*Manuscript
Click here to view linked References

Optimal unified approach for rare variant association testing with application to small sample case-control whole-exome sequencing studies

Seunggeun Lee[1], Mary Emond[2], Michael Bamshad[3], Kathleen Barnes[4], Mark Rieder[5], Deborah Nickerson[5], NHLBI GO Exome Sequencing Project – ESP Lung Project Team[6], David C. Christiani[7,8], Mark M. Wurfel[9], and Xihong Lin[1]

1 Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA

2 Department of Biostatistics, University of Washington, Seattle, WA 98101, USA

3 Department of Pediatrics, University of Washington, Seattle, WA 98101, USA

4 Department of Medicine, Johns Hopkins University, Baltimore, MD 21224, USA

5 Department of Genome Science, University of Washington, Seattle, WA 98101, USA

6 ESP Banner available in the supplemental materials

7 Department of Environmental Health, Harvard School of Public Health, Boston, MA 02115, USA

8 Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA

9 Division of Pulmonary and Critical Care Medicine, The University of Washington, Seattle, WA 98104, USA

Corresponding Author:

Xihong Lin

Department of Biostatistics

Harvard School of Public Health

655 Huntington Ave, Building 2, 4th Floor

Boston, MA 02115

Tel: (617) 432-2914

Fax: (617) 432-5619

Email: xlin@hsph.harvard.edu

Abstract

We propose in this paper a unified approach for testing the association between rare variants and phenotypes in sequencing association studies. This approach maximizes power by adaptively using the data to optimally combine the burden test and non-burden sequence kernel association test (SKAT). Burden tests are more powerful when most variants in a region are causal and the effects are in the same direction, while SKAT is more powerful when a large fraction of the variants in a region are non-causal or the effects of causal variants are in different directions. The proposed unified test maintains the power in both scenarios. We show that the unified test corresponds to the optimal test in an extended family of SKAT tests, which we refer to as SKAT-O. The second goal of this paper is to develop a small sample adjustment procedure for the proposed methods to correct conservative type I error rates of SKAT-family tests when the trait of interest is dichotomous and the sample size is small. Both small sample adjusted SKAT and the optimal unified test (SKAT-O) are computationally efficient and can be easily applied to genome-wide sequencing association studies. We evaluate the finite sample performance of the proposed methods using extensive simulation studies and illustrate their application using the acute lung injury exome sequencing data of the NHLBI exome sequencing project.

1 Introduction

Array-based genotyping technologies have been successfully used in hundreds of genome-wide association studies in the last few years for identifying over a thousand common genetic variants associated with many complex diseases. The recent advance of massively parallel sequencing technologies [1, 2] has transformed human genetic research. These emerging sequencing technologies provide a rich opportunity to study the association between rare variants and complex traits. Rare variants, which have minor allele frequencies (MAFs) less than $0.01 \sim 0.05$, might play an important role in the etiology of complex traits and account for missing heritability unexplained by common variants [3, 4]. Several complex traits have been found to be associated with rare variants [5, 6, 7].

In recent years, significant efforts have been devoted to developing powerful and computationally efficient statistical methods to test associations between rare variants and complex traits. While single variant tests are typically conducted to investigate associations of common variants and phenotypes, the same approach has little power for testing for rare variant effects due to their low frequencies and large numbers. Instead, the statistical development of rare variants analysis has been focused on testing cumulative effects of rare variants in genetic regions or SNP sets, such as genes. These tests can be broadly classified as burden and non-burden tests.

Burden tests collapse rare variants in a genetic region into a single burden variable, and then regress the phenotype on the burden variable to test for the cumulative effects of rare variants in the region. Examples of the burden tests include the cohort allelic sum test (CAST) [8], the combