

Article

Detecting the Common and Individual Effects of Rare Variants on Quantitative Traits by Using Extreme Phenotype Sampling

Ya-Jing Zhou ^{1,2}, Yong Wang ^{1,*} and Li-Li Chen ^{1,2}

Received: 29 September 2015; Accepted: 5 January 2016; Published: 14 January 2016
Academic Editor: Montserrat Corominas

¹ Department of Mathematics, School of Science, Harbin Institute of Technology, Harbin 150001, China; zhouyajing_0321@163.com (Y.-J.Z.); chenlili_02_06@163.com (L.-L.C.)

² School of Mathematical Sciences, Heilongjiang University, Harbin 150080, China

* Correspondence: mathwy@hit.edu.cn; Tel./Fax: +86-451-8641-4208

Abstract: Next-generation sequencing technology has made it possible to detect rare genetic variants associated with complex human traits. In recent literature, various methods specifically designed for rare variants are proposed. These tests can be broadly classified into burden and nonburden tests. In this paper, we take advantage of the burden and nonburden tests, and consider the common effect and the individual deviations from the common effect. To achieve robustness, we use two methods of combining p -values, Fisher's method and the minimum- p method. In rare variant association studies, to improve the power of the tests, we explore the advantage of the extreme phenotype sampling. At first, we dichotomize the continuous phenotypes before analysis, and the two extremes are treated as two different groups representing a dichotomous phenotype. We next compare the powers of several methods based on extreme phenotype sampling and random sampling. Extensive simulation studies show that our proposed methods by using extreme phenotype sampling are the most powerful or very close to the most powerful one in various settings of true models when the same sample size is used.

Keywords: association study; extreme sampling; random sampling; rare variants

1. Introduction

Hundreds of common genetic variants associated with many complex diseases and human traits have been successfully identified by the genome-wide association studies (GWAS). However, these common genetic variants with minor allele frequencies (MAF) $>3\%$ have small to moderate effects, and explain only a small fraction of disease heritability for common disease [1–4]. Thus, it has been hypothesized that rare variants with MAF $<3\%$ may account for some of the missing heritability [5–10]. Next-generation sequencing technology will soon sequence the whole genome of large groups of individuals and thus will make testing rare variants possible. Unfortunately, rare variants are difficult to detect even with large sample size. Thus, we need to develop powerful study designs.

Various methods have been proposed to detect association between rare variants and complex diseases. These tests can be broadly classified into burden and nonburden tests. Burden tests collapse multiple rare variants in a genetic region into a single variant, and then test the association between the single variant and the trait of interest. Burden tests include the cohort allelic sums test (CAST) [11], the combined multivariate and collapsing (CMC) method [12], the weighted methods [13], and the variable minor allele frequency threshold method [14]. The same strategy is used in many methods [15–18]. In fact, burden tests detect the common effect of all rare variants in a region.

Thus, burden tests are powerful when the effects of all rare variants in a region are in the same direction and all variants are causal variants. However, these tests will suffer great power loss when these assumptions are violated. Nonburden tests, which are called “variance component tests”, use the kernel machine regression framework. In this framework, the effects of variants are assumed to be independently and identically distributed with a mean 0 and variance τ^2 . To test whether a set of variants is associated with the phenotype, it is equivalent to test whether the variance $\tau^2 = 0$. Examples of nonburden tests include C-alpha [19], the sequence kernel association test (SKAT) [20], the optimal SKAT (SKAT-O) [21], the mixed effects test (MiST) [22], and an optimally weighted combination of variants (TOW) [23]. Variance component tests are more powerful than burden tests when a genetic region has both protective and deleterious variants or many noncausal variants.

From the nonburden tests, we can see that the average association across variants is zero. However, unless the effect of all rare variants are in opposite directions with the same strength, and they cancel out, the average effect will not be zero. Thus, a model restricting the average effect to be zero may lose power. Thus, we use a more flexible model proposed by Wang *et al.* [24]. In this model, we take advantage of the burden and nonburden tests, and consider the common effect and the individual deviations from the common effect. In order to increase the power, we consider Fisher’s method and minimum- p method of combining p -values.

In this paper, we also explore the advantage of the extreme phenotype sampling in rare variant analysis and refine this design framework for future large-scale association studies on quantitative traits. Sampling individuals with extreme phenotypes can enrich the frequency of rare variants and therefore lead to an increase in power compared to random sampling. Recently, several statistical methods have been proposed for rare variants association study when extreme phenotypes are sampled [25–29]. Here, we use random sampling and extreme phenotype sampling, and compare the powers of different methods by the two sample techniques in the same sample size. In extreme phenotype sampling, we sample the individuals with higher trait value as cases and sample the individuals with lower trait value as controls. A logistic model is used for these “case-control” data. We conduct a large number of simulations and then analyze the type I error rates and powers of several methods.

2. Materials and Methods

2.1. Materials

We only consider the quantitative traits. Consider a sample of n individuals and p rare variants in a genomic region. For $i = 1, 2, \dots, n$, let Y_i denote the trait value of the i th individual; for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, p$, let G_{ij} denote the number of the minor alleles that the i th subject carries at the j th variant site.

2.2. Methods

Consider a linear regression model

$$Y_i = \beta_0 + \beta_1 G_{i1} + \dots + \beta_p G_{ip} + \varepsilon_i, \quad (1)$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ is the regression coefficients for $G_i = (G_{i1}, G_{i2}, \dots, G_{ip})'$, ε_i is an error term with a mean of zero and a variance of σ^2 . Testing whether there is effect of a set of rare variants on a trait is equivalent to testing the null hypothesis $H_0 : \beta = 0$, that is, $\beta_1 = \beta_2 = \dots = \beta_p = 0$. For rare variants, the likelihood ratio test with p degrees of freedom has low power.

