RESEARCH ARTICLE

Quantitative Trait Loci Identification by Estimating the Genetic Model based on the Extremal Samples

Zining Yang[1], Yaning Yang[1], Xu Steven Xu[2] and Min Yuan[3,*]

[1]*Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, China;* [2]*Genmab US, Inc, Princeton, NJ 08540, USA;* [3]*Center for Data Science in Health, School of Public Health Administration, Anhui Medical University, Hefei 230032, China*

Abstract: ***Background*:** In genetic association studies with quantitative trait loci (QTL), the association between a candidate genetic marker and the trait of interest is commonly examined by the omnibus F test or by the t-test corresponding to a given genetic model or mode of inheritance. It is known that the t-test with a correct model specification is more powerful than the F test. However, since the underlying genetic model is rarely known in practice, the use of a model-specific t-test may incur substantial power loss. Robust-efficient tests, such as the Maximin Efficiency Robust Test (MERT) and MAX3 have been proposed in the literature.

***Methods*:** In this paper, we propose a novel two-step robust-efficient approach, namely, the genetic model selection (GMS) method for quantitative trait analysis. GMS selects a genetic model by testing Hardy-Weinberg disequilibrium (HWD) with extremal samples of the population in the first step and then applies the corresponding genetic model-specific t-test in the second step.

***Results*:** Simulations show that GMS is not only more efficient than MERT and MAX3, but also has comparable power to the optimal t-test when the genetic model is known.

***Conclusion*:** Application to the data from Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort demonstrates that the proposed approach can identify meaningful biological SNPs on chromosome 19.

1. INTRODUCTION

The clinical importance and health relevance of many quantitative traits, such as diabetes, obesity, cholesterol level and blood pressure, has been recognized. Although quantitative trait data is widely used in genetic association studies [1, 2], there are relatively few studies on robust-efficient methods in this area. For genetic association studies with a quantitative trait locus (QTL), a linear regression model can be used to link the genotypic value and the phenotypic value and an omnibus F-test is applied to test the null hypothesis that the QTL is unliked. When the genetic model (or mode of inheritance) is known, the genotypic value can be coded according to the underlying genetic model and effectively, the number of interesting parameters is reduced. This results

in a t-test for the null hypothesis, which is more powerful than the omnibus F-test due to the inclusion of extra information on the mode of inheritance. However, it is also widely known that misspecification of genetic model may lead to substantial power loss of the t-test. Therefore, powerful tests that are robust to model misspecification, referred to as robust-efficient tests, are desired.

Robust-efficient procedures have been well studied for qualitative trait analysis in case-control design. Common robust-efficient approaches to test the association between a qualitative trait and a candidate marker include the Pearson's chi-square test with 2 degrees of freedom, Maximin Efficiency Robust Test, MAX3 and genetic model selection [3-6]. The last three methods are based on the Cochran-Armitage's trend test (CATT) [6-10]. The CATT is derived for a specified genetic model or mode of inheritance and is proved to be optimal in power when the genetic model is correctly specified. However, the CATT suffers from substantial power loss when the genetic model is misspecified,

*Address correspondence to this author at the Center for Data Science in Health, School of Public Health Administration, Anhui Medical University, Hefei 230032, China; E-mail: myuan@ustc.edu.cn

especially when the dominant model is misspecified as a recessive model or *vice versa*. Maximin Efficiency Robust Test (MERT) linearly combines the extreme pair of CATT by considering the correlation between CATTs and has the highest minimum efficiency relative to the optimal test [11, 12] among all linear combinations of the normally distributed optimal test statistics. MAX3 takes the maximum over the CATTs under dominant, additive and recessive models. The third type of robust approach is a two-step procedure known as genetic model selection (GMS). Hardy-Weinberg disequilibrium coefficient has been examined and used to test association for the case-control design [13, 14]. It has been shown that the Hardy-Weinberg disequilibrium (HWD) contains the genetic model information and can be used to estimate the underlying genetic model for case-control design. The GMS test applies an adaptive procedure to select a genetic model before applying an appropriate CATT. The MERT, MAX3 and GMS are shown to have good power performance across the common genetic models.

However, there are relatively few studies on robust-efficient methods in QTL study. Deng *et al.* [15] proposed a QTL mapping method by measuring and testing for Hardy-Weinberg and linkage disequilibrium at a series of linked marker loci in extreme samples of populations. Inspired by how the genetic model information is extraction from data based on HWD in case-control design, we propose a GMS procedure, which is a two-step procedure similar to the GMS method in case control study. In the first step, we estimate the genetic model by using the HWD calculated from extreme individuals as in the study by Deng *et al.* [15]. In the second step, we apply the t-test corresponding to the genetic model estimated in the first step. We derive the asymptotic correlation between the t-test for recessive model and dominant model. We also derive and adjust the asymptotic correlation between HWD test and t-tests for different genetic models so that the GMS has approximately correct size. MERT and MAX3 for quantitative trait are another two candidate choices, which can be readily used for testing association for a quantitative trait once the model-specific t-test statistics are obtained. Extensive simulation studies are conducted to compare the GMS procedure with existing approaches. Real data analysis is also performed to illustrate the proposed approach.

The rest of this paper is organized as follows. Section 2 introduces notations, models and the theoretical results. Extensive simulation studies are implemented to examine the performance of the proposed methods in Section 3. In Section 4, we apply the proposed methods to real data from a subsample of the Alzheimer's Disease Neuroimaging Initiative (ADNI). Conclusions and discussions are given in the last section.

2. METHODS

2.1. Notations and Models

Consider a QTL with two alleles: A_1 and A_2. Allele A_1 is assumed to be the risk allele and A_2 is the reference allele. Allele A_2 may also be regarded as all of the non-A_1 alleles with similar genetic effects. Let p be the frequency of the allele A_1 and $q = 1 - p$ be the frequency of allele A_2. Denote the three genotypes A_1A_1, A_1A_2 and A_2A_2 by g_i, $i = 0, 1, 2$ respectively. Let μ_0 be the mean (genotypic value) for individuals with genotype A_1A_1 and μ_1 and μ_2 be the genotypic value of individuals with A_1A_2 and A_2A_2, respectively. The null hypothesis is established that there is no association between the locus and the trait, *i.e.* H_0: $\mu_0 = \mu_1 = \mu_2$. The alternative hypothesis is H_1: $\mu_0 \geq \mu_1 \geq \mu_2$ and at least one inequality holds. If we can determine the underlying genetic model, the H_1 will reduced to two-dimensional plane, *i.e.*, H_d: $\mu_0 = \mu_1 > \mu_2$ under the dominant model; H_r: $\mu_0 > \mu_1 = \mu_2$ under the recessive model; H_a: $2\mu_1 = \mu_0 + \mu_2$, $\mu_0 > \mu_2$ under the additive model.

The general linear model for a QTL is described as follows [16]. The phenotypic value of the jth individual with genotype g_i in the population $y_{ij} = \mu + G_i + s_{ij}$, $i = 0, 1, 2, j = 1, 2, \cdots, n_i$. μ is the mean baseline of the quantitative trait, G_i is the genotypic value at the QTL for the ith genotype, and s_{ij} represents a random variable for the combined effects of all the rest of the polymorphic QTLs and all random environmental effects. Without loss of generality, we assume that $\mu = 0$. Thus the genotypic value for the ith genotype is equal to μ_0, μ_1 and μ_2, respectively, for genotypes A_1A_1, A_1A_2, and A_2A_2. We also assume that s_{ij} follows a normal tion $N(0, \sigma_e^2)$.

2.2. Common Test Statistics

To test the association between the candidate genetic marker and a quantitative trait, we can either use the F test with the alternative hypothesis H_1 or t-test $t.r$, $t.a$, $t.d$ for a given genetic model with alternative hypothesis H_r, H_a or H_d, respectively. The F test is derived by calculating the ratio of between-group variance and within-group variance, that is,

$$F = \frac{\sum_{i=0}^{2} n_i (\bar{y}_i - \bar{y})^2 / 2}{\sum_{i=0}^{2} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / (n-3)} \qquad (2.1)$$

Under the null hypothesis, the F test follows an F distribution with degrees of freedom 2 and $n - 3$. When the genetic model is specified as one of the dominant, additive and recessive models, the corresponding t-test can be derived as follows.

$$t.d = \frac{\frac{n_0 \bar{y}_0 + n_1 \bar{y}_1}{n_0 + n_1} - \bar{y}_2}{\sqrt{\left(\frac{1}{n_0 + n_1} + \frac{1}{n_2}\right) s_d^2}} \qquad (2.2)$$

$$t.a = \frac{n_0(n_1 + 2n_2)\bar{y}_0 + n_1(n_2 - n_0)\bar{y}_1 - n_2(n_1 + 2n_0)\bar{y}_2}{\sqrt{\frac{n^2(n_0 n_1 + n_1 n_2 + 4n_0 n_2)^2}{n_0(n_1 + 2n_2)^2 + n_1(n_2 - n_0)^2 + n_2(n_1 + 2n_0)^2} s_a^2}} \qquad (2.3)$$

$$t.r = \frac{\bar{y}_0 - \frac{n_1 \bar{y}_1 + n_2 \bar{y}_2}{n_1 + n_2}}{\sqrt{\left(\frac{1}{n_0} + \frac{1}{n_1 + n_2}\right) s_r^2}} \qquad (2.4)$$

with

$$s_d^2 = \left(\sum_{i=0}^{2}\sum_{j=0}^{n_i}(y_{ij}-\bar{y})^2 + \frac{n_0 n_1}{n_0+n_1}(\bar{y}_0-\bar{y}_1)^2\right)/(n-2)$$

$$s_a^2$$

$$= \left(\sum_{i=0}^{2}\sum_{j=0}^{n_i}(y_{ij}-\bar{y})^2 \right.$$
$$\left. + \frac{n_0 n_1 n_2}{n_0 n_1 + 4 n_0 n_2 + n_1 n_2}(\bar{y}_0+\bar{y}_2-2\bar{y}_1)^2\right)/(n-2)$$

$$s_r^2 = \left(\sum_{i=0}^{2}\sum_{j=0}^{n_i}(y_{ij}-\bar{y})^2 + \frac{n_1 n_2}{n_1+n_2}(\bar{y}_1-\bar{y}_2)^2\right)/(n-2)$$

Under the null hypothesis, $t.r$, $t.a$, $t.d$ follow the t distribution with degrees of freedom $n-2$. When the sample size is sufficiently large, we can use a standard normal distribution to approximate the t distribution. It is obvious that $t.r$, $t.a$, $t.d$ are the correspondingly optimal tests for a known genetic model. As the above tests incorporate risk trend in the genotypic value, we call them trend tests as that in case control studies. In practice, the underlying genetic model is rarely known and the use of the trend test specified for a certain model could suffer from substantial loss of power if the genetic model is misspecified. Therefore, it is essential to consider efficiency robust methods to protect against genetic model mis-specifications. The F test, MERT and MAX3 are three common robust tests. MERT linearly combines $t.r$ and $t.d$ by taking into account the correlation between them. The correlation efficient between $t.r$ and $t.d$ denoted as ρ_{rd} equals $\sqrt{n_0 n_2/((n_0+n_1)(n_1+n_2))}$ asymptotically. Therefore, The MERT can be derived as follows,

$$Z_{mert} = \frac{t.r+t.d}{\sqrt{2(1+\rho_{rd})}} \tag{2.5}$$

The MAX3, taking the maximum of three optimal tests under different genetic models, is given by,

$$Z_{max} = \max\{|t.r|, |t.a|, |t.d|\} \tag{2.6}$$

Thresholds to control type I error at level α for Z_{mert} and Z_{max} are determined through equations $P_{H_0}(|Z_{mert}| > c_{mert}) = \alpha$ and $P_{H_0}(Z_{max} > c_{max}) = \alpha$ respectively. As Z_{mert} is asymptotically normal distributed, we can use the theoretical quantile as the threshold when sample size is large enough. However, it is hard to obtain the asymptotical distribution of Z_{max}. In this article, we will use Monto-Carlo method to obtain all these thresholds.

2.3. Genetic Model Selection and Related Robust Test

We assume that HWE is present in the population. The quantitative trait in the population then follows a mixture distribution of three normal distributions, each of which is weighted by the respective genotype frequencies in the population. The probability density function is $p(x) = p^2\phi(x,\mu_0,\sigma_e^2) + 2pq\phi(x,\mu_1,\sigma_e^2) + q^2\phi(x,\mu_2,\sigma_e^2)$ where $\phi(x,\mu,\sigma^2)$ is the density function for the normal random variable with mean μ and σ^2.

For a given upper threshold U, extreme individuals are defined as those people with phenotype greater than U, that is $y > U$. Denote $\psi_{11} = P(y > U|A_1A_1)$, $\psi_{12} = P(y > U|A_1A_2)$, $\psi_{22} = P(y > U|A_2A_2)$ be the probabilities to obtain an extreme sample conditioning on the genotypes, then:

$$\boldsymbol{\psi_{11}} = \boldsymbol{\Phi}(\tfrac{\mu_0-U}{\sigma_e}), \boldsymbol{\psi_{12}} = \boldsymbol{\Phi}(\tfrac{\mu_1-U}{\sigma_e}), \boldsymbol{\psi_{22}} = \boldsymbol{\Phi}(\tfrac{\mu_2-U}{\sigma_e}), \tag{2.7}$$

where $\boldsymbol{\Phi(x,\mu,\sigma^2)}$ is the cumulative probability function for the normal random variable with mean μ and σ^2. The proportion of the extreme individuals denoted as $\boldsymbol{\psi_U}$, can be computed as follows:

$$P(y > U) = p^2\Phi(\frac{\mu_0-U}{\sigma_e}) + 2pq\Phi(\frac{\mu_1-U}{\sigma_e})$$
$$+ q^2\Phi(\frac{\mu_2-U}{\sigma_e})$$
$$= p^2\psi_{11} + 2pq\psi_{12} + q^2\psi_{22}.$$

Similarly, for a lower threshold T, we can compute the proportion of the population with $y < T$ by using the following relationship,

$$\psi_T = P(y < T) =$$

$$p^2\Phi(\tfrac{T-\mu_0}{\sigma_e}) + 2pq\Phi(\tfrac{T-\mu_1}{\sigma_e}) + q^2\Phi(\tfrac{T-\mu_2}{\sigma_e}).$$

It can be easily shown that the probabilities of a certain genotype and the risk allele given that the individual comes from the extreme population with $y > U$ are respectively,

$$P(A_1A_1|y > U) = \frac{P(y>U|A_1A_1)P(A_1A_1)}{P(y>U)} = \frac{p^2\psi_{11}}{\psi_U} \tag{2.8}$$

and

$$P(A_1|y > U) = \frac{P(y>U|A_1)P(A_1)}{P(y>U)}$$
$$= \frac{pP(y>U|A_1A_1)P(A_1A_1|A_1)+pP(y>U|A_1A_2)P(A_1A_2|A_1)}{P(y>U)} \tag{2.9}$$
$$= \frac{p^2\psi_{11}+pq\psi_{12}}{\psi_U}.$$

The Hardy-Weinberg disequilibrium coefficient in the extreme population can be defined by the difference between the genotype frequency and the allele as follows,

$$D_{A_1A_1|\{y>U\}} = P(A_1A_1|y > U) - P^2(A_1|y >$$

$$U) = \frac{p^2 q^2(\psi_{11}\psi_{22}-\psi_{12}^2)}{\psi_U^2}. \tag{2.10}$$

When the null hypothesis is true, *i.e.*, $\mu_0 = \mu_1 = \mu_2$, $\psi_{11} = \psi_{12} = \psi_{22}$, we have $\psi_{11}\psi_{22} - \psi_{12}^2 = 0$. When the underlying genetic model is dominant, *i.e.*, $\mu_0 = \mu_1 > \mu_2$, $\psi_{11} = \psi_{12} > \psi_{22}$, thus $\psi_{11}\psi_{22} - \psi_{12}^2 = \psi_{12}(\psi_{22} - \psi_{11}) < 0$. When the underlying genetic model is recessive, *i.e.*, $\mu_1 = \mu_2 < \mu_0$, $\psi_{12} = \psi_{22} < \psi_{11}$ and $\psi_{11}\psi_{22} - \psi_{12}^2 =$

$\psi_{22}(\psi_{11} - \psi_{22}) > 0$. When the genetic model is additive model, *i.e.*, $2\mu_1 = \mu_0 + \mu_2$, then $\psi_{11}\psi_{22} - \psi_{12}^2 = \Phi(\frac{\mu_0 - U}{\sigma_e})\Phi(\frac{\mu_2 - U}{\sigma_e}) - (\Phi(\frac{\mu_1 - U}{\sigma_e}))^2$ can either be positive or negative and is close to 0 if both μ_0 and μ_2 are close to zero (small or moderate effects). The signs of $D_{A_1A_1|\{y>U\}}$ under the dominant and recessive are opposite, indicating that $D_{A_1A_1|\{y>U\}}$ can be used to select a genetic model.

Denote $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}$, then $\bar{y}_i$ follows a normal distribution with mean μ_i and variance σ_e^2/n_i, $i = 0, 1, 2$. We use $\hat{\mu}_i = \bar{y}_i$ to estimate μ_i. The allele frequency of A_1 can be estimated by $\hat{p} = (2n_0 + n_1)/(2n_0 + 2n_1 + 2n_2)$. Denote $\hat{q} = 1 - \hat{p}$. $\hat{\psi}_{11}, \hat{\psi}_{12}, \hat{\psi}_{22}, \hat{\psi}_U$ are defined by estimating the unknown parameters μ_i, $i = 0, 1, 2$ and p in the expressions of $\psi_{11}, \psi_{12}, \psi_{22}$ and ψ_U by $\hat{\mu}_i$, $i = 0, 1, 2$ and $\hat{p}$ respectively. Therefore, Hardy-Weinberg disequilibrium coefficient $D_{A_1A_1|\{y>U\}}$ can be estimated by,

$$\hat{D}_{A_1A_1|\{y>U\}} = \frac{\hat{p}^2\hat{q}^2(\hat{\psi}_{11}\hat{\psi}_{22} - \hat{\psi}_{12}^2)}{\hat{\psi}^2}.$$

The Hardy-Weinberg disequilibrium test can be constructed by standardizing the disequilibrium coefficient as follows,

$$Z_{hwd} = \frac{\hat{D}_{A_1A_1|\{y>U\}} - E_{H_0}(\hat{D}_{A_1A_1|\{y>U\}})}{\sqrt{\hat{V}ar_{H_0}(\hat{D}_{A_1A_1|\{y>U\}})}}. \quad (2.11)$$

Under the null hypothesis, $E_{H_0}(\hat{D}_{A_1A_1|\{y>U\}}) = 0$ and $\hat{V}ar_{H_0}(\hat{D}_{A_1A_1|\{y>U\}})$ can be calculated by the delta method. Details are provided in the Appendix (A1). The expectation of Z_{hwd} has the same sign as the Hardy-Weinberg disequilibrium coefficient $\hat{D}_{A_1A_1|\{y>U\}}$, therefore, we select the recessive model if $Z_{hwd} > c_1$; the dominant model if $Z_{hwd} < c_2$ and additive model otherwise. The thresholds c_1 and c_2 are two constants, which can be either prespecified or determined based on the data. Once the underlying model is determined,

the correspondingly optimal trend test could be applied. Such procedure is a two-step method, which extracts the model information from the data and cooperates them into the association test. This data driven procedure is referred the GMS in this article. The choice of thresholds of c_1 and c_2 affects the accuracy of model selection. For example, if c_1 is relatively large and c_2 is relatively small, it is more likely to select the dominant (recessive) model as the recessive (dominant) model, which is the situation we should always avoid. Therefore, one needs to choose c_1 and c_2 carefully when applying the GMS method. In this paper, we propose to determine the thresholds c_1 and c_2 by a data driven method. The main idea of the data driven procedure is to calculate the mean of the HWD statistics under different assumptions and use the midpoints as the thresholds to distinguish different genetic models. Denote,

$$\xi(x, y, z) =$$

$$\frac{\Phi(\frac{(n_0x + n_1y + n_2z)/n - U}{\sigma_e})(\Phi(\frac{x-U}{\sigma_e})\Phi(\frac{z-U}{\sigma_e}) - (\Phi(\frac{y-U}{\sigma_e}))^2)}{\psi_U^2 \Phi(\frac{(n_0x + n_1y + n_2z)/n - U}{\sigma_e})\sqrt{\frac{1}{n_0} + \frac{4}{n_1} + \frac{1}{n_2}}} \quad (2.12)$$

where

$$\psi_U =$$

$$(\frac{2n_0 + n_1}{2n})^2 \Phi(\frac{x-U}{\sigma_e}) + \frac{(2n_0 + n_1)(n_1 + 2n_2)}{2n^2}\Phi(\frac{y-U}{\sigma_e}) +$$

$$(\frac{n_1 + 2n_2}{2n})^2 \Phi(\frac{z-U}{\sigma_e}).$$

For a given genotypic value μ, it is obvious that $\xi(\mu, \mu, \mu)$, $\xi(\mu, \mu, -\mu)$, $\xi(\mu, 0, -\mu)$ and $\xi(\mu, -\mu, -\mu)$ are the asymptotic mean of Z_{hwd} under the null hypothesis and three different genetic models. Estimators of these four values are denoted by $\xi_{null}, \xi_{dom}, \xi_{add}$ and ξ_{rec} respectively. It is easy to show that $\xi_{dom} < \xi_{add} < \xi_{rec}$. Therefore, thresholds for model selection can be sensibly determined as the midpoints of points ξ_{dom} and ξ_{add}, of points ξ_{add} and ξ_{rec} respectively. Specifically, $c_1 = (\xi_{rec} + \xi_{add})/2$ and $c_2 = (\xi_{add} + \xi_{dom})/2$. To control type I error at level α, we need to find the threshold c_α under H_0 such that:

$$P_{H_0}(Z_{hwd} > c_2, t.d > c_\alpha) + P_{H_0}(Z_{hwd} < c_1, t.r$$
$$> c_\alpha) + P_{H_0}(c_1 < Z_{hwd} < c_2, t.a > c_\alpha)$$
$$= \alpha$$

