

RESEARCH ARTICLE

# Joint Analysis of Multiple Traits Using "Optimal" Maximum Heritability Test

Zhenchuan Wang, Qiuying Sha, Shuanglin Zhang\*

Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan, 49931, United States of America

\* [shuzhang@mtu.edu](mailto:shuzhang@mtu.edu)

## Abstract

The joint analysis of multiple traits has recently become popular since it can increase statistical power to detect genetic variants and there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases. Currently, most of existing methods use all of the traits for testing the association between multiple traits and a single variant. However, those methods for association studies may lose power in the presence of a large number of noise traits. In this paper, we propose an "optimal" maximum heritability test (MHT-O) to test the association between multiple traits and a single variant. MHT-O includes a procedure of deleting traits that have weak or no association with the variant. Using extensive simulation studies, we compare the performance of MHT-O with MHT, Trait-based Association Test uses Extended Simes procedure (TATES), SUM\_SCORE and MANOVA. Our results show that, in all of the simulation scenarios, MHT-O is either the most powerful test or comparable to the most powerful test among the five tests we compared.



## OPEN ACCESS

**Citation:** Wang Z, Sha Q, Zhang S (2016) Joint Analysis of Multiple Traits Using "Optimal" Maximum Heritability Test. PLoS ONE 11(3): e0150975. doi:10.1371/journal.pone.0150975

**Editor:** Yun Li, University of North Carolina, UNITED STATES

**Received:** December 10, 2015

**Accepted:** February 22, 2016

**Published:** March 7, 2016

**Copyright:** © 2016 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information file.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Increasing evidence shows that pleiotropy, the effect of one variant on multiple traits, is a widespread phenomenon in complex diseases [1]. Furthermore, in genetic association studies of complex diseases, multiple related traits are usually measured. For example, hyperuricemia is usually present in patients with gout [2]; coronary heart disease is predicted by cytokine interleukin-6, C-reactive protein, interleukin-1, tumor necrosis factor- $\alpha$  and fibrinogen [3, 4]; and neuropsychiatric disorders depend on a range of overlapping clinical characteristics [5]. Although most published genome-wide association studies (GWASs) analyze each of the related traits separately, joint analysis of multiple traits may increase statistical power to detect genetic variants [6–9]. Thus, joint analysis of multiple traits has recently become popular.

Several statistical methods have been developed for joint analysis of multiple traits. These methods can be roughly divided into three groups: combining the univariate analysis results, regression methods, and dimension reduction methods. For combining univariate analysis results, one first conducts the univariate test by performing an association test for each trait individually and then combines the univariate test statistics or combines the p-values of the

univariate tests [2, 10–12]. Regression methods include mixed effect models [9, 13, 14], generalized estimating equation (GEE) methods [15, 16], and reverse regression methods [5, 17]. Mixed effect models can account for relatedness, population structure, and polygenic background effect, but it is computationally challenging. The GEE methods, based on a marginal regression model, allow the variant having different effect sizes and effect directions on different traits. These methods can also accommodate covariates and different types of traits. Reverse regression methods take genotypes as the response variable and multiple traits as independent predictors, therefore, reverse regression models do not need to know the complex distributions of traits and can be applied to a large number of mixed types of traits. Dimension reduction methods include canonical correlation analysis (CCA) [18], principal components of traits (PCT) [19], and principal components of heritability (PCH) [20–23]. CCA is to seek a linear combination of multiple variants and a linear combination of multiple traits such that the correlation between the two linear combinations reaches its maximum. The PCT methods are usually based on the first PC or first few PCs of the traits [22, 24]. However, as Aschard et al. [2014] showed that testing only the first few PCs often has low power, whereas combining signals across all PCs can have greater power. Nevertheless, it is not clear how many PCs are needed, and how robust these methods are when there exists noise traits. PCH is to find a linear combination of multiple traits such that this linear combination has the maximum heritability.

In this article, we first propose a maximum heritability test (MHT). Based on MHT, we develop an “optimal” maximum heritability test (MHT-O) to test the association between multiple traits and a single variant. In each step of MHT-O, we delete one trait that has the weakest association with the variant. Then, we find the optimal number of traits and use MHT to test the association between the optimal number of traits and the variant. Using extensive simulation studies, we compare the performance of MHT-O with MHT, Trait-based Association Test uses Extended Simes procedure (TATES) [11], SUM\_SCORE and MANOVA [8]. Our results show that, in all of the simulation scenarios, MHT-O is either the most powerful test or comparable to the most powerful test among the five tests we compared.

## Method

We consider a sample with  $n$  unrelated individuals. Each individual has  $K$  (potentially correlated) traits and has been genotyped at one variant. Let  $Y = (Y_1, \dots, Y_K)^T$  denote the random vector of  $K$  traits and  $X$  denote the random variable of the genotype score at a variant. Let  $y_i = (y_{i1}, \dots, y_{iK})^T$  denote the values of  $K$  traits and  $x_i$  denote the genotype score of the  $i^{th}$  individual, where  $x_i$  is the number of minor alleles that the  $i^{th}$  individual has at the variant. We can consider that  $y_1, \dots, y_n$  is a random sample from  $Y$  and  $x_1, \dots, x_n$  is a random sample from  $X$ .

Now, let us consider linear models

$$Y_k = \alpha_k + \beta_k X + \varepsilon_k \quad (k = 1, \dots, K).$$

We partition the total phenotypic covariance of  $Y$  as  $V_P = V_G + V_R$  [25];  $V_G = \text{var}[\beta_1 X, \dots, \beta_K X] = \text{var}(X)\beta\beta^T$  is the genetic variance due to the genotype scores  $X$ , where  $\beta = (\beta_1, \dots, \beta_K)^T$ ;  $V_R = \text{var}[\varepsilon_1, \dots, \varepsilon_K]$  is the residual covariance after removing the genetic effect.  $\text{var}(X)$  can be estimated by  $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ ,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .  $\beta$  and  $V_R$  can be estimated from the linear models

$$y_{ik} = \alpha_k + \beta_k x_i + \varepsilon_{ik} \quad (k = 1, \dots, K; i = 1, \dots, n).$$

$\beta_k$  is estimated by the least square estimator. Let  $r_{ik}$  denote the estimates of residuals  $\varepsilon_{ik}$ . Then, the  $(j, k)^{th}$  element of  $V_R$  is estimated by  $\frac{1}{n} \sum_{i=1}^n r_{ij} r_{ik}$ .

Let us consider a linear combination of  $Y$ ,  $w^T Y = \sum_{k=1}^K w_k Y_k$ , where  $w = (w_1, \dots, w_K)^T$ . The heritability of  $w^T Y$  can be written as

$$h_w^2 = \frac{w^T V_G w}{w^T V_P w}.$$

If we define  $W = V_P^{-1/2} w$ , we can write  $h_w^2$  as

$$h_w^2 = \frac{W^T V_P^{-1/2} V_G V_P^{-1/2} W}{W^T W} = \frac{W^T V W}{W^T W},$$

where  $V = V_P^{-1/2} V_G V_P^{-1/2}$ . The heritability of  $w^T Y$  depends on  $w$  and we can find a linear combination of  $w^T Y$  that has the largest heritability among all linear combinations of  $Y$ . We define the maximum heritability as the test statistic to test the association between these  $K$  traits and the variant. We denote this test as maximum heritability test (MHT). The MHT statistic can be written as

$$T_{MHT} = \max_w h_w^2 = \lambda_{\max}(V_G V_P^{-1}) = \text{var}(X) \lambda_{\max}(\beta \beta^T V_P^{-1}) = \text{var}(X) \beta^T V_P^{-1} \beta,$$

where  $\lambda_{\max}(A)$  denotes the largest eigenvalue of matrix  $A$ .

However, the test statistic  $T_{MHT}$  may lose power in the presence of a large number of noise traits. Therefore, we propose an ‘‘optimal’’ maximum heritability test (MHT-O) to test the association between multiple traits and the variant. MHT-O includes a procedure of deleting traits that have weak or no association with the variant. It has the following steps:

- Step 1. Given traits  $Y = (Y_1, \dots, Y_K)$ , initialize  $r = K$  and  $Y^{(r)} = Y$ . Denote  $T_{MHT, r}$  as  $T_{MHT}$  based on  $Y^{(r)}$ .
- Step 2. Denote  $T_{MHT, r}^{-i}$  as  $T_{MHT}$  based on  $Y^{(r)}$  with the  $i^{th}$  trait deleted for  $i = 1, \dots, r$ ; denote  $I = \arg \max_i T_{MHT, r}^{-i}$  and  $T_{MHT, r-1} = T_{MHT, r}^{-I}$ . Let  $Y^{(r-1)}$  denote  $Y^{(r)}$  with the  $I^{th}$  trait deleted and update  $r = r - 1$ .
- Step 3. Repeat step 2 until  $r = 1$ .

Denote  $p_r$  as the p-value of  $T_{MHT, r}$ . The test statistic of MHT-O is defined as

$$T_{MHT-O} = \min_{1 \leq r \leq K} p_r.$$

We use a permutation test to evaluate the p-value of  $T_{MHT-O}$ . Intuitively, two layers of permutations are needed to estimate  $p_r$  and the overall p-value for the test statistic  $T_{MHT-O}$ . Ge et al. [26] proposed that one layer of permutation can be used to estimate these p-values. We use the permutation procedure of Ge et al. to estimate  $p_r$  and the overall p-value for the test statistic  $T_{MHT-O}$ . In details, we randomly shuffle the genotypes in each permutation. Suppose we perform  $B$  times of permutations. Let  $T_{MHT, r}^{(b)}$  denote the value of  $T_{MHT, r}$  based on the  $b^{th}$  permuted data, where  $b = 0$  represents the original data. Then, we transfer  $T_{MHT, r}^{(b)}$  to  $p_r^{(b)}$  by

$$p_r^{(b)} = \frac{\#\{d : T_{MHT, r}^{(d)} > T_{MHT, r}^{(b)} \text{ for } d = 0, 1, \dots, B\}}{B}.$$

Let  $p^{(b)} = \min_{1 \leq r \leq K} p_r^{(b)}$ , then, the p-value of  $T_{MHT-O}$  is given by

$$\frac{\#\{b : p^{(b)} < p^{(0)} \text{ for } b = 1, 2, \dots, B\}}{B}.$$

The R code of MHT-O is available at Shuanglin Zhang’s homepage <http://www.math.mtu.edu/~shuzhang/software.html>.

### Comparisons of Methods

We compare our proposed method with MHT, TATES [11], MANOVA [8], and SUM\_SCORE. TATES combines p-values obtained in a standard univariate GAWS to acquire one trait-based p-value, while correcting for correlations between components. SUM\_SCORE performs an association test for each trait individually to obtain the univariate score test statistic for each trait. Then, the test statistic of SUM\_SCORE is the summation of the univariate score test statistics. We use asymptotic distributions to evaluate the p-values of SUM\_SCORE, TATES and MANOVA.

### Simulation

To evaluate the type I error rates and powers of MHT and MHT-O, we generate genotypes according to minor allele frequency (MAF) and assume Hardy Weinberg equilibrium. Then, we generate  $K$  traits by the factor model [11, 19]

$$y = \lambda x + c\gamma f + \sqrt{1 - c^2} \times \varepsilon, \tag{1}$$

where  $y = (y_1, \dots, y_K)^T$ ;  $x$  is the genotype score at the variant of interest;  $\lambda = (\lambda_1, \dots, \lambda_K)$  is the vector of effect sizes of the genetic variant on the  $K$  traits;  $f = (f_1, \dots, f_R)^T \sim MVN(0, \Sigma)$ ,  $\Sigma = (1 - \rho)I + \rho A$ ,  $A$  is a matrix with elements of 1,  $I$  is the identity matrix, and  $\rho$  is the correlation between factors;  $\gamma$  is a  $K$  by  $R$  matrix;  $c$  is a constant number; and  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_K)^T$  is a vector of residuals, and  $\varepsilon_1, \dots, \varepsilon_K$  are independent, and  $\varepsilon_k \sim N(0, 1)$  for  $k = 1, \dots, K$ .

Based on Eq (1), we consider five models:

Model 1: There is only one factor and genotypes impact on all traits with the same effect size. That is,  $R = 1$ ,  $\lambda = (\beta, \dots, \beta)^T$ , and  $\gamma = (1, \dots, 1)^T$ .

Model 2: There are five factors and genotypes impact on one factor. That is,

$$R = 5, \lambda = \left(0, \dots, 0, \underbrace{\beta, \dots, \beta}_{K/5}\right)^T, \text{ and } \gamma = \text{diag}(D_1, D_2, D_3, D_4, D_5), \text{ where } D_i = \left(\underbrace{1, \dots, 1}_{K/5}\right)^T$$

for  $i = 1, \dots, 5$ .

Model 3: There are two factors and genotypes impact on one factor. That is,

$$R = 2, \lambda = \left(0, \dots, 0, \underbrace{\beta, \dots, \beta}_{K/2}\right)^T, \text{ and } \gamma = \text{diag}(D_1, D_2), \text{ where } D_i = \left(\underbrace{1, \dots, 1}_{K/2}\right)^T \text{ for } i = 1, 2.$$

Model 4: There are five factors and genotypes impact on one trait. That is,  $R = 5$ ,  $\lambda = (0, \dots, 0, \beta)^T$ ,

and  $\gamma = \text{diag}(D_1, D_2, D_3, D_4, D_5)$ , where  $D_i = \left(\underbrace{1, \dots, 1}_{K/5}\right)^T$  for  $i = 1, \dots, 5$ .

Model 5: There is only one factor and genotypes impact on one trait. That is,  $R = 1$ ,

$$\lambda = (0, \dots, 0, \beta)^T, \text{ and } \gamma = (1, \dots, 1)^T.$$

To evaluate type I error rates of MHT and MHT-O, we let  $\beta = 0$ . To evaluate powers, we let  $\beta > 0$ . In the simulation studies for evaluation of type I error rates and powers, we set  $MAF = 0.3$  and  $\rho = 0.2$ .

### Results

To evaluate the type I error rates of the two proposed tests (MHT and MHT-O), we consider 20 quantitative traits. We also consider different sample sizes, different significance levels, and different models. In each simulation scenario, the p-values of MHT and MHT-O are estimated by 1,000 permutations and the type I error rates of the two tests are evaluated using 10,000 replicated samples. For 10,000 replicated samples, the 95% confidence intervals (CIs) for estimated type I error rates of nominal levels 0.05 and 0.01 are (0.046, 0.054) and (0.008, 0.012), respectively (see [Appendix](#) for details). The estimated type I error rates of the two tests are summarized in [Table 1](#). From this table, we can see that 58 out of 60 (greater than 95%) estimated type I error rates are within the 95% CIs and the two estimated type I error rates (0.05415 and 0.0126) not within the 95% CIs are very close to the bound of the corresponding 95% CI, which indicates that the two tests are all valid.

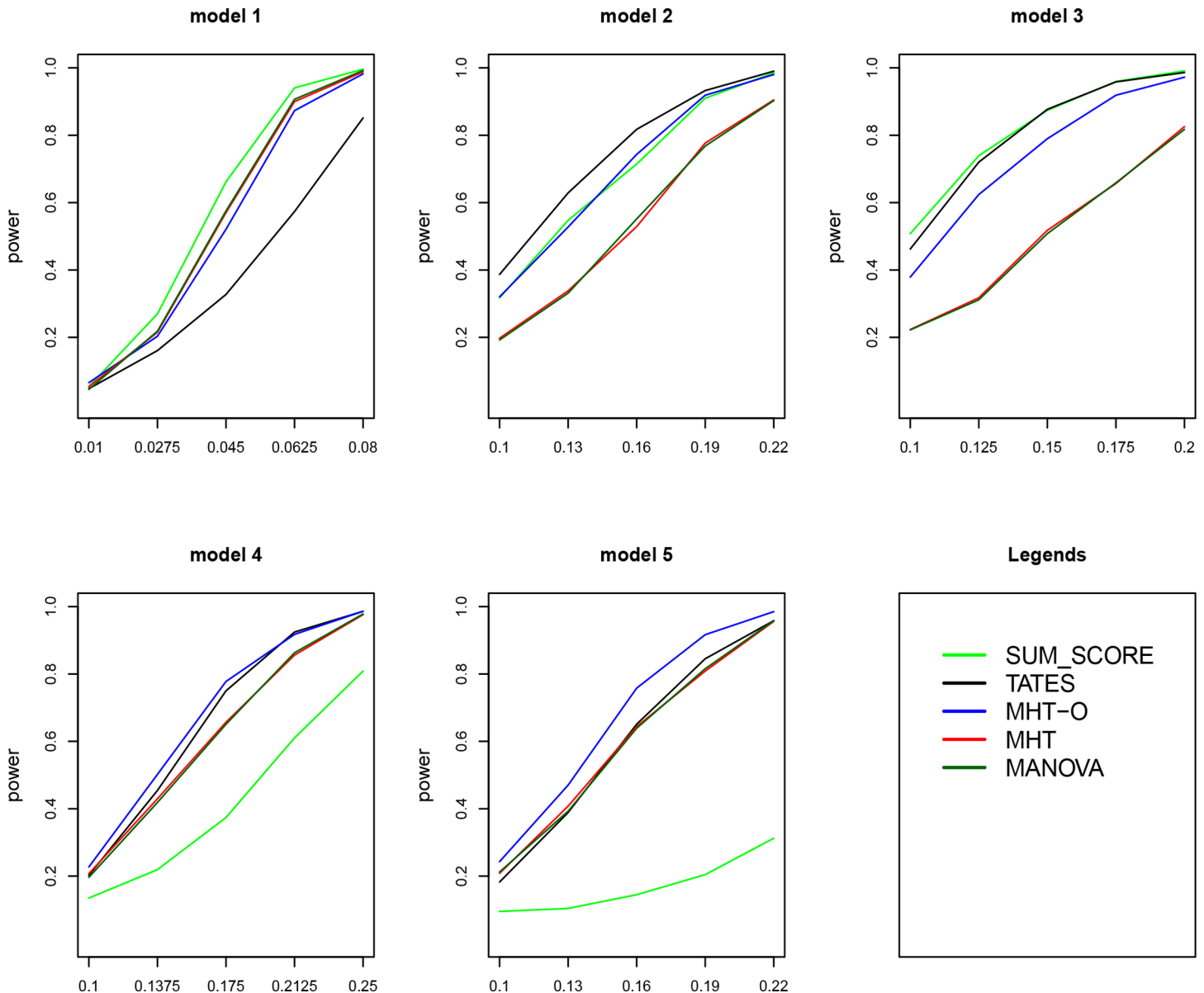
For power comparisons, we consider different values of the effect size, different models, and different numbers of traits. Sample size is 1,000 for all the cases. In each of the simulation scenarios, the p-values of MHT and MHT-O are estimated using 1,000 permutations and the p-values of SUM\_SCORE, TATES and MANOVA are estimated using their asymptotic distributions. The powers of all of the five tests are evaluated using 500 replicated samples at a significance level of 0.05.

