Association analysis of rare and common variants with multiple traits based on variable reduction method

LILI CHEN[1,2]*, YONG WANG[1] AND YAJING ZHOU[2]

[1]*Department of Mathematics, School of Science, Harbin Institute of Technology, Harbin 150001, China*
[2]*School of Mathematical Sciences, Heilongjiang University, Harbin 150080, China*

Summary

Pleiotropy, the effect of one variant on multiple traits, is widespread in complex diseases. Joint analysis of multiple traits can improve statistical power to detect genetic variants and uncover the underlying genetic mechanism. Currently, a large number of existing methods target one common variant or only rare variants. Increasing evidence shows that complex diseases are caused by common and rare variants. Here we propose a region-based method to test both rare and common variant associated multiple traits based on variable reduction method (abbreviated as MULVR). However, in the presence of noise traits, the MULVR method may lose power, so we propose the MULVR-O method, which jointly analyses the optimal number of traits associated with genetic variants by the MULVR method, to guard against the effect of noise traits. Extensive simulation studies show that our proposed method (MULVR-O) is applied to not only multiple quantitative traits but also qualitative traits, and is more powerful than several other comparison methods in most scenarios. An application to the two genes (*SHBG* and *CHRM3*) and two phenotypes (systolic blood pressure and diastolic blood pressure) from the GAW19 dataset illustrates that our proposed methods (MULVR and MULVR-O) are feasible and efficient as a region-based method.

1. Introduction

Genome-wide association studies (GWAS) aim to detect genetic variants associated with complex traits. Though GWAS have successfully uncovered a large number of common genetic variants in human complex diseases, these common variants can only explain a small proportion of disease heritability (Bansal *et al.*, 2010). Research has shown that rare variants are actually responsible for part of the heritability of complex disease (Manolio *et al.*, 2009). Because of low minor allele frequency (MAF) of rare variants, many methods for single common variants are underpowered to detect a single rare variant. To improve the power of rare variant association analysis, many methods test the collective effect of rare variants in a genomic region, including burden tests and non-burden tests (Li & Leal, 2008; Madsen & Browning, 2009; Price *et al.*, 2010; Neale *et al.*, 2011; Wu *et al.*, 2011).

However, almost all of the aforementioned methods have primarily focused on a single trait. In the study of complex diseases, pleiotropy is a widespread phenomenon (Sivakumaran *et al.*, 2011), and multiple correlated traits are usually measured, for example, hypertension is diagnosed by systolic blood pressure (SBP) and diastolic blood pressure (DBP); coronary heart disease is evaluated using cytokine interleukin-6, C-reactive protein, interleukin-1, tumor necrosis factor-α and fibrinogen. Joint analysis of multiple traits can improve statistical power and provide additional insights into the genetic architecture of the complex disease (Aschard *et al.*, 2014). Currently, there are many methods to jointly analyse multiple traits, for example, regression methods (Korte *et al.*, 2012; O'Reilly *et al.*, 2012; Zhou & Stephens, 2014), combining test statistics from univariate analysis (Yang *et al.*, 2010; Van Der Sluis *et al.*, 2013), and variable reduction methods (Klei *et al.*, 2008; Tang & Ferreira, 2012; Aschard *et al.*, 2014).

Though these methods can test association between one common variant and multiple traits, for rare variant association studies, they may suffer loss of power.

---

* Corresponding author: Tel: +86 451 86608282. E-mail: chen-lili_02_06@163.com

In addition, there is increasing evidence to show that complex diseases are caused by both common and rare variants (Bodmer & Bonilla, 2008; Ng *et al.*, 2009; Teer & Mullikin, 2010). In this article, we propose a region-based method to detect both rare and common variants associated with multiple traits by variable reduction method (abbreviated as MULVR). We first used the optimal weights method (TOW) proposed by Sha *et al.* (2012) to test association between each trait and genetic variants in a genomic region. Then we took the aforementioned single TOW statistic as weight to combine the original traits. Finally, we still used the TOW method to test association between the linear combination of traits and multiple variants in a genomic region. But in the presence of noise traits, our method may lose power. To guard against the effect of the noise traits, we propose the MULVR-O method which uses the optimal number of traits to detect both rare and common variants by the MULVR method. Extensive simulation studies show that our proposed method (MULVR-O) is more powerful than several other comparison methods in most scenarios. In addition, analysing two genes (*SHBG* and *CHRM3*) and two phenotypes (SBP and DBP) from the GAW19 dataset illustrates that our proposed methods (MULVR and MULVR-O) are feasible and efficient as region-based methods.

2. Materials and methods

Consider n unrelated individuals. Each individual has either K correlated quantitative traits or K correlated qualitative traits, and has been genotyped at M variants (rare or common variants) in a genomic region (a gene or a pathway). For the ith individual, y_{ik} denotes the kth trait value, $g_{im} \in \{0, 1, 2\}$ denotes the number of minor alleles at the mth variant ($i = 1, 2, \ldots, n, m = 1, 2, \ldots, M, k = 1, \ldots, K$). We propose a MULVR method to test association between M variants and K traits. The detailed steps of our method are given as follows.

First, we respectively tested association between each trait and M variants. For the kth trait, we considered generalized linear models

$$g(E(y_{ik})) = \beta_{k0} + \beta_{k1}g_{i1} + \beta_{k2}g_{i2} + \cdots + \beta_{kM}g_{iM}, i = 1, \ldots, n, \quad (1)$$

where $g(\cdot)$ is a link function, the logit function, $g(Pr(y_{ik} = 1)) = \log \frac{Pr(y_{ik} = 1)}{Pr(y_{ik} = 0)}$, for qualitative trait; and the identity function, $g(E(y_{ik})) = E(y_{ik})$, for quantitative trait. We used the TOW method proposed by Sha *et al.* (2012) to test the null hypothesis H_0: $\beta_{k1} = \beta_{k2} = \cdots = \beta_{kM} = 0$. Then the test statistic is

given by

$$T_k = \sum_{m=1}^{M} \frac{\left(\sum_{i=1}^{n}(y_{ik} - \overline{y}_k)(g_{im} - \overline{g}_m) \right)^2}{(n-1) \sum_{i=1}^{n}(g_{im} - \overline{g}_m)^2}, k$$
$$= 1, 2, \ldots, K, \quad (2)$$

where $\overline{y}_k = \frac{1}{n} \sum_{i=1}^{n} y_{ik}, \overline{g}_m = \frac{1}{n} \sum_{i=1}^{n} g_{im}$.

Second, let $y_k = (y_{1k}, y_{2k}, \ldots, y_{nk})^{\mathrm{T}}, k = 1, 2, \ldots, K$. We combined $y_1, y_2, \ldots, y_K$ with weights $T_1, T_2, \ldots, T_K$. The test statistic T_k reflects the association between the kth trait and genotypes. The stronger the association, the greater the value of statistic T_k, and the larger the weight of the kth trait y_k. Let $Y_i = \sum_{k=1}^{K} T_k y_{ik}, i = 1, 2, \ldots, n, Y = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$.

Finally, we detected association between Y and M variants. We considered generalized linear models

$$g(E(Y_i)) = \alpha_0 + \alpha_1 g_{i1} + \alpha_2 g_{i2} + \cdots + \alpha_M g_{iM}, i = 1, \ldots, n, \quad (3)$$

and still used the TOW method to test the null hypothesis H_0:$\alpha_1 = \alpha_2 = \cdots = \alpha_M = 0$, and obtained the test statistic

$$TT = \sum_{m=1}^{M} \frac{\left(\sum_{i=1}^{n}(Y_i - \overline{Y})(g_{im} - \overline{g}_m) \right)^2}{(n-1) \sum_{i=1}^{n}(g_{im} - \overline{g}_m)^2} \quad (4)$$

where $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$.

However, in the presence of noise traits, the MULVR method may lose power. To guard against the effect of noise traits, we propose the following MULVR-O method which uses the optimal number of traits to test the genetic variants by the MULVR method. In detail, we sorted test statistics $T_1, \ldots, T_k, \ldots, T_K$ in descending order and used T'_k to denote the kth largest test statistic, and accordingly denote y'_k as the trait which is used for calculating $T'_k, k = 1, 2, \ldots, K$. let $Y^{(k)} = (y'_1, \ldots, y'_k)$ denote the first k traits of $y'_1, \ldots, y'_k, \ldots, y'_K$. For each $Y^{(k)}$, we used the MULVR method to obtain the test statistic TT_k, and denoted the according P-value of TT_k as $P_{TT_k}, k = 1, \ldots, K$. The overall statistic was defined as $TP = \min_{1 \le k \le K} \{P_{TT_k}\}$. We used a permutation process to evaluate the P-values of TT_k and TP. In each permutation, we randomly shuffled the genotypes and recalculated $T_1, \ldots, T_K$ and $TT_1, \ldots, TT_K$. Suppose we perform B times of permutations. For the bth permutation, let $TT_k^{(b)}$ denote the value of $TT_k, k = 1, \ldots, K$,

Table 1. *The four compared methods.*

Methods	Definition	Pros and cons	Applications
MULVR-O	It jointly analyses the optimal number of traits associated with genetic variants by variable reduction method	Pros: it can simultaneously detect both rare and common variants and can be applied to not only multiple quantitative traits but also qualitative traits; Cons: when there is correlation between any two traits, it suffers loss of power for multiple quantitative traits in the presence of a large number of noise traits	It can simultaneously detect both rare and common variants associated with multiple quantitative traits or qualitative traits
AWRR	It uses a reverse regression model to test association between collapse genotypes and multiple traits	Pros: it does not need to know the complex distributions of the traits; Cons: for qualitative traits, it suffers loss of power; when causal variants impact on all the traits, it may lose power	It may analyse rare variants associated with multiple quantitative traits, qualitative traits or mixed types of traits
CCA	It is a multivariate approach for analysing correlation between two groups of variables by calculating Wilk's statistic	Pros: the asymptotical distribution of the CCA statistic works well for common variants without the need for permutation testing; Cons: for qualitative traits, it suffers substantial loss of power; when causal variants impact on all the traits, it may lose power	It may separately analyse exclusively uncommon or common variants associated with multiple quantitative traits
WSRR	It regresses the weighted sum of the mutation counts on multiple traits	Pros: it does not need to know the complex distributions of the traits; Cons: it is not robust to the percentage of protective variants	It may test rare variants associated with multiple quantitative traits, qualitative traits or mixed types of traits

$b = 0, 1, \ldots, B$, where $b = 0$ represents the original data. Then, we obtained P-values by

$$P^{(b)}_{TT_k} = \frac{\sum_{b'=1}^{B} I_{\{TT_k^{(b')} > TT_k^{(b)}\}}}{B}, \quad b = 0, 1, \ldots, B, \text{ and } k = 1, \ldots, K. \quad (5)$$

