Hindawi Publishing Corporation
BioMed Research International
Volume 2015, Article ID 461593, 10 pages
http://dx.doi.org/10.1155/2015/461593

Research Article

Robust Association Tests for the Replication of Genome-Wide Association Studies

Jungnam Joo,[1] Ju-Hyun Park,[2] Bora Lee,[1] Boram Park,[1] Sohee Kim,[1] Kyong-Ah Yoon,[3] Jin Soo Lee,[3] and Nancy L. Geller[4]

[1] *Biometric Research Branch, Research Institute and Hospital, National Cancer Center, Gyeonggi-do, Goyang-si 410-769, Republic of Korea*
[2] *Department of Statistics, Dongguk University, Seoul 100-715, Republic of Korea*
[3] *Lung Cancer Branch, Research Institute and Hospital, National Cancer Center, Gyeonggi-do, Goyang-si 410-769, Republic of Korea*
[4] *Office of Biostatistics Research, National Heart, Lung and Blood Institute, Bethesda, MD 20892-7938, USA*

Correspondence should be addressed to Jungnam Joo; jooj@ncc.re.kr

Received 14 November 2014; Revised 14 February 2015; Accepted 14 February 2015

Academic Editor: Taesung Park

In genome-wide association study (GWAS), robust genetic association tests such as maximum of three CATTs (MAX3), each corresponding to recessive, additive, and dominant genetic models, the minimum p value of Pearson's Chi-square test with 2 degrees of freedom, and CATT based on additive genetic model (MIN2), genetic model selection (GMS), and genetic model exclusion (GME) methods have been shown to provide better power performance under wide range of underlying genetic models. In this paper, we demonstrate how these robust tests can be applied to the replication study of GWAS and how the overall statistical significance can be evaluated using the combined test formed by p values of the discovery and replication studies.

1. Introduction

With the advance of biotechnology and substantial reduction of genotyping costs, a genome-wide association study (GWAS) using hundred thousand markers in several thousand individuals is now increasingly utilized and has been successful in detecting genetic associations across the entire genome with complex human traits [1–6]. Among many challenges this application holds; development of more efficient and robust statistical methodologies with higher power to detect an association with a single marker has been one of the most important statistical issues, given that effects of individual markers are usually characterized as being small to moderate. One attempt to overcome this challenge is focused on developing efficient tests that are robust against underlying genetic model misspecification.

Two most frequently used association tests are the allele-based test (ABT) and the genotype-based test (GBT). ABT compares the allele frequencies between cases and controls, while GBT compares the genotype distributions of cases and controls. The Cochran-Armitage trend test (CATT) [7, 8] is a popular GBT which takes into account the underlying genetic model. It is well known, however, that the ABT may inflate type I error when Hardy-Weinberg equilibrium (HWE) does not hold in the samples [9]. Even under HWE, when the genetic model is recessive or dominant, the ABT may suffer from serious power loss. On the other hand, the CATT does not depend on HWE, but to apply the CATT the choice of scores optimal for the underlying genetic model needs to be specified. For complex diseases, the genetic model is usually unknown and robust tests such as the maximum of three CATTs (MAX3) [10] and the maximum efficiency robust test (MERT) [11, 12] are preferable. Alternatively, Zheng and Ng [13] and Joo et al. [14] proposed a two-phase analysis based on the genetic model selection (GMS) and genetic model exclusion (GME). Moreover, an alternative approach was proposed by the Wellcome trust case-control consortium (WTCCC) [5] which used a minimum p value of Pearson's Chi-square test and additive CATT, and the asymptotic properties of this approach were studied in detail by Joo et al. [15]. These methods provide better or comparable power performance than some of the robust tests such as MAX3.

In this paper, we illustrate how these robust tests can be applied to a replication study of GWAS and how overall statistical significance can be evaluated using the combined test formed by p values of the discovery and replication studies. The importance of replication or validation in GWAS has been well recognized [16, 17], and joint analysis in a two-stage design of GWAS has been proved to be more powerful than replication-based analysis and has been widely conducted in GWAS with a variety of phenotypes of interest [18, 19].

The paper is organized as follows. We first describe the data structures and notation and review existing robust association tests for a single data set. Then we describe how to obtain the p value for the replication data set, given the significant result of the discovery stage, using robust tests. In the next section, a combined test of the p values of the discovery and replication data sets is proposed, together with the way to evaluate the statistical significance for the combined test. Simulation studies are conducted to compare the type I error rates and powers of various analytical strategies. For illustration purposes, the summarized methods are applied to a non-small-cell lung cancer data set and at the end there is a discussion.

2. Methods

2.1. Data and Notation. For a marker with two alleles A and B, let the frequencies of B in cases and controls be $p = P(B \mid \text{case})$ and $q = P(B \mid \text{control})$. Denote three genotypes by $G_0 = AA$, $G_1 = AB$, and $G_2 = BB$. In case-control association studies, r cases and s controls are independently sampled from each population. The observed genotype counts for (G_0, G_1, G_2) are (r_0, r_1, r_2) in the cases and (s_0, s_1, s_2) in the controls. Disease prevalence is denoted by $k = P(\text{disease})$ and penetrance by $f_i = P(\text{disease} \mid G_i)$ for $i = 0, 1, 2$. Two genotype relative risks (GRRs) are denoted by $\lambda_1 = f_1/f_0$ and $\lambda_2 = f_2/f_0$ using $f_0 > 0$ as baseline penetrance. Under the null hypothesis of no association $H_0 : f_0 = f_1 = f_2 = k$ or alternatively $H_0 : \lambda_2 = \lambda_1 = 1$. Genetic model is recessive (REC), additive (ADD), multiplicative (MUL), and dominant (DOM) when $\lambda_1 = 1$, $\lambda_1 = (1+\lambda_2)/2$, $\lambda_1 = \lambda_2^{1/2}$, and $\lambda_2 = \lambda_1$, respectively.

2.2. Review of Association Tests for a Single Data Set. The association in case-control studies can be tested using various methods which have been extensively studied. The general association between the disease status and the SNP can be tested using Pearson's Chi-square test which has an asymptotic Chi-square distribution with 2 degrees of freedom under H_0. The test is given by

$$T_{\text{chi2}} = \sum_{j=0}^{2} \frac{\left(r_j - n_j r/n\right)^2}{n_j r/n} + \sum_{j=0}^{2} \frac{\left(s_j - n_j s/n\right)^2}{n_j s/n}, \tag{1}$$

where $n_i = r_i + s_i$ for $i = 0, 1, 2$ and $n = r + s$. Under Hardy-Weinberg equilibrium (HWE), an allele-based test (ABT) and CATT with scores $(0, x, 1)$ for (G_0, G_1, G_2), where $0 \le x \le 1$, are given by

$$\begin{aligned} Z_{\text{ABT}} &= \frac{n^{1/2}\left\{2r\left(2s_0 + s_1\right) - 2s\left(2r_0 + r_1\right)\right\}}{\left\{2rs\left(2n_0 + n_1\right)\left(n_1 + 2n_2\right)\right\}^{1/2}}, \\ Z_x &= \frac{n^{1/2}\sum_{i=0}^{2} x_i\left(sr_i - rs_i\right)}{\left[rsn\left\{n\sum_{i=0}^{2} x_i^2 n_i - \left(\sum_{i=0}^{2} x_i n_i\right)^2\right\}\right]^{1/2}}, \end{aligned} \tag{2}$$

where $(x_0, x_1, x_2) = (0, x, 1)$ [9]. The optimal choices of x for the recessive (REC), additive/multiplicative (ADD/MUL), and dominant (DOM) models are $x = 0, 1/2$ and 1, respectively [9, 20]. Both Z_x and Z_{ABT} asymptotically follow a standard normal distribution under H_0. Z_x can be used even when HWE does not hold. However, without the HWE assumption, Z_{ABT} does not follow a standard normal distribution due to the correlation between two alleles.