[Fig 1](#) gives the power comparisons of the five tests (SUM\_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size based on the five models for 20 traits. This figure shows that (1) MHT-O is either the most powerful one (genotypes directly impact on a single trait: models 4–5) or comparable to the most powerful one (genotypes directly impact on all or a portion of the traits: models 1–3) among the five tests; (2) MHT and

**Table 1. The estimated type I error rates of MHT and MHT-O.** 10,000 replicates used.

		Sample size			
		500	1000	2000	
Model 1	$\alpha = 0.05$	MHT-O	0.05415	0.0494	0.04875
		MHT	0.05235	0.05005	0.0501
	$\alpha = 0.01$	MHT-O	0.01035	0.012	0.0091
		MHT	0.00985	0.01195	0.01105
Model 2	$\alpha = 0.05$	MHT-O	0.0499	0.0515	0.0526
		MHT	0.04815	0.05175	0.05285
	$\alpha = 0.01$	MHT-O	0.01045	0.01175	0.01135
		MHT	0.0117	0.0118	0.0126
Model 3	$\alpha = 0.05$	MHT-O	0.05015	0.0517	0.05315
		MHT	0.04875	0.0507	0.0529
	$\alpha = 0.01$	MHT-O	0.00995	0.0109	0.012
		MHT	0.0104	0.01035	0.012
Model 4	$\alpha = 0.05$	MHT-O	0.04815	0.0516	0.05255
		MHT	0.04875	0.05275	0.0507
	$\alpha = 0.01$	MHT-O	0.00975	0.0118	0.00975
		MHT	0.00855	0.012	0.01
Model 5	$\alpha = 0.05$	MHT-O	0.04865	0.0499	0.04975
		MHT	0.05095	0.05195	0.04755
	$\alpha = 0.01$	MHT-O	0.012	0.0119	0.00915
		MHT	0.01075	0.01115	0.0096

doi:10.1371/journal.pone.0150975.t001



**Fig 1. Power comparisons of the five tests (SUM\_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size.** Sample size is 1000. Total number of traits is 20.

doi:10.1371/journal.pone.0150975.g001

MANOVA have very similar powers; (3) MHT and MANOVA are much less powerful than other methods when genotypes directly impact on only a portion of the traits (models 2–3); (4) TATES is much less powerful than other methods when genotypes directly impact on all the traits (model 1); and (5) SUM\_SCORE is much less powerful than other methods when genotypes directly impact on a single trait (models 4–5).

Power comparisons of the five tests for 30 and 40 traits are given in Figs 2 and 3, respectively. The patterns of power comparisons for 30 and 40 traits (Figs 2 and 3) are similar to that for 20 traits (Fig 1). We also give power comparisons of the five tests using a significance level of  $5 \times 10^{-8}$  with  $10^8$  permutations and 500 replicates for 20 traits under model 1 (S1 Fig). S1 Fig shows that the patterns of the power comparisons using significance level  $5 \times 10^{-8}$  are similar to

that using a significance level of 0.05 in Fig 1 (model 1). In summary, MHT-O is either the most powerful test or comparable to the most powerful test among all the tests we compared. Therefore, our MHT-O is a robust test to a variety of models.

## Discussion

We propose MHT-O to perform joint analysis of multiple traits in association studies based on the following reasons: (1) multiple related traits are usually measured in genetic association studies of complex diseases; (2) there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases; and (3) the power of existing methods decreases in the presence of non-associated traits. The proposed MHT-O includes a procedure of deleting traits that have weak or no association with the variant. Therefore, it can be robust to the existence and the number of non-associated traits. By deleting one trait that has the weakest association with the variant in each step, MHT-O can maintain high power in the presence of a large number of non-associated traits. This feature is essentially important when there exist a large number of correlated traits but there are no guidelines to select relevant traits. Our results show that MHT-O has correct type I error rates and is either the most powerful test or comparable to the most powerful test among the five tests we compared. No other methods in the simulation studies show consistent good performance.

