Annals of human genetics

doi: 10.1111/ahg.12071

Likelihood Ratio Tests in Rare Variant Detection for Continuous Phenotypes

Ping Zeng[1,2], Yang Zhao[1], Jin Liu[1], Liya Liu[1], Liwei Zhang[1], Ting Wang[2], Shuiping Huang[2] and Feng Chen[1]*

[1]*Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 211166, P. R. China*

[2]*Department of Epidemiology and Biostatistics, School of Public Health, Xuzhou Medical College, Xuzhou, Jiangsu 221004, P. R. China*

Summary

It is believed that rare variants play an important role in human phenotypes; however, the detection of rare variants is extremely challenging due to their very low minor allele frequency. In this paper, the likelihood ratio test (LRT) and restricted likelihood ratio test (ReLRT) are proposed to test the association of rare variants based on the linear mixed effects model, where a group of rare variants are treated as random effects. Like the sequence kernel association test (SKAT), a state-of-the-art method for rare variant detection, LRT and ReLRT can effectively overcome the problem of directionality of effect inherent in the burden test in practice. By taking full advantage of the spectral decomposition, exact finite sample null distributions for LRT and ReLRT are obtained by simulation. We perform extensive numerical studies to evaluate the performance of LRT and ReLRT, and compare to the burden test, SKAT and SKAT-O. The simulations have shown that LRT and ReLRT can correctly control the type I error, and the controls are robust to the weights chosen and the number of rare variants under study. LRT and ReLRT behave similarly to the burden test when all the causal rare variants share the same direction of effect, and outperform SKAT across various situations. When both positive and negative effects exist, LRT and ReLRT suffer from few power reductions compared to the other two competing methods; under this case, an additional finding from our simulations is that SKAT-O is no longer the optimal test, and its power is even lower than that of SKAT. The exome sequencing SNP data from Genetic Analysis Workshop 17 were employed to illustrate the proposed methods, and interesting results are described.

Keywords: Rare variants, sequencing data, likelihood ratio test, restricted likelihood ratio test, variance component test, mixed effects model, association analysis

Introduction

In the last few years a large number of genetic susceptibility loci underlying many complex phenotypes have been discovered by genome-wide association studies (GWAS) (Hindorff et al., 2009). Current GWAS relies essentially on the so-called common disease-common variant hypothesis (Reich & Lander, 2001), here the common variant is generally defined as a locus with minor allele frequency (MAF) greater than 1%. However, for most common diseases the common variants that have been identified explain only a remarkably small part of the heritability (Manolio et al., 2009). The contribution from rare variants (MAF $<$ 1%) is thought to be one of the potential sources of the missing heritability, that is, the so-called common disease-rare variant (CDRV) hypothesis (Schork et al., 2009).

Although there is still some controversy between these two hypotheses (Gibson, 2012), it is widely believed that rare variants contribute substantially to common complex diseases (Cirulli & Goldstein, 2010). For example, according to the odds ratio (OR) distribution, it has been shown that very few common variants have values above 2 and

*Corresponding author: F. Chen, Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, 818 East Tianyuan Road, Nanjing, Jiangsu 211166, P. R. China. Tel: +86 25 86862754; Fax: +86 25 86527613; E-mail: fengchen@njmu.edu.cn

the mean OR is 1.36; while most rare variants have values above 2 and the mean OR is 3.74 (Bodmer & Bonilla, 2008). The CDRV hypothesis is further supported by the growing evidence from recent studies (Romeo et al., 2009; Kang et al., 2013; Ladouceur et al., 2013; Peng et al., 2013).

However, detecting causal rare variants is extremely challenging and the existing statistical methods for common variants are not feasible due to the rather low MAF of rare variants (Bansal et al., 2010). Therefore, there is an urgent need to develop much more powerful statistical approaches especially for rare variants. A variety of approaches have been proposed recently (e.g., Han & Pan, 2010; Basu & Pan, 2011; Wu et al., 2011; Yi & Zhi, 2011; Zhang et al., 2012; Chen et al., 2013; Derkach et al., 2013; Lee et al., 2013 Preston & Dudbridge, 2014). At present the burden test is one of the widely used methods for rare variants. The basic idea of the burden test is to first collapse the rare variants within a functional region (e.g., gene and pathway) into one variant, and then to examine this pooled variant. The cohort allelic sum test (Morgenthaler & Thilly, 2007), the combined multivariate and collapsing test (Li & Leal, 2008), and the group wise weighted sum test (Madsen & Browning, 2009) are the primary representatives.

The reasonable logic behind the burden test depends on two characteristics: (1) the statistical power increases significantly due to enriching association signals by collapsing; (2) it has been observed that multiple rare variants are often collectively related to the diseases (Bansal et al., 2010; Daye et al., 2012). However, the burden test explicitly assumes that all the rare variants are in the same direction of effect, that is, always contributing either protective or deleterious effects. Such an assumption may be not appropriate, and in practice the effects can be in both directions, under which the burden test can suffer from a remarkable loss of power (Wu et al., 2010; Basu & Pan, 2011).

Price et al. (2010) developed a variable-threshold method for rare variant identification. In that study a z-score for each allele-frequency threshold was calculated, the maximum z-score was used as the statistic. Neale et al. (2011) proposed another novel method called C-alpha which can address both risk and protective effects of rare variants. However, the P-values in both methods are obtained via the permutation procedure; as a result, it is not permissible to adjust covariates in the data analysis. For instance, a common phenomenon encountered in GWAS is population stratification, the usual way to control population stratification is to include the top principal components into the model. Accordingly, the variable-threshold method and C-alpha test cannot overcome this difficulty.

More recently the sequence kernel association test (SKAT) has turned out to be a powerful approach for rare variant identification, and has been shown to outperform the burden test under most situations (Wu et al., 2011). SKAT treats a group of rare variants as random effects and conducts a score-based variance component test (Lin, 1997). It avoids the directionality of effect and consequently enhances the power when both protective and deleterious effects exist. It was also proved that the C-alpha test is a special case of SKAT. Now SKAT is typically seen as the state-of-the-art method for the detection of rare variants.

However, due to being essentially a score test, SKAT itself has shortcomings. First, SKAT does not involve any parameter estimation for the alternative model; thus there is a lack of effect measurement such as OR for the logistic regression in single locus association analysis. As a result, it is difficult to evaluate the relative importance across various sets of rare variants by SKAT. Second, a large score in SKAT does not necessarily mean that the effect of a set of rare variants is also great, as it may result from a large number of variants with weak effects (Zhan & Xu, 2012).

In this paper we propose the likelihood ratio test (LRT) and restricted likelihood ratio test (ReLRT) for rare variant association analysis. In contrast to SKAT, LRT and ReLRT need to estimate both the null and alternative models. Doing this efficiently overcomes the aforementioned issues and results in at least two advantages compared to SKAT: (1) LRT and ReLRT offer a measurement analogous to OR to assess the genetic contribution of multiple rare variants, which is useful for researchers in practice; (2) the statistical powers of LRT and ReLRT are expected to improve since information about the random effects of rare variants is explicitly employed in the procedures. These advantages are our main motivation in applying LRT and ReLRT to the identification of rare variants by examining the variance component.

Based on extensive simulations and comparisons, our major conclusions are presented as follows. Like the burden test, SKAT and SKAT-O (the optimal version of SKAT), LRT and ReLRT can correctly control the type I error, and the controls are robust to the weights chosen and the number of rare variants in a set. LRT and ReLRT outperform SKAT across various situations, and behave similarly to the burden test when all the causal rare variants share the same direction of effect. When both positive and negative effects exist, LRT and ReLRT suffer from fewer power reductions compared to the burden test, SKAT and SKAT-O; under this case, an additional finding from our simulations is that SKAT-O is no longer the optimal test, and its power is even lower than that of SKAT. An application is also given. The simulation results as well as the practical application demonstrate that the proposed LRT and ReLRT are promising and powerful for rare variant detection.

Materials and Methods

Here we mainly focus on continuous phenotypes (e.g., BMI, blood pressure, and triglyceride level). Denote the phenotype by y_i for $i = 1, 2, \ldots, n$, where n is the sample size. Denote covariates (e.g., gender, age, and height) and the first several principal components for population structure correction (Price et al., 2006) by $X_i = [x_{i1}, \ldots, x_{im}]$ and genotypes for rare variants by $G_i = [g_{i1}, g_{i2}, \ldots, g_{ip}]$. The additive genetic model is used, so that $g = 0$, 1, and 2 represent the number of copies of the minor allele.

In the following paper, an introduction to the linear mixed effects model for rare variant detection as well as the burden test and SKAT is first given, and then LRT and ReLRT are given. The suitable weight choices for rare variants are discussed at length. We conduct numerical studies to evaluate LRT and ReLRT and compare them with the burden test, SKAT and SKAT-O. Finally, the exome sequencing SNP data from Genetic Analysis Workshop 17 (GAW17) is employed to illustrate the proposed methods.

Linear Mixed Effects Model

Consider the linear mixed effects model (Laird & Ware, 1982)

$$y_i = X_i'\boldsymbol{\alpha} + G_i'\boldsymbol{\beta} + \varepsilon_i, \tag{1}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_m]$ are the fixed effects for covariates, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_p]$ are the random effects for rare variants, and each β_j is assumed to be normally distributed with mean zero and variance τw_j, where τ is a variance component and w is a prespecified weight related to MAF. We will discuss how to choose w for LRT and ReLRT later in detail. Here ε is a normal error with mean zero and variance σ^2, and independent of the random effects $\boldsymbol{\beta}$.

Under these conditions, we can obtain

$$\begin{aligned} E(\boldsymbol{y}) &= \boldsymbol{X\alpha}, \\ Var(\boldsymbol{y}) &= \tau \boldsymbol{GWG}' + \sigma^2 \boldsymbol{I}_n = \sigma^2 \boldsymbol{V}_\lambda, \end{aligned} \tag{2}$$

where $\mathbf{V}_\lambda = \lambda\mathbf{GWG}' + \mathbf{I}_n$ and $\lambda = \tau/\sigma^2$. Here $\mathbf{G}$ is the $n \times P$ genotype matrix, $\mathbf{W}$ is a diagonal matrix of order P with its elements being w, and $\mathbf{I}_n$ is an identity matrix of order n. Examining whether a group of rare variants are collectively associated with the phenotype is equivalent to testing the null hypothesis for random effects H_0: $\beta_1 = \ldots = \beta_p = 0$, or to testing the null hypothesis for the variance component H_0: $\tau = 0$ (or $\lambda = 0$).