To decrease degrees of freedom and increase the power, burden test collapses the p rare variants into a single variant. Then, the model is simplified as

$$Y_i = \theta_0 + \theta_1 C_i + \varepsilon_i, \quad (2)$$

where $C_i = \sum_{j=1}^p G_{ij}$, θ_1 represents the common effect across all rare variants. The null hypothesis of no association is $H_0 : \theta_1 = 0$. Burden score test statistic of $\theta_1 = 0$ is

$$T_B = \left(\frac{1}{\sigma^2} \sum_{i=1}^n C_i (Y_i - \bar{Y}) \right)^2, \tag{3}$$

where $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, and σ^2 is estimated by the variance of Y . When these individuals are randomly sampled, the burden test is denoted as RS_burden.

In practice, it may be more likely that the very extremes of phenotype distribution may consist of unknown genetic heterogeneity due to genes with large effects (*i.e.*, Mendelian disorders). In such cases, the corresponding variants will be enriched in the extreme sample. Thus, we think that extreme phenotype sampling will be more powerful than random sampling. For example, in clinical practice, diseases such as hypertension and obesity, are dichotomized by setting a threshold for quantitative traits. When individuals with extreme phenotypes are sampled, the high phenotypic extremes are regarded as cases and the low phenotypic extremes are regarded as controls. The logistic model for these “case-control” data is

$$\text{logit}P(Y_i = 1) = \theta_0 + \theta_1 C_i. \tag{4}$$

The test statistic of $H_0 : \theta_1 = 0$ is

$$T'_B = \left(\sum_{i=1}^n C_i (Y_i - \bar{Y}) \right)^2. \tag{5}$$

This method of using extreme phenotype sampling is denoted as ES_burden. We note that the two models assume that all rare variants have the same magnitude and the same direction effects on the phenotypes. When these assumptions are violated, the two methods can suffer from power loss.

Here, we use the following model [24]

$$Y_i = \theta_0 + \theta_1 C_i + \beta_1 G_{i1} + \dots + \beta_p G_{ip} + \varepsilon_i, \\ E(\beta_j) = 0, \text{cov}(\beta) = \sigma_\beta^2 I_p, \varepsilon_i \sim N(0, \sigma^2), \tag{6}$$

where θ_1 represents the common effect across all rare variants and is regarded as fixed effect, $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ represents the vector of individual effect deviations from the common effect and is regarded as random effect. Under this model, testing whether the rare variants influence the phenotype corresponds to testing the null hypothesis

$$H_0 : \theta_1 = 0, \sigma_\beta^2 = 0. \tag{7}$$

In the Appendix, we show that the score statistic for the null hypothesis H_0 is given by

$$S = (S_1, S_2)' = \left(\frac{1}{\sigma^2} \sum_{i=1}^n C_i (Y_i - \bar{Y}), \frac{1}{2} U'U - \frac{1}{2\sigma^2} \text{tr}(G'G) \right)' \tag{8}$$

where $U = (U_1, \dots, U_p)'$, $U_k = \frac{1}{\sigma^2} \sum_{i=1}^n G_{ik} (Y_i - \bar{Y})$ and $G = (G_1, G_2, \dots, G_n)'$. Let p_1 denote the two-sided p -value of S_1 , and p_2 denote the p -value of S_2 . To combine the p -values obtained from S_1 and S_2 , we propose to use Fisher’s method, and the test statistic of $H_0 : \theta_1 = 0, \sigma_\beta^2 = 0$ is defined as

$$T_1 = -2 \log p_1 - 2 \log p_2. \tag{9}$$

In addition, we consider other methods for combining p -values, such as the minimum- p approach. In the minimum- p approach, the test statistic of H_0 is defined as

$$T_2 = \min\{p_1, p_2\}. \quad (10)$$

The two methods of combining P-values have been studied by many authors [22,30]. When we consider random sampling, the two methods are respectively called RS_Fisher and RS_min-p.

Next, we consider extreme phenotype sampling. Dichotomizing the higher and the lower phenotypic extremes as cases and controls, the logistic regression model for these “case-control” data is

$$\begin{aligned} \text{logit}\{P(Y_i = 1)\} &= \theta_0 + \theta_1 C_i + \beta_1 G_{i1} + \cdots + \beta_p G_{ip}, \\ E(\beta_j) &= 0, \text{cov}(\beta) = \sigma_\beta^2 I_p. \end{aligned} \quad (11)$$

In the Appendix, we show that the score statistic for the null hypothesis $H_0 : \theta_1 = 0, \sigma_\beta^2 = 0$ is

$$S = (S_1, S_2)' = \left(\sum_{i=1}^n (Y_i - \bar{Y}) C_i, \frac{1}{2} U' U - \frac{1}{2} \text{tr}(\bar{Y}(1 - \bar{Y}) G' G) \right)', \quad (12)$$

where $U = (U_1, \dots, U_p)'$ and $U_k = \sum_{i=1}^n G_{ik} (Y_i - \bar{Y})$. p_3 and p_4 are respectively the p -values obtained from S_1 and S_2 . We also use the two methods of combining p -values: Fisher’s method and minimum- p method. The test statistics of $H_0 : \theta_1 = 0, \sigma_\beta^2 = 0$ is

$$T_3 = -2 \log p_3 - 2 \log p_4, \quad \text{or} \quad T_4 = \min\{p_3, p_4\}. \quad (13)$$

We denote the two methods as ES_Fisher and ES_min-p, respectively. We use permutation approach to obtain the p -value of the statistic T_j , for $j = 1, 2, 3, 4$. The permutation process is the same as that of Lin *et al.* [31].