The threshold c_α can be obtained analytically by the delta method and the asymptotic joint multivariate normal distribution of Z_{hwd} and *t.r*, *t.d* or *t.a* when the sample size is moderately large (details to derive the expression of the correlations between Z_{hwd} and the trend tests t.r, (t.a, t.d) are provided in Appendix A2). Our simulations generate 10000 independent datasets under the null hypothesis by resampling method and using the empirical quantile as the critical value.

3. RESULTS

3.1. Simulation

We first evaluate the accuracy of the model selection procedure under various scenarios. We also check the performance of the GMS and compare it with the other methods, including the F test in (2.1) (denoted by F in the figures and tables), the trend tests in (2.2), MERT, Z_{mert} in (2.5) and the MAX3 test, Z_{max} in (2.6). Data are generated under 54 simulation settings. The population allele frequency p is taken to be 0.1, 0.3, or 0.5. Sample size is set at $n = 200, 500,$ or 1000. The genotypic value is $\mu = 0.1, 0.5,$ or 1. Three genetic models are considered in the simulation: the dominant (DOM), the additive (ADD) and the recessive (REC). Effective sample proportion (ESP) defined as the proportion of extreme individuals is $\psi = 5\%,$ or 10%. Each simulation is replicated 10000 times. Simulation results show that the model selection accuracy increases with the allele frequency,
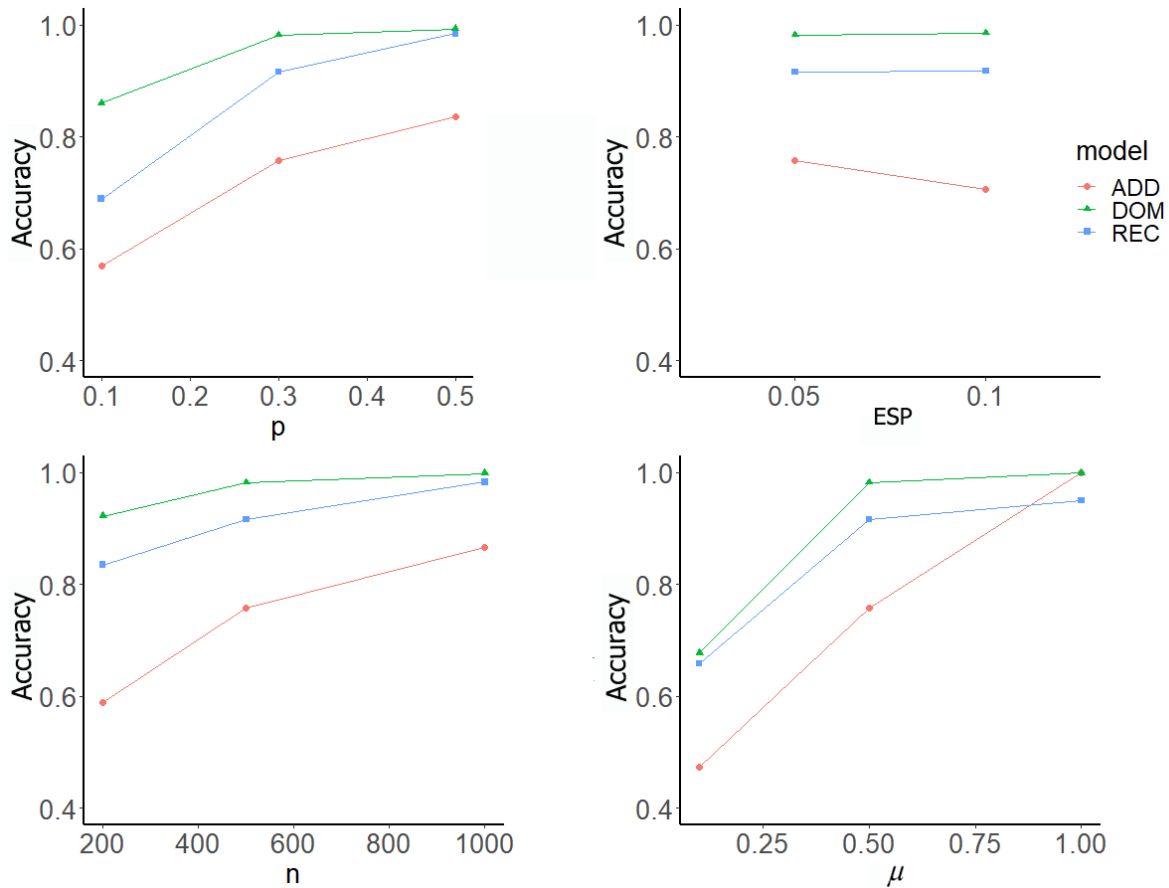
Fig. (1). Relationship between accuracy and *p*, *n*, ESP, *μ* when other parameters are fixed (allele frequency *p* = 0.3, effective sample proportion ESP = 0.05, sample size *n* = 500, genotypic value *μ* = 0.5). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

the genotypic value, the sample size and effective sample size proportion (Fig. **1**).

Table **1** reports the accuracy of the genetic model selection for a moderate allele frequency and sample size, *i.e. p* = 0.3 and *n* = 500 for different genotypic values. From Table **1** we can see the model selection accuracy is high for a moderate genotypic value. For example, when the genotypic value *μ* = 0.5, the probabilities of selecting the true model are about 92%, 76% and 98%, respectively.

Type I error rates are provided in Table **2**. All of the seven methods can control type I error rate well at the nominal level 0.05. Power comparisons are presented in Fig. (**2**). The trend tests (*i.e.*, *t.d*, *t.a*, and *t.r*) are sensitive to the model assumption. For example, when the true model is recessive, the dominant trend test t_d is least powerful among all the tests. Compared to the trend tests, the three robust-efficient procedures are generally more robust against model misspecification with a higher power. In addition, the proposed method, GMS, is generally more powerful than MAX3 and MERT under different genetic models. We also compared the performance of seven different methods in the case of rare variants (results are provided in Appendix A3). When the allele frequency is low (*i.e.*, p=0.01, 0.05, 0.1), the genetic effect is moderate (*μ* = 0.2) and the effective sample size is small

(*i.e.*, 100), the powers of all tests are less than 50%. When the sample size is gradually increased to 4000 (effective sample size 400), the powers of robust tests, GMS, MERT and MAX3, can reach the power of more than 80%. For example, when the genetic model is DOM and *p* = 0.05, the powers of these tests are 83.53% (MERT), 89.50% (MAX3) and 95.34% (GMS) respectively and the corresponding optimal power is 97.76% (*t.d*). The simulation results under rare variants scenarios show that GMS, MERT and MAX3 tests can find statistical differences when the sample size is sufficiently large. Whether it is common variants or rare variants, GMS always has the highest power among these three methods and is comparable to the optimal test.