A robust test, MAX3 proposed by Friedlin et al. [10], can be obtained by taking the maximum of three CATTs under the three genetic models as $\text{MAX3} = \max(|Z_0|, |Z_{1/2}|, |Z_1|)$. Parametric bootstrap or permutation methods can be used to find the p value of MAX3 [4].

Let the p values of Pearson's Chi-square test and CATT under the additive genetic model $Z_{1/2}$ be P_{chi2} and $P_{1/2}$, respectively. WTCCC [5] proposed an alternative robust test $\text{MIN2} = \min(P_{\text{chi2}}, P_{1/2})$. Joo et al. [15] derived the asymptotic null distribution of MIN2 and using their result the p value of MIN2 can be obtained as

$$\begin{aligned} P_{\text{MIN2}} &= \frac{1}{2}\exp\left\{-\frac{1}{2}H_1^{-1}\left(1 - \text{MIN2}\right)\right\} + \frac{1}{2}\text{MIN2} \\ &\quad - \frac{1}{2\pi}\int_{H_1^{-1}(1-\text{MIN2})}^{-2\log(\text{MIN2})} e^{-v/2}\arcsin\left(\frac{2H_1^{-1}\left(1 - \text{MIN2}\right)}{v - 1}\right)dv, \end{aligned} \tag{3}$$

where H_1 and H_2 are the cumulative distributions of Chi-square distributions with 1 and 2 degrees of freedom.

On the other hand, Song and Elston [21] considered a Hardy-Weinberg disequilibrium trend test (HWDTT) given by

$$Z_H = \frac{(rs/n)^{1/2}\left(\widehat{\Delta}_p - \widehat{\Delta}_q\right)}{\left\{1 - n_2/n - n_1/(2n)\right\}\left\{n_2/n + n_1/(2n)\right\}}, \tag{4}$$

where $\widehat{\Delta}_p = \widehat{p}_2 - (\widehat{p}_2 + \widehat{p}_1/2)^2$ and $\widehat{\Delta}_q = \widehat{q}_2 - (\widehat{q}_2 + \widehat{q}_1/2)^2$ are the estimates of Δ_p and Δ_q, where $\widehat{p}_i = r_i/r$ and $\widehat{q}_i = s_i/s$. Here, Δ denotes the Hardy-Weinberg disequilibrium (HWD) coefficient defined by $Pr(BB) - \{Pr(AB)/2 + Pr(BB)\}^2$ and Δ_p and Δ_q denote the HWD coefficient in cases and controls, respectively.

Zheng and Ng [13] used the information contained in the signs of (Δ_p, Δ_q) to determine the genetic models in their two-phase method. Their two-phase statistic Z_{GMS} is given by $Z_{\text{GMS}} = Z_0$ if $Z_H > c$, Z_1 if $Z_H < -c$, and $Z_{1/2}$ otherwise, where $c = \Phi^{-1}(1 - \alpha_H)$ for $\alpha_H = 0.05$. The asymptotic

correlations between Z_H and three CATTs under HWE were derived and the significance level was adjusted accordingly to control the desired type I error. Based on the observation that this method assumes B is the risk allele, Joo et al. [14] studied the behavior of Z_{GMS} when either one of the alleles can be a risk allele. They chose the risk allele based on the sign of $Z_{1/2}$; that is, if $Z_{1/2} > 0$, B is the risk allele, and Z_0, $Z_{1/2}$, and Z_1 are chosen for REC, ADD, and DOM models, respectively. If $Z_{1/2} < 0$, the respective test statistics are chosen to be $-Z_1$, $-Z_{1/2}$, and $-Z_0$. They incorporate this property in defining the test statistic for genetic model selection (Z_{GMS}) and calculating the p value. Let $\Theta_0(z) = \{z : z > c\}$, $\Theta_{1/2}(z) = \{z : |z| < c\}$, and $\Theta_1(z) = \{z : z < -c\}$. Then, the p value of this method can be obtained by

$$
\begin{aligned}
&p_{\mathrm{GMS}} \\
&= 2\left\{\sum_{x=0}^{1}\int_{\Theta_x(z_H)}\int_0^{\infty}\int_t^{\infty}\phi_x\left(z_x, z_{1/2}, z_H\right)dz_x dz_{1/2}dz_H\right\} \\
&\quad + 2\left\{\int_{\Theta_{1/2}(z)}\Phi\left(\frac{(-t\wedge 0)+\rho_{1/2}z}{\left(1-\rho_{1/2}^2\right)^{1/2}}\right)d\Phi\left(z\right)\right\},
\end{aligned}
\tag{5}
$$

where $\rho_x = \mathrm{Corr}(Z_x, Z_H)$ in (5) and $\rho_{x,1/2} = \mathrm{Corr}(Z_x, Z_{1/2})$ ($x = 0, 1$) are replaced by their consistent estimates. Here, $t = z_{\mathrm{GMS}}$ and $(-t \wedge 0) = \min(-t, 0)$. Moreover, z_{GMS} and $z_{1/2}$ are the observed values of Z_{GMS} and $Z_{1/2}$, respectively.

While studying the properties of GMS, Joo et al. [14] noticed that the probability of selecting the true recessive or dominant models using Z_H is very low especially for low to moderate GRRs, but the unlikely genetic model can be successfully excluded. This led to genetic model exclusion method Z_{GME} which is the same as the Z_{GMS} described above except Z_x for $x = 0, 1/2, 1$ is replaced by Z_x^* where $Z_x^* = (Z_x + Z_{1/2})/\{2(1+\widehat{\rho}_{x,1/2})\}^{1/2}$. And the p value of GME can be obtained as

$$
\begin{aligned}
&p_{\mathrm{GME}} \\
&= 2\left\{\sum_{x=0}^{1}\int_{\Theta_x(z_H)}\int_0^{\infty}\int_L^{\infty}\phi_x\left(z_x, z_{1/2}, z_H\right)dz_x dz_{1/2}dz_H\right\} \\
&\quad + 2\left\{\int_{\Theta_{1/2}(z)}\Phi\left(\frac{(-t\wedge 0)+\rho_{1/2}z}{\left(1-\rho_{1/2}^2\right)^{1/2}}\right)d\Phi\left(z\right)\right\},
\end{aligned}
\tag{6}
$$

where $L = t\{2(1+\widehat{\rho}_{x,1/2})\}^{1/2} - z_{1/2}$ for $t = z_{\mathrm{GME}}$.