Let $P^{(b)}_{TT} = \min_{1 \le k \le K} P^{(b)}_{TT_k}$ for $b = 0, 1, \ldots, B$. Then the P-value of TP is given by

$$\frac{\sum_{b=1}^{B} I_{\{P^{(b)}_{TT} < P^{(0)}_{TT}\}}}{B}. \quad (6)$$

3. Simulation studies

(i) Simulation design

For simulation studies, we used the GAW17 dataset, which contains genotypes of 697 unrelated individuals on 3205 genes. Based on the simulation procedure of Sha *et al.* (2012), we chose four genes: *ELAVL4* (gene 1), *MSH4* (gene 2), *PDE4B* (gene 3) and *ADAMTS4* (gene 4) with 10, 20, 30 and 40 variants, and merged the four genes to form a super gene (Sgene) with 100 variants. According to the genotypes

of 697 individuals in the Sgene, we generated genotypes of n individuals.

To evaluate the type I error rate and power, we generated K quantitative traits by the factor model (Wang *et al.*, 2016)

$$Y = \Lambda G + \sqrt{\rho}\gamma f + \sqrt{1 - \rho}\varepsilon, \quad (7)$$

where $Y = (y_1, y_2, \ldots, y_K)^T$, $G = (g_1, \ldots, g_{N_c})^T$ is the vector of the genotype scores at the causal variants, N_c is the number of causal variants, $\Lambda = (\beta_1, \ldots, \beta_k, \ldots, \beta_K)^T_{K \times N_c}$, $\beta_k = (\beta_{k1}, \ldots, \beta_{kN_c})^T$, $f = (f_1, \ldots, f_R)^T \sim MVN(0, I)$ is a vector of R independent standard normal latent variables, I is the identity matrix, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)^T \sim MVN(0, I)$ is a vector of errors, γ is a $K \times R$ loading matrix, and ρ is a constant number. Therefore, $Y \sim MVN(\Lambda G, \Sigma)$, where $\Sigma = \rho\gamma\gamma^T + (1 - \rho)I$. According to eqn (7), we considered two models: (1) there is only one factor ($R = 1$), $\gamma = (1, \ldots, 1)^T$, and Σ is a $K \times K$ matrix whose main diagonal elements are 1 and off-diagonal elements are 0.5; (2) there are two factors ($R = 2$), $\gamma = diag(D_1, D_2)$, where $D_1 = (1, \ldots, 1)^T_{[K/2] \times 1}$, $D_2 = (1, \ldots, 1)^T_{(K-[K/2]) \times 1}$ and $\Sigma = diag(\Sigma_1, \Sigma_2)$, where Σ_1 is a $[K/2] \times [K/2]$ matrix whose main diagonal elements are 1, and off-diagonal elements are 0.5; Σ_2 is a $(K - [K/2]) \times (K - [K/2])$ matrix whose elements are similar to those of Σ_1.

Table 2. *The type I error rates.*

				Sample size		
				500	1000	2000
Quantitative traits	Model 1	$\alpha = 0.01$	MULVR	0.014	0.008	0.012
			MULVR-O	0.008	0.008	0.010
		$\alpha = 0.05$	MULVR	0.032	0.058	0.058
			MULVR-O	0.036	0.044	0.048
	Model 2	$\alpha = 0.01$	MULVR	0.006	0.008	0.016
			MULVR-O	0.012	0.006	0.002
		$\alpha = 0.05$	MULVR	0.038	0.048	0.044
			MULVR-O	0.044	0.052	0.036
Qualitative traits	Model 1	$\alpha = 0.01$	MULVR	0.006	0.014	0.006
			MULVR-O	0.010	0.010	0.006
		$\alpha = 0.05$	MULVR	0.050	0.034	0.036
			MULVR-O	0.048	0.052	0.054
	Model 2	$\alpha = 0.01$	MULVR	0.016	0.012	0.008
			MULVR-O	0.006	0.012	0.010
		$\alpha = 0.05$	MULVR	0.048	0.052	0.038
			MULVR-O	0.036	0.054	0.046