Here we note that the parameter λ has a close relationship with the heritability H, that is, the proportion of phenotypic variance explained by variants. According to the classical definition of heritability (Yang et al., 2010), H can be expressed as $\lambda/(1+\lambda)$. Therefore, λ is considered to be a measurement analogous to the OR in the logistic regression or to the regression coefficient in the linear regression to evaluate the relative importance (or effect strength) of a set of rare variants. A larger value of λ is a stronger indication of the association of rare variants with the phenotype.

SKAT and the Burden Test

SKAT uses a score statistic to test the variance component H_0,

$$Q_{SKAT} = (\boldsymbol{y} - \hat{\boldsymbol{y}})' \boldsymbol{K} (\boldsymbol{y} - \hat{\boldsymbol{y}}) = \sum_{j=1}^{p} \left\{ w_j \left[\sum_{i=1}^{n} g_{ij} (y_i - \hat{y}_i) \right]^2 \right\}, \tag{3}$$

where $\hat{\boldsymbol{y}}$ is the predicted value of $\boldsymbol{y}$ under H_0, in which the mixed effects model degenerates to a standard general linear model, and $\mathbf{K} = \mathbf{GWG}'$ is called the kernel function. Under H_0, the score statistic Q_{SKAT} follows a mixture of χ^2 distributions, and the P-value is obtained by the method of Davies (1980) or other methods (Wu et al., 2011).

The burden test is conducted by pooling multiple rare variants into a single variant

$$y_i = \boldsymbol{X}_i'\boldsymbol{\alpha} + \left(\sum_{j=1}^{p} w_j g_{ij} \right) \beta_c + \varepsilon. \tag{4}$$

Thus only one common coefficient β_c is required to be tested. The resulting score statistic is

$$Q_{burden} = \left[\sum_{i=1}^{n} (y_i - \hat{y}_i) \left(\sum_{j=1}^{p} w_j g_{ij} \right) \right]^2. \tag{5}$$

Provided that all the rare variants have the same direction of effect, the burden test achieves a significant improvement of power since the degrees of freedom are reduced by pooling a class of rare variants. But in practical data analysis such an assumption is often violated (Han & Pan, 2010; Basu & Pan, 2011; Wu et al., 2011).

In fact the burden test and SKAT can be studied within a unified framework if taking into account the correlation structure of the random effects, this test is referred to as the optimal SKAT (SKAT-O). SKAT corresponds to an independent correlation and the burden test corresponds to a perfect correlation. The reader is referred to Lee et al. (2012a, b) for more information.

LRT

When examining variance component in the linear mixed effects model, LRT and ReLRT becomes a natural alternative to the score test (Stram & Lee, 1994; Crainiceanu & Ruppert, 2004). For convenience we here focus on λ. Note that $\lambda = 0$ if and only if $\tau = 0$, and the parameter space for λ is $\Omega = [0, \infty)$. The null and alternative hypotheses are respectively

$$H_0 : \lambda = 0,\ H_1 : \lambda > 0. \tag{6}$$

The test in (6) is nonstandard since under H_0 λ is on the boundary of the parameter space. Under some regularity assumptions, Stram and Lee (1994) proved that the LRT statistic for (6) followed a 50:50 mixture of $\mathcal{X}_0^2$ and $\mathcal{X}_1^2$, where $\mathcal{X}_0^2$ is a point probability mass at zero and $\mathcal{X}_1^2$ is a χ^2 distribution with single degree of freedom. However, Crainiceanu and Ruppert (2004) argued that the conditions assumed in Stram and Lee (1994) are restrictive and do not hold for most mixed effects models, therefore, in practice the 50:50 mixture approximation may be very poor.

We use the techniques of profile likelihood and spectral decomposition (Crainiceanu & Ruppert, 2004) to build LRT and ReLRT for rare variant detection. The log-likelihood function for model (1) up to a constant independent of the parameters is,

$$L(\boldsymbol{\alpha}, \lambda, \sigma^2) = -\frac{1}{2}\left\{n\log(\sigma^2) + \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha})' \boldsymbol{V}_\lambda^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) + \log|\boldsymbol{V}_\lambda|\right\}. \tag{7}$$

For a given value of λ, the maximum likelihood (ML) estimators for $\boldsymbol{\alpha}$ and σ^2 are,

$$\hat{\boldsymbol{\alpha}}(\lambda) = (\boldsymbol{X}' \boldsymbol{V}_\lambda^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{V}_\lambda^{-1} \boldsymbol{y},$$
$$\hat{\sigma}^2(\lambda) = \frac{1}{n}\left[\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\alpha}}(\lambda)\right]' \boldsymbol{V}_\lambda^{-1}\left[\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\alpha}}(\lambda)\right]. \tag{8}$$

Replacing $\boldsymbol{\alpha}$ and σ^2 in (7) by $\hat{\boldsymbol{\alpha}}(\lambda)$ and $\hat{\sigma}^2(\lambda)$ yields the profile log-likelihood function up to a constant

$$L(\lambda) = -\frac{1}{2}\left\{n\log(\boldsymbol{y}' \boldsymbol{P}_\lambda' \boldsymbol{V}_\lambda^{-1} \boldsymbol{P}_\lambda \boldsymbol{y}) + \log|\boldsymbol{V}_\lambda|\right\}, \tag{9}$$

where,

$$\boldsymbol{P}_\lambda = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}' \boldsymbol{V}_\lambda^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{V}_\lambda^{-1}. \tag{10}$$

The LRT statistic is defined as

$$\begin{aligned} \mathrm{LRT_n} &= 2[\sup_{\lambda\in\Omega} L(\lambda) - L(\lambda = 0)], \\ &= \sup_{\lambda\in\Omega}\{-n\log(\boldsymbol{y}' \boldsymbol{P}_\lambda' \boldsymbol{V}_\lambda^{-1} \boldsymbol{P}_\lambda \boldsymbol{y}) - \log|\boldsymbol{V}_\lambda| \\ &\quad + n\log(\boldsymbol{y}' \boldsymbol{P}_0 \boldsymbol{y})\}, \end{aligned} \tag{11}$$

Where,

$$\boldsymbol{P}_0 = \boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'. \tag{12}$$

Note that $n\log(\boldsymbol{y}'\mathbf{P}_0\boldsymbol{y})$ in equation (11) is a constant.

To obtain the finite sample distribution of the LRT statistic, we give the spectral representation of LRT_n,

$$f_n(\lambda) = \sup_{\lambda\in\Omega}\left\{\mathrm{n}\log\left[1 + \frac{\mathrm{N_n}(\lambda)}{\mathrm{D_n}(\lambda)}\right] - \sum_{\mathrm{j}=1}^{\mathrm{p}}\log\left(1 + \lambda\xi_{\mathrm{j}}\right)\right\}, \tag{13}$$

where ξ_j's are the eigenvalues of matrix $\mathbf{W}^{1/2}\mathbf{G}'\mathbf{G}\mathbf{W}^{1/2}$, and

$$\begin{aligned} N_n(\lambda) &= \sum_{j=1}^{p}\frac{\lambda\mu_j}{1 + \lambda\mu_j}u_j^2, \\ D_n(\lambda) &= \sum_{j=1}^{p}\frac{1}{1 + \lambda\mu_j}u_j^2 + \sum_{j=p+1}^{n-m}u_j^2, \end{aligned} \tag{14}$$

where μ_j's are the eigenvalues of matrix $\mathbf{W}^{1/2}\mathbf{G}'\mathbf{P}_0\mathbf{G}\mathbf{W}^{1/2}$, and u_j's are independently standard normal random variables. Then LRT_n is equal to $f_n(\lambda)$ in distribution.

By taking full advantage of the spectral representation used in equation (14), a simulation-based algorithm for the finite sample distribution of LRT_n was described in Crainiceanu and Ruppert (2004).

ReLRT

It is well-known that the ML estimator of σ^2 is biased downward since it does not take into account the loss in degrees of freedom due to the estimation of $\boldsymbol{\alpha}$, while the restricted ML method provides an unbiased estimator for σ^2 by using a set of n–m linearly independent error contrasts (Harville, 1974; Corbeil & Searle, 1976; Harville, 1977).

The profile restricted log-likelihood function up to a constant independent of the parameters is,

$$L_{\mathrm{Re}}(\lambda) = \frac{1}{2}\left\{-\log|\boldsymbol{V}_\lambda| - (n-m)\log\left(\boldsymbol{y}' \boldsymbol{P}_\lambda' \boldsymbol{V}_\lambda^{-1} \boldsymbol{P}_\lambda \boldsymbol{y}\right) - \log|\boldsymbol{X}' \boldsymbol{V}_\lambda^{-1} \boldsymbol{X}|\right\}. \tag{15}$$

The ReLRT statistic is defined as,

$$\mathrm{ReLRT}_n = 2\left[\sup_{\lambda\in\Omega}\mathrm{L_{Re}}(\lambda) - \mathrm{L_{Re}}(\lambda = 0)\right]. \tag{16}$$

By using similar reasoning to that used for LRT_n, ReLRT_n is equal to,

$$f_n(\lambda) = \sup_{\lambda\in\Omega}\left\{(\mathrm{n} - \mathrm{m})\log\left[1 + \frac{\mathrm{N_n}(\lambda)}{\mathrm{D_n}(\lambda)}\right]\right.$$

$$-\sum_{j=1}^{p}\log\left(1+\lambda\mu_j\right)\Bigg\}, \tag{17}$$

in distribution. The simulation algorithm similar to LRT_n can also be applied to ReLRT_n.

It has been shown that there are nonzero probabilities of LRT_n and ReLRT_n having mass at zero even under the null hypothesis (Crainiceanu & Ruppert, 2004). Therefore, the traditional approximation derived under the large sample theory that LRT_n (or ReLRT_n) follows a χ^2 distribution with single degree of freedom is also not suitable.

Weight Choice

We now discuss how to choose weight for LRT and ReLRT. The simplest choice is $w = 1$, that is, the matrix $\mathbf{W}$ is an identity matrix, implying that all rare variants act equally. However, Madsen and Browning (2009), Wu et al. (2011) and others stated that an appropriate choice of weight can increase the statistical power for rare variant detection. Under the assumption that a rarer variant makes a larger contribution to the phenotype, Wu et al. (2011) suggested using $w = [\mathrm{Beta}(\mathrm{MAF};\, a_1,\, a_2)]^2$, where Beta is the beta density function, and setting $a_1 = 1$ and $a_2 = 25$ which places more weight on a rarer variant and less weight on a common variant.

The weight selection is not problematic for SKAT because it is a type of score test whose type I error is protected for the weight used (Wu et al., 2011). However, this property does not hold for LRT and ReLRT since the calculation of an alternative model explicitly involves the weight. It is important to determine reasonable weights for rare variants in LRT and ReLRT procedures. For example, if using the default values of $a_1 = 1$ and $a_2 = 25$ and setting $\mathrm{MAF}_j = 0.002$, the weight is about 567 which gives too large a scale for the variance of rare variant j. More specifically, for a given value of σ^2, a larger weight will lead to much smaller estimates of τ and λ. This may cause numerical inaccuracy and instability when calculating LRT_n and ReLRT_n. For instance, assume that the true variance for the rare variant j is 0.14, and the error variance $\sigma^2 = 1$, then theoretically τ is about 2.5×10^{-4} (0.14/567), and λ is also about 2.5×10^{-4} which is a very small value. As a result, a slight change of λ can have a substantial influence on the statistics. These values used above are chosen in terms of our simulation results.

The statements above have shown that care should be taken regarding the weight choice of rare variants in LRT and ReLRT. To avoiding specifying a relatively large scale for weight, in the present study we divide the weight used in SKAT by S,

$$w_j = \frac{\left[\mathrm{Beta}\left(\mathrm{MAF}_j; a_1, a_2\right)\right]^2}{S}, \tag{18}$$

where $S \geq 1$ is a prespecified constant. That is, we increase the scale of λ by S times. To select a reasonable S for LRT and ReLRT, we perform simulations to evaluate different selection of $S = 1$, 1000, 500, 100, sum($\boldsymbol{w}$), and max($\boldsymbol{w}$), where sum and max indicate the summation and maximum over all the w_js, respectively.

Our simulations demonstrated that the direct application of the default weight (i.e., $S = 1$) of SKAT to LRT and ReLRT is underpowered, while other selections of S are more powerful, and lead to nearly the same results (Table S1). The results show that increasing the scale of λ can avoid the numerical issue efficiently, thus the power of LRT and ReLRT rises. It is easy to see that the P-values of the burden test, SKAT and SKAT-O are invariant under any transformation of weight as done in equation (18). In the following paper $S = \max(\boldsymbol{w})$ is used.

Simulations

We generate genotypes based on a coalescent model for European population by using the package COSI (Schaffner et al., 2005). The simulation parameters for COSI and implementation details are given in the supplementary materials.

The covariates X are set to $[1, x_1, x_2]$, where x_1 and x_2 are the standard normal variable and the binary variable with a probability of 0.5, respectively, and are mutually independent. For the type I error simulation the phenotype is generated as

$$y \sim N(1.0 + 0.5x_1 + 0.5x_2, 1),$$

where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. The number of runs is 2000. The type I error rate is estimated as the proportion of P-values less than the significance level α. In the simulations we use $\alpha = 0.05$ and 0.01. In the following only the results for $\alpha = 0.01$ are shown, those for $\alpha = 0.05$ are presented in the supplementary materials.

In the power simulation, 20% of rare variants are set to being causal variants, the effect size $|\beta|$ is specified to $0.3|\log_{10}\mathrm{MAF}|$, leading to a size of 1.2 for MAF = 0.0001 and 0.6 for MAF = 0.01. Among these causal rare variants, 0%, 30% or 50% are specified to have negative effects, that is, their effect sizes are set to $-0.3|\log_{10}\mathrm{MAF}|$ if we claim that M% of rare variants have negative effects. For the power simulation the phenotype is generated as

$$y \sim N\left(1.0 + 0.5x_1 + 0.5x_2 + \sum\nolimits_{j=1}^{s} g_j^c \beta_j^c, 1\right),$$

Table 1 Type I error and power simulation under different weights with $\alpha = 0.01$.

		Type I error					Power				
a_1	a_2	Burden	SKAT	SKAT-O	LRT	ReLRT	Burden	SKAT	SKAT-O	LRT	ReLRT
1	25	0.009	0.011	0.012	0.011	0.013	0.051	0.211	0.173	0.233	0.239
0.5	0.5	0.014	0.007	0.006	0.008	0.009	0.035	0.158	0.150	0.173	0.186
1	1	0.011	0.008	0.011	0.009	0.010	0.053	0.189	0.161	0.215	0.223
1	50	0.013	0.011	0.011	0.013	0.014	0.047	0.200	0.160	0.217	0.228

Table 2 Type I error control under varying number of rare variants with $\alpha = 0.01$.

		Type I error				
Subregions	No	Burden	SKAT	SKAT-O	LRT	ReLRT
0.1	16	0.012	0.011	0.014	0.012	0.013
0.2	31	0.009	0.008	0.009	0.009	0.010
0.3	47	0.010	0.008	0.008	0.008	0.009
0.4	63	0.010	0.009	0.012	0.007	0.008
0.5	78	0.010	0.009	0.010	0.011	0.012
0.6	93	0.012	0.009	0.010	0.012	0.013
0.7	109	0.011	0.009	0.012	0.009	0.010
0.8	125	0.012	0.011	0.008	0.010	0.013
0.9	141	0.012	0.009	0.008	0.010	0.011

Note: The "No" column indicates the average number of rare variants used in the simulation.

where s is the number of selected causal rare variants, g^c's are the genotypes and β^c's are the effect sizes given above. The number of runs is 1000. The power is estimated as the proportion of P-values less than α. The simulation characteristics under these specifications are displayed in Table S2.

In the present study we consider the following simulation cases, and implement five methods, that is, the burden test, SKAT, SKAT-O, LRT, and ReLRT.

1. **Simulation under different weights**
   Four groups of values for a_1 and a_2 in equation (18) are investigated. Here only the power simulation that 30% of causal rare variants have negative effects is considered, and $n = 400$.
2. **Simulation under varying number of rare variants**
   The subregions are set to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. Only the power simulation that 0% of causal rare variants have negative effects is considered, that is, all the selected causal rare variants share the same positive direction of effect, and $n = 400$.
3. **Simulation under varying sample size and presence of both positive and negative effects**
   Here we set $n = 400$, 600, 800, and 1000. Power simulations with 0%, 30%, and 50% of causal rare variants having negative effects are considered. The subregions are set to 30%.
4. **Simulation for the estimation of parameter λ**
   In this simulation the effect size of causal rare variants is assumed to follow a normal distribution with mean zero, and the variance τ is chosen so that the parameter λ is respectively equal to 0.25, 1, and 4 given a fixed value of σ^2. In each run a total of 20 causal rare variants are generated and $n = 400$.

Application of GAW17 data

The GAW17 data (Almasy et al., 2011; Ghosh et al., 2011) consists of three covariates (age, gender, and smoke), three continuous phenotypes (Q1, Q2, and Q3) and a binary phenotype (affected), and 24,487 autosomal SNPs belonging to 3205 genes on 697 individuals. The MAF ranges from 0.07% to 25.8% with the median 0.002; 74% of SNPs have MAF less than 0.01 and 12.8% of SNPs have MAF more than 0.05. Further descriptions about these data were presented in Almasy, et al. (2011). Here we use Q1 in the first replicate of GAW17 data as the phenotype, and perform a genome-wide analysis for these data.

Results

Simulation Results

Tables 1 and S3 show that all the five tests can control the type I error correctly at the given α level and are relatively robust to the weights used. However, slight differences of power are observed under different weights. For example, the weight with $a_1 = a_2 = 0.5$ gives the lowest power for all of the five tests, the weight with $a_1 = 1$ and $a_2 = 25$ gives the highest powers for SKAT, SKAT-O, LRT, and ReLRT, and the weight with $a_1 = a_2 = 1$ gives the highest power for the burden test. The following results are based on the weight with $a_1 = 1$ and $a_2 = 25$, which is also the choice recommended in Wu et al. (2011).

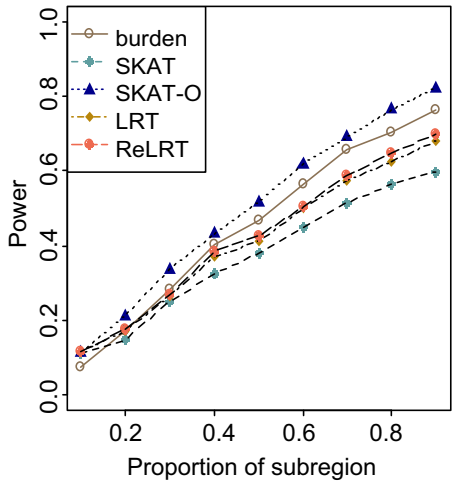


Figure 1 Estimated power under varying number of rare variants with $\alpha = 0.01$.

Table 3 Type I error control under varying sample size with $\alpha = 0.01$.

Sample size	Burden	SKAT	SKAT-O	LRT	ReLRT
400	0.010	0.008	0.008	0.008	0.009
600	0.010	0.008	0.015	0.010	0.010
800	0.010	0.008	0.009	0.009	0.011
1000	0.011	0.011	0.010	0.010	0.010

Tables 2 and S4 show that the number of rare variants seems to have little impact on the type I error control for all five approaches. Note that although the number of causal rare variants becomes large as the subregions grow wide, the number of noncausal rare variants also increases. Nonetheless, it is observed that all of the tests enhance power as the subregions increase (Figs. 1 and S1). The burden test and SKAT-O consistently have the highest power since in this simulation where all the causal effects are set to be positive, under this situation the burden test and SKAT-O have been proven to be considerably powerful (Han & Pan, 2010; Basu & Pan, 2011; Wu et al., 2011; Lee et al., 2012a, 2012b). It is also shown that LRT and ReLRT always outperform SKAT regardless of the proportion of subregions.

Tables 3 and S5 show the estimated type I error under varying sample size. Figures 2 and S2 show the estimated power. Tables 3 and S5 demonstrate that all the five tests can control the type I error correctly at the given α level regardless of sample size. When all the causal rare variants have the same direction of effect and the sample size is not too large (e.g., $n = 400$ and 600), SKAT-O is the most powerful procedure; the burden test is slightly better than SKAT, while LRT and ReLRT behave similarly to the burden test. Thus, when the sample size is sufficient (e.g., $n = 800$ and 1000), SKAT-O, SKAT, LRT, and ReLRT become better than the burden test even when all the effects are in the same direction.

The results have shown that LRT and ReLRT are uniformly slightly better than SKAT regardless of sample size and the proportion of negative causal rare variants. The advantage of LRT and ReLRT over SKAT seems more evident when there are more causal rare variants with negative effects. For example, when $n = 1000$ and $\alpha = 0.01$, and 30% of causal rare variants have negative effects, the powers of ReLRT, LRT, and SKAT are 0.748, 0.741, and 0.634, respectively. Furthermore, we also discover that ReLRT consistently outperforms LRT although the difference is rather small. This advantage of ReLRT may stem from its unbiased estimation of variance component.

Compared to the situation of zero causal rare variants having negative effects, the power losses for the situations of 30% and 50% of causal rare variants having negative effects are shown in Figures 3 and S3. When there are both positive and negative causal rare variants, the burden test has a dramatic reduction of power. For example, when $n = 1000$ and $\alpha = 0.01$, the power of the burden test decreases from 0.683 for zero causal rare variants having negative effects to 0.035 for 50% of causal rare variants having negative effects. In this case SKAT, LRT, and ReLRT without exception suffer from power losses, and it is much more interesting that SKAT-O has substantially reduced power relative to SKAT, LRT, and ReLRT, suggesting that when both positive and negative effects are possible SKAT-O is suboptimal and the application of SKAT-O is not a reasonable choices. Similar results were also observed in Sun et al. (2013). The reason may be that when the effects of rare variants have mixed directions the correlation structure used in SKAT-O cannot capture the truly complicated relationships among rare variants. However, the decreases for these four tests are much lower compared with the burden test. For example, when $n = 1000$ and $\alpha = 0.01$, the power of ReLRT is 0.773 for zero causal rare variants having negative effects and is 0.715 for 50% of causal rare variants having negative effects. Furthermore, it is shown that LRT and ReLRT on average lose slightly less power than SKAT and much less than SKAT-O.

Figure 4 shows the observed quantiles of LRT_n against the expected values of the 50:50 χ^2 mixture. This plot shows that the finite sample distribution of LRT_n (or ReLRT_n) is different from the 50:50 χ^2 mixture; the latter usually gives larger values and hence is conservative. In fact the finite sample probability of LRT_n (or ReLRT_n) equalling zero under H_0 is significantly greater than 50% (Crainiceanu & Ruppert, 2004; Pinheiro & Bates, 2009). See also Table S2.

The estimated values of λ are shown in Figure 5. It is found that LRT and ReLRT yield almost the same results,

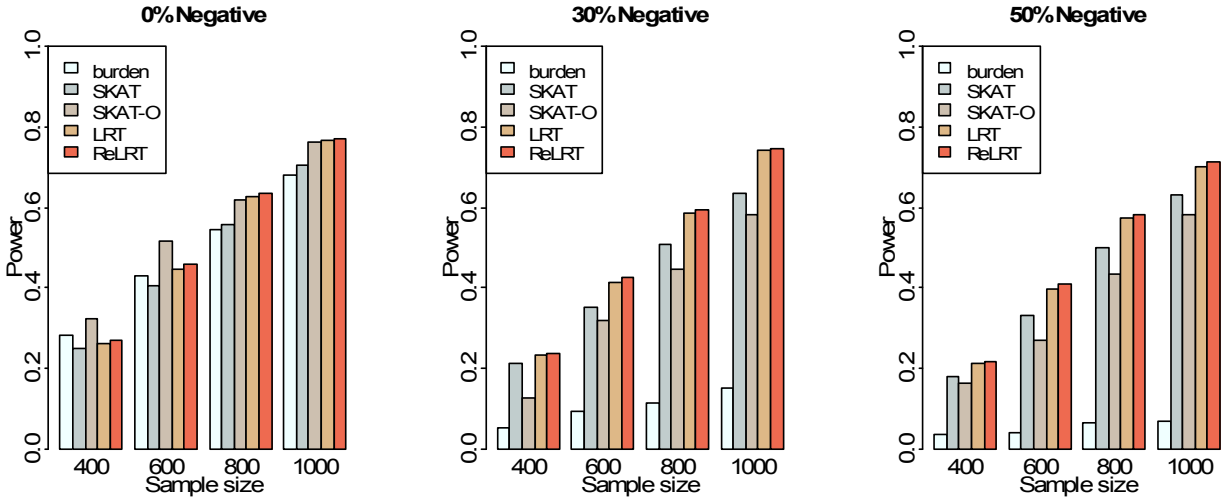


Figure 2 Power under varying sample size and presence of both positive and negative effects with $\alpha = 0.01$. M% (M = 0, 30, and 50) negative means that in these associated rare variants M% have effects $-0.3|\log_{10}\text{MAF}|$ and the rest (100–M)% are $0.3|\log_{10}\text{MAF}|$.

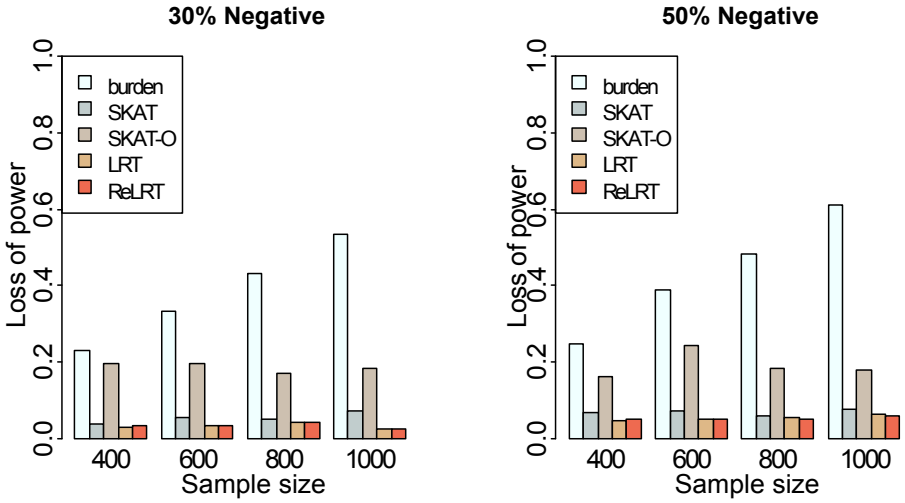


Figure 3 Losses of the power between the situations of 30% and 50% of rare variants having negative effects versus the situation of none (0%) of the rare variants having negative effects $\alpha = 0.01$.

thus only the estimates of LRT are reported here. When the true value of λ is relatively large (e.g., $\lambda = 1$ and 4), LRT obtains approximately unbiased estimates; the averages for $\lambda = 1$ and $\lambda = 4$ are 1.004 ($se = 0.017$) and 4.103 ($se = 0.053$), respectively. However, if λ is small (e.g., $\lambda = 0.25$), LRT tends to overestimate the value slightly; the average for $\lambda = 0.25$ is 0.302 ($se = 0.007$). But we note that that the estimated median for $\lambda = 0.25$ is 0.267, thus this overestimation may be not a concern considering that the estimated λ is often asymmetrically distributed as displayed in Figure 5. These observations are expected since λ can be considered as a signal-to-noise ratio, a larger λ indicates more information and thus results in good estimation, while a smaller λ leads to poor estimation.

Data Analysis Results

The genome-wide association results for the rare variants of GAW17 data are given in Figure S4, and QQ plots for *P*-values of the burden test, SKAT, SKAT-O, LRT, and ReLRT are shown in Figure S5. It is observed that the patterns of Manhattan plot are very similar across the testing procedures.

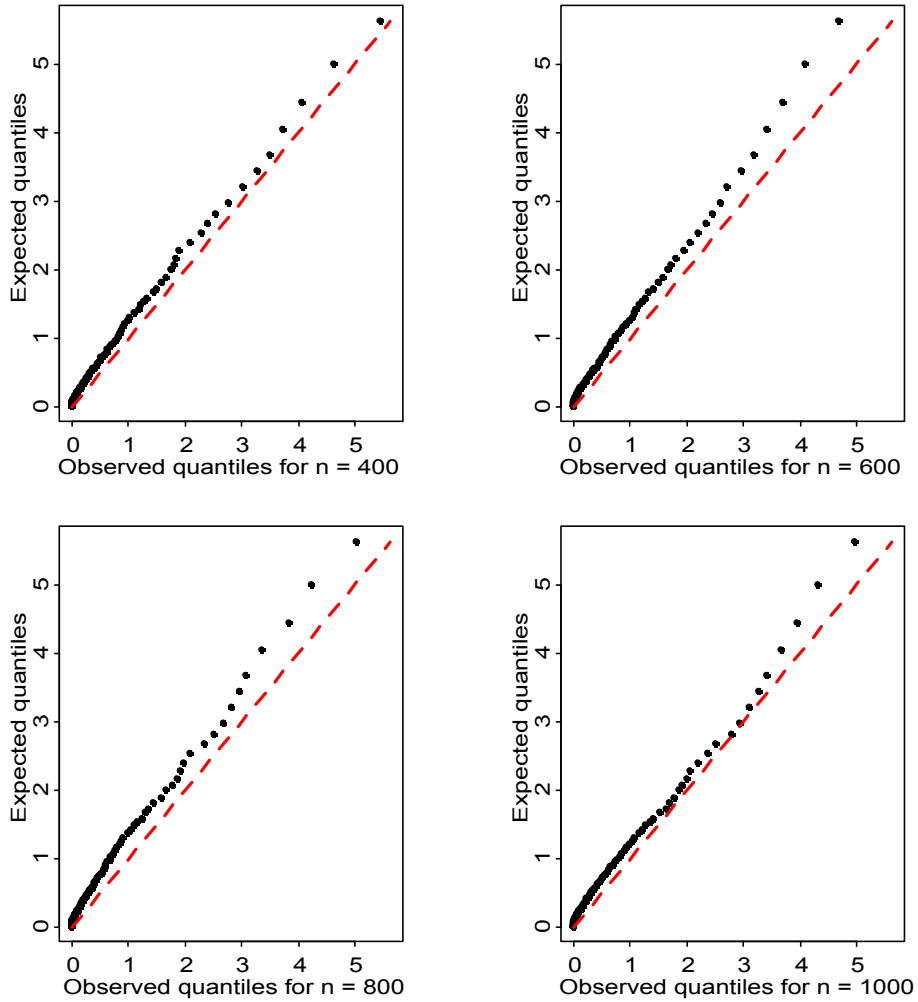


Figure 4 Observed quantiles of LRT_n against expected values of the 50:50 χ^2 mixture. The plots for $ReLRT_n$ are similar to these of LRT_n, so are not shown.

Note that since many of the *P*-values generated by LRT and ReLRT are equal to one, the QQ plots for LRT and ReLRT do not go along the diagonal linearly. The top ten genes with the smallest *P*-values are listed in Table S6, and it is observed that the burden test, SKAT-O, LRT and ReLRT share more communality and identify three true signals, whereas only one true signal appears in the list for SKAT.