2.3. p Value of Replication Data Using the Robust Method. In the discovery stage, the p value of robust association tests, including MAX3, MIN2, Z_{GMS}, and Z_{GME}, can be obtained as described in Section 2.2. For the p value of replication data using the robust method, we use the same analytic method that was used for discovery and the risk allele identified by it [16]. This means that when the best test statistic or genetic model is selected in the discovery stage, the replication stage will adopt the discovery stage selection and the direction of association.

Suppose that, for simplicity of notation, our interest is in GWAS with two stages, one for discovery and the other for replication, although the methodology described below can be extended to multistages for replication. Let $Z_x^{(i)}$ for $x = 0, 1/2, 1$ be the CATT optimal for recessive, additive, and dominant models and let $P_x^{(i)}$ be corresponding p value for ith stage ($i = 1$ for discovery and $i = 2$ for replication stages). Also, denote $Z_x^{(i)*} = (Z_x^{(i)} + Z_{1/2}^{(i)})/\{2(1+\widehat{\rho}_{x,1/2}^{(i)})\}^{1/2}$ for $x = 0, 1/2, 1$. Then, for CATT with a preselected genetic model, $P_x^{(2)} = 1 - \Phi(\mathrm{sign}(Z_x^{(1)} \cdot Z_x^{(2)}) \cdot |Z_x^{(2)}|)$ using a one-sided p value given the direction of association from the discovery stage, and $P_x^{(2)*} = 1 - \Phi(\mathrm{sign}(Z_x^{(1)*} \cdot Z_x^{(2)*}) \cdot |Z_x^{(2)*}|)$. Moreover, denote the test statistics and p values using Pearson's Chi-square test from the ith stage as $T_{\mathrm{chi2}}^{(i)}$ and $P_{\mathrm{chi2}}^{(i)}$. Further, let HWDTT from the ith stage be $Z_H^{(i)}$. Then, the second stage p values, using MAX3, MIN2, Z_{GMS}, and Z_{GME}, denoted as $P_{\mathrm{MAX3}}^{(2)}$, $P_{\mathrm{MIN2}}^{(2)}$, $P_{\mathrm{GMS}}^{(2)}$, and $P_{\mathrm{GME}}^{(2)}$, can be obtained as follows:

$$
\begin{aligned}
P_{\mathrm{MAX3}}^{(2)} &= P_{x^*}^{(2)}, \quad \text{where } x^* = \arg\min_{x\in\{0,1/2,1\}} P_x^{(1)}, \\
P_{\mathrm{MIN2}}^{(2)} &= P_{1/2}^{(2)} \cdot I\left(P_{1/2}^{(1)} \le P_{\mathrm{chi2}}^{(2)}\right) + P_{\mathrm{chi2}}^{(2)} \cdot I\left(P_{1/2}^{(1)} > P_{\mathrm{chi2}}^{(2)}\right), \\
P_{\mathrm{GMS}}^{(2)} &= P_0^{(2)} I\left(Z_H^{(1)} > c\right) + P_1^{(2)} I\left(Z_H^{(1)} < -c\right) \\
&\quad + P_{1/2}^{(2)} I\left(\left|Z_H^{(1)}\right| \le c\right), \quad \text{if } Z_{1/2}^{(1)} > 0; \\
&= P_0^{(2)} I\left(Z_H^{(1)} < -c\right) + P_1^{(2)} I\left(Z_H^{(1)} > c\right) \\
&\quad + P_{1/2}^{(2)} I\left(\left|Z_H^{(1)}\right| \le c\right), \quad \text{if } Z_{1/2}^{(1)} \le 0, \\
P_{\mathrm{GME}}^{(2)} &= P_0^{(2)*} I\left(Z_H^{(1)} > c\right) + P_1^{(2)*} I\left(Z_H^{(1)} < -c\right) \\
&\quad + P_{1/2}^{(2)*} I\left(\left|Z_H^{(1)}\right| \le c\right), \quad \text{if } Z_{1/2}^{(1)} > 0; \\
&= P_0^{(2)*} I\left(Z_H^{(1)} < -c\right) + P_1^{(2)*} I\left(Z_H^{(1)} > c\right) \\
&\quad + P_{1/2}^{(2)*} I\left(\left|Z_H^{(1)}\right| \le c\right), \quad \text{if } Z_{1/2}^{(1)} \le 0.
\end{aligned}
\tag{7}
$$

It is important to note that even though the direction of the test statistics and the selected genetic models are used to obtain the second stage p values, the p values from the two stages are independent under the null hypothesis. This is because, under the null hypothesis, the probability of $Z_{1/2}$ being positive or negative is simply 1/2, and the probability of the selection of a certain genetic model is also a constant (α_H for the recessive and dominant models and $1 - 2\alpha_H$ for the additive model).

2.4. Combined Test Using p Values and Its Statistical Significance. For a given robust test, we can consider the joint analysis by combining p values from the discovery and replication stages of GWAS. We consider using p values rather than the test statistics because test statistics can have

Table 1: Type I error rates of three approaches—replication-based (REP) test, Fisher's combination (Z_{FC}), and linear combination of test (Z_{LC})—based on the CATT with an additive model ($Z_{1/2}$), χ^2, MAX3, MIN2, GMS, and GME. The disease prevalence $K = 0.1$, $M = 10$ markers, $r = 1,500$ cases, and $s = 1,500$ controls are considered based on 20,000 simulations.

MAF	π_s	α_D		$F = 0$					
				$Z_{1/2}$	χ^2	MAX3	MIN2	GMS	GME
0.3	0.5	0.05	REP	0.00530	0.00455	0.00505	0.00485	0.0050	0.00490
			Z_{FC}	0.00500	0.00535	0.00495	0.00510	0.00515	0.00460
			Z_{LC}	0.00535	0.00525	0.00485	0.00510	0.0050	0.00485
		0.1	REP	0.00510	0.00560	0.00565	0.00485	0.00525	0.00545
			Z_{FC}	0.00565	0.00535	0.00565	0.00545	0.00565	0.00540
			Z_{LC}	0.00520	0.00565	0.00525	0.00520	0.00530	0.00525
0.3	0.6	0.05	REP	0.00510	0.00485	0.00480	0.00515	0.00480	0.00500
			Z_{FC}	0.00445	0.00455	0.00450	0.00455	0.00450	0.00460
			LC	0.00500	0.00515	0.00495	0.00520	0.00475	0.00480
		0.1	REP	0.00500	0.00485	0.00485	0.00535	0.00530	0.00505
			Z_{FC}	0.00465	0.00490	0.00485	0.00500	0.00455	0.00460
			Z_{LC}	0.00480	0.00470	0.00515	0.00490	0.00485	0.00475
0.4	0.5	0.05	REP	0.00590	0.00505	0.00530	0.00565	0.00505	0.00510
			Z_{FC}	0.00575	0.00430	0.00460	0.00535	0.00460	0.00500
			Z_{LC}	0.00600	0.00445	0.00500	0.00540	0.00490	0.00490
		0.1	REP	0.00525	0.00470	0.00535	0.00450	0.00480	0.00515
			Z_{FC}	0.00515	0.00510	0.00495	0.00475	0.00540	0.00500
			Z_{LC}	0.00530	0.00500	0.00485	0.00475	0.00495	0.00510
0.4	0.6	0.05	REP	0.00475	0.00585	0.00480	0.00500	0.00515	0.00495
			Z_{FC}	0.00460	0.00470	0.00420	0.00490	0.00455	0.00440
			Z_{LC}	0.00525	0.00550	0.00520	0.00580	0.00510	0.00510
		0.1	REP	0.00550	0.00490	0.00515	0.00535	0.00555	0.00540
			Z_{FC}	0.00520	0.00370	0.00495	0.00450	0.00515	0.00510
			Z_{LC}	0.00565	0.00485	0.00570	0.00530	0.00610	0.00580

complex forms and obtaining the distribution of the joint test can be difficult. On the other hand, calculating a p value for each data set might be relatively simple, and the distribution of p values under the null hypothesis of no association is easy to handle.