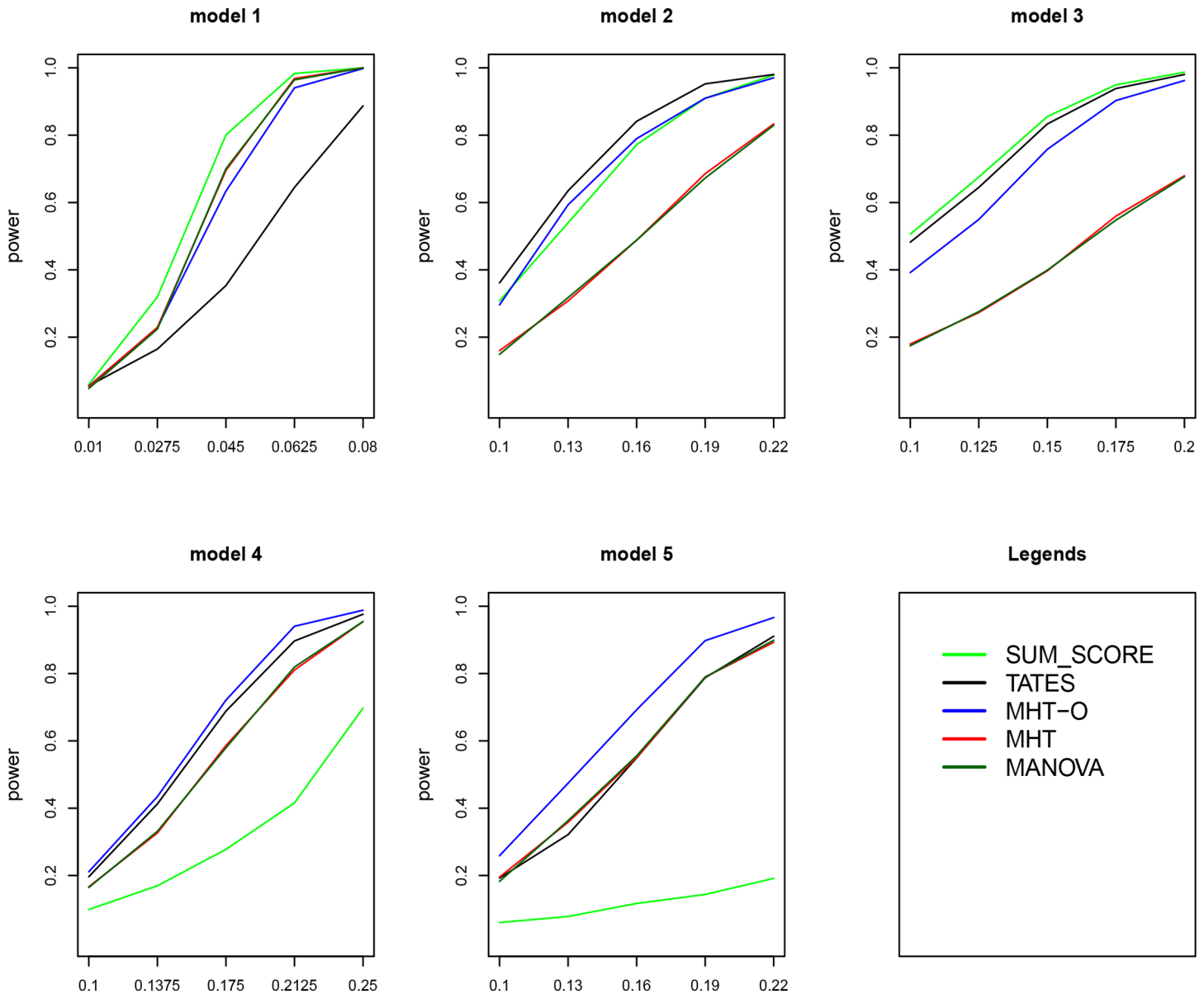
Due to the allelic heterogeneity and the extreme rarity of individual variants in rare variant association studies, the variant-by-variant methods for common variant association studies may not be optimal [27]. It has been shown by recent studies that complex diseases are caused by both common and rare variants [28–34]. Statistical methods including burden tests [27, 35–38], quadratic tests [39–41], and combined tests [42–44] have been developed for rare variant association studies with a single trait. Currently, there are limited researches on rare variant association studies for joint analysis of multiple traits [14, 45]. MHT-O can be extended to rare variant association studies by extending Eq (1) to include multiple variants. MHT-O can also be extended to family-based studies by extending Eq (1) to mixed linear model. However, the performance of MHT-O in rare variant association studies and in family-based association studies needs further investigation.

The fact that population stratification can seriously confound association results has been long recognized in association studies based on unrelated individuals [46, 47]. Several methods to control for population stratification have been developed for association studies based on unrelated individuals. These methods include principal component (PC) approach [48–52], genomic control (GC) approach [53–55], and mixed linear model (MLM) approach [29, 56]. Like most association tests based on unrelated individuals, MHT-O subjects to bias due to population stratification. To make MHT-O robust to population stratification, we can use the PC approach. Let  $P_i = (p_{i1}, \dots, p_{iL})^T$  denote the first  $L$  PCs of the genotypes at a set of genomic markers for the  $i^{th}$  individual. Let  $y_{ik}^*$  and  $x_i^*$  denote the residuals of the regressions  $y_{ik} = \alpha_{0k} + \alpha_k^T P_i + \epsilon_{ik}$  and the residuals of the regression  $x_i = \alpha_0 + \alpha^T P_i + \epsilon_i$ , respectively. Using  $y_{ik}^*$  and  $x_i^*$  to replace  $y_{ik}$  and  $x_i$ , we can make MHT-O robust to population stratification. However, the performance of using the PC approach to control for population stratification in MHT-O needs further investigations.

## Appendix

Let  $p$  denote the p-value of the test and denote a random variable

$$\xi = \begin{cases} 1, & p \leq \alpha \\ 0, & p > \alpha \end{cases},$$



**Fig 2. Power comparisons of the five tests (SUM\_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size.** Sample size is 1000. Total number of traits is 30.