We present in Table 4 the results of three typical genes; *FLT4*, *FLT1*, and *KDR*. All the SNPs within these genes are positively related to the response, thus from Table 4 it is seen that the burden test and SKAT-O always have the smallest *P*-values. As expected, the *P*-values for LRT and ReLRT are larger than that for SKAT. *FLT1* and *KDR* are therefore considered to be significant, while *FLT4* is not due to its very weak effects (Almasy et al., 2011) and to the relatively small sample size of the GAW17 data.

According to Almasy, et al. (2011), *FLT1* can be viewed as being weakly associated with the response while *KDR* appears to be strongly associated with the response. In view of this, it is interesting to observe that the *P*-values of *FLT1* for all the methods are much smaller than those of *KDR*. This phenomenon may express a misleading message that *FLT1* is more associated with the response than *KDR*. However, the burden test, SKAT, and SKAT-O will not show any additional information regarding this problem. On the other hand, the estimates of λ from LRT and ReLRT explicitly demonstrate that the effect of *KDR* is much stronger than that of *FLT1*. Accordingly the results of LRT and ReLRT correctly reflect the true situation designed in Almasy, et al. (2011).

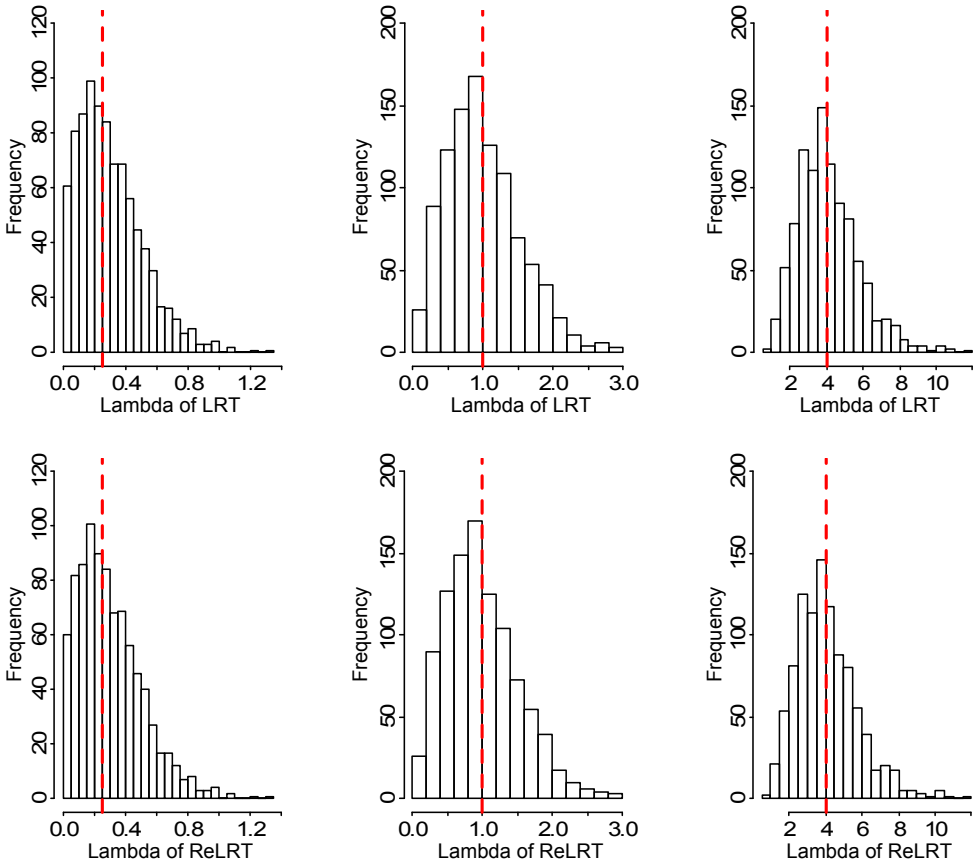


Figure 5 Estimated values of λ for LRT and ReLRT. The reference lines represent the true values of λ.

Table 4 Results of the genes *FLT4*, *FLT1*, and *KDR*.

	P-value					λ	
Gene	Burden	SKAT	SKAT-O	LRT	ReLRT	LRT	ReLRT
FLT4	0.0131	0.1589	0.0209	0.1642	0.1585	0.144	0.145
FLT1	6.129×10^{-8}	9.008×10^{-7}	1.032×10^{-9}	6.280×10^{-7}	5.439×10^{-7}	0.750	0.748
KDR	9.268×10^{-7}	1.291×10^{-3}	2.781×10^{-6}	4.991×10^{-5}	4.826×10^{-5}	1.778	1.767

Discussion

In this paper we have proposed the use of the LRT and ReLRT methods for rare variant identification instead of the score test employed in SKAT. Extensive simulations have shown that across a range of situations LRT and ReLRT are powerful and consistently outperform SKAT. An additional finding from our simulations is that, when both deleterious and protective rare variants exist, SKAT-O is no longer the optimal test and its power is even lower than that of SKAT. This may be because the true relationships are much more complex than the correlation used in SKAT-O. The simulations have also shown that LRT and ReLRT are robust to the weight chosen and the number of rare variants under study. The higher power for LRT and ReLRT may result from two features: estimation of both the null and alternative models and use of the exact finite sample distributions for LRT_n and $ReLRT_n$ instead of the asymptotical mixture distribution.

However, it should be mentioned that LRT and ReLRT are consequently slower than SKAT due to these computations. For example, using a 3.09 GHz personal computer with

3.16 Gb memory, to finish 100 runs with the sample size equal to 500 and the number of rare variants being 50, SKAT only requires about 50 s, while LRT (or) ReLRT require about 1100 s, and more time will be needed if the sample size increases.

A attractive aspect of LRT and ReLRT is that they provide not only a test for rare variant association but also an estimate of λ, which can be thought to be an indirect estimate of heritability explained by rare variants provided that suitable weights are used (Yang et al., 2010). The latter is a great advantage of the proposed approaches over the existing methods. To the best of our knowledge, little work has been done to measure the importance of a set of variants in association analysis. Considering the popularity of set-based strategy in GWAS (Wang et al., 2007; Medina et al., 2009), the proposed methods provide a flexible alternative. We also wish to encourage more studies regarding this topic based on our work.

As has been shown, LRT and ReLRT can improve power by using appropriate weight. We have found that the scale of weight adopted has a substantial impact on LRT and ReLRT in our simulations. To avoid numerical inaccuracy, we empirically reduce the scale. This reduction has no influence on SKAT but enhances the powers of LRT and ReLRT. However, choosing $S = \max(\boldsymbol{w})$ is to some extent arbitrary and has no explicitly theoretical justification, although our simulations have shown that this choice works very well. The choice of weight for rare variants is important and is based on some assumptions as well as some genetic evidence, but whether or not the chosen weight is reasonable is difficult to validate in practice. It is not clear how to choose the optimal weight for rare variants in LRT and ReLRT, and this may be an interesting topic for further investigation.

As mentioned before, although LRT and ReLRT can increase the statistical power, this comes at the expense of more computational time. The consumption of time is mainly due to the estimation of λ in the alternative model and the simulation algorithm for obtaining the null distributions of LRT_n and ReLRT_n. To streamline the proposed approaches for use as a routine tool in large scale genome-wide application, more efforts are required, such as accelerating the estimation, designing more effective computation procedures and approximating the null distributions of the LRT and ReLRT statistics. Work regarding these aspects is now in progress.

Acknowledgments

We are grateful to Stephen Schaffner and Michael Wu for helpful suggestions about the rare variants data simulation. We appreciate the support of the editor and we thank two anonymous reviewers for their valuable suggestions and comments on our manuscript. This work was supported in part by the National Natural Science Foundation of China (No. 81072389, 81373102), the Research Fund for the Doctoral Program of Higher Education of China (No. 20113234110002), the Key Grant of Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 10KJA330034), the College Philosophy and Social Science Foundation from Education Department of Jiangsu Province of China (No. 2013SJD790032, 2013SJB790059), the Research Foundation from Xuzhou Medical College (No. 2012KJ02), the Research and Innovation Project for College Graduates of Jiangsu province of China (No. CXLX13_574) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

Almasy, L., Dyer, T., Peralta, J., Kent, J., Charlesworth, J., Curran, J. & Blangero, J. (2011) Genetic analysis workshop 17 mini-exome simulation. *BMC Proc* **5**, S2.

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

Basu, S. & Pan, W. (2011) Comparison of statistical tests for disease association with rare variants. *Genet Epidemiol* **35**, 606–619.

Bodmer, W. & Bonilla, C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* **40**, 695–701.

Chen, H., Meigs, J. B., & Dupuis, J. (2013) Sequence kernel association test for quantitative traits in family samples. *Genet Epidemiol* **37**, 196–204.

Cirulli, E. T. & Goldstein, D. B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**, 415–425.

Corbeil, R. R. & Searle, S. R. (1976) Restricted maximum likelihood (REML) estimation of variance omponents in the mixed model. *Technometrics* **18**, 31–38.

Crainiceanu, C. M. & Ruppert, D. (2004) Likelihood ratio tests in linear mixed models with one variance component. *J Roy Stat Soc, B* **66**, 165–185.

Davies, R. B. (1980) Algorithm AS 155: The distribution of a linear combination of chi-2 random variables. *J Roy Stat Soc, C* **29**, 323–333.

Daye, Z. J., Li, H., & Wei, Z. (2012) A powerful test for multiple rare variants association studies that incorporates sequencing qualities. *Nucleic Acids Res* **40**, e60–e60.

Derkach, A., Lawless, J. F., & Sun, L. (2013) Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet Epidemiol* **37**, 110–121.

Ghosh, S., Bickeboller, H., Bailey, J., Bailey-Wilson, J., Cantor, R., Culverhouse, R., Daw, W., Destefano, A., Engelman, C., Hinrichs, A., Houwing-Duistermaat, J., Konig, I., Kent, J., Laird, N., Pankratz, N., Paterson, A., Pugh, E., Suarez, B., Sun, Y., Thomas, A., Tintle, N., Zhu, X., Ziegler, A., Maccluer, J., & Almasy, L. (2011) Identifying rare variants from exome scans: The GAW17 experience. *BMC Proc* **5**, S1.

Gibson, G. (2012) Rare and common variants: Twenty arguments. *Nat Rev Genet* **13**, 135–145.

Han, F. & Pan, W. (2010) A data-adaptive sum test for disease association with multiple common or rare variants. *Hum Hered* **70**, 42–54.

Harville, D. A. (1974) Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383–385.

Harville, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* **72**, 320–338.

Hindorff, L., Sethupathy, P., Junkins, H., Ramos, E., Mehta, J., Collins, F., & Manolio, T. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362–9367.