There are several methods for combining test statistics from two stages [22], and two most commonly used forms are based on Fisher's combination and a linear combination after inverse normal transformation [23]. Fisher's combination (FC) directly sums p values after $-2\log$ transformation; that is, $Z_{FC} = -2w_1 \log(P^{(1)}) - 2w_2 \log(P^{(2)})$, where $P^{(i)}$ is p value from $i = 0$ for discovery and $i = 1$ for replication stages using a given robust test. A specification of $w_1 = w_2 = 1$ gives the same weight for discovery and replication stages, and one can consider $w_1 = 2\pi_s$ and $w_2 = 2(1 - \pi_s)$ where $\pi_s = N_D/(N_D + N_R)$, and N_D and N_R are sample sizes of the discovery and replication data sets. A linear combination of two P values after taking the inverse of the standard normal cumulative distribution is given by $Z_{LC} = \{w_1\Phi^{-1}(1-P^{(1)}/2)+w_2\Phi^{-1}(1-P^{(2)})\}/\sqrt{w_1^2 + w_2^2}$ with a natural choice of $w_1 = \sqrt{\pi_s}$ and $w_2 = \sqrt{1 - \pi_s}$. Let the significance level of the discovery stage be α_D, which means that markers with $P^{(1)} < \alpha_D$ are selected and replicated in the replication stage. The p value of combined test can then be obtained as $p_{FC} = P_{H_0}(P^{(1)} < \alpha_D, Z_{FC} > z_{FC})$ where the observed value of Z_{FC} is z_{FC}. The p_{FC} are calculated as $e^{-z_{FC}/2}(1+z_{FC}/2+\log\alpha_D)$ for equal weights where $z_{FC} > -2\log\alpha_D$ and $(w_1/(w_1 - w_2))e^{-z_{FC}/2w_1} - (w_2/(w_1-w_2))e^{-z_{FC}/2w_2}\alpha_d^{-(w_1-w_2)/w_2}$ for unequal weights where $z_{FC} > -2w_1\log\alpha_D$. Detailed derivations are described in the Appendix. Equivalently, for an overall type I error threshold for a single marker of α, one may obtain the threshold C_{FC} of Z_{FC} that satisfies $P_{H_0}(P^{(1)} < \alpha_D, Z_{FC} > C_{FC}) \leq \alpha$. Similarly, for the Z_{LC}, the p value is calculated as $p_{LC} = P_{H_0}(P^{(1)} < \alpha_D, Z_{LC} > z_{LC}) = \int_{z_{1-\alpha_D/2}}^{\infty} \phi(z)[1 - \Phi((\sqrt{w_1^2 + w_2^2}z_{LC} - w_1 z)/w_2)]dz$ for $z_{LC} > z_{1-\alpha_D/2}$ where the observed value of $Z_{LC} = z_{LC}$.

3. Simulation Results

3.1. Type I Error. Table 1 provides the type I errors under different scenarios. A disease prevalence of 10% is assumed, and a total of 1500 cases and 1500 controls were divided into two stages. The proportions of samples in the first stage (π_s) of 0.5 and 0.6 were considered for the minor allele frequency (MAF) of 0.3 and 0.4. We considered $M = 10$ markers to control the genome-wide false positive rate at $\alpha = 0.05$ with the Bonferroni correction. We did not consider a larger M

such as 300,000 or 500,000 because this will require more than 10 million simulations to show a stable estimate of the type I error rate. With $M = 10$, we performed 20,000 simulations which result in less than 10% of a coefficient of variation for a significance level $0.05/M = 0.005$ for each marker [24]. The test statistics considered are $Z_{1/2}$, Pearson's Chi-square test, MIN2, MAX3, GMS, and GME. For the second stage analysis, we considered a replication-based analysis, Z_{FC}, and Z_{LC} as proposed above. The results are based on the situation under HWE (HWE coefficient $F = 0$). As expected, all tests control the type I error reasonablly well, and similar results were obtained when a slight deviation from HWE is present with $F = 0.05$ (results not shown).

3.2. Empirical Power. We examined the empirical powers of different tests considered above. In Figure 1, we considered $M = 10$ markers, a disease prevalence of 10%, the same genotype relative risk for two stages ($r_1 = 1.4$ and $r_2 = 1.4$), and 1,000 cases and 1,000 controls. 2,000 simulations were performed under HWE ($F = 0$) to control the genome-wide false positive rate at $\alpha = 0.05$. The recessive, additive, and dominant models were assumed for the first, second, and third rows. Both joint analyses showed better power performances compared to the replication-based analysis (up to 15.9% in scenarios considered in Figure 1), and LC and FC have comparable powers with less than 2% difference. The power gain of using the joint analysis is not as much as that observed in Skol et al. [18]. However, as reported by Skol et al. [18], when the between-stage heterogeneity exists and the risk allele has a larger effect in the first stage than that in the second stage, much improved power is observed by using the joint test. Figure 2 shows results under this scenario with $r_1 = 1.6$ and $r_2 = 1.4$, and the observed increase in power using the joint test is as high as 33.9%. Again, the difference between LC and FC is minor with less than 3% difference. As for comparison between different robust methods, MAX3, GMS, and GME perform well under the recessive model, while $Z_{1/2}$, χ^2, and MIN2 are less powerful. Under the additive model, $Z_{1/2}$ is most powerful, as expected, and χ^2 is least powerful. Other robust methods perform well with a slight decrease in power compared to $Z_{1/2}$. Under the dominant model, MAX3, GMS, and GME perform the best even though all tests show good power performances, and the difference is minor. Similar patterns were observed when a slight deviation from the HWE is present (results not shown).

4. Real Data Application

The GWAS on non-small-cell lung cancer (NSCLC) by Yoon et al. [25] studied 621 NSCLC patients and 1541 control subjects in the discovery stage. After stringent quality control steps, a total of 246,758 SNPs were tested for the association with NSCLC based on $Z_{1/2}$. In the replication stage, 168 SNPs with p value less than 1×10^{-4} in the first stage based on $Z_{1/2}$ were tested using 804 patients and 1470 control samples. We identified additional 234 SNPs using MIN2 in the first stage which could be studied in the replication stage if MIN2 was used instead of $Z_{1/2}$ since MIN2 produces stronger evidence for the additional SNPs than $Z_{1/2}$ does. The Manhattan plots of using MIN2 and $Z_{1/2}$ are presented in Figure 3. One example is *rs*385272 located in chromosome 2, which had a p value of 1.37×10^{-7} which reached significance level at Bonferroni correction in discovery samples alone, whereas $Z_{1/2}$ yielded a p value greater than 1×10^{-4}. Even though there is possibility of false positive findings, these SNPs could have been selected for replication if robust methods were used.