3. Simulation and Results

3.1. Simulation Design

The GAW17 provides the Mini-Exome genotype data for simulation studies. This dataset contains genotypes and phenotypes for 697 unrelated individuals on 3205 genes. We follow the simulation set-up of Sha *et al.* [23]. Specifically, we select a gene (ADAMTS4) with 40 variants, and infer its haplotypic phases for the 697 individuals. To generate the genotypes with 40 variants for N individuals, we randomly combine two haplotypes of the 697 individuals.

To evaluate type I error rate, we generate quantitative trait values by using the model:

$$Y_i = \beta_0 + \varepsilon_i, \quad (14)$$

where $\beta_0 = 0.1$, ε_i follows a standard normal distribution. We estimate the empirical type I error rate as the proportion of p -values less than $\alpha = 0.01$ or 0.05 .

To evaluate power, we generate phenotypes for the N individuals by using the following model:

$$Y_i = \beta_0 + \beta_1 G_{i1} + \beta_2 G_{i2} + \cdots + \beta_p G_{ip} + \varepsilon_i, \quad (15)$$

where $\beta_0 = 0.1$, ε_i follows a standard normal distribution. Effects of causal variants depend on minor allele frequencies (MAF), *i.e.*, $|\beta_j| = -0.2 \log_{10}(\text{MAF}_j)$. The percentages of causal variants with $\text{MAF} < 0.03$ are assigned three values: 40%, 60%, and 80%. The percentages of causal variants with positive effect are assigned three values: 50%, 80%, and 100%. We also consider different sample sizes ($n = 500, 1000, \text{and } 2000$) and the proportions of extreme phenotypes (10% and 20%).

After the genotype and phenotype data for N individuals are simulated, for random sampling, n individuals are arbitrarily selected from the N individuals. For extreme phenotype sampling, we denote the highest $n/2$ extremes from the N individuals as cases and the lowest $n/2$ extremes as controls. The method of Wang *et al.* [24] is denoted as JOINT. The RS_Fisher, the RS_min-p, the RS_burden and the JOINT use random sampling, the ES_Fisher, the ES_min-p, and the ES_burden use extreme phenotype sampling.

3.2. Evaluation on Type I Error Rates

For type I error rates, we consider different sample sizes, different proportions of extreme phenotypes, and different significance levels. In each simulation setting, p -values are estimated by 500 permutations and type I error rates are evaluated by 1000 replications. The estimated type I error rates of the seven methods (JOINT, RS_Fisher, RS_min-p, ES_Fisher, ES_min-p, RS_burden and ES_burden) are summarized in Table 1. From Table 1, we can see that the estimated type I error rates are not significantly different from the nominal levels. Thus, all tests are valid tests.

Table 1. The estimated type I error rates for all tests.

Tails	Sample Size	α	JOINT	RS_Fisher	RS_min-p	ES_Fisher	ES_min-p	RS_Burden	ES_Burden
0.1	500	0.01	0.015	0.013	0.013	0.014	0.013	0.010	0.016
	1000	0.01	0.011	0.010	0.018	0.008	0.003	0.015	0.008
	2000	0.01	0.011	0.012	0.012	0.011	0.013	0.012	0.016
	500	0.05	0.047	0.051	0.045	0.043	0.043	0.048	0.047
	1000	0.05	0.057	0.048	0.052	0.050	0.052	0.042	0.047
	2000	0.05	0.050	0.049	0.050	0.050	0.050	0.049	0.052
0.2	500	0.01	0.008	0.005	0.007	0.009	0.011	0.009	0.015
	1000	0.01	0.012	0.012	0.012	0.011	0.010	0.010	0.013
	2000	0.01	0.009	0.013	0.013	0.014	0.011	0.018	0.018
	500	0.05	0.053	0.049	0.054	0.041	0.044	0.054	0.037
	1000	0.05	0.046	0.040	0.039	0.052	0.061	0.036	0.052
	2000	0.05	0.041	0.049	0.052	0.050	0.045	0.050	0.049

Note: "tails" represents 10% or 20% high/low extreme phenotype sampling; α represents the significance level.

3.3. Power Comparisons

For power comparisons, we consider different sample sizes, different proportions of extreme phenotypes, different percentages of causal variants, and different percentages of causal variants with positive effects. In each simulation scenario, p -values are estimated by 500 permutations and powers are evaluated using 500 replications at a significance level of 0.05. In all cases, the threshold value of rare variants is selected as 0.03.

Power comparisons of the seven tests for half risk variants and half protective variants are given in Figure 1. As shown in Figure 1, the three tests with extreme phenotype sampling are more powerful than the other four tests with random sampling. The ES_Fisher and the ES_min-p have similar powers, and they are much more powerful than the ES_burden. Among the other four tests (JOINT, RS_Fisher, RS_min-p, and RS_burden), the JOINT is the least powerful one. The RS_Fisher and RS_min-p are slightly better than the RS_burden. The powers of all tests increase with the increase of the sample size. All tests show an increase in power with the increase of the percentage of causal variants given the same sample size. In particular, as the percentage of extreme sample increases from 10% to 20%, the powers of all tests decrease, but the power difference among all tests reduces. This is because a big percentage of extreme sampling is similar to the random sampling, which makes minor allele frequencies decrease, so that the test powers suffer loss.