Kang, H., Quigley, D., Kim, I., Wakabayashi, Y., Ferguson-Smith, M., D'alessandro, M., Birgitte Lane, E., Akhurst, R., Goudie, D., & Balmain, A. (2013) Multiple self-healing squamous epithelioma (MSSE): rare variants in an adjacent region of chromosome 9q22.3 to known TGFBR1 mutations suggest a digenic or multilocus etiology. *J Invest Dermatol* **133**, 1907–1910.

Ladouceur, M., Zheng, H. F., Greenwood, C. M. T., & Richards, J. B. (2013) Empirical power of very rare variants for common traits and disease: Results from sanger sequencing 1998 individuals. *Eur J Hum Genet* **21**, 1027–1030.

Laird, N. M. & Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.

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. (2012a) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* **91**, 224–237.

Lee, S., Teslovich, T. M., Boehnke, M., & Lin, X. (2013) General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet* **93**, 42–53.

Lee, S., Wu, M. C., & Lin, X. (2012b) Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**, 762–775.

Li, B. & Leal, S. M. (2008) Methods for detecting associations with rare ariants for common diseases: Application to analysis of sequence data. *Am J Hum Genet* **83**, 311–321.

Lin, X. (1997) Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309–326.

Madsen, B. & Browning, S. (2009) A group-wise association test for rare mutations using a weighted sum statistic. *PLoS Genet* **5**, e1000384.

Manolio, T., Collins, F., Cox, N., Goldstein, D., Hindorff, L., Hunter, D., Mccarthy, M., Ramos, E., Cardon, L., & Chakravarti, A. (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.

Medina, I., Montaner, D., & Bonifaci, N. (2009) Gene set-based analysis of polymorphisms: Finding pathways or biological processes associated to traits in genome-wide association studies. *Nucleic Acids Res* **37**, W340-W344.

Morgenthaler, S. & Thilly, W. G. (2007) A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). *Mutat Res* **615**, 28–56.

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 Genet* **7**, e1001322.

Peng, G., Fan, Y., Palculict, T. B., Shen, P., Ruteshouser, E. C., Chi, A.-K., Davis, R. W., Huff, V., Scharfe, C., & Wang, W. (2013) Rare variant detection using family-based sequencing analysis. *Proc Natl Acad Sci U S A* **110**, 3985–3990.

Pinheiro, J. C. & Bates, D. (2009) *Mixed-effects models in S and S-PLUS*. New York: Springer.

Preston, M. D. & Dudbridge, F. (2014) Utilising family-based designs for detecting rare variant disease associations. *Ann Hum Genet* **78**, 129–140.

Price, A., Kryukov, G., De Bakker, P., Purcell, S., Staples, J., Wei, L., & Sunyaev, S. (2010) Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* **86**, 832–838.

Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N., & Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909.

Reich, D. E. & Lander, E. S. (2001) On the allelic spectrum of human disease. *Trends Genet* **17**, 502–510.

Romeo, S., Yin, W., Kozlitina, J., Pennacchio, L. A., Boerwinkle, E., Hobbs, H. H., & Cohen, J. C. (2009) Rare loss-of-function mutations in ANGPTL family members contribute to plasma triglyceride levels in humans. *J Clin Invest* **119**, 70–79.

Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., & Altshuler, D. (2005) Calibrating a coalescent simulation of human genome sequence variation. *Genome Res* **15**, 1576–1583.

Schork, N. J., Murray, S. S., Frazer, K. A., & Topol, E. J. (2009) Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* **19**, 212–219.

Stram, D. O. & Lee, J. W. (1994) Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177.

Sun, J., Zheng, Y., & Hsu, L. (2013) A unified mixed-effects model for rare-variant association in sequencing studies. *Genet Epidemiol* **37**, 334–344.

Wang, K., Li, M., & Bucan, M. (2007) Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* **81**, 1278–1283.

Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., & Lin, X. (2010) Powerful SNP-Set analysis for case-control genome-wide association studies. *Am J Hum Genet* **86**, 929–942.

Wu, M., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011) Rare variant association testing for sequencing data using the sequence kernel association test. *Am J Hum Genet* **89**, 82–93.

Yang, J., Benyamin, B., Mcevoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565–569.

Yi, N. & Zhi, D. (2011) Bayesian analysis of rare variants in genetic association studies. *Genet Epidemiol* **35**, 57–69.

Zhan, H. & Xu, S. (2012) Adaptive ridge regression for rare variant detection. *PLoS One* **7**, e44173.

Zhang, L., Pei, Y. F., Hai, R., Lin, Y., & Deng, H. W. (2012) Testing rare variants for association with diseases: A Bayesian marker selection approach. *Ann Hum Genet* **76**, 74–85.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

1. Implementation details

2. Simulation setting

Table S1 Results for different S in the weight selection.

Table S1 Simulation characteristics.

Table S3 Type I error and power simulation under different weights with $\alpha = 0.05$.

Table S4 Type I error control under varying number of rare variants with $\alpha = 0.05$.

Table S5 Type I error control under varying sample size with $\alpha = 0.05$.

Table S6 Top ten genes with the smallest *P*-values.

Figure S1 Estimated power under varying number of rare variants with $\alpha = 0.05$.

Figure S2 Power under varying sample size and presence of both positive and negative effects with $\alpha = 0.05$. M% (M = 0, 30, and 50) negative means that in these associated rare variants M% have effects $-0.3|\log_{10}\text{MAF}|$ and the rest (100–M)% are $0.3|\log_{10}\text{MAF}|$.

Figure S3 Losses of power between the situations of 30% and 50% of rare variants having negative effects versus the situation of none (0%) of the rare variants having negative effects with $\alpha = 0.05$.

Figure S4 Genome-wide association results for the rare variants of GAW17 data. The *P*-values of genes are shown as the Manhattan plot widely used in GWAS, where the *P*-values are in $-\log_{10}$ scale in the vertical axis and the genes along chromosomes in the horizontal axis. Different chromosomes are distinguished with colors.

Figure S5 QQ plot for *P*-values generated by the five methods: burden test, SKAT, SKAT-O, LRT, and ReLRT.

Received: 27 December 2013
Accepted: 22 April 2014

Supporting Information for "Likelihood ratio tests in rare variant detection for continuous phenotypes"

Ping Zeng[1,2], Yang Zhao[1], Jin Liu[1], Liya Liu[1], Liwei Zhang[1], Ting Wang[2], Shuiping Huang[2], Feng Chen[1*]

[1]*Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, 211166, P. R. China*
[2]*Department of Epidemiology and Biostatistics, School of Public Health, Xuzhou Medical College, Xuzhou, Jiangsu, 221004, P. R. China*

1. Implementation details

All the analyses are finished under the R statistical environment (R Core Team, 2013), the burden test, SKAT and SKAT-O are implemented through the SKAT package provided in Wu, et al. (2011). The optimizations in Equations (11) and (16) are conducted via the L-BFGS-B method (Byrd et al., 1995, Zhu et al., 1997) with a linear constraint of $\lambda \geq 0$, which is implemented in the optim function. The p values of LRT_n and ReLRT_n are calculated according to the simulation algorithm described in Crainiceanu and Ruppert (2004) with the simulation number 10^5. This algorithm is implemented with the LRTSim and RLRTSim functions in the RLRsim package (Scheipl et al., 2008). R functions and an example for the proposed LRT and ReLRT are provided.

2. Simulation setting

We generate genotypes based on a coalescent model for European population by using the package COSI (Schaffner et al., 2005), which can be available at http://www.broadinstitute.org/~sfs/cosi/. The parameters for the coalescent model are specified as follows. The genomic length is set to 10^5, the mutation rate is set to $1.5 \times$

* Corresponding author: Feng Chen, Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, 818 East Tianyuan Road, Nanjing, Jiangsu, 211166, P. R. China. Tel: +86 25 86862754, Fax: +86 25 86527613; E-mail: fengchen@njmu.edu.cn

10^{-8}, the recombination rate is set to 1.0×10^{-8}, the population size is set to 10^4. Randomly selected continuous 30% subregions of the simulated genotypes are used. Variants with MAF less than 0.01 are viewed as rare variants.

3. Table

Table S1 Results for different *S* in the weight selection

Table S2 Simulation characteristics

Table S3 Type I error and power simulation under different weights with α = 0.05

Table S4 Type I error simulation under varying number of rare variants with α = 0.05

Table S5 Type I error under varying sample size with α = 0.05

4. Figure

Figure S1 Estimated power under varying number of rare variants with α = 0.05

Figure S2 Power under varying sample size and presence of both positive and negative effects with α = 0.05. M% (M = 0, 30, and 50) negative means that in these associated rare variants M% have effects $-0.3|\log_{10}\text{MAF}|$ and the rest (100-M)% are $0.3|\log_{10}\text{MAF}|$.

Figure S3 Losses of the power between the situations of 30% and 50% rare variants having negative effects versus the situation of none (0%) the rare variants having negative effects with α = 0.05.

Figure S4 Genome-wide association results for the rare variants of GAW17 data. The p values of genes are shown as the Manhattan plot widely used in GWAS, where the p values are shown in $-\log_{10}$ scale in the vertical axis and the genes along chromosomes in the horizontal axis. Different chromosomes are distinguished with colors.

Figure S5 QQ plot for p values generated by the five methods: burden test, SKAT, SKAT-O, LRT, and ReLRT

Table S1 Results for different S in the weight selection

	1	1000	500	100	sum($\boldsymbol{w}$)	max($\boldsymbol{w}$)
power1	0.239629	0.265707	0.266512	0.266412	0.266110	0.266412
power2	0.313532	0.462747	0.462042	0.462042	0.462243	0.462143
			Summary for estimates of λ			
prop.0	0.663008	0.124849	0.125856	0.131998	0.124346	0.125252
min	0	0	0	0	0	0
q5	0	0	0	0	0	0
q10	0	0	0	0	0	0
q25	0	0.084059	0.042037	0.008413	2.030339	0.049482
mean	0.000215	0.243406	0.121709	0.024344	6.107197	0.143271
sd	0.000366	0.200848	0.100421	0.020089	5.096704	0.118211
median	0	0.213393	0.106700	0.021349	5.327681	0.125601
q75	0.000344	0.359398	0.179702	0.035947	9.032177	0.211534
q90	0.000937	0.509568	0.254781	0.050963	12.874344	0.299916
q95	0.001056	0.618385	0.309193	0.061841	15.581858	0.363973
max	0.001723	1.588619	0.794310	0.158863	35.165323	0.935046

Note: In these simulations all the causal rare variants are assumed to have positive effects, set a_1 = 1 and a_2 = 25, and n = 400, the other simulation settings are described in the paper. Only the results for LRT are presented since the results for ReLRT are similar. In the table prop.0 is the proportion of λ equaling zero, and power1 and power2 are the powers estimated at α = 0.01 and 0.05, respectively.