3.2. Application

We applied our method to the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The data was downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.usc.edu/ADNI). The ADNI is an ongoing longitudinal multicenter study aimed at detecting and monitoring the early stage of AD by investigating magnetic resonance imaging, positron emission tomography, genetic, biochemical biomarkers, and neuropsychological and clinical assessment. The ADNI data collected 2074 observations and 113 variables, including clinical outcomes and biomarker variables for

Table 1. **Proportion of selecting of different genetic models (*p*=0.3 *n*=500, ESP=0.05).**

-	True Model	Selected Model		
		REC	ADD	DOM
$\mu = 0.1$	NULL	0.4933	0.2702	0.2365
	REC	0.6581	0.3068	0.0351
	ADD	0.3265	0.4742	0.1993
	DOM	0.0933	0.2288	0.6779
$\mu = 0.5$	NULL	0.3477	0.3170	0.3353
	REC	0.9158	0.0842	0.0000
	ADD	0.0345	0.7582	0.2073
	DOM	0.0000	0.0179	0.9821
$\mu = 1$	NULL	0.2353	0.6985	0.0662
	REC	0.9495	0.0505	0.0000
	ADD	0.0007	0.9993	0.0000
	DOM	0.0000	0.0009	0.9991
$\mu = 2$	NULL	0.3417	0.6472	0.0111
	REC	1.0000	0.0000	0.0000
	ADD	0.0000	1.0000	0.0000
	DOM	0.0000	0.0002	0.9998

Table 2. **Type I error rate of the seven methods with various parameters.**

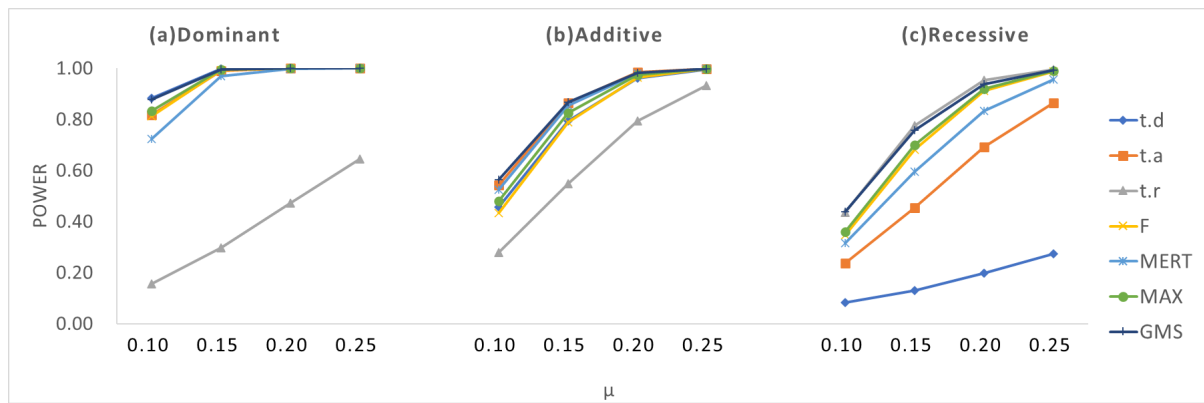| ESP	*p*	*n*	Seven Methods						
			t.d	*t.a*	*t.r*	*F*	MERT	MAX3	GMS
0.05	0.1	200	0.0487	0.0503	0.0492	0.0500	0.0493	0.0500	0.0489
		500	0.0514	0.0519	0.0492	0.0500	0.0514	0.0500	0.0437
		1000	0.0505	0.0505	0.0501	0.0500	0.0503	0.0500	0.0517
	0.3	200	0.0495	0.0498	0.0502	0.0500	0.0502	0.0499	0.0480
		500	0.0496	0.0515	0.0461	0.0500	0.0514	0.0500	0.0498
		1000	0.0514	0.0539	0.0530	0.0501	0.0501	0.0500	0.0476
	0.5	200	0.0502	0.0495	0.0511	0.0500	0.0493	0.0500	0.0472
		500	0.0540	0.0520	0.0523	0.0500	0.0517	0.0500	0.0456
		1000	0.0477	0.0492	0.0500	0.0500	0.0477	0.0500	0.0498
0.1	0.1	200	0.0501	0.0506	0.0491	0.0500	0.0506	0.0500	0.0488
		500	0.0542	0.0482	0.0520	0.0500	0.0484	0.0500	0.0461
		1000	0.0520	0.0509	0.0502	0.0500	0.0509	0.0502	0.0485
	0.3	200	0.0491	0.0498	0.0498	0.0500	0.0498	0.0500	0.0491
		500	0.0504	0.0499	0.0493	0.0498	0.0500	0.0500	0.0509
		1000	0.0496	0.0476	0.0505	0.0500	0.0473	0.0505	0.0505
	0.5	200	0.0496	0.0497	0.0498	0.0500	0.0499	0.0500	0.0478
		500	0.0477	0.0489	0.0468	0.0500	0.0490	0.0500	0.0513
		1000	0.0505	0.0481	0.0540	0.0500	0.0479	0.0500	0.0466

Fig. (2). The power comparison among seven methods with $p = 0.3$, $n = 1000$, ESP=0.1 under (a) dominant model; (b) additive model and (c) recessive model. *t.a*, *t.d*, and *t.r*: trend tests corresponding to additive, dominant, recessive genetic models. F: *F*-test; MERT: Z_{mert} in (2.5); MAX3: the MAX3 test Z_{max} in (2.6); GMS: two step approach by estimating genetic model with Z_{hwd} in (2.11) and corresponding trend test in (2.2)-(2.4). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).
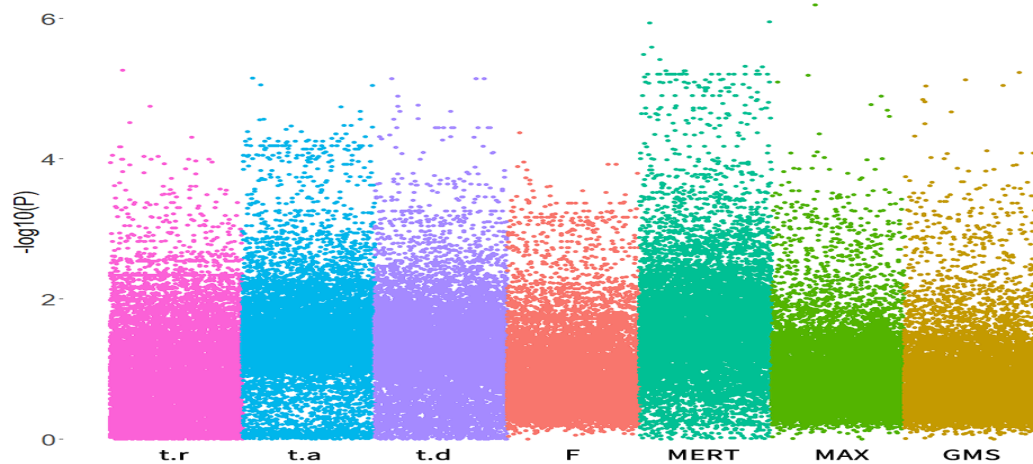