Since we do not have replication data for these additional SNPs selected using MIN2 because the first stage selection was based on $Z_{1/2}$ in Yoon et al. [25], just for illustration purpose of the proposed methods, we present the results of three SNPs including *rs*2131877 that was reported by Yoon et al. [25]. When the significance level in the discovery stage is set at $\alpha_D = 5 \times 10^{-5}$ so that all these exemplary SNPs can be selected in the discovery stage; the p value of combined test based on four robust methods (MAX3, MIN2, GMS, and GME) as well as $Z_{1/2}$ and Pearson's Chi-square test is presented in Table 2. Fisher's combination was used for the joint test in the second stage. Only *rs*2131877 was found to be significant with Bonferroni correction (p value $< 2.03 \times 10^{-7}$) by all except MAX3 method.

5. Discussion

In genetic association studies, efficiency robust tests whose performance does not depend on the underlying genetic model have been extensively studied, and their power benefit over a wide range of genetic models has been well recognized. In this paper, we described how the idea of these robust association tests can be applied to the replication studies and further how overall statistical significance can be evaluated using the combined test formed by p values of the discovery and replication studies.

When the robust tests are used, the test statistic of each stage can have a complex form and thus dealing with the distribution of the joint test can be difficult, whereas calculating the p value of each stage might be relatively simple. Because the asymptotic distribution of the p value under the null hypothesis of no association is easy to handle, the combined test using p values rather than the test statistics themselves can provide computational convenience.

There are several methods for combining test statistics from two stages and Won et al. [22] compared the performances of various choices. Two most commonly used forms are based on Fisher's combination and the linear combination after the inverse normal transformation [23], and we presented the test statistics and p values of these two methods. In our limited experience, the linear combination and Fisher's combination are fairly comparable. Fisher's combination seems to perform slightly better than the linear combination when there exists some heterogeneity between stages in terms of the genotype relative risk, while the linear combination seems to perform slightly better in most of other situations. However, the difference is extremely minor. Further research is required for the thorough comparison of various methods of combining p values in the application of efficiency robust tests to the replication of genetic association studies.

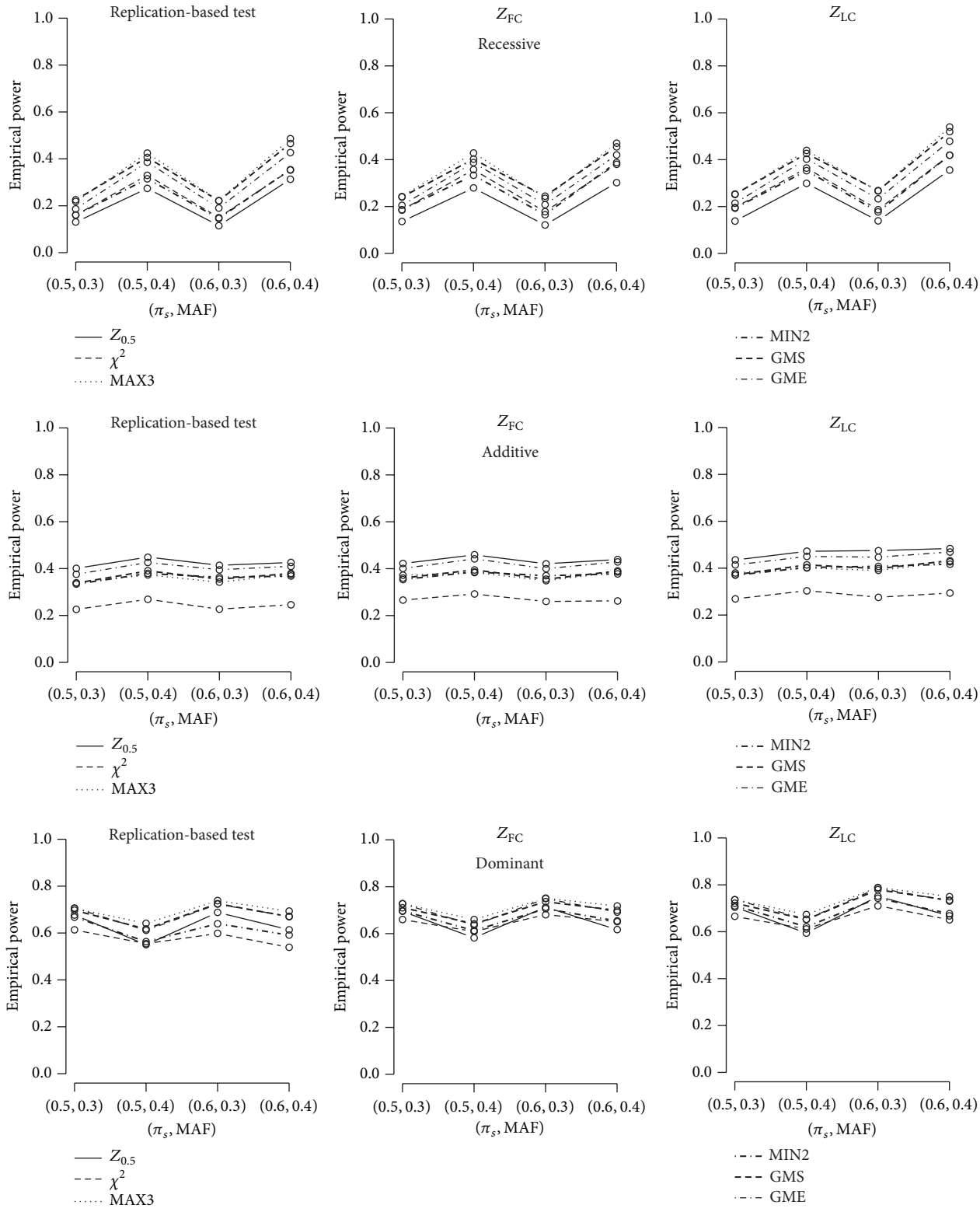

Figure 1: Empirical powers based on 2,000 simulations for $M = 10$ markers, genotype relative risks of both stages = 1.4, and disease prevalence $K = 0.1$ under the recessive, additive, and dominant models. 1,000 cases and 1,000 controls are considered to control $\alpha = 0.05$. The first stage type I error rate for discovery is $\alpha_D = 0.05$. Six test statistics, $Z_{1/2}$, χ^2, MAX3, MIN2, GMS, and GME, are considered. The first, second, and third columns depict powers using the replication-based test, Z_{FC}, and Z_{LC}, respectively.