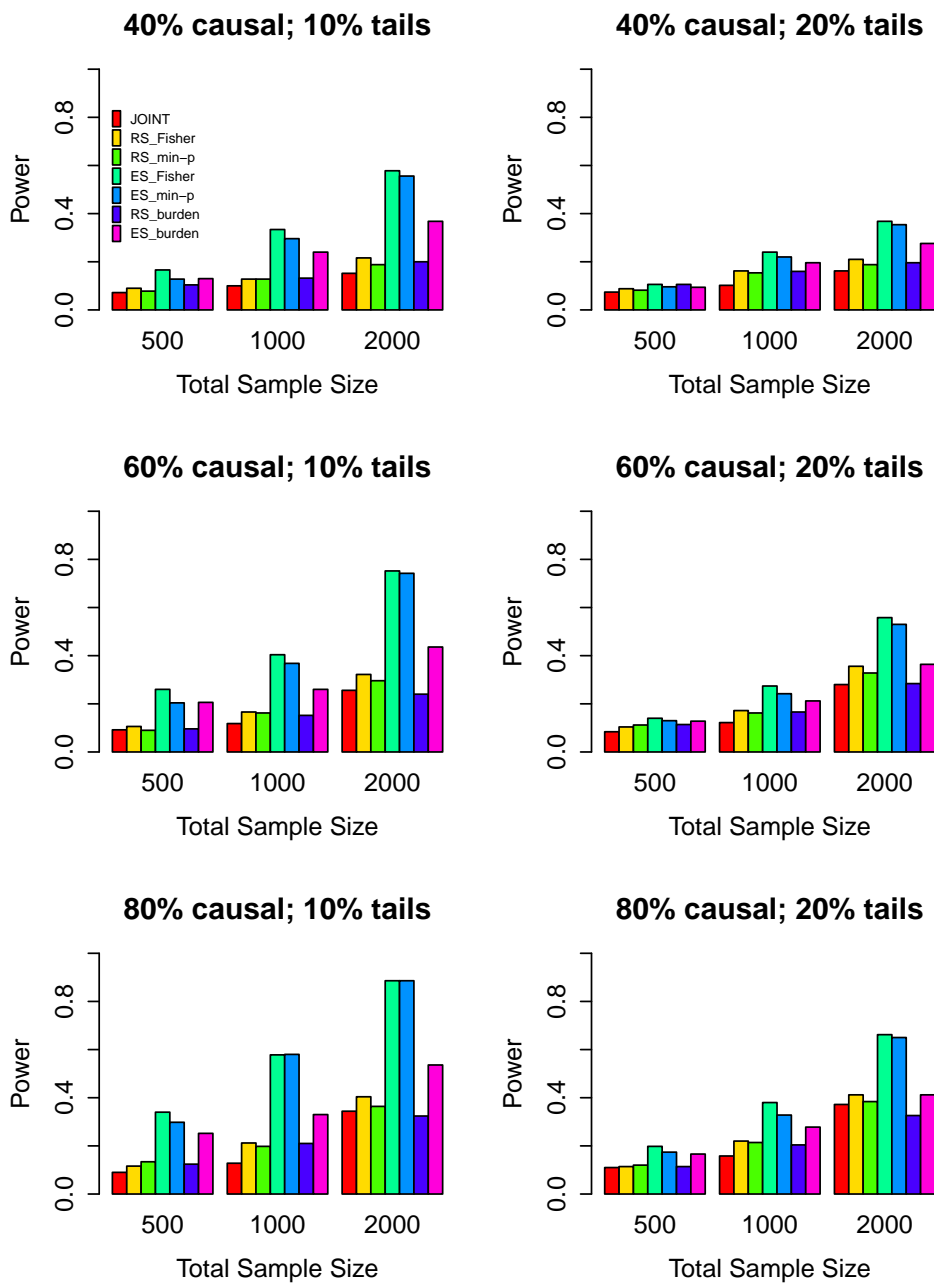


Figure 1. Power comparisons of seven tests when 50% causal variants have a positive effect on phenotype while the remaining 50% have a negative effect. The left panel considers 10% high/low extreme phenotype sampling with the three rows corresponding to 40%, 60%, and 80% causal variants. The right panel considers 20% high/low extreme phenotype sampling. Three sample sizes are considered: $n = 500, 1000, 2000$. Powers are estimated at the 0.05 significance level.

Power comparisons of the seven tests for 80% risk variants and 20% protective variants are given in Figure 2. By comparing Figure 2 with Figure 1, we can see that the powers of all tests increase uniformly and patterns of power comparisons is very similar. The difference of the ES_Fisher and the ES_burden decreases.