Note: α represents the significance level.
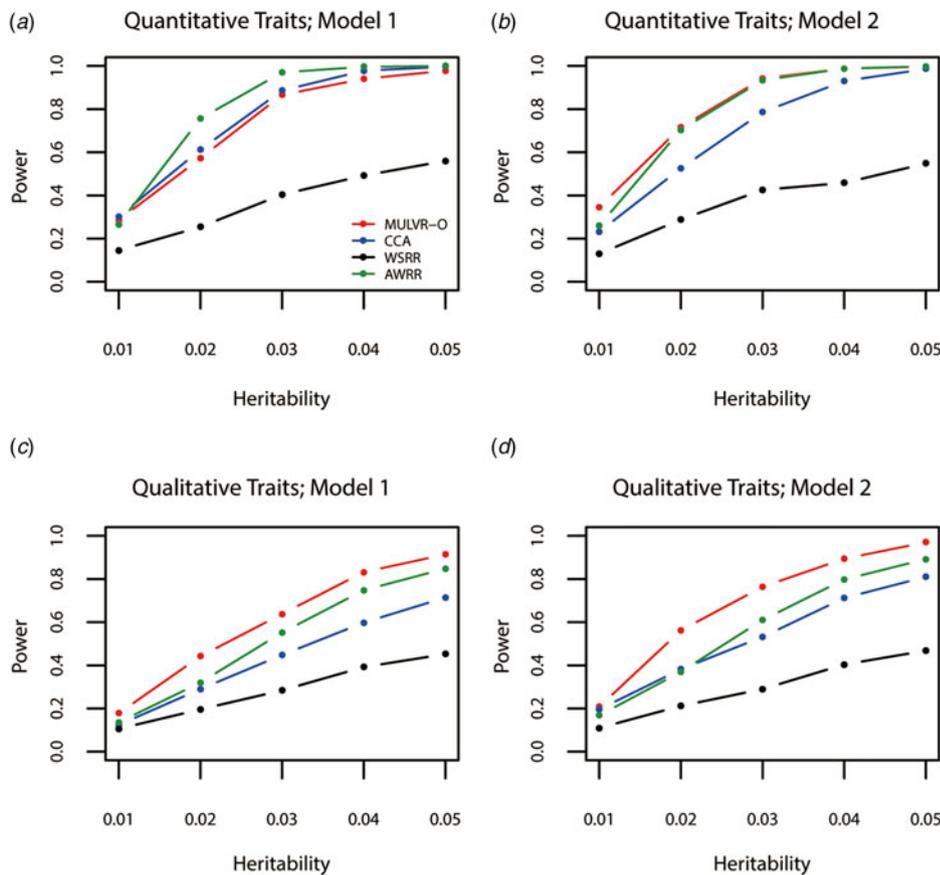
Fig. 1. Power comparisons for different values of the total heritability in two models. Total number of traits is six and causal variants impact on four traits. One common variant and 10% of rare variants are causal, and 20% of rare causal variants are protective variants. The sample size is 1000 and $\rho = 0.5$. (*a*) Multiple quantitative traits under model 1 of simulation design. (*b*) Multiple quantitative traits under model 2 of simulation design. (*c*) Multiple qualitative traits under model 1 of simulation design. (*d*) Multiple qualitative traits under model 2 of simulation design.

Based on a quantitative trait, an individual is defined as affected if the individual's corresponding quantitative trait value is at least one standard deviation larger than the phenotypic mean. We supposed a prevalence of 16% for the simulated disease in the general population. Therefore, we could generate multiple qualitative traits.

For evaluating the type I error rate, let $\beta_{kj} = 0$, $k = 1$, …, K, $j = 1$, …, N_c; for comparing power, we considered that causal variants contain both rare and common variants, and β_{kj} is a constant and its value depends on the total heritability and the ratio of the heritability of rare causal variants to the heritability of common causal variants. Suppose that the heritability of each rare causal variant is not always equal, and there is one common causal variant. Without loss of generality, our method is still applied to multiple common causal variants. We compared our proposed method (MULVR-O) with canonical correlation analysis (CCA) (Tang & Ferreira, 2012), adaptive weighting reverse regression (AWRR) (Wang *et al.*, 2016), and the weighted sum reverse regression (WSRR) (Madsen & Browning, 2009; Wang *et al.*, 2016). The definitions, pros and cons and applications for the four comparison methods are summarized in Table 1. Based on the research of Wang *et al.* (2016), we used permutation procedure to evaluate the *P*-value of the CCA statistic instead of the asymptotical distribution of the CCA statistic. The AWRR method was implemented with its R script.

(ii) *Evaluation on type I error rates*

For evaluating type I error rates, *P*-values were estimated by 1000 permutations and type I error rates were evaluated by 500 replications. Table 2 summarizes the estimated type I error rates for different types of traits, different sample sizes, different significance levels and two different models, and shows that the MULVR and MULVR-O methods can control type I error rate.

(iii) *Power comparisons*

For power comparisons, we considered two different types of traits and models. For each type of trait and each model, we considered different values of heritability, different percentages of protective variants,
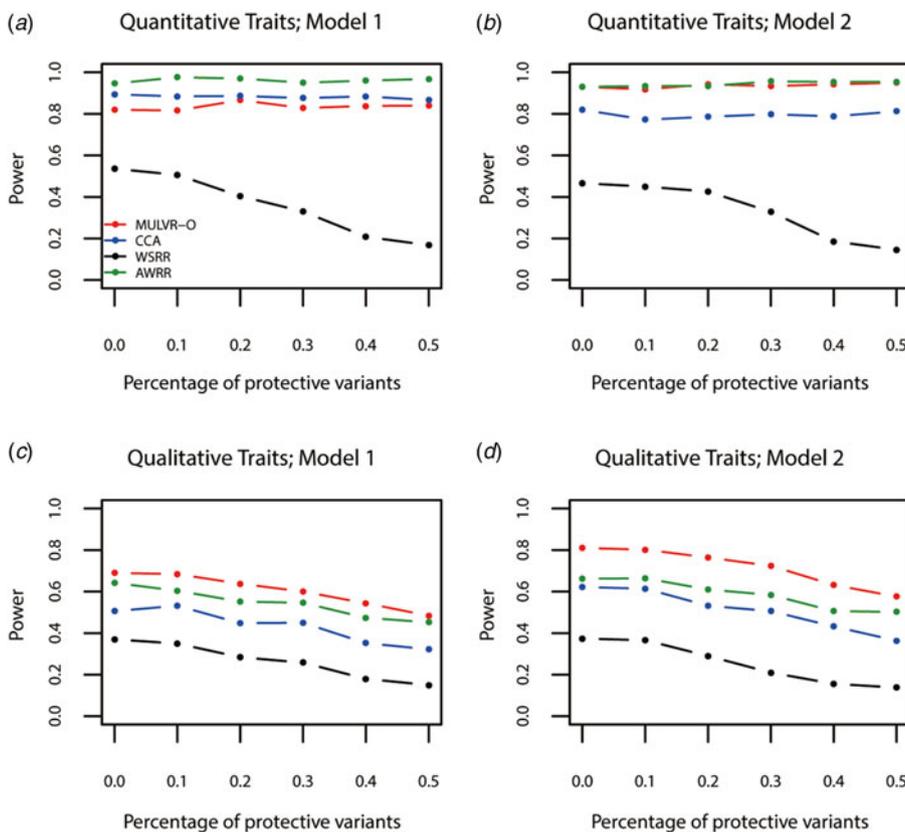
Fig. 2. Power comparisons for different percentages of protective variants in two models. Total number of traits is six and causal variants impact on four traits. One common variant and 10% of rare variants are causal, and the total heritability of all causal variants is 0.03. The sample size is 1000 and $\rho = 0.5$. (*a*) Multiple quantitative traits under model 1 of simulation design. (*b*) Multiple quantitative traits under model 2 of simulation design. (*c*) Multiple qualitative traits under model 1 of simulation design. (*d*) Multiple qualitative traits under model 2 of simulation design.