Table S2 Simulation characteristics

n	total SNPs	selected SNPs	used RV	prop.I	prop.0	prop.30	prop.50
400	435	131	47	0.564/0.557	0.120/0.117	0.133/0.127	0.154/0.147
600	458	137	55	0.567/0.554	0.067/0.066	0.072/0.066	0.068/0.068
800	476	143	60	0.575/0.561	0.027/0.026	0.033/0.031	0.034/0.033
1000	489	147	64	0.569/0.557	0.023/0.023	0.014/0.011	0.015/0.015

Note: The column of total SNPs indicates the simulated rare and common variants within the total region. The column of selected SNPs indicates the randomly selected rare and common variants within the 30% subregions. The column of used RV indicates the rare variants used in the simulations, among which 20% are causal for power simulations. The columns of prop.I, prop.0, prop.30, and prop.50 indicate the proportions that LRT/ReLRT is equal to zero for the type I error simulation and power simulations with 0%, 30% and 50% casual rare variants having negative effects, respectively.

Table S3 Type I error and power simulation under different weights with $\alpha = 0.05$

a_1	a_2	type I error					power				
		Burden	SKAT	SKAT-O	LRT	ReLRT	Burden	SKAT	SKAT-O	LRT	ReLRT
1	25	0.048	0.053	0.049	0.051	0.056	0.138	0.420	0.329	0.448	0.462
0.5	0.5	0.054	0.041	0.049	0.042	0.046	0.115	0.366	0.328	0.371	0.384
1	1	0.051	0.055	0.044	0.049	0.056	0.145	0.394	0.335	0.431	0.444
1	50	0.058	0.056	0.066	0.053	0.059	0.137	0.394	0.335	0.414	0.429

Table S4 Type I error simulation under varying number of rare variants with $\alpha = 0.05$

Sub-regions	No	type I error				
		Burden	SKAT	SKAT-O	LRT	ReLRT
0.1	16	0.046	0.054	0.048	0.051	0.053
0.2	31	0.045	0.045	0.048	0.049	0.053
0.3	47	0.048	0.053	0.046	0.045	0.049
0.4	63	0.052	0.052	0.047	0.052	0.057
0.5	78	0.046	0.050	0.049	0.047	0.054
0.6	93	0.043	0.043	0.049	0.045	0.051
0.7	109	0.053	0.042	0.051	0.047	0.054
0.8	125	0.051	0.051	0.047	0.042	0.047
0.9	141	0.053	0.047	0.049	0.048	0.057

Note: The column of no indicates the average number of rare variants used in the simulation.

Table S5 Type I error under varying sample size with $\alpha = 0.05$

Sample size	Burden	SKAT	SKAT-O	LRT	ReLRT
400	0.048	0.053	0.046	0.045	0.049
600	0.048	0.046	0.054	0.049	0.054
800	0.055	0.049	0.056	0.050	0.055
1000	0.056	0.055	0.052	0.057	0.060

Table S6 The top ten genes with the smallest p values

Burden	SKAT	SKAT-O	LRT	ReLRT
RGPD8	RGPD8	RGPD8	RGPD8	RGPD8
KDR	JAK1	***KDR***	***FLT1***	***FLT1***
FLT1	HSZFP36	***FLT1***	HSZFP36	HSZFP36
ZNF91	BRCA1	JAK1	JAK1	JAK1
PSG2	RUNX2	PSG2	***VEGFC***	***VEGFC***
HSZFP36	TAS2R48	ZNF91	ZNF91	ZNF91
EPHA5	VEGFC	HSZFP36	***KDR***	BRCA1
LYK5	GRK1	LYK5	BRCA1	***KDR***
KIT	***FLT1***	***VEGFC***	BRCA2	BRCA2
VEGFC	PINX1	BRCA1	ZNF77	TAS2R48

Note: The true gene signals are denoted in bold and italic.

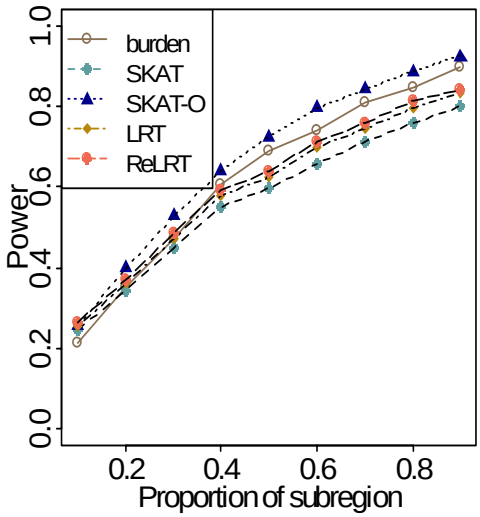


Figure S1 Estimated power under varying number of rare variants with $\alpha = 0.05$

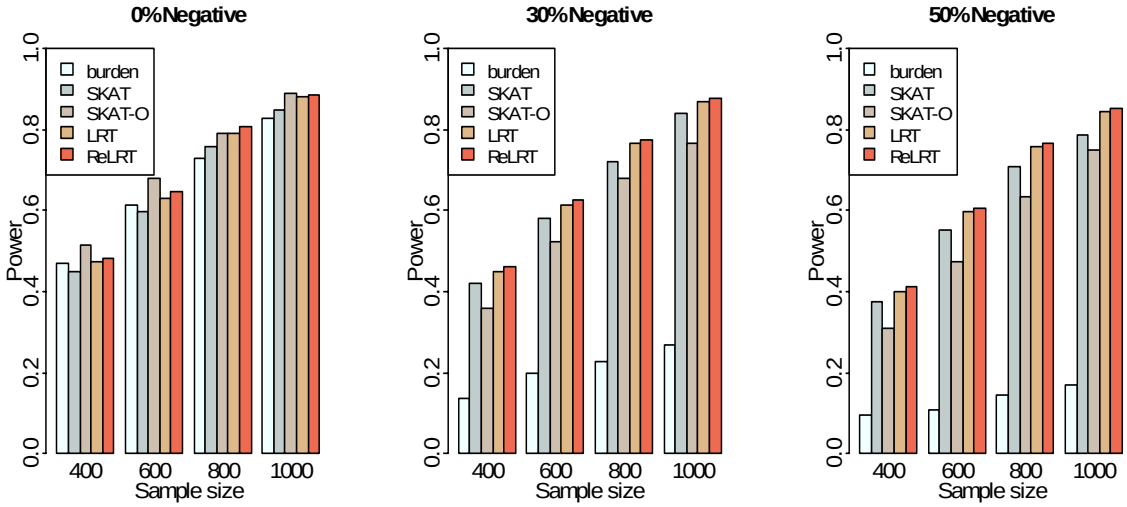


Figure S2 Power under varying sample size and presence of both positive and negative effects with $\alpha = 0.05$. M% (M = 0, 30, and 50) negative means that in these associated rare variants M% have effects $-0.3|\log_{10}\text{MAF}|$ and the rest (100-M)% are $0.3|\log_{10}\text{MAF}|$.

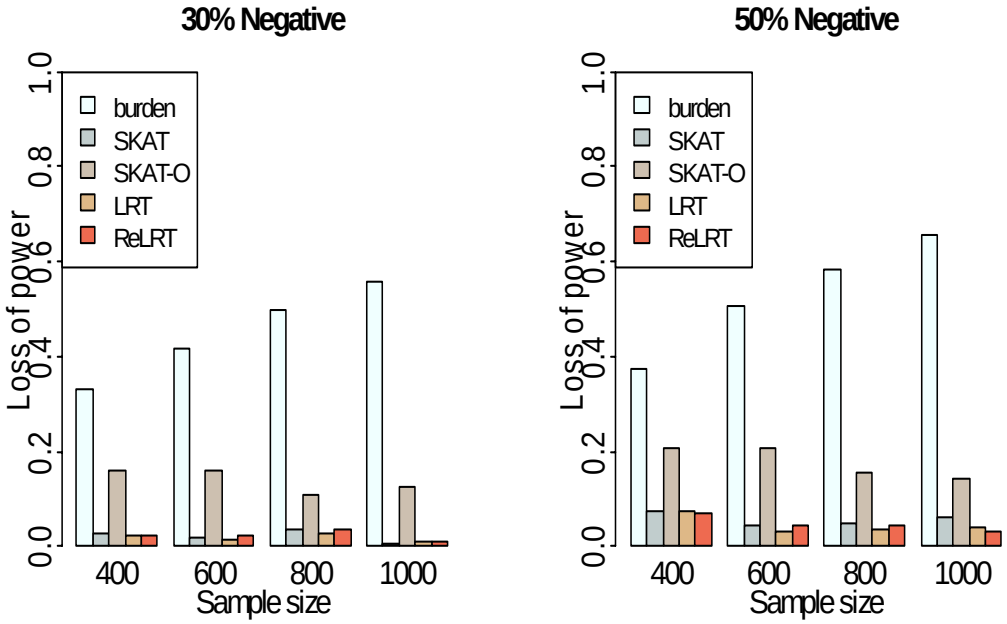


Figure S3 Losses of the power between the situations of 30% and 50% rare variants having negative effects versus the situation of none (0%) the rare variants having negative effects with $\alpha = 0.05$.