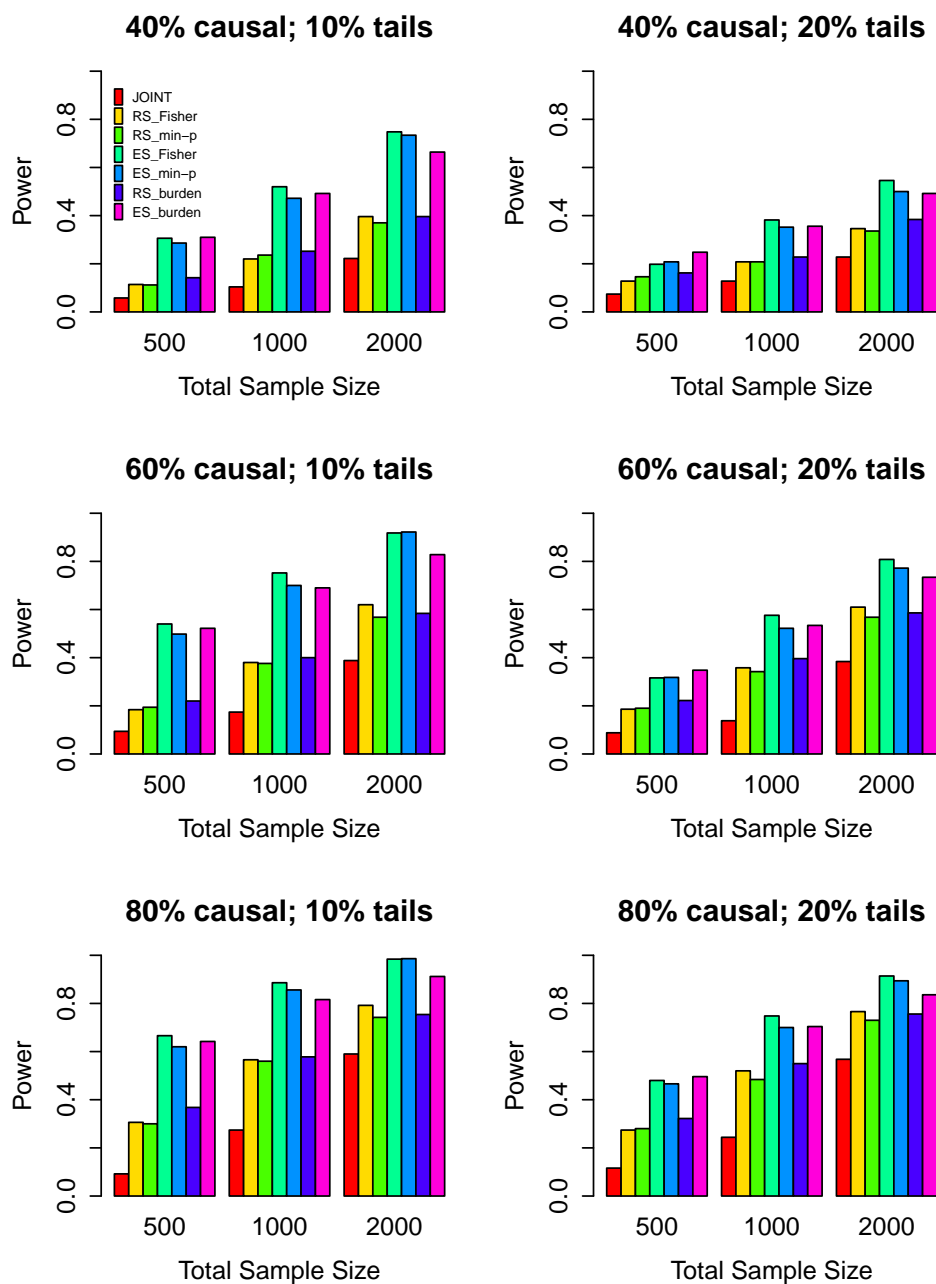


Figure 2. Power comparisons of seven tests when 80% causal variants have a positive effect on phenotype while the remaining 20% have a negative effect. The left panel considers 10% high/low extreme phenotype sampling with the three rows corresponding to 40%, 60%, and 80% causal variants. The right panel considers 20% high/low extreme phenotype sampling. Three sample sizes are considered: $n = 500, 1000, 2000$. Powers are estimated at the 0.05 significance level.

Power comparisons of the seven tests for the same direction effect of variants are given in Figure 3. The ES_Fisher, the ES_min-p and the ES_burden have similar powers. The ES_burden are slightly better than the ES_Fisher. From Figures 1–3, we can see that the difference of the ES_Fisher and the ES_burden decreases gradually. This is because the burden tests assume that variants have the same direction effects and all variants are causal, but our proposed methods allow for different direction effects of variants and also allow for the inclusion of noncausal variants. Thus, when risk and protective variants are present, the burden tests suffer substantial loss of power.