different percentages of causal variants, different numbers of associated traits and different sample sizes. In each simulation, *P*-values were estimated by 1000 permutations and powers were evaluated by 500 replications at a significance level of 0.05. We first considered the performances of the four methods (MULVR-O, CCA, WSRR and AWRR) in the presence of noise traits.

Power comparisons for different values of heritability are given by Fig. 1. As shown in Fig. 1, powers of all methods increase with the increasing heritability. Figure 1(*a*) shows that AWRR is the most powerful, followed by CCA and MULVR-O, and MULVR-O is closely comparable to CCA. Model 1 indicates that there is correlation between any two traits. Because the MULVR-O method uses the weighted combination of original traits, for quantitative traits, this phenotypic correlation affects MULVR-O to exclude the noise traits, and thus it suffers loss of power. Figure 1(*b*) shows that MULVR-O and AWRR perform similarly with power larger than CCA. Figures 1(*c*) and (*d*) show that, for qualitative traits, MULVR-O performs the best, followed by AWRR, and CCA suffers loss of power, because it

is designed for quantitative traits. Without considering the direction of effect of variants, WSRR performs the worst.

Figure 2 shows power comparisons for different percentages of protective variants. For quantitative traits, except for the WSRR method, the other methods are robust to the percentage of protective variants. For qualitative traits, the powers of all methods decrease with the increase of the percentage of protective variants. According to the researches of Wu *et al.* (2011) and Wang *et al.* (2016), the reason is that protective variants lower MAFs in cases and thus make observing rare variants in cases more difficult. The decrease of power of WSRR is due to the sensitivity to the direction of the effect of variants.

Power comparisons for different percentages of causal variants are shown in Fig. 3. As seen in Fig. 3, the three methods (MULVR-O, CCA and AWRR) are relatively robust to the percentage of causal variants, while the power of the WSRR method increases with the increasing percentage of causal variants in most situations.