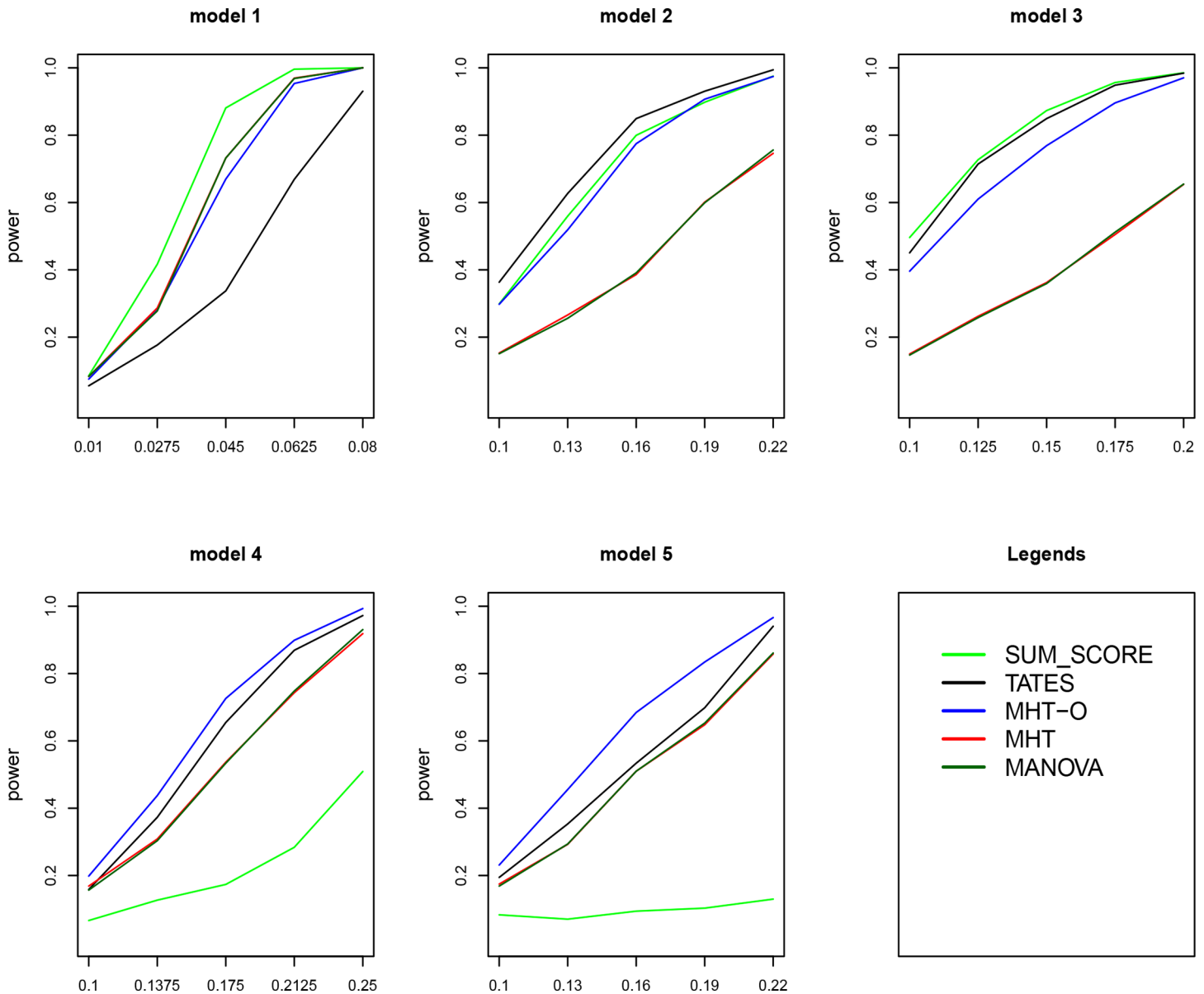
doi:10.1371/journal.pone.0150975.g002

where  $\alpha$  is the significance level. Then,  $\Pr(\xi = 1) = \alpha$  and  $\Pr(\xi = 0) = 1 - \alpha$  because  $p$  follows a uniform distribution between 0 and 1 under the null hypothesis. Suppose there are  $R$  replicates. Let  $\xi_i$  denote the value of  $\xi$  for the  $i^{th}$  replicate, where  $i = 1, \dots, R$ . Then, the estimated type I error rate is given by  $\bar{\xi} = \frac{1}{R} \sum_{i=1}^R \xi_i$  that asymptotically follows a normal distribution  $N(\alpha, \frac{\alpha(1-\alpha)}{R})$ . Thus,

$$\Pr\left(\left|\frac{\bar{\xi} - \alpha}{\sqrt{\alpha(1-\alpha)/R}}\right| \leq 1.96\right) = \Pr(\alpha - 1.96\sqrt{\alpha(1-\alpha)/R} \leq \bar{\xi} \leq \alpha + 1.96\sqrt{\alpha(1-\alpha)/R}) = 0.95.$$

We define  $(\alpha - 1.96\sqrt{\alpha(1-\alpha)/R}, \alpha + 1.96\sqrt{\alpha(1-\alpha)/R})$  as the 95% confidence interval for the estimated type I error rate for the nominal level  $\alpha$ .





**Fig 3. Power comparisons of the five tests (SUM\_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size.** Sample size is 1000. Total number of traits is 40.

doi:10.1371/journal.pone.0150975.g003

### Supporting Information

**S1 Fig. Power comparisons of the five tests (SUM\_SCORE, TATES, MHT, MHT-O and MANOVA) for the power as a function of the effect size (model 1).** Sample size is 1000. Total number of traits is 20. The significance level is  $5 \times 10^{-8}$ . The number of replicates is 500. The number of permutations is  $10^8$ . (EPS)

## Author Contributions

Conceived and designed the experiments: SZ. Performed the experiments: ZW SZ. Analyzed the data: SZ ZW. Wrote the paper: SZ ZW QS.