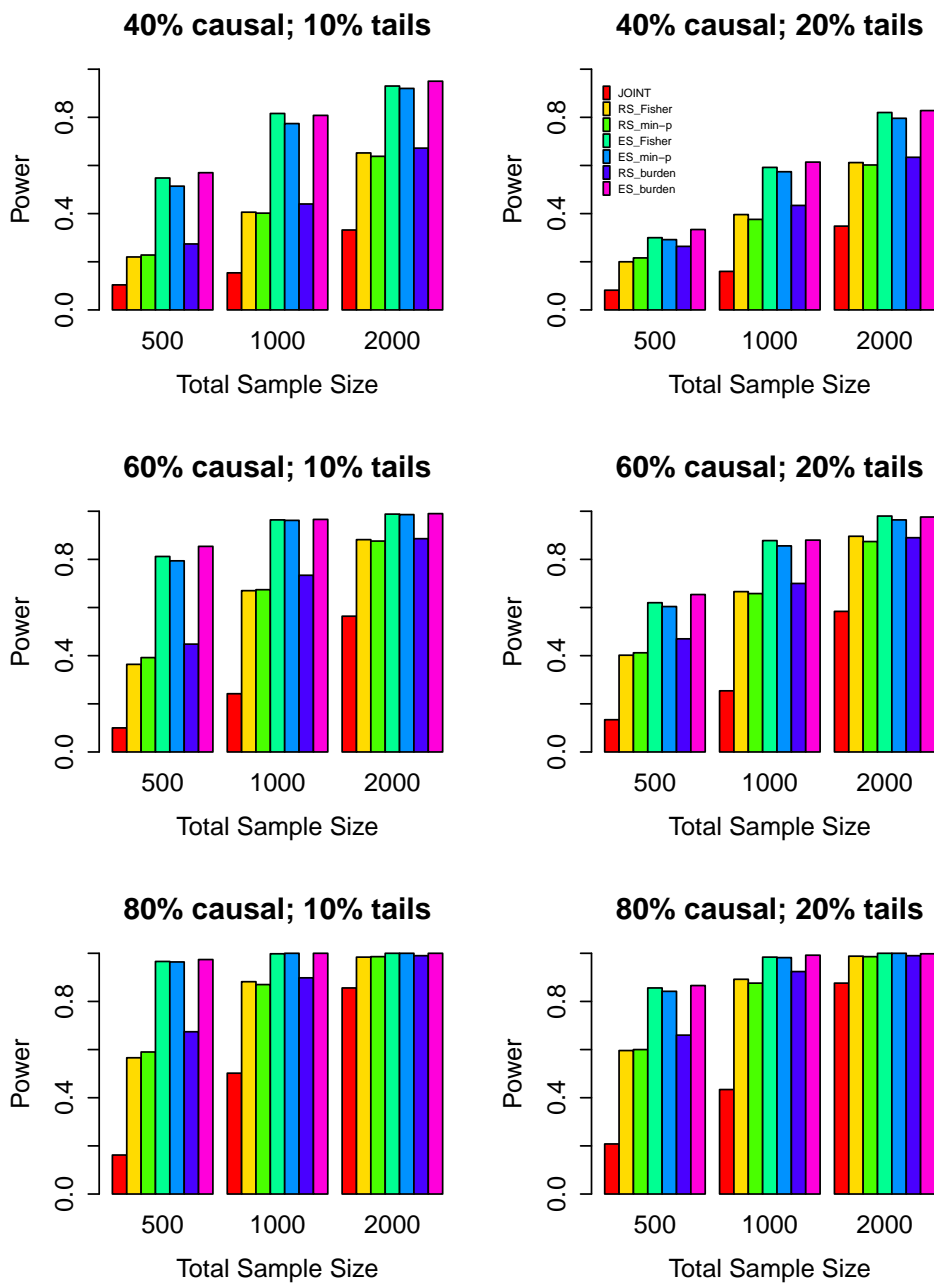


Figure 3. Power comparisons of seven tests when all causal variants have the same effect direction. The left panel considers 10% high/low extreme phenotype sampling with the three rows corresponding to 40%, 60%, and 80% causal variants. The right panel considers 20% high/low extreme phenotype sampling. Three sample sizes are considered: $n=500, 1000, 2000$. Powers are estimated at the 0.05 significance level.

In summary, the ES_Fisher and the ES_min-p are either the most powerful tests or have similar powers to the most powerful one in each setting. The powers of the ES_Fisher and the ES_min-p are relatively robust to the increase of protective variants and neutral variants. It means that in rare variants association studies, extreme phenotype sampling is superior to random sampling in the same sample size.

4. Discussion

GWAS have identified many genetic variants associated with many multifactorial diseases. However, most GWAS approaches do not consider the disease heterogeneity and the follow up functional analysis of risk variants. Recently, a new field of ‘molecular pathological epidemiology (MPE)’ has emerged as an interdisciplinary integration of ‘molecular pathology’ and epidemiology” [32]. The MPE research approach mainly examines the relationships between potential etiological factors and disease subtypes based on molecular signatures [33]. In addition, MPE also assesses the interactive effects of environmental influences and disease molecular signatures on disease progression. MPE can be one of the next steps of GWAS. Thus, the GWAS-MPE approach was proposed, to take disease heterogeneity into account following GWAS analyses [34]. In the traditional GWAS, a disease of interest is regarded as a single entity without consideration of heterogeneity. By employing the MPE approach, molecular disease classification can help to identify a specific disease subtype that is more strongly associated with a given risk variant than other subtypes of the same disease. A basic approach of MPE is a case-case approach, where diseases are classified into subtypes according to a molecular feature and then distributions of an exposure variable of interest among different subtypes are compared. Thus, in this paper, we may classify into subtypes according to a molecular feature, and then compare the distributions of an exposure variable of interest among different subtypes. We may also examine how lifestyle or genetic factors interact with the molecular features to influence prognosis or clinical outcome. This is something the authors are working on for the future.

The idea of sampling the extremes was initially proposed in linkage analysis as a way to increase efficiency [35]. However, the potential gain by sampling the extremes and technical details of this design has not been well established. For planning future large-scale association studies, we explored the advantage of extreme phenotype sampling for rare variants. In fact, Li *et al.* [28] have demonstrated the potential cost advantages of this design. In this paper, we have demonstrated that with the higher information content in the extreme sample, the performance of our proposed methods can be substantially improved in comparison with traditional designs. While clear advantages exist in applying extreme phenotype sampling for a quantitative trait, the realization of such advantages depends greatly on the underlying diseases mechanism. However, cancer or cardiovascular disease might have a more complex underlying mechanism, the use of extreme phenotype sampling may be limited, and the investigators need to evaluate the appropriateness of using underlying quantitative traits as a proxy for these disease mechanisms.

Our proposed methods easily adjust for covariates, such as age, gender, and principal components for population stratifications. When considering covariates, we use the following model

$$g(E(Y_i)) = \theta_0 + \theta_1 C_i + \beta' G_i + \alpha' X_i, \quad (16)$$

where $G_i = (G_{i1}, G_{i2}, \dots, G_{ip})'$ and $X_i = (X_{i1}, X_{i2}, \dots, X_{im})'$ are respectively the genotype and covariate of the i th subject. $g(\cdot)$ is a link function: $g(P(Y_i = 1)) = \log\{P(Y_i = 1)/P(Y_i = 0)\}$ for extreme phenotype sampling; $g(E(Y_i)) = E(Y_i)$ for random sampling.

5. Conclusions

In this paper, we propose two methods for testing whether a set of variants is associated with continuous phenotypes. We use the same model with the JOINT method, in which common effects of all rare variants and individual effect deviations from the common effect are jointly considered. However, the SKAT assumes that the average effect is zero. In fact, the average effect will not be zero unless the effects of all rare variants are in opposite directions with the same strength.