Figure 4 shows power comparisons for different numbers of traits impacted by causal variants. As
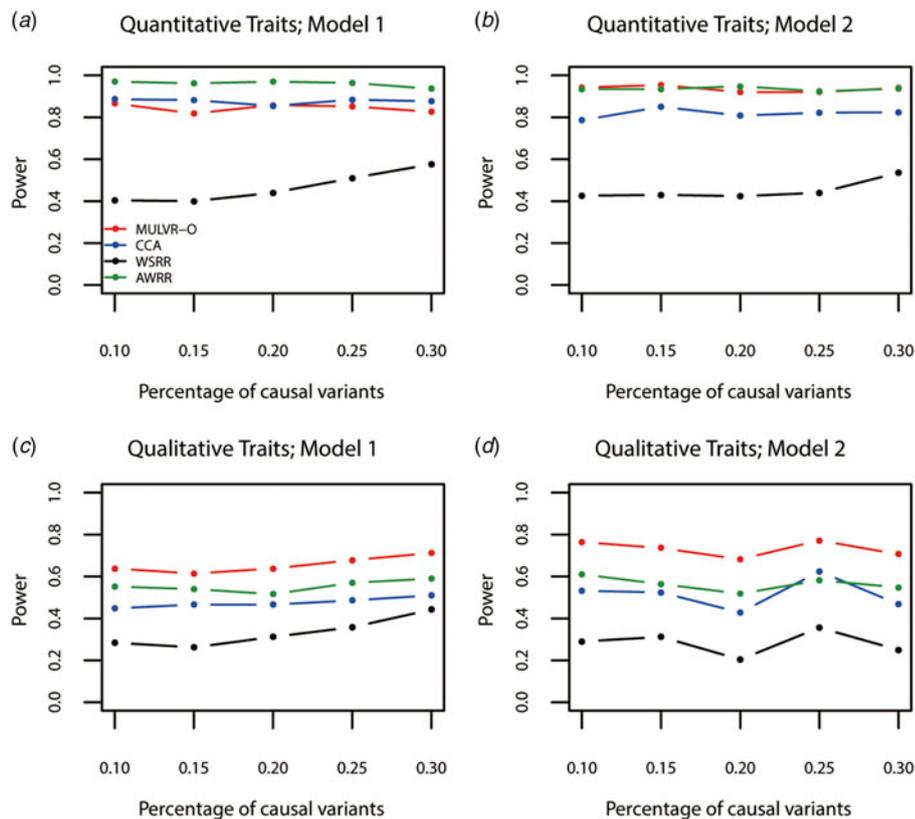
Fig. 3. Power comparisons for different percentages of rare causal variants in two models. Total number of traits is six and causal variants impact on four traits. One common variant is causal, 20% of rare causal variants are protective variants, and the total heritability of all causal variants is 0.03. The sample size is 1000 and $\rho = 0.5$. (*a*) Multiple quantitative traits under model 1 of simulation design. (*b*) Multiple quantitative traits under model 2 of simulation design. (*c*) Multiple qualitative traits under model 1 of simulation design. (*d*) Multiple qualitative traits under model 2 of simulation design.
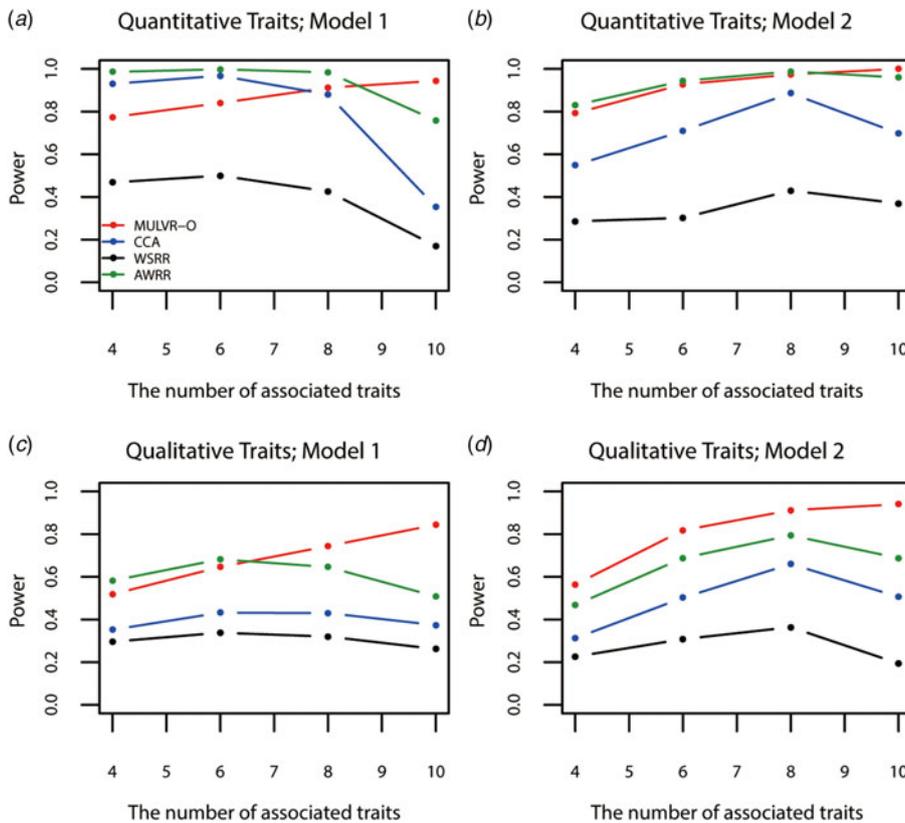
Fig. 4. Power comparisons for different numbers of traits influenced by causal variants in two models. Total number of traits is ten. One common variant and 10% of rare variants are causal, 20% of rare causal variants are protective variants, and the total heritability of all causal variants is 0.03. The sample size is 1000 and $\rho = 0.5$. (*a*) Multiple quantitative traits under model 1 of simulation design. (*b*) Multiple quantitative traits under model 2 of simulation design. (*c*) Multiple qualitative traits under model 1 of simulation design. (*d*) Multiple qualitative traits under model 2 of simulation design.

shown in Fig. 4, the powers of all methods increase with increasing number of traits associated with causal variants, until causal variants impact on all traits, MULVR-O reaches maximum power, while the other three methods suffer loss of power. This performance of CCA is consistent with previous reports (Allison *et al.*, 1998; Evans & Duffy, 2004; Ferreira & Purcell, 2009). AWRR and WSRR use reverse regression method in common. When causal variants affect all traits, the performances of losing power of the two methods coincide with that of reverse regression method reported by Kim & Pan (2017). Figure 4(*a*) shows that in model 1, AWRR and CCA are more powerful than MULVR-O in the presence of a large number of noise traits. For quantitative traits, the correlation between any two traits affects the MULVR-O method to exclude the noise traits, and thus it suffers loss of power. AWRR regresses the collapse genotypes on multiple traits, so it depends on those traits associated with causal variants, and is robust to the inclusion of noise traits. According to Tang & Ferreira (2012), when the causal variants influence only a subset of all traits, CCA has larger power. The other three figures show that MULVR-O is either the most powerful test or comparable to the most powerful test.

Power comparisons for different sample sizes are given in Fig. 5. This figure shows that powers of all methods increase with increasing sample sizes.

When all traits are associated with causal variants, we also compare the powers of the four methods (MULVR, CCA, WSRR and AWRR) for quantitative traits and qualitative traits. These results are given by Supplementary Figures S1–S5. Figure S1 shows that powers of all methods increase with the increasing heritability, MULVR is consistently more powerful than the other three methods, CCA and AWRR suffer loss of power and WSRR is the least powerful. The variation trend of powers in Fig. S2, S3 and S5 are similar to that in Fig. 2, 3 and 5. As shown in Fig. S4, the power of the CCA method decreases with the increasing number of traits, while the powers of the other three methods do not change relatively.