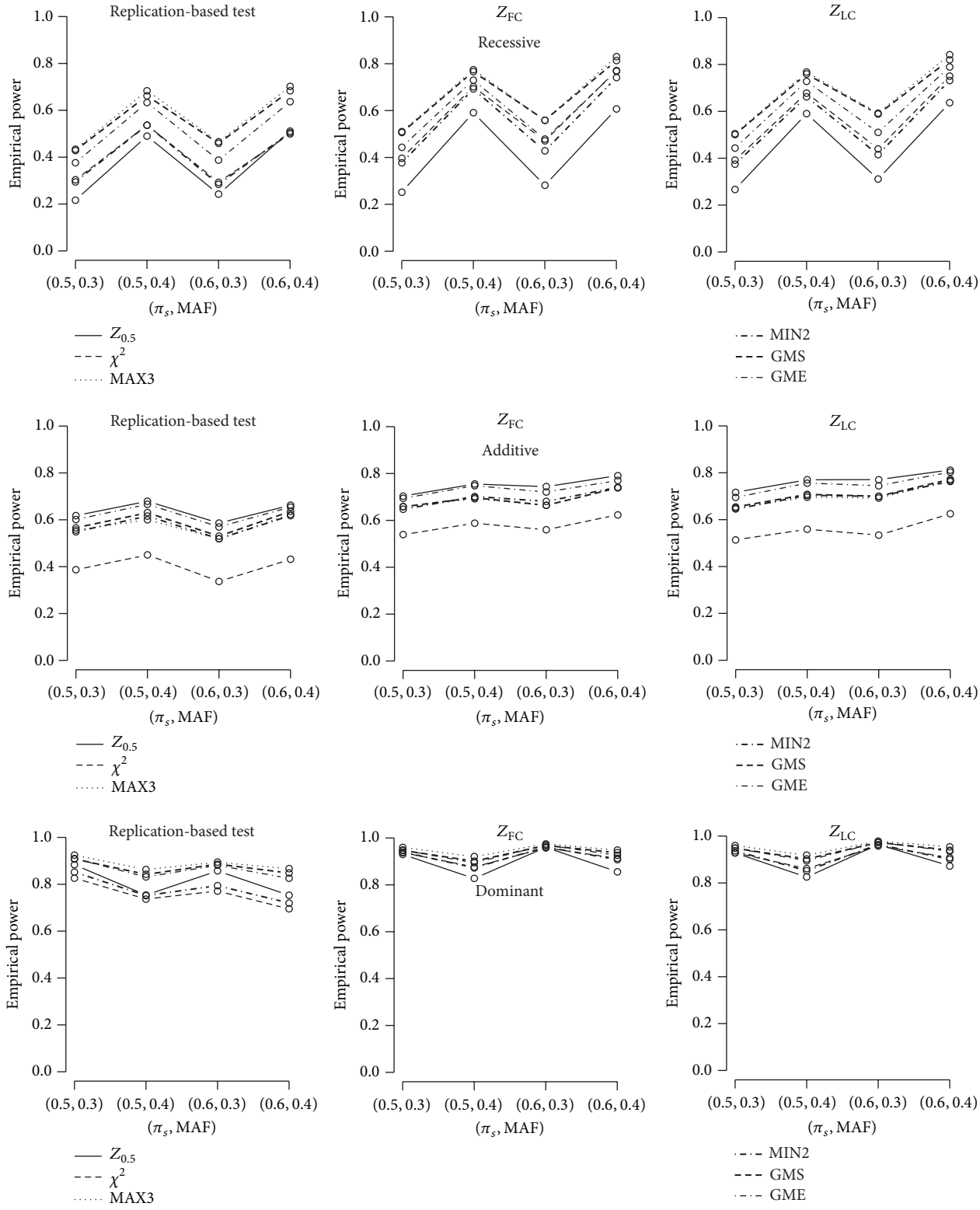

FIGURE 2: Empirical powers based on 2,000 simulations for $M = 10$ markers; genotype relative risks of two stages are different ($r_1 = 1.6$, $r_2 = 1.4$); disease prevalence $K = 0.1$ under the recessive, additive, and dominant models. 1,000 cases and 1,000 controls are considered to control $\alpha = 0.05$. The first stage type I error rate for discovery is $\alpha_D = 0.05$. Six test statistics, $Z_{1/2}$, χ^2, MAX3, MIN2, GMS, and GME, are considered. The first, second, and third columns depict powers using the replication-based test, Z_{FC}, and Z_{LC}, respectively.

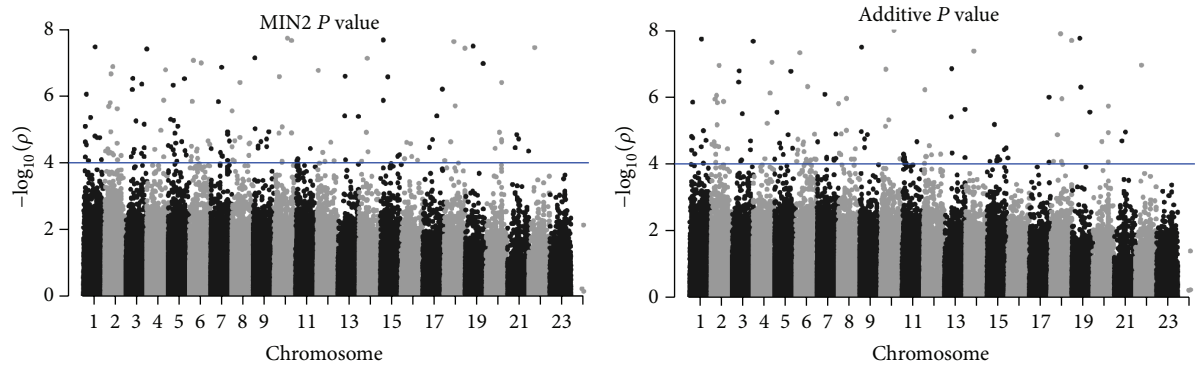

Figure 3: Manhattan plots of 246,758 SNPs from Yoon et al. [25] based on MIN2 (a) and $Z_{1/2}$ (b). The x axis is chromosomal location and the y axis is the significance ($-\log_{10}P$) of association. The horizontal line corresponds to the significance level 10^{-4}.

Table 2: For selected exemplary three SNPs for testing association with NSCLC, p value of combined test using additive CATT ($Z_{1/2}$), Pearson's Chi-square test (T_{chi2}), MAX3, MIN2, Z_{GMS}, and Z_{GME}.