## References

1. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. *American journal of human genetics*. 2011; 89(5):607–18. doi: [10.1016/j.ajhg.2011.10.004](https://doi.org/10.1016/j.ajhg.2011.10.004) PMID: [22077970](https://pubmed.ncbi.nlm.nih.gov/22077970/); PubMed Central PMCID: PMC3213397.
2. Yang Q, Wu H, Guo CY, Fox CS. Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genetic epidemiology*. 2010; 34(5):444–54. doi: [10.1002/gepi.20497](https://doi.org/10.1002/gepi.20497) PMID: [20583287](https://pubmed.ncbi.nlm.nih.gov/20583287/); PubMed Central PMCID: PMC3090041.
3. Rifai N, Ridker PM. Inflammatory markers and coronary heart disease. *Current opinion in lipidology*. 2002; 13(4):383–9. PMID: [12151853](https://pubmed.ncbi.nlm.nih.gov/12151853/).
4. Yudkin JS, Kumari M, Humphries SE, Mohamed-Ali V. Inflammation, obesity, stress and coronary heart disease: is interleukin-6 the link? *Atherosclerosis*. 2000; 148(2):209–14. PMID: [10657556](https://pubmed.ncbi.nlm.nih.gov/10657556/).
5. O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FC, Elliott P, Jarvelin MR, et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS one*. 2012; 7(5):e34861. doi: [10.1371/journal.pone.0034861](https://doi.org/10.1371/journal.pone.0034861) PMID: [22567092](https://pubmed.ncbi.nlm.nih.gov/22567092/); PubMed Central PMCID: PMC3342314.
6. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nature reviews Genetics*. 2013; 14(7):483–95. doi: [10.1038/nrg3461](https://doi.org/10.1038/nrg3461) PMID: [23752797](https://pubmed.ncbi.nlm.nih.gov/23752797/); PubMed Central PMCID: PMC4104202.
7. Stephens M. A unified framework for association analysis with multiple related phenotypes. *PLoS one*. 2013; 8(7):e65245. doi: [10.1371/journal.pone.0065245](https://doi.org/10.1371/journal.pone.0065245) PMID: [23861737](https://pubmed.ncbi.nlm.nih.gov/23861737/); PubMed Central PMCID: PMC3702528.
8. Yang Q, Wang Y. Methods for analyzing multivariate phenotypes in genetic association studies. *J Probab Stat*. 2012; 2012:652569. doi: [10.1155/2012/652569](https://doi.org/10.1155/2012/652569) PMID: [24748889](https://pubmed.ncbi.nlm.nih.gov/24748889/); PubMed Central PMCID: PMC3989935.
9. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature methods*. 2014; 11(4):407–9. doi: [10.1038/nmeth.2848](https://doi.org/10.1038/nmeth.2848) PMID: [24531419](https://pubmed.ncbi.nlm.nih.gov/24531419/); PubMed Central PMCID: PMC4211878.
10. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984; 40(4):1079–87. PMID: [6534410](https://pubmed.ncbi.nlm.nih.gov/6534410/).
11. van der Sluis S, Posthuma D, Dolan CV. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS genetics*. 2013; 9(1):e1003235. doi: [10.1371/journal.pgen.1003235](https://doi.org/10.1371/journal.pgen.1003235) PMID: [23359524](https://pubmed.ncbi.nlm.nih.gov/23359524/); PubMed Central PMCID: PMC3554627.
12. Kim J, Bai Y, Pan W. An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics. *Genetic epidemiology*. 2015; 39(8):651–63. doi: [10.1002/gepi.21931](https://doi.org/10.1002/gepi.21931) PMID: [26493956](https://pubmed.ncbi.nlm.nih.gov/26493956/)
13. Korte A, Vilhjalmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature genetics*. 2012; 44(9):1066–71. doi: [10.1038/ng.2376](https://doi.org/10.1038/ng.2376) PMID: [22902788](https://pubmed.ncbi.nlm.nih.gov/22902788/); PubMed Central PMCID: PMC3432668.
14. Casale FP, Rakitsch B, Lippert C, Stegle O. Efficient set tests for the genetic analysis of correlated traits. *Nature methods*. 2015; 12(8):755–8. doi: [10.1038/nmeth.3439](https://doi.org/10.1038/nmeth.3439) PMID: [26076425](https://pubmed.ncbi.nlm.nih.gov/26076425/).
15. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986; 42(1):121–30. PMID: [3719049](https://pubmed.ncbi.nlm.nih.gov/3719049/).
16. Zhang Y, Xu Z, Shen X, Pan W, Alzheimer's Disease Neuroimaging I. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *Neuroimage*. 2014; 96:309–25. doi: [10.1016/j.neuroimage.2014.03.061](https://doi.org/10.1016/j.neuroimage.2014.03.061) PMID: [24704269](https://pubmed.ncbi.nlm.nih.gov/24704269/); PubMed Central PMCID: PMC34043944.
17. Yan T, Li Q, Li Y, Li Z, Zheng G. Genetic association with multiple traits in the presence of population stratification. *Genetic epidemiology*. 2013; 37(6):571–80. doi: [10.1002/gepi.21738](https://doi.org/10.1002/gepi.21738) PMID: [23740720](https://pubmed.ncbi.nlm.nih.gov/23740720/).
18. Tang CS, Ferreira MA. A gene-based test of association using canonical correlation analysis. *Bioinformatics*. 2012; 28(6):845–50. doi: [10.1093/bioinformatics/bts051](https://doi.org/10.1093/bioinformatics/bts051) PMID: [22296789](https://pubmed.ncbi.nlm.nih.gov/22296789/).
19. Aschard H, Vilhjalmsson BJ, Greliche N, Morange PE, Tregouet DA, Kraft P. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *American journal of human genetics*. 2014; 94(5):662–76. doi: [10.1016/j.ajhg.2014.03.016](https://doi.org/10.1016/j.ajhg.2014.03.016) PMID: [24746957](https://pubmed.ncbi.nlm.nih.gov/24746957/); PubMed Central PMCID: PMC4067564.

20. Ott J, Rabinowitz D. A principal-components approach based on heritability for combining phenotype information. *Human heredity*. 1999; 49(2):106–11. PMID: [10077732](#).
21. Lange C, van Steen K, Andrew T, Lyon H, DeMeo DL, Raby B, et al. A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Statistical applications in genetics and molecular biology*. 2004; 3:Article17. doi: [10.2202/1544-6115.1067](#) PMID: [16646795](#).
22. Klei L, Luca D, Devlin B, Roeder K. Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genetic epidemiology*. 2008; 32(1):9–19. doi: [10.1002/gepi.20257](#) PMID: [17922480](#).
23. Zhou JJ, Cho MH, Lange C, Lutz S, Silverman EK, Laird NM. Integrating multiple correlated phenotypes for genetic association analysis by maximizing heritability. *Hum Hered*. 2015; 79(2):93–104. doi: [10.1159/000381641](#) PMID: [26111731](#); PubMed Central PMCID: PMCPCMC4508328.
24. Feng T, Zhang S, Sha Q. A method dealing with a large number of correlated traits in a linkage genome scan. *BMC proceedings*. 2007; 1 Suppl 1:S84. PMID: [18466587](#); PubMed Central PMCID: PMC2367490.
25. Falconer DS, Mackay TFC. *Introduction to quantitative genetics*. 4th ed. Essex, England: Longman; 1996. xiii, 464 p. p.
26. Ge Y, Dudoit S, Speed TP. Resampling-based multiple testing for microarray data analysis. *Test*. 2003; 12(1):1–77.
27. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*. 2008; 83(3):311–21. doi: [10.1016/j.ajhg.2008.06.024](#) PMID: [18691683](#); PubMed Central PMCID: PMC2842185.
28. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*. 2008; 40(6):695–701. doi: [10.1038/ng.f.136](#) PMID: [18509313](#); PubMed Central PMCID: PMC2527050.
29. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*. 2010; 42(4):348–54. doi: [10.1038/ng.548](#) PMID: [20208533](#); PubMed Central PMCID: PMC3092069.
30. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*. 2001; 69(1):124–37. doi: [10.1086/321272](#) PMID: [11404818](#); PubMed Central PMCID: PMC1226027.
31. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant... or not? *Hum Mol Genet*. 2002; 11(20):2417–23. PMID: [12351577](#).
32. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet*. 2008; 40(1):17–22. doi: [10.1038/ng.2007.53](#) PMID: [18163131](#).
33. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet*. 2010; 19(R2):R145–51. doi: [10.1093/hmg/ddq333](#) PMID: [20705737](#); PubMed Central PMCID: PMC2953745.
34. Walsh T, King MC. Ten genes for inherited breast cancer. *Cancer Cell*. 2007; 11(2):103–5. doi: [10.1016/j.ccr.2007.01.010](#) PMID: [17292821](#).
35. Morgenthaler S, Thilly WG. 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(1–2):28–56. doi: [10.1016/j.mrfmmm.2006.09.003](#) PMID: [17101154](#).
36. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS genetics*. 2009; 5(2):e1000384. doi: [10.1371/journal.pgen.1000384](#) PMID: [19214210](#); PubMed Central PMCID: PMC2633048.
37. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, et al. Pooled association tests for rare variants in exon-resequencing studies. *American journal of human genetics*. 2010; 86(6):832–8. doi: [10.1016/j.ajhg.2010.04.005](#) PMID: [20471002](#); PubMed Central PMCID: PMC3032073.
38. Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. *American journal of human genetics*. 2010; 87(5):604–17. doi: [10.1016/j.ajhg.2010.10.012](#) PMID: [21070896](#); PubMed Central PMCID: PMC2978957.
39. Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, et al. Testing for an unusual distribution of rare variants. *PLoS genetics*. 2011; 7(3):e1001322. doi: [10.1371/journal.pgen.1001322](#) PMID: [21408211](#); PubMed Central PMCID: PMC3048375.
40. Sha Q, Wang X, Wang X, Zhang S. Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genetic epidemiology*. 2012; 36(6):561–71. doi: [10.1002/gepi.21649](#) PMID: [22714994](#).

41. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *American journal of human genetics*. 2011; 89(1):82–93. doi: [10.1016/j.ajhg.2011.05.029](https://doi.org/10.1016/j.ajhg.2011.05.029) PMID: [21737059](https://pubmed.ncbi.nlm.nih.gov/21737059/); PubMed Central PMCID: PMC3135811.
42. Sha Q, Zhang S. A rare variant association test based on combinations of single-variant tests. *Genetic epidemiology*. 2014; 38(6):494–501. doi: [10.1002/gepi.21834](https://doi.org/10.1002/gepi.21834) PMID: [25065727](https://pubmed.ncbi.nlm.nih.gov/25065727/); PubMed Central PMCID: PMC4127117.
43. Derkach A, Lawless JF, Sun L. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genetic epidemiology*. 2013; 37(1):110–21. doi: [10.1002/gepi.21689](https://doi.org/10.1002/gepi.21689) PMID: [23032573](https://pubmed.ncbi.nlm.nih.gov/23032573/).
44. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American journal of human genetics*. 2012; 91(2):224–37. doi: [10.1016/j.ajhg.2012.06.007](https://doi.org/10.1016/j.ajhg.2012.06.007) PMID: [22863193](https://pubmed.ncbi.nlm.nih.gov/22863193/); PubMed Central PMCID: PMC3415556.
45. Wang Y, Liu A, Mills JL, Boehnke M, Wilson AF, Bailey-Wilson JE, et al. Pleiotropy analysis of quantitative traits at gene level by multivariate functional linear models. *Genetic epidemiology*. 2015; 39(4):259–75. doi: [10.1002/gepi.21895](https://doi.org/10.1002/gepi.21895) PMID: [25809955](https://pubmed.ncbi.nlm.nih.gov/25809955/); PubMed Central PMCID: PMC4443751.
46. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *American journal of human genetics*. 1988; 43(4):520–6. PMID: [3177389](https://pubmed.ncbi.nlm.nih.gov/3177389/); PubMed Central PMCID: PMC1715499.
47. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994; 265(5181):2037–48. PMID: [8091226](https://pubmed.ncbi.nlm.nih.gov/8091226/).
48. Bauchet M, McEvoy B, Pearson LN, Quillen EE, Sarkisian T, Hovhannesyan K, et al. Measuring European population stratification with microarray genotype data. *American journal of human genetics*. 2007; 80(5):948–56. doi: [10.1086/513477](https://doi.org/10.1086/513477) PMID: [17436249](https://pubmed.ncbi.nlm.nih.gov/17436249/); PubMed Central PMCID: PMC1852743.
49. Chen HS, Zhu X, Zhao H, Zhang S. Qualitative semi-parametric test for genetic associations in case-control designs under structured populations. *Annals of human genetics*. 2003; 67(Pt 3):250–64. PMID: [12914577](https://pubmed.ncbi.nlm.nih.gov/12914577/).
50. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*. 2006; 38(8):904–9. doi: [10.1038/ng1847](https://doi.org/10.1038/ng1847) PMID: [16862161](https://pubmed.ncbi.nlm.nih.gov/16862161/).
51. Zhang S, Zhu X, Zhao H. On a semiparametric test to detect associations between quantitative traits and candidate genes using unrelated individuals. *Genetic epidemiology*. 2003; 24(1):44–56. doi: [10.1002/gepi.10196](https://doi.org/10.1002/gepi.10196) PMID: [12508255](https://pubmed.ncbi.nlm.nih.gov/12508255/).
52. Zhu X, Zhang S, Zhao H, Cooper RS. Association mapping, using a mixture model for complex traits. *Genetic epidemiology*. 2002; 23(2):181–96. doi: [10.1002/gepi.210](https://doi.org/10.1002/gepi.210) PMID: [12214310](https://pubmed.ncbi.nlm.nih.gov/12214310/).
53. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; 55(4):997–1004. PMID: [11315092](https://pubmed.ncbi.nlm.nih.gov/11315092/).
54. Devlin B, Roeder K, Wasserman L. Genomic control, a new approach to genetic-based association studies. *Theoretical population biology*. 2001; 60(3):155–66. doi: [10.1006/tpbi.2001.1542](https://doi.org/10.1006/tpbi.2001.1542) PMID: [11855950](https://pubmed.ncbi.nlm.nih.gov/11855950/).
55. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genetic epidemiology*. 2001; 20(1):4–16. doi: [10.1002/1098-2272\(200101\)20:1<4::AID-GEPI2>3.0.CO;2-T](https://doi.org/10.1002/1098-2272(200101)20:1<4::AID-GEPI2>3.0.CO;2-T) PMID: [11119293](https://pubmed.ncbi.nlm.nih.gov/11119293/).
56. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nature genetics*. 2010; 42(4):355–60. doi: [10.1038/ng.546](https://doi.org/10.1038/ng.546) PMID: [20208535](https://pubmed.ncbi.nlm.nih.gov/20208535/); PubMed Central PMCID: PMC2931336.