Throughout the simulations, we observed that no method can maintain the highest power across all scenarios, because the performance of a method depends on the type of traits, the number of associated traits, the phenotypic correlation, the percentage of protective variants and the percentage of causal variants. In summary, our proposed method (MULVR-O) remains
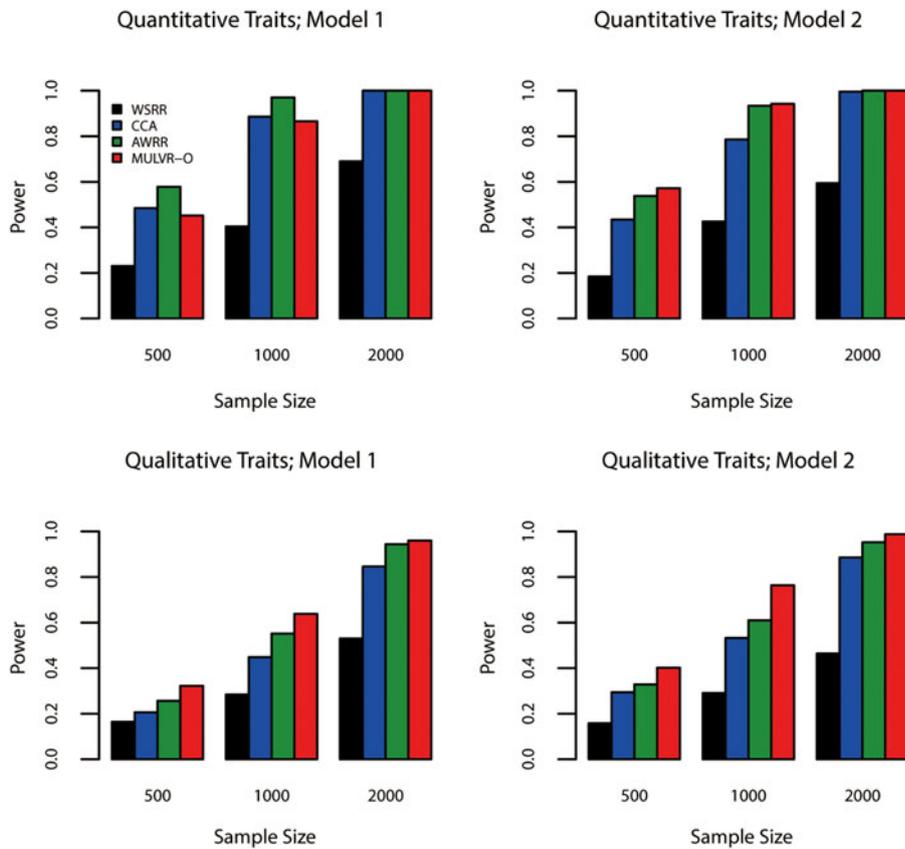
Fig. 5. Power comparisons for different sample sizes in two models. Total number of traits is six and causal variants impact on four traits. One common variant and 10% of rare variants are causal, 20% of rare causal variants are protective variants, and the total heritability of all causal variants is 0.03. $\rho = 0.5$.

Table 3. *The results of real data analysis.*

Chr	Gene	Position	SNPs	MULVR	MULVR-O	CCA	WSRR	AWRR
1	CHRM3	239792372-240072717	12	0.0281	0.0164	0.0276	0.4698	0.0273
17	SHBG	7517381-7536701	44	0.0245	0.0109	0.0396	0.0049	0.0471

Note: *P*-values were estimated based on 10^4 permutations.

powerful across a wide range of situations, and in particular, it shows better performance for qualitative traits.

4. Real data analysis

To explore the performance of the five methods (MULVR, MULVR-O, CCA, WSRR and AWRR), we respectively used them to analyse the GAW19 dataset, which includes 1943 Hispanic individuals with whole-exome sequence data, and two phenotypes, SBP and DBP, age, sex and anti-hypertensive medication status. We selected the *CHRM3* and *SHBG* genes in GAW19, which have been reported to be associated with the two phenotypes (SBP and DBP) (Sun *et al.*, 2016). We used the hg19 reference

as the annotation file to obtain the start and end positions of the two genes, and then used PLINK to extract genotypes of SNPs from the GAW19 dataset. Because many variants in the two genes are very rare, possibly observed only once or twice, we restricted analysis to only the variants that have four or more carriers. Missing genotype values were imputed by the corresponding variant's MAFs. We considered a total of 1851 individuals after removing subjects who had one or both missing phenotypes, and applied a log transformation to SBP and DBP so as to eliminate skewness. Because there were too many missing values for anti-hypertensive medication status, we only used age and sex as covariates. To guard against confounders caused by covariates, the logSBP and logDBP were adjusted for age and sex

with a linear regression. The residuals of logSBP and logDBP were treated as new phenotypes. With a significance level of 0.05, significance of the *SHBG* gene was identified by all the methods. Except for the WSRR method, all methods showed significant association of the *CHRM3* gene with SBP and DBP (Table 3).

5. Discussion

In genetic association studies, joint analysis of multiple traits can increase statistical power to detect genetic variants. Currently, the majority of methods are usually suitable for a single common variant or only rare variants. So, in this paper, we analysed association of both rare and common variants with multiple traits by variable reduction method. Extensive simulation studies show that no method can maintain the highest power across all scenarios. When there is correlation between any two traits, the MULVR-O method suffers loss of power for multiple quantitative traits in the presence of a large number of noise traits; when causal variants impact on all traits, AWRR and CCA lose power; CCA and AWRR suffer loss of power for qualitative traits. In summary, our proposed method (MULVR-O) remains powerful across a wide range of situations.

