison for  $IGOF_{combined}$ , SKAT, and FGLM at  $\alpha = 1\%$  for Scenario 2, where only common variant pairs had interaction effects under the Additive, XOR, Color, and Classical models.

**Figure S3** | Power comparison for  $IGOF_{combined}$ , SKAT, FGLM, and KBAC at  $\alpha = 1\%$  for Scenario 2, where only LF variant pairs had interaction effects under the Additive, XOR, Color, and Classical models.

**Figure S4** | Power comparison for  $IGOF_{combined}$ , SKAT, FGLM, and KBAC at  $\alpha = 1\%$  for Scenario 2, where interaction effects were simulated for both LF and common variant pairs under the Additive, XOR, Color, and Classical models.

**Figure S5** | Power comparison at  $\alpha = 1\%$  for  $IGOF_{combined}$ , SKAT, and FGLM for Scenario 3, where pure epistasis among common variants was simulated.

**Table S1** | Power simulation models.

**Table S2** | Type I error rates for the  $IGOF$  tests at the 0.1% significance level.

**Table S3** | Genetic variants and their minor allele frequencies in the 10 most significant gene pairs from the ASD analysis.

## REFERENCES

- Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T. R., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* 19, 4072–4082. doi: 10.1093/hmg/ddq307
- Autism Genome Project, C., Szatmari, P., Paterson, A. D., Zwaigenbaum, L., Roberts, W., Brian, J., et al. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.* 39, 319–328. doi: 10.1038/ng1985
- Baranzini, S. E., Galwey, N. W., Wang, J., Khankhanian, P., Lindberg, R., Pelletier, D., et al. (2009). Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum. Mol. Genet.* 18, 2078–2090. doi: 10.1093/hmg/ddp120
- Box, G. E. P. (1954). Some Theorems on quadratic forms applied in the study of analysis of variance. *Ann. Mat. Stat.* 25, 290–302. doi: 10.1214/aoms/1177728786
- Che, R., Jack, J. R., Motsinger-Reif, A. A., and Brown, C. C. (2014). An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use. *BioData Min.* 7:9. doi: 10.1186/1756-0381-7-9
- Chung, R. H., Tsai, W. Y., and Martin, E. R. (2014). Family-based association test using both common and rare variants and accounting for directions of effects for sequencing data. *PLoS ONE* 9:e107800. doi: 10.1371/journal.pone.0107800
- Chung, R. H., Tsai, W. Y., Hsieh, C. H., Hung, K. Y., Hsiung, C. A., and Hauser, E. R. (2015). SeqSIMLA2: simulating correlated quantitative traits accounting for shared environmental effects in user-specified pedigree structure. *Genet. Epidemiol.* 39, 20–24. doi: 10.1002/gepi.21850
- Cordell, H. J. (2009). Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* 10, 392–404. doi: 10.1038/nrg2579
- Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287. doi: 10.1038/ng.3656
- Emily, M., Mailund, T., Hein, J., Schausser, L., and Schierup, M. H. (2009). Using biological networks to search for interacting loci in genome-wide association studies. *Eur. J. Hum. Genet.* 17, 1231–1240. doi: 10.1038/ejhg.2009.15
- Ionita-Laza, I., Buxbaum, J. D., Laird, N. M., and Lange, C. (2011). A new testing strategy to identify rare variants with either risk or protective effect on disease. *PLoS Genet.* 7:e1001289. doi: 10.1371/journal.pgen.1001289
- Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., and Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. *Am. J. Hum. Genet.* 92, 841–853. doi: 10.1016/j.ajhg.2013.04.015