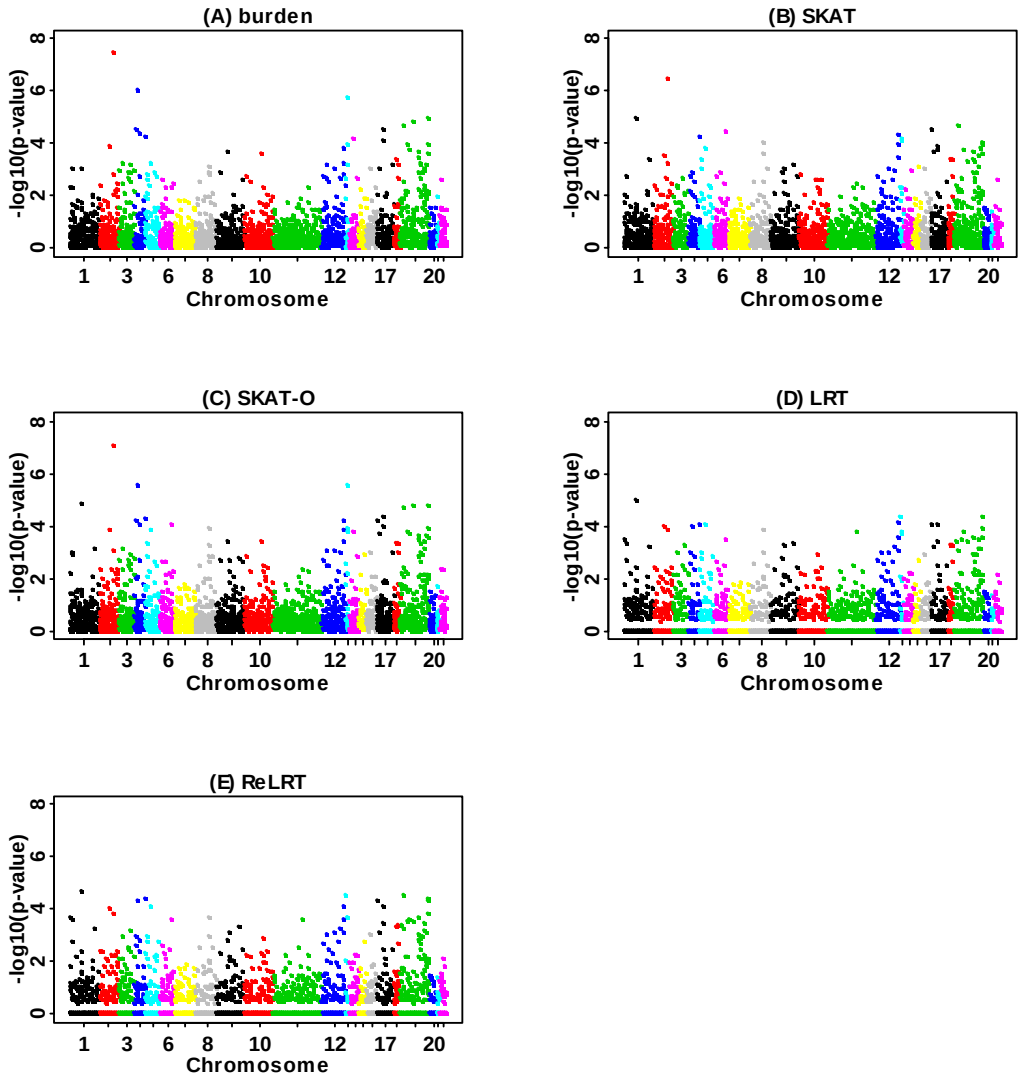
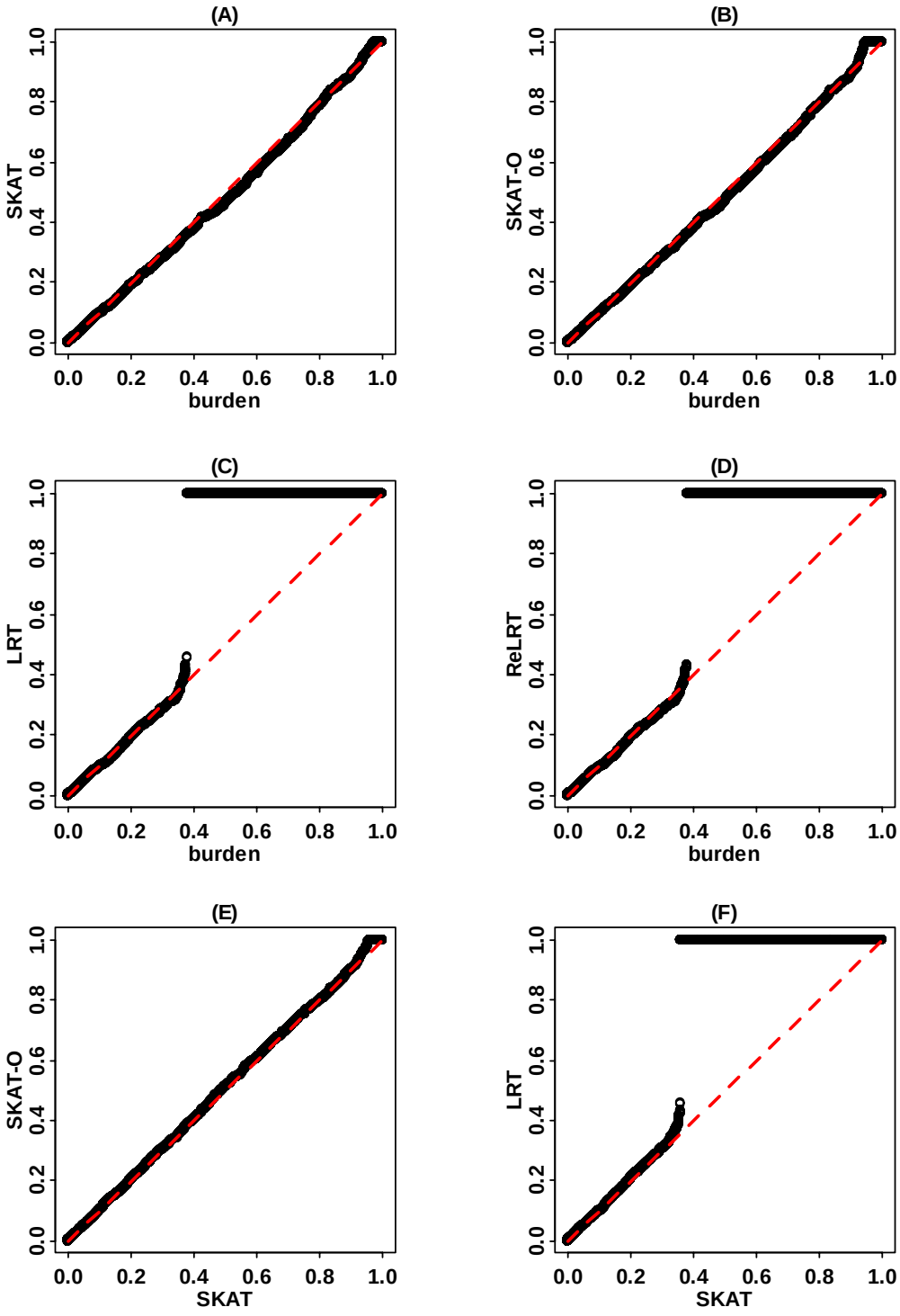


Figure S4 Genome-wide association results for the rare variants of GAW17 data. The p values of genes are shown as the Manhattan plot widely used in GWAS, where the p values are in $-\log_{10}$ scale in the vertical axis and the genes along chromosomes in the horizontal axis. Different chromosomes are distinguished with colors.


(A)
SKAT
burden
(B)
SKAT-O
burden
(C)
LRT
burden
(D)
ReLRT
burden
(E)
SKAT-O
SKAT
(F)
LRT
SKAT
0.0
0.2
0.4
0.6
0.8
1.0

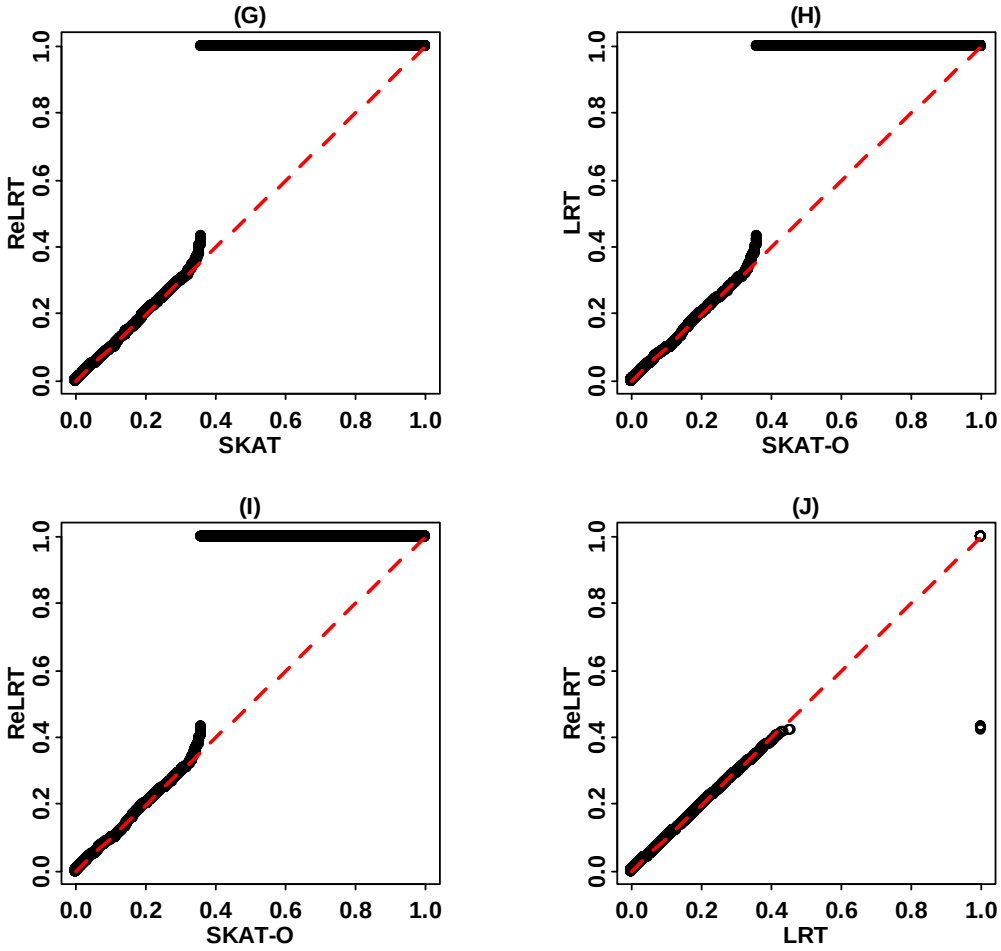


Figure S5 QQ plot for p values generated by the five methods: burden test, SKAT, SKAT-O, LRT, and ReLRT.

References

Byrd, R., Lu, P., Nocedal, J. & Zhu, C. (1995) A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J Sci Comput* **16**, 1190-1208.

Crainiceanu, C.M. & Ruppert, D. (2004) Likelihood ratio tests in linear mixed models with one variance component. *J R Stat Soc Series B* **66**, 165-185.

R Core Team (2013) R: A language and environment for statistica computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Schaffner, S.F., Foo, C., Gabriel, S., Reich, D., Daly, M.J. & Altshuler, D. (2005) Calibrating a coalescent simulation of human genome sequence variation. *Genome Res* **15**, 1576-1583.

Scheipl, F., Greven, S. & Küchenhoff, H. (2008) Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput Stat Data Anal* **52**, 3283-3299.

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. *Am J Hum Genet* **89**, 82-93.

Zhu, C., Byrd, R.H., Lu, P. & Nocedal, J. (1997) Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans Math Softw* **23**, 550-560.