Compared with Fisher’s method and the minimum- p method, the JOINT method is the sum of standardized S_1 and standardized S_2 , but the Fisher’s method and the minimum- p method combine the p -values of S_1 and S_2 . So the Fisher’s and minimum- p methods are more powerful when only all

rare variants have common effect on the trait or when only rare variants have individual effects on the trait. When the true underlying disease model includes risk variants and protective variants, the Fisher's and minimum- p methods are more powerful than burden tests. In the same sample size, each of the three methods (Fisher, minimum- p , and burden) uses random sampling and extreme phenotype sampling. Our simulation results show that sampling from extreme phenotypes outperforms random sampling methods when the same size is used.

Acknowledgments: The authors would like to thank the joint Editor and referees for comments that greatly improved the presentation of the paper. This research was supported by the National Natural Science Foundation of China (No. 11201129). The Genetic Analysis Workshops (GAW) are supported by GAW grant R01 GM031575 from the National Institute of General Medical Sciences. Preparation of the Genetic Analysis Workshop 17 Simulated Exome Dataset was supported in part by National Institutes of Health NIH R01 MH059490 and used sequencing data from the 1000 Genomes Project (<http://www.1000genomes.org>).

Author Contributions: Ya-Jing Zhou and Yong Wang designed the study and prepared the manuscript. Ya-Jing Zhou and Li-Li Chen prepared the material of the study and performed the genotype experiments. Yong Wang revised the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix. Score Vector

Use the notation in the Methods section. Under the linear model

$$\begin{aligned} Y_i &= \theta_0 + \theta_1 C_i + \beta_1 G_{i1} + \cdots + \beta_p G_{ip} + \varepsilon_i, \\ E(\beta_j) &= 0, \text{cov}(\beta) = \sigma_\beta^2 I_p, \varepsilon_i \sim N(0, \sigma^2), \end{aligned} \quad (\text{A1})$$

the log-likelihood is given by

$$\log L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \theta_0 - \theta_1 C_i - \beta_1 G_{i1} - \cdots - \beta_p G_{ip})^2. \quad (\text{A2})$$

Then,

$$\begin{aligned} \frac{\partial \log L}{\partial \theta_1} &= \frac{1}{\sigma^2} \sum_{i=1}^n C_i (Y_i - \theta_0 - \theta_1 C_i - \beta_1 G_{i1} - \cdots - \beta_p G_{ip}), \\ \frac{\partial \log L}{\partial \beta_j} &= \frac{1}{\sigma^2} \sum_{i=1}^n G_{ij} (Y_i - \theta_0 - \theta_1 C_i - \beta_1 G_{i1} - \cdots - \beta_p G_{ip}), \\ \frac{\partial^2 \log L}{\partial \beta_j \partial \beta_l} &= -\frac{1}{\sigma^2} \sum_{i=1}^n G_{ij} G_{il}. \end{aligned} \quad (\text{A3})$$

Let $\hat{\theta}_0$ and $\hat{\sigma}^2$ denote the maximum likelihood estimates of θ_0 and σ^2 under null hypothesis $H_0: \theta_1 = 0, \sigma_\beta^2 = 0$. Then, $\hat{\theta}_0 = \bar{Y}$, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$.

Using the results in Lemma 3 of Goeman *et al.* [36], we obtain

$$\frac{\partial \log L}{\partial \sigma_\beta^2} \Big|_{H_0} = \frac{1}{2} U' U - \frac{1}{2} \text{tr}(I) \quad (\text{A4})$$

where

$$\begin{aligned} U &= \frac{\partial \log L}{\partial \beta} \Big|_{H_0} = \frac{1}{\sigma^2} \left(\sum_{i=1}^n G_{i1} (Y_i - \bar{Y}), \dots, \sum_{i=1}^n G_{ip} (Y_i - \bar{Y}) \right)', \\ I &= -\frac{\partial^2 \log L}{\partial \beta \partial \beta'} \Big|_{H_0} = \frac{1}{\sigma^2} G' G. \end{aligned} \quad (\text{A5})$$

So the score vector of H_0 is

$$S = (S_1, S_2)' = \left(\frac{\partial \log L}{\partial \theta_1}, \frac{\partial \log L}{\partial \sigma_\beta^2} \right)' \Big|_{H_0} = \left(\frac{1}{\widehat{\sigma^2}} \sum_{i=1}^n C_i (Y_i - \bar{Y}), \frac{1}{2} U' U - \frac{1}{2\widehat{\sigma^2}} \text{tr}(G'G) \right)'. \quad (\text{A6})$$

The score vector under the logistic model is similar to the score vector under the linear model.