- Li, B., and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* 83, 311–321. doi: 10.1016/j.ajhg.2008.06.024
- Li, M. X., Gui, H. S., Kwan, J. S., and Sham, P. C. (2011). GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am. J. Hum. Genet.* 88, 283–293. doi: 10.1016/j.ajhg.2011.01.019
- Li, M. X., Kwan, J. S., and Sham, P. C. (2012). HYST: a hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. *Am. J. Hum. Genet.* 91, 478–488. doi: 10.1016/j.ajhg.2012.08.004
- Lin, P. L., Yu, Y. W., and Chung, R. H. (2016). Pathway analysis incorporating protein-protein interaction networks identified candidate pathways for the seven common diseases. *PLoS ONE* 11:e0162910. doi: 10.1371/journal.pone.0162910
- Liu, D. J., and Leal, S. M. (2010). A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet.* 6:e1001156. doi: 10.1371/journal.pgen.1001156
- Liu, L., Sabo, A., Neale, B. M., Nagaswamy, U., Stevens, C., Lim, E., et al. (2013). Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genet.* 9:e1003443. doi: 10.1371/journal.pgen.1003443
- Ma, L., Clark, A. G., and Keinan, A. (2013). Gene-based testing of interactions in association studies of quantitative traits. *PLoS Genet.* 9:e1003321. doi: 10.1371/journal.pgen.1003321
- Madsen, B. E., and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 5:e1000384. doi: 10.1371/journal.pgen.1000384
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753. doi: 10.1038/nature08494
- Matsunami, N., Hadley, D., Hensel, C. H., Christensen, G. B., Kim, C., Frackelton, E., et al. (2013). Identification of rare recurrent copy number variants in high-risk autism families and their prevalence in a large ASD population. *PLoS ONE* 8:e52239. doi: 10.1371/journal.pone.0052239
- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283. doi: 10.1038/ng.3643
- Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466, 368–372. doi: 10.1038/nature09146
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika* 6, 309–316. doi: 10.1007/BF02288586
- Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., and Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* 15, 1576–1583. doi: 10.1101/gr.3709305
- Urbanowicz, R. J., Kiralis, J., Sinnott-Armstrong, N. A., Heberling, T., Fisher, J. M., and Moore, J. H. (2012). GAMETES: a fast, direct algorithm for generating pure, strict, epistatic models with random architectures. *BioData Min.* 5:16. doi: 10.1186/1756-0381-5-16
- von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., et al. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 33, D433–D437. doi: 10.1093/nar/gki005
- Wan, X., Yang, C., Yang, Q., Xue, H., Fan, X., Tang, N. L., et al. (2010). BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am. J. Hum. Genet.* 87, 325–340. doi: 10.1016/j.ajhg.2010.07.021
- Wang, Y. T., Sung, P. Y., Lin, P. L., Yu, Y. W., and Chung, R. H. (2015). A multi-SNP association test for complex diseases incorporating an optimal P-value threshold algorithm in nuclear families. *BMC Genomics* 16:381. doi: 10.1186/s12864-015-1620-3
- Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika* 29, 350–362. doi: 10.1093/biomet/29.3-4.350
- Wiens, D. P., Beaulieu, N. C., and Loskot, P. (2006). On the exact distribution of the sum of the largest n-k out of n normal random variables with differing mean values. *Statistics* 40, 165–173. doi: 10.1080/02331880500484762
- Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., et al. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *Am. J. Hum. Genet.* 86, 929–942. doi: 10.1016/j.ajhg.2010.05.002
- Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* 89, 82–93. doi: 10.1016/j.ajhg.2011.05.029
- Yates, F. (1934). Contingency tables involving small numbers and the X<sup>2</sup> test. *Suppl. J. R. Stat. Soc.* 1, 217–235. doi: 10.2307/2983604
- Zaykin, D. V., Zhivotovskiy, L. A., Westfall, P. H., and Weir, B. S. (2002). Truncated product method for combining P-values. *Genet. Epidemiol.* 22, 170–185. doi: 10.1002/gepi.0042
- Zhao, J., Zhu, Y., and Xiong, M. (2016). Genome-wide gene-gene interaction analysis for next-generation sequencing. *Eur. J. Hum. Genet.* 24, 421–428. doi: 10.1038/ejhg.2015.147
- Zheng, J., Li, Y., Abecasis, G. R., and Scheet, P. (2011). A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genet. Epidemiol.* 35, 102–110. doi: 10.1002/gepi.20552

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Chung and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.