Fig. (3). Manhattan plot for all p-values (on –log 10 scale) of seven methods for all 10683 SNPs on chromosome 19. *t.a*, *t.d*, and *t.r*: trend tests corresponding to additive, dominant, recessive genetic models. F: *F*-test; MERT: Z_{mert} in (2.5); MAX3: the MAX3 test Z_{max} in (2.6); GMS: two step approach by estimating genetic model with Z_{hwd} in (2.11) and corresponding trend test in (2.2)-(2.4). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Alzheimer's Disease (AD) for 784 individuals. In this section, we used baseline Rey's Auditory Verbal Learning Test (RAVLT) score as the response variable to examine the association between SNPs on chromosome 19 and AD. The RAVLT is a powerful neuropsychological tool for testing episodic memory, which is widely used for cognitive assessment in dementia and pre-dementia conditions [17]. Several studies have shown that impairment in RAVLT scores reflects well the underlying pathology caused by Alzheimer's disease (AD), thus making RAVLT an effective early marker to detect AD in persons with memory complaints [18]. The higher score indicates the severity of the auditory verbal learning ability impairment. We selected individuals with RAVLT score greater than 4 as the extreme samples (*i.e.*, U>4). After quality control, missing values are removed and 783 individuals genotyped at 10683 SNPs on chromosome 19 are included in the analysis.

We apply the proposed GMS method to the ADNI data and compare it with the other six methods. We permuted data 10,000 times to determine the thresholds of the MAX3 and GMS. Thresholds for the other 5 methods are determined by the asymptotic normal distribution. After FDR correction, we find 282 significant SNPs at significance level $\alpha = 0.05$. Manhattan plot for the ADNI data is presented in Fig. (**3**), which shows the p-values of all SNPs with seven methods on the –log10 scale. In Table **3**, we report the top 10 significant SNPs selected by GMS, as well as p-values of the other six methods. All three robust methods (MERT, MAX3 and GMS) exhibit robustness to genetic model specification. For ADNI data, it seems that MERT and GMS have comparable power, while MAX3 is slightly less powerful than MERT and GMS with relatively large p-values. Existing studies have shown that zinc plays an important role in the development of Alzheimer's disease [19, 20]. Among the 10 significant

Table 3.   The p-value of significant SNPs selected by GMS.

SNP	t.r*	t.a*	t.d*	F*	MERT*	MAX*	GMS*
rs143663113	2.06e-04	2.30e-06	2.53e-04	7.09e-04	1.80e-05	4.09e-04	5.81e-06
rs62131791	3.01e-06	5.15e-06	1.45e-02	1.69e-04	8.76e-05	3.14e-04	7.45e-06
rs73026154	1.10e-05	6.07e-06	6.46e-03	1.21e-04	2.61e-05	1.21e-05	8.95e-06
rs71839901	5.26e-06	9.11e-05	1.49e-03	1.69e-04	8.76e-05	5.07e-05	9.14e-06
rs111321022	6.30e-05	1.73e-05	3.42e-04	6.58e-04	1.67e-05	3.14e-04	1.07e-05
rs113397810	2.74e-05	6.02e-06	2.64e-05	3.40e-04	7.97e-06	1.48e-04	1.09e-05
rs7258002	4.34e-04	1.17e-05	2.53e-04	2.68e-04	6.20e-06	2.10e-05	1.20e-05
rs10409896	1.44e-03	6.06e-06	2.81e-05	2.67e-04	6.20e-06	1.62e-04	1.37e-05
rs7257286	1.82e-03	1.02e-05	1.07e-06	5.52e-05	1.15e-06	8.23e-05	1.65e-05
rs117810408	7.40e-04	1.58e-05	2.62e-03	3.14e-04	7.31e-06	1.84e-04	2.15e-05

*t.a, t.d, and *t.r*: trend tests corresponding to additive, dominant, recessive genetic models. F: *F*-test; MERT: Z_{mert} in (2.5); MAX3: the MAX3 test Z_{max} in (2.6); GMS: two step approach by estimating genetic model with Z_{hwd} in (2.11) and corresponding trend test in (2.2)-(2.4).

SNPs, five SNPs (rs143663113, rs113397810, rs7258002, rs10409896, rs117810408) are screened and reported to be related to zinc finger protein 578 (Gene ZNF578), and two SNPs (rs143663113, rs111321022) be related to Gene ZNF808.

CONCLUSION

In this article, we proposed a two-step procedure to test the genetic association between an SNP and a quantitative trait. The underlying genetic model is determined in the first step based on the difference of HWD coefficients under different genetic models using extreme individuals, and in the second step, the t-test corresponding to the genetic model selected in the first step is then applied to test the association between the SNP and the trait.

The proposed method is shown to be more efficient than the previously reported approaches (MERT and MAX3) by simulation studies. Under three genetic models, GMS method has similar power to the t-test corresponding to the true model with a well-controlled type I error rate at the given significant level. In our study, we used a data-driven threshold in selecting genetic model. Simulation results show that this method works well in selecting the underlying genetic model. We would like to mention that in the literature of QTL, robustness often refers to nonparametric methods that are robust to distribution specification of the error term in linear models [21, 22]. Robustness in this work refers to the insensitivity of the tests to genetic model specification.

The proposed GMS has high efficiency in power for common variants. For the rare variants with low allele frequency, even the best test has very low power. However, GMS is the closest to the optimal test among the existing robust methods for either common or rare variants. In practical applications, higher efficiency could be achieved by increasing the sample size for rare variants.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the fundings of the study are available within the article.

FUNDING

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

APPENDIX

Appendix A1: Derivation of the estimated variance of $(\widehat{\varphi}_{11}\widehat{\varphi}_{22} - \widehat{\varphi}_{12}^2)/\widehat{\varphi}^2$.

Let $f(\mu_0, \mu_1, \mu_2) = (\psi_{11}\psi_{22} - \psi_{12}^2)/\psi^2$, because

$$\frac{\partial \psi_{11}}{\partial \mu_0} = \frac{1}{\sigma_e}\phi\left(\frac{\mu_0 - U}{\sigma_e}\right), \frac{\partial \psi_{11}}{\partial \mu_1} = \frac{\partial \psi_{11}}{\partial \mu_2} = 0$$

$$\frac{\partial \psi_{12}}{\partial \mu_1} = \frac{1}{\sigma_e}\phi\left(\frac{\mu_1 - U}{\sigma_e}\right), \frac{\partial \psi_{12}}{\partial \mu_0} = \frac{\partial \psi_{12}}{\partial \mu_2} = 0$$

$$\frac{\partial \psi_{22}}{\partial \mu_2} = \frac{1}{\sigma_e}\phi\left(\frac{\mu_2 - U}{\sigma_e}\right), \frac{\partial \psi_{22}}{\partial \mu_0} = \frac{\partial \psi_{22}}{\partial \mu_1} = 0$$

$$\frac{\partial \psi}{\partial \mu_0} = p^2 \frac{\partial \psi_{11}}{\partial \mu_0}, \frac{\partial \psi}{\partial \mu_1} = 2pq \frac{\partial \psi_{12}}{\partial \mu_1}, \frac{\partial \psi}{\partial \mu_2} = q^2 \frac{\partial \psi_{22}}{\partial \mu_2},$$

$$\frac{\partial f}{\partial \mu_0} = \frac{1}{\psi^3} \frac{\partial \psi_{11}}{\partial \mu_0} [\psi_{22}\psi - 2p^2(\psi_{11}\psi_{22} - \psi_{12}^2)]$$

$$\frac{\partial f}{\partial \mu_1} = -\frac{2}{\psi^3} \frac{\partial \psi_{12}}{\partial \mu_1} [\psi_{12}\psi + 2pq(\psi_{11}\psi_{22} - \psi_{12}^2)]$$

$$\frac{\partial f}{\partial \mu_2} = \frac{1}{\psi^3} \frac{\partial \psi_{22}}{\partial \mu_2} [\psi_{11}\psi - 2q^2(\psi_{11}\psi_{22} - \psi_{12}^2)]$$