SNP	p value of $Z_{1/2}$			p value of T_{chi2}		
	Discovery	Replication	Combined test	Discovery	Replication	Combined test
rs2131877	7.88×10^{-5}	1.04×10^{-4}	7.97×10^{-8}	1.40×10^{-4}	1.49×10^{-4}	1.84×10^{-7}
rs905551	1.83×10^{-5}	7.02×10^{-3}	7.70×10^{-6}	8.06×10^{-5}	4.89×10^{-2}	1.40×10^{-5}
rs1695109	2.48×10^{-4}	3.46×10^{-2}	2.17×10^{-6}	4.56×10^{-5}	1.53×10^{-1}	2.07×10^{-5}
SNP	p value of MAX3			p value of MIN2		
	Discovery	Replication	Combined test	Discovery	Replication	Combined test
rs2131877	1.53×10^{-4}	4.05×10^{-2}	1.92×10^{-5}	1.32×10^{-4}	1.04×10^{-4}	1.26×10^{-7}
rs905551	4.50×10^{-5}	7.02×10^{-3}	1.92×10^{-6}	1.34×10^{-4}	4.89×10^{-2}	1.99×10^{-5}
rs1695109	3.54×10^{-5}	2.63×10^{-2}	4.64×10^{-6}	2.36×10^{-5}	2.63×10^{-2}	3.35×10^{-6}
SNP	p value of Z_{GMS}			p value of Z_{GME}		
	Discovery	Replication	Combined test	Discovery	Replication	Combined test
rs2131877	1.86×10^{-4}	1.04×10^{-4}	1.71×10^{-7}	1.03×10^{-4}	1.04×10^{-4}	1.02×10^{-7}
rs905551	5.19×10^{-5}	7.02×10^{-3}	2.16×10^{-6}	7.35×10^{-5}	8.01×10^{-3}	3.20×10^{-6}
rs1695109	6.89×10^{-4}	1.27×10^{-1}	3.85×10^{-5}	2.69×10^{-5}	4.19×10^{-2}	5.40×10^{-6}

In a genetic study where the purpose of considering a replication stage is to validate or replicate the genetic findings from the discovery stage, which is the case considered in this paper, the analysis in the replication stage utilized the test statistic or genetic model that is selected as being the best in the discovery stage and also the direction of the risk allele, following guidelines for exact replication in genetic association studies. If the purpose is to simply combine the evidence from different data sources such as in meta-analysis, other strategies may be devised. Further research, again, is required to provide fully detailed properties of such methods.

Power gain of a joint analysis over the conventional replication-based analysis was thoroughly studied by Skol et al. [18, 19]. In our simulation, the amount of power increase using a joint test compared to the replication-based analysis was much minor than what was observed by Skol et al. [18, 19]. The exact reason is not known, but we suspect this might be due to the power advantages of robust methods and also due to the fact that the optimal choice from the first stage is used when calculating the second stage p values. However, even though it was minor in some situations, the joint anlysis presented better power performance than the replication-based analysis in our study. This type of joint analysis raised concerns about the exact meaning of replication [17]. However, McCarthy et al. [26] mentioned that joint analyses "blur the boundaries of where exactly replication starts, but whichever analytical approach is taken, confirmation in many independent samples is important and it is the overall strength of the evidence of association that matters." Purpose of the current study was to present how the overall strength of the evidence of association can be evaluated when robust tests are used in GWAS replication studies.

We illustrated how the proposed methods can be applied in the real data that studied the association of SNPs with non-small-cell lung cancer (NSCLC) in discovery and replication stages. In the original study reported by Yoon et al. [25], SNPs were selected in the discovery data set not based on the robust tests but based on additive CATT. Therefore, we found that some SNPs could have been selected by one

of the robust methods but they were not included in the replication data set. For these SNPs, we were not able to perform the joint analysis that we propose, and it was not possible to examine whether there are other SNPs that could have been found to be associated with NSCLC by proposed methods in the replication study. For this reason, we merely presented how many additional SNPs could have been further followed in the replication stage when robust methods were used. In many GWASs, it is a common practice to report the summary test statistics and p values of the SNPs under a specific genetic model, usually an additive model, which were further genotyped in the replication stage and were finally defined to be significantly associated with a phenotype of interest. As emphasized in this paper, one may have a better chance of finding many missing SNPs by applying more powerful and robust methods that consider different genetic models simultaneously. Therefore, we urge the community to share test results under not only an additive model but also other genetic models, although they were not significant at a stringent significance level, so that future research may have enriched data resources, to which robust tests can be applied in association studies.

Appendix

p value of Fisher's Combination for Equal and Unequal Weights

Equal Weights. Assume $w_1 = w_2 = 1$. Under the null hypothesis of no association, $X_1 = -2\log P^{(1)}$ and $X_2 = -2\log P^{(2)}$ are independent and each asymptotically follows a χ^2 distribution with 2 degrees of freedom Let $f_k(x)$ and $F_k(x)$ be the probability and cumulative density functions of χ^2 random variable with k degrees of freedom Then $f_2(x) = \exp(-x/2)/2$, $f_4(x) = x\exp(-x/2)/4$, $F_2(x) = 1-\exp(-x/2)$, and $F_4(x) = 1-\exp(-x/2)-x\exp(-x/2)/2$. Denote the cutoff of the discovery stage based on α_D as C_D; that is, $F_2(C_D) = 1-\alpha_D$. For observed value $z_{\mathrm{FC}} > C_D$ of $X_1 + X_2$, the p value is written as

$$\begin{aligned} &P_{H_0}\left(X_1 > C_D, X_1 + X_2 > z_{\mathrm{FC}}\right) \\ &= \alpha_D - \int_{C_D}^{z_{\mathrm{FC}}} f_2(x)\, F_2\left(z_{\mathrm{FC}} - x\right) dx \\ &= \alpha_D + \exp\left(-\frac{z_{\mathrm{FC}}}{2}\right) - \alpha_D + \frac{1}{2}\exp\left(-\frac{z_{\mathrm{FC}}}{2}\right)\left(z_{\mathrm{FC}} - C_D\right) \\ &= \exp\left(-\frac{z_{\mathrm{FC}}}{2}\right)\left(1 + \frac{z_{\mathrm{FC}}}{2} + \log\alpha_D\right). \end{aligned} \tag{A.1}$$

Unequal Weights. When different proportions of samples are used in the discovery and replication stages, it may be more appropriate to assign weights proportional to the sample sizes for each stage. For example, when only a small portion is used in the discovery stage, to prevent Fisher's combination test from being dominated by the significant result in the discovery stage, one may want to assign a small weight to the discovery stage result.

When π_s is the proportion of samples used in the discovery stage, one selection for weights is $w_1 = 2\pi_s$ and $w_2 = 2(1-\pi_s)$ for discovery and replication stages, which simplifies to equal weights when $\pi_s = 0.5$. Based on these weights, we consider unequal-weighted Fisher's combination as $-2\log P^{(1)w_1}P^{(2)w_2} = w_1X_1 + w_2X_2$ [27]. Its density function is given by

$$\begin{aligned} f_w(x) = &\frac{1}{2\left(w_1 - w_2\right)}\exp\left(-\frac{x}{\left(2w_1\right)}\right) \\ &- \frac{1}{2\left(w_1 - w_2\right)}\exp\left(-\frac{x}{\left(2w_2\right)}\right), \end{aligned} \tag{A.2}$$

and the probability distribution function is

$$\begin{aligned} F_w(x) = 1 - &\left\{\frac{w_1}{\left(w_1 - w_2\right)}\exp\left(-\frac{x}{2w_1}\right)\right. \\ &\left. - \frac{w_2}{\left(w_1 - w_2\right)}\exp\left(-\frac{x}{2w_2}\right)\right\}, \quad w_1 \neq w_2. \end{aligned} \tag{A.3}$$