In our proposed method, we used the TOW method proposed by Sha *et al.* (2012) to detect rare and common variants associated with a single trait. The TOW method has three important advantages. First, TOW is suitable for detecting both rare and common variants. Second, TOW is robust to the different directions of variants and percentage of neutral variants. Third, TOW can adjust for covariates. Then our method can adjust for covariates according to the TOW method.

It is known that population stratification (PS) often causes spurious associations based on unrelated individuals. Our method is subject to bias in the presence of PS. So we can use principal component approach to guard against the effect of PS, which is one of the issues that continues to need consideration. In addition, we considered applying our method to family-based design, which is robust to PS, and efficient in detecting associations of rare variants. Of course, this issue needs to be further investigated in the future.

Because the asymptotical distribution of the CCA statistic is very conservative for rare variants (Wang *et al.*, 2016), based on the research of Wang *et al.* (2016), we used permutation procedure to evaluate the *P*-value of the CCA statistic. Thus all methods (MULVR, MULVR-O, CCA, WSRR and AWRR) use the permutation procedure to calculate the *P*-values of test statistics. It is time-consuming for

these permutation-based methods to perform genome-wide association studies. Hence, in consideration of computation time, we did not carry out genome-wide association analysis of the GAW19 dataset.

Declaration of interest

None.

Supplementary material

The online supplementary material can be found available at https://doi.org/10.1017/S0016672317000052.

References

Allison, D. B., Thiel, B., Jean, P. S., Elston, R. C., Infante, M. C. & Schork, N. J. (1998). Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. *American Journal of Human Genetics* **63**, 1190–1201.

Aschard, H., Vilhjalmsson, B. J., Greliche, N., Morange, P. E., Tregouet, D. A. & Kraft, P. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *American Journal of Human Genetics* **94**, 662–676.

Bansal, V., Libiger, O., Torkamani, A. & Schork, N. J. (2010). Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics* **11**, 773–785.

Bodmer, W. & Bonilla, C. (2008). Resampling-based multiple testing for microarray data analysis. *Nature Genetics* **40**, 695–701.

Evans, D. M. & Duffy, D. L. (2004). A simulation study concerning the effect of varying the residual phenotypic correlation on the power of bivariate quantitative trait loci linkage analysis. *Behavior Genetics* **34**, 135–141.

Ferreira, M. A. & Purcell, S. M. (2009). A multivariate test of association. *Bioinformatics* **25**, 132–133.

Kim, J. & Pan, W. (2017). Adaptive testing for multiple traits in a proportional odds model with applications to detect SNP-brain network associations. *Genetic Epidemiology* **41**, 259–277.

Klei, L., Luca, D., Devlin, B. & Roeder, K. (2008). Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genetic Epidemiology* **32**, 9–19.

Korte, A., Vilhjalmsson, B. J., Segura, V., Platt, A., Long, Q. & Nordborg, M. (2012). A mixed-model approach

for genome-wide association studies of correlated traits in structured populations. *Nature Genetics* **44**, 1066–1071.

Lange, C., Van Steen, K., Andrew, T., Lyon, H., Demeo, D. L., Raby, B., Li, B. & Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *American Journal of Human Genetics* **83**, 311–321.

Madsen, B. E. & Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genetics* **5**, e1000384.

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., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A. & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.

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. (2011). Testing for an unusual distribution of rare variants. *PLoS Genetics* **7**, e1001322.

Ng, S. B., Turner, E. H. & Robertson, P. D. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276.

O'Reilly, P. F., Hoggart, C. J., Pomyen, Y., Calboli, F. C., Elliott, P., Jarvelin, M. R. & Coin, L. J. (2012). MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One* **7**(5), e34861.

Price, A. L., Kryukov, G. V., Bakker, P. I., Purcell, S. M., Staples, J., Wei, L. J. & Sunyaev, S. R. (2010). Pooled association tests for rare variants in exon-resequencing studies. *American Journal of Human Genetics* **86**, 832–838.

Sha, Q., Wang, X., Wang, X. & Zhang, S. (2012). Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genetic Epidemiology* **36**, 561–571.

Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T. & Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. *American Journal of Human Genetics* **89**(5), 607–618.

Sun, J., Bhatnagar, S. R., Oualkacha, K., Ciampi, A. & Greenwood, C. M. (2016). Joint analysis of multiple blood pressure phenotypes in GAW19 data by using a multivariate rare-variant association test. *BioMed Central* **10**, 309.

Tang, C. S. & Ferreira, M. A. (2012). A gene-based test of association using canonical correlation analysis. *Bioinformatics* **28**, 845–850.

Teer, J. K. & Mullikin, J. C. (2010). Exome sequencing: the sweet spot before whole genomes. *Human Molecular Genetics* **19**(R2), R145–R151.

Van Der Sluis, S., Posthuma, D. & Dolan, C. V. (2013). TATES: efficient multivariate genotype–phenotype analysis for genomewide association studies. *PLoS Genetics* **9**, e1003235.

Wang, Z., Wang, X., Sha, Q. & Zhang, S. (2016). Joint analysis of multiple traits in rare variant association studies. *Annals of Human Genetics* **80**(3), 162–171.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M. & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics* **89**, 82–93.

Yang, Q., Wu, H., Guo, C. Y. & Fox, C. S. (2010). Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genetic Epidemiology* **34**, 444–454.

Zhou, X. & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods* **11**, 407–409.