Therefore, by applying the delta-method, we have,

$$Varf(\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2) = \begin{pmatrix} \frac{\partial f}{\partial \mu_0} \\ \frac{\partial f}{\partial \mu_1} \\ \frac{\partial f}{\partial \mu_2} \end{pmatrix}' \begin{pmatrix} \frac{\sigma_e^2}{n_0} & 0 & 0 \\ 0 & \frac{\sigma_e^2}{n_1} & 0 \\ 0 & 0 & \frac{\sigma_e^2}{n_2} \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial \mu_0} \\ \frac{\partial f}{\partial \mu_1} \\ \frac{\partial f}{\partial \mu_2} \end{pmatrix}.$$

When the null hypothesis is true, $\mu_0 = \mu_1 = \mu_2 = \mu^*$, $\psi_{11} = \psi_{12} = \psi_{22} = \psi \triangleq \psi^* = \Phi(\frac{\mu^*-U}{\sigma_e})$ and $\frac{\partial \psi_{11}}{\partial \mu_0} = \frac{\partial \psi_{12}}{\partial \mu_1} = \frac{\partial \psi_{22}}{\partial \mu_2} = \frac{1}{\sigma_e}\phi(\frac{\mu^*-U}{\sigma_e})$, further we have

$$\frac{\partial f}{\partial \mu_0}\Big|_{H_0} = \frac{1}{\psi^*}\frac{1}{\sigma_e}\phi\left(\frac{\mu^* - U}{\sigma_e}\right),$$

$$\frac{\partial f}{\partial \mu_1}\Big|_{H_0} = -\frac{2}{\psi^*}\frac{1}{\sigma_e}\phi\left(\frac{\mu^* - U}{\sigma_e}\right),$$

$$\frac{\partial f}{\partial \mu_2}\Big|_{H_0} = \frac{1}{\psi^*}\frac{1}{\sigma_e}\phi\left(\frac{\mu^* - U}{\sigma_e}\right).$$

Therefore

$$Var_{H_0} f(\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2) = (\frac{\phi\left(\frac{\mu^*-U}{\sigma_e}\right)}{\Phi\left(\frac{\mu^*-U}{\sigma_e}\right)})^2 \left(\frac{1}{n_0} + \frac{4}{n_1} + \frac{1}{n_2}\right),$$

and

$$\hat{Var}_{H_0} f(\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2) = (\frac{\phi\left(\frac{\bar{y}-U}{\sigma_e}\right)}{\Phi\left(\frac{\bar{y}-U}{\sigma_e}\right)})^2 \left(\frac{1}{n_0} + \frac{4}{n_1} + \frac{1}{n_2}\right),$$

where $\bar{y} = \sum_{i,j} y_{ij} / n$.

Appendix A2: Details to derive the expression of the expectation of the Hardy-Weinberg disequilibrium test Z_{hwd} and the correlations between Z_{hwd} and the trend tests $t.r(t.a, t.d)$ respectively.

The Hardy-Weinberg disequilibrium test Z_{hwd} is

$$Z_{hwd} = \frac{(\hat{\psi}_{11}\hat{\psi}_{22}-\hat{\psi}_{12}^2)/(\hat{p}^2\hat{\psi}_{11}+2\hat{p}\hat{q}\hat{\psi}_{12}+\hat{q}^2\hat{\psi}_{22})^2}{\frac{\phi(\frac{\bar{y}-U}{\sigma_e})}{\Phi(\frac{\bar{y}-U}{\sigma_e})}\sqrt{\frac{1}{n_0}+\frac{4}{n_1}+\frac{1}{n_2}}}$$

So

$$EZ_{hwd} = \frac{(\psi_{11}\psi_{22}-\psi_{12}^2)/(p^2\psi_{11}+2pq\psi_{12}+q^2\psi_{22})^2}{\frac{\phi(\frac{\bar{y}-U}{\sigma_e})}{\Phi(\frac{\bar{y}-U}{\sigma_e})}\sqrt{\frac{1}{n_0}+\frac{4}{n_1}+\frac{1}{n_2}}}$$

$$= \frac{\Phi(\frac{\bar{y}-U}{\sigma_e})(\psi_{11}\psi_{22}-\psi_{12}^2)}{(p^2\psi_{11}+2pq\psi_{12}+q^2\psi_{22})^2\phi(\frac{\bar{y}-U}{\sigma_e})\sqrt{\frac{1}{n_0}+\frac{4}{n_1}+\frac{1}{n_2}}}$$

Under specific genetic model, $\psi_{11} = \Phi(\frac{\mu_0-U}{\sigma_e})$, $\psi_{12} = \Phi(\frac{\mu_1-U}{\sigma_e})$, $\psi_{22} = \Phi(\frac{\mu_2-U}{\sigma_e})$, $p = (2n_0 + n_1)/2n$, $q = 1 - p$, $\bar{y} = (n_0\mu_0 + n_1\mu_1 + n_2\mu_2)/n$, so $EZ_{hwd} = \xi(\mu_0, \mu_1, \mu_2)$. Under the null hypothesis and three different genetic models, EZ_{hwd} equal to $\xi(\mu, \mu, \mu), \xi(\mu, \mu, -\mu)$, $\xi(\mu, 0, -\mu)$ and $\xi(\mu, -\mu, -\mu)$. We also calculate the correlation between Z_{hwd} and $t.r(t.a, t.d)$ by delta method.

$$\rho_{hr} = \sqrt{\frac{\frac{1}{n_0+n_1}+\frac{1}{n_2}}{\frac{1}{n_0}+\frac{4}{n_1}+\frac{1}{n_2}}}, \rho_{ha} = 0, \rho_{hd} = -\sqrt{\frac{\frac{1}{n_0}+\frac{1}{n_1+n_2}}{\frac{1}{n_0}+\frac{4}{n_1}+\frac{1}{n_2}}}$$

Appendix A3:

Table A1: Powers with different allele frequencies (sample size: 4000; μ = 0.2; ESP=10%).

	p	t.d	t.a	t.r	F	MERT	MAX3	GMS
	0.01	0.8189	0.8098	0.0695	0.6825	0.6136	0.7465	0.7688
DOM	0.05	0.9775	0.9689	0.1213	0.8936	0.8352	0.8950	0.9534
	0.1	0.9898	0.9803	0.4966	0.9351	0.9180	0.9465	0.9634
	0.01	0.4183	0.4242	0.0687	0.3306	0.3027	0.3424	0.3725
ADD	0.05	0.7697	0.7709	0.1335	0.7433	0.6864	0.7520	0.7560
	0.1	0.8992	0.9294	0.5889	0.8884	0.8972	0.9089	0.9193
	0.01	0.0520	0.0532	0.0683	0.0622	0.0602	0.0618	0.0784
REC	0.05	0.0512	0.0548	0.1445	0.1159	0.1050	0.1133	0.1451
	0.1	0.0758	0.1664	0.6741	0.5644	0.4585	0.5730	0.6389

REFERENCES

[1] Li, Y.; Levran, O.; Kim, J.; Zhang, T.; Chen, X.; Suo, C. Extreme sampling design in genetic association mapping of quantitative trait loci using balanced and unbalanced case-control samples. *Sci. Rep.,* **2019**, *9*(1), 15504.
http://dx.doi.org/10.1038/s41598-019-51790-w PMID: 31664079

[2] Vandiedonck, C. Genetic association of molecular traits: A help to identify causative variants in complex diseases. *Clin. Genet.,* **2018**, *93*(3), 520-532.
http://dx.doi.org/10.1111/cge.13187 PMID: 29194587

[3] Bagos, P.G. Genetic model selection in genome-wide association studies: robust methods and the use of meta-analysis. *Stat. Appl. Genet. Mol. Biol.,* **2013**, *12*(3), 285-308.
http://dx.doi.org/10.1515/sagmb-2012-0016 PMID: 23629457

[4] Dimou, N.L.; Tsirigos, K.D.; Elofsson, A.; Bagos, P.G. GWAR: robust analysis and meta-analysis of genome-wide association studies. *Bioinformatics,* **2017**, *33*(10), 1521-1527.
http://dx.doi.org/10.1093/bioinformatics/btx008 PMID: 28108451

[5] Zheng, G.; Freidlin, B.; Gastwirsh, J.L. Comparison of robust tests for genetic association using case-control studies. *IMS. Lect. Notes Monogr. Ser.,* **2006**, *49*, 253-265.
http://dx.doi.org/10.1214/074921706000000491

[6] Zheng, G.; Ng, H.K.T. Genetic model selection in two-phase analysis for case-control association studies. *Biostatistics,* **2008**, *9*(3), 391-399.
http://dx.doi.org/10.1093/biostatistics/kxm039 PMID: 18003629

[7] Armitage, P. Tests for linear trends in proportions and frequencies. *Biometrics,* **1955**, *11*, 375-386.
http://dx.doi.org/10.2307/3001775

[8] Cochran, W.G. Some methods for strengthening the common chi-square tests. *Biometrics,* **1954**, *10*, 417-451.
http://dx.doi.org/10.2307/3001616

[9] Joo, J.; Kwak, M.; Zheng, G. Improving power for testing genetic association in case-control studies by reducing the alternative space. *Biometrics,* **2010**, *66*(1), 266-276.
http://dx.doi.org/10.1111/j.1541-0420.2009.01241.x PMID: 19397584

[10] Sasieni, P.D. From genotypes to genes: doubling the sample size. *Biometrics,* **1997**, *53*(4), 1253-1261.
http://dx.doi.org/10.2307/2533494 PMID: 9423247

[11] Gastwirth, J.L. On robust procedures. *J. Am. Stat. Assoc.,* **1966**, *61*, 929-948.
http://dx.doi.org/10.1080/01621459.1966.10482185

[12] Gastwirth, J.L. The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *J. Am. Stat. Assoc.,* **1985**, *80*, 380-384.
http://dx.doi.org/10.1080/01621459.1985.10478127

[13] Feder, J.N.; Gnirke, A.; Thomas, W.; Tsuchihashi, Z.; Ruddy, D.A.; Basava, A.; Dormishian, F.; Domingo, R., Jr; Ellis, M.C.; Fullan, A.; Hinton, L.M.; Jones, N.L.; Kimmel, B.E.; Kronmal, G.S.; Lauer, P.; Lee, V.K.; Loeb, D.B.; Mapa, F.A.; McClelland, E.; Meyer, N.C.; Mintier, G.A.; Moeller, N.; Moore, T.; Morikang, E.; Prass, C.E.; Quintana, L.; Starnes, S.M.; Schatzman, R.C.; Brunke, K.J.; Drayna, D.T.; Risch, N.J.; Bacon, B.R.; Wolff, R.K. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat. Genet.,* **1996**, *13*(4), 399-408.
http://dx.doi.org/10.1038/ng0896-399 PMID: 8696333

[14] Nielsen, D.M.; Ehm, M.G.; Weir, B.S. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.,* **1998**, *63*(5), 1531-1540.
http://dx.doi.org/10.1086/302114 PMID: 9867708

[15] Deng, H.W.; Chen, W.M.; Recker, R.R. QTL fine mapping by measuring and testing for Hardy-Weinberg and linkage disequilibrium at a series of linked marker loci in extreme samples of populations. *Am. J. Hum. Genet.,* **2000**, *66*(3), 1027-1045.
http://dx.doi.org/10.1086/302804 PMID: 10712216

[16] Falconer, D.S. *Introduction to Quantitative Genetics,* 3rd ed; Longman Scientific & Technical: Harlow, **1989**.

[17] Rey, A. *L'examen Clinique en psychologie*; Presses Universitaires de France: Paris, **1964**.

[18] Estévez-González, A.; Kulisevsky, J.; Boltes, A.; Otermín, P.; García-Sánchez, C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: comparison with mild cognitive impairment and normal aging. *Int. J. Geriatr. Psychiatry,* **2003**, *18*(11), 1021-1028.
http://dx.doi.org/10.1002/gps.1010 PMID: 14618554

[19] Wang, S.M.; Lee, Y.C.; Ko, C.Y.; Lai, M.D.; Lin, D.Y.; Pao, P.C.; Chi, J.Y.; Hsiao, Y.W.; Liu, T.L.; Wang, J.M. Increase of zinc finger protein 179 in response to CCAAT/enhancer binding protein delta conferring an antiapoptotic effect in astrocytes of Alzheimer's disease. *Mol. Neurobiol.,* **2015**, *51*(1), 370-382.
http://dx.doi.org/10.1007/s12035-014-8714-9 PMID: 24788683

[20] Zhang, X.; Li, H.; Mao, Y.; Li, Z.; Wang, R.; Guo, T.; Jin, L.; Song, R.; Xu, W.; Zhou, N.; Zhang, Y.; Hu, R.; Wang, X.; Huang, H.; Lei, Z.; Niu, G.; Irwin, D.M.; Tan, H. An over expression APP model for anti-Alzheimer disease drug screening created by zinc finger nuclease technology. *PLoS One,* **2013**, *8*(11), e75493.
http://dx.doi.org/10.1371/journal.pone.0075493 PMID: 24223114

[21] Kozlitina, J.; Schucany, W.R. A robust distribution-free test for genetic association studies of quantitative traits. *Stat. Appl. Genet. Mol. Biol.,* **2015**, *14*(5), 443-464.
http://dx.doi.org/10.1515/sagmb-2014-0050 PMID: 26426896

[22] Wenjun, X.; You, S.; Juan, D. Empirical likelihood-based robust tests for genetic association analysis with quantitative traits. *J. Appl. Stat.,* **2017**, *44*(16), 2923-2935.
http://dx.doi.org/10.1080/02664763.2016.1266469