References

1. Bansal, V.; Libiger, O.; Torkamani, A.; Schork, N.J. Statistical analysis strategies for association studies involving rare variants. *Nat. Rev. Genet.* **2010**, *11*, 773–785.
2. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **2008**, *456*, 18–21.
3. McCarthy, M.I.; Abecasis, G.R.; Cardon, L.R.; Goldstein, D.B.; Little, J.; Ioannidis, J.P.; Hirschhorn, J.N. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* **2008**, *9*, 356–369.
4. Schork, N.J.; Murray, S.S.; Frazer, K.A.; Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **2009**, *19*, 212–219.
5. Bodmer, W.; Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* **2008**, *40*, 695–701.
6. Gorlov, I.P.; Gorlova, O.Y.; Sunyaev, S.R.; Spitz, M.R.; Amos, C.I. Shifting paradigm of association studies: Value of rare single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **2008**, *82*, 100–112.
7. Ji, W.; Foo, J.N.; OafRoak, B.J.; Zhao, H.; Larson, M.G.; Simon, D.B.; Newton-Cheh, C.; State, M.W.; Levy, D.; Lifton, R.P. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat. Genet.* **2008**, *40*, 592–599.
8. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; et al. Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747–753.
9. Nejentsev, S.; Walker, N.; Riches, D.; Egholm, M.; Todd, J.A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **2009**, *324*, 387–389.
10. Pritchard, J.K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **2001**, *69*, 124–137.
11. Morgenthaler, S.; Thilly, W.G. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). *Mutat. Res.* **2007**, *615*, 28–56.
12. Li, B.; Leal, S.M. Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am. J. Hum. Genet.* **2008**, *83*, 311–321.
13. Madsen, B.E.; Browning, S.R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **2009**, *5*, e1000384.
14. Price, A.L.; Kryukov, G.V.; de Bakker, P.I.; Purcell, S.M.; Staples, J.; Wei, L.J.; Sunyaev, S.R. Pooled association tests for rare variants in exon-resequencing studies. *Am. J. Hum. Genet.* **2010**, *86*, 832–838.
15. Basu, S.; Pan, W. Comparison of statistical tests for disease association with rare variants. *Genet. Epidemiol.* **2011**, *35*, 606–619.
16. Fang, S.; Sha, Q.; Zhang, S. Two adaptive weighting methods to test for rare variant associations in family-based designs. *Genet. Epidemiol.* **2012**, *36*, 499–507.
17. Feng, T.; Elston, R.C.; Zhu, X. Detecting rare and common variants for complex traits: Sibpair and odds ratio weighted sum statistics (SPWSS, ORWSS). *Genet. Epidemiol.* **2011**, *35*, 398–409.
18. Lin, D.Y.; Tang, Z.Z. A general framework for detecting disease associations with rare variants in sequencing studies. *Am. J. Hum. Genet.* **2011**, *89*, 354–367.
19. Neale, B.M.; Rivas, M.A.; Voight, B.F.; Altshuler, D.; Devlin, B.; Orho-Melander, M.; Kathiresan, S.; Purcell, S.M.; Roeder, K.; Daly, M.J. Testing for an unusual distribution of rare variants. *PLoS Genet.* **2011**, *7*, e1001322.
20. Wu, M.C.; Lee, S.; Cai, T.; Li, Y.; Boehnke, M.; Lin, X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **2011**, *89*, 82–93.

21. Lee, S.; Emond, M.J.; Bamshad, M.J.; Barnes, K.C.; Rieder, M.J.; Nickerson, D.A.; Christiani, D.C.; Wurfel, M.M.; Lin, X. Optimal unified approach for rare variant association testing with application to small sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* **2012**, *91*, 224–237.
22. Sun, J.; Zheng, Y.; Li Hsu. A unified mixed-effects model for rare-variant association in sequencing studies. *Genet. Epidemiol.* **2013**, *37*, 334–344.
23. Sha, Q.; Wang, X.; Wang, X.; Zhang, S. Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genet. Epidemiol.* **2012**, *36*, 561–571.
24. Wang, Y.; Chen, Y.-H.; Yang, Q. Joint rare variant association test of the average and individual effects for sequencing studies. *PLoS ONE* **2012**, *7*, e32485.
25. Barnett, I. J.; Lee, S.; Lin, X. Detecting rare variant effects using extreme phenotype sampling in sequencing association studies. *Genet. Epidemiol.* **2013**, *37*, 142–151.
26. Hu, S.; Zhong, Y.; Hao, Y.; Luo, M.; Zhou, Y.; Guo, H.; Liao, W.; Wan, D.; Wei, H.; Gao, Y.; *et al.* Novel rare alleles of ABCA1 are exclusively associated with extreme high-density lipoprotein-cholesterol levels among the Han Chinese. *Clin. Chem. Lab. Med.* **2009**, *47*, 1239–1245.
27. Huang, B.E.; Lin, D.Y. Efficient association mapping of quantitative trait loci with selective genotyping. *Am. J. Hum. Genet.* **2007**, *80*, 567–576.
28. Li, D.; Lewinger, J.P.; Gauderman, W.J.; Murcray, C.E.; Conti, D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet. Epidemiol.* **2011**, *35*, 790–799.
29. Wallace, C.; Chapman, J.M.; Clayton, D.G. Improved power offered by a score test for linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. *Am. J. Hum. Genet.* **2006**, *78*, 498–504.
30. Derkach, A.; Lawless, J.F.; Sun, L. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet. Epidemiol.* **2013**, *37*, 110–121.
31. Lin, W.-Y.; Lou, X.-Y.; Gao, G.; Liu, N. Rare variant association testing by adaptive combination of p -values. *PLoS ONE* **2014**, *9*, e85728.
32. Ogino, S.; Chan, A.T.; Fuchs, C.S.; Giovannucci, E. Molecular pathological epidemiology of colorectal neoplasia: An emerging transdisciplinary and interdisciplinary field. *Gut* **2011**, *60*, 397–411.
33. Ogino, S.; Lochhead, P.; Chan, A.T.; Nishihara, R.; Cho, E.; Wolpin, B.M.; Meyerhardt, J.A.; Meissner, A.; Schernhammer, E.S.; Fuchs, C.S.; *et al.* Molecular pathological epidemiology of epigenetics: Emerging integrative science to analyze environment, host, and disease. *Mod. Pathol.* **2013**, *26*, 465–484.
34. Ogino, S.; Campbell, P.T.; Nishihara, R.; Phipps, A.I.; Beck, A.H.; Sherman, M.E.; Chan, A.T.; Troester, M.A.; Bass, A.J.; Fitzgerald, K.C.; *et al.* Proceedings of the second international molecular pathological epidemiology (MPE) meeting. *Cancer Causes Control* **2015**, *26*, 959–972.
35. Risch, N.; Zhang, H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* **1995**, *268*, 1584–1589.
36. Goeman, J.J.; van de Geer, S.A.; van Houwelingen, H.C. Testing against a high dimensional alternative. *J. R. Stat. Soc. B* **2006**, *68*, 477–493.