Using the previous notation, we have the following form of p value:

$$\begin{aligned} &P_{H_0}\left(X_1 > C_D, w_1X_1 + w_2X_2 > z_{\mathrm{FC}}\right) \\ &= \alpha_D - \int_{C_D}^{z_{\mathrm{FC}}/w_1} f_2(x)\, F_2\left(\frac{z_{\mathrm{FC}} - w_1x}{w_2}\right) dx \\ &= \exp\left(-\frac{z_{\mathrm{FC}}}{2w_1}\right) + \frac{w_2}{w_1 - w_2}\exp\left(-\frac{z_{\mathrm{FC}}}{2w_2}\right) \\ &\quad \times \left\{\exp\left(\frac{w_1 - w_2}{2w_1w_2}z_{\mathrm{FC}}\right) - \exp\left(\frac{w_1 - w_2}{2w_1w_2}w_1C_D\right)\right\} \\ &= \frac{w_1}{w_1 - w_2}\exp\left(-\frac{z_{\mathrm{FC}}}{2w_1}\right) \\ &\quad - \frac{w_2}{w_1 - w_2}\exp\left(-\frac{z_{\mathrm{FC}}}{2w_2}\right)\alpha_D^{-(w_1 - w_2)/w_2} \\ &= \frac{w_1}{w_1 - w_2}\exp\left(-\frac{z_{\mathrm{FC}}}{2w_1}\right) \\ &\quad - \frac{w_2}{w_1 - w_2}\exp\left(-\frac{z_{\mathrm{FC}}}{2w_2}\right)\alpha_D^{-(w_1 - w_2)/w_2}, \end{aligned} \tag{A.4}$$

where $z_{\mathrm{FC}}/w_1 > C_D$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are indebted to late Dr. Gang Zheng for his inspiration and support on their work. This work was supported by grant of the National Cancer Center (no. NCC-1210060).

References

[1] R. J. Klein, C. Zeiss, E. Y. Chew et al., "Complement factor H polymorphism in age-related macular degeneration," *Science*, vol. 308, no. 5720, pp. 385–389, 2005.

[2] R. H. Duerr, K. D. Taylor, S. R. Brant et al., "A genome-wide association study identifies *IL 23R* as an inflammatory bowel disease gene," *Science*, vol. 314, no. 5804, pp. 1461–1463, 2006.

[3] A. Herbert, N. P. Gerry, M. B. McQueen et al., "A common genetic variant is associated with adult and childhood obesity," *Science*, vol. 312, no. 5771, pp. 279–283, 2006.

[4] R. Sladek, G. Rocheleau, J. Rung et al., "A genome-wide association study identifies novel risk loci for type 2 diabetes," *Nature*, vol. 445, no. 7130, pp. 881–885, 2007.

[5] P. R. Burton, D. G. Clayton, L. R. Cardon et al., "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, no. 7145, pp. 661–678, 2007.

[6] S. H. Kwak, S. H. Kim, Y. M. Cho et al., "A genome-wide association study of gestational diabetes mellitus in Korean women," *Diabetes*, vol. 61, no. 2, pp. 531–541, 2012.

[7] W. G. Cochran, "Some methods for strengthening the common χ^2 tests," *Biometrics*, vol. 10, no. 4, pp. 417–451, 1954.

[8] P. Armitage, "Tests for linear trends in proportions and frequencies," *Biometrics*, vol. 11, no. 3, pp. 375–386, 1955.

[9] P. D. Sasieni, "From genotypes to genes: doubling the sample size," *Biometrics*, vol. 53, no. 4, pp. 1253–1261, 1997.

[10] B. Freidlin, G. Zheng, Z. Li, and J. L. Gastwirth, "Trend tests for case-control studies of genetic markers: power, sample size and robustness," *Human Heredity*, vol. 53, no. 3, pp. 146–152, 2002.

[11] J. L. Gastwirth, "On robust procedures," *Journal of the American Statistical Association*, vol. 61, no. 316, pp. 929–948, 1966.

[12] J. L. Gastwirth, "The use of maximin efficiency robust tests in combining contingency tables and survival analysis," *Journal of the American Statistical Association*, vol. 80, no. 390, pp. 380–384, 1985.

[13] G. Zheng and H. K. T. Ng, "Genetic model selection in two-phase analysis for case-control association studies," *Biostatistics*, vol. 9, no. 3, pp. 391–399, 2008.

[14] J. Joo, M. Kwak, and G. Zheng, "Improving power for testing genetic association in case-control studies by reducing the alternative space," *Biometrics*, vol. 66, no. 1, pp. 266–276, 2010.

[15] J. Joo, M. Kwak, K. Ahn, and G. Zheng, "A robust genome-wide scan statistic of the Wellcome Trust Case-Control Consortium," *Biometrics*, vol. 65, no. 4, pp. 1115–1122, 2009.

[16] P. Kraft, E. Zeggini, and J. P. A. Ioannidis, "Replication in genome-wide association studies," *Statistical Science*, vol. 24, no. 4, pp. 561–573, 2009.

[17] D. C. Thomas, G. Casey, D. V. Conti, R. W. Haile, J. P. Lewinger, and D. O. Stram, "Methodological issues in multistage genome-wide association studies," *Statistical Science*, vol. 24, no. 4, pp. 414–429, 2009.

[18] A. D. Skol, L. J. Scott, G. R. Abecasis, and M. Boehnke, "Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies," *Nature Genetics*, vol. 38, no. 2, pp. 209–213, 2006.

[19] A. D. Skol, L. J. Scott, G. R. Abecasis, and M. Boehnke, "Optimal designs for two-stage genome-wide association studies," *Genetic Epidemiology*, vol. 31, no. 7, pp. 776–788, 2007.

[20] G. Zheng, B. Freidlin, Z. Li, and J. L. Gastwirth, "Choice of scores in trend tests for case-control studies of candidate-gene associations," *Biometrical Journal*, vol. 45, no. 3, pp. 335–348, 2003.

[21] K. Song and R. C. Elston, "A powerful method of combining measures of association and Hardy-Weinberg Disequilibrium for fine-mapping in case-control studies," *Statistics in Medicine*, vol. 25, no. 1, pp. 105–126, 2006.

[22] S. Won, N. Morris, Q. Lu, and R. C. Elston, "Choosing an optimal method to combine *P*-values," *Statistics in Medicine*, vol. 28, no. 11, pp. 1537–1553, 2009.

[23] F. Begum, D. Ghosh, G. C. Tseng, and E. Feingold, "Comprehensive literature review and statistical considerations for GWAS meta-analysis," *Nucleic Acids Research*, vol. 40, no. 9, pp. 3777–3784, 2012.

[24] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall/CRC Press, 1993.

[25] K. A. Yoon, J. H. Park, J. Han et al., "A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population," *Human Molecular Genetics*, vol. 19, no. 24, pp. 4948–4954, 2010.

[26] M. I. McCarthy, G. R. Abecasis, L. R. Cardon et al., "Genome-wide association studies for complex traits: consensus, uncertainty and challenges," *Nature Reviews Genetics*, vol. 9, no. 5, pp. 356–369, 2008.

[27] I. J. Good, "On the weighted combination of significance tests," *Journal of the Royal Statistical Society, Series B*, vol. 17, pp. 264–265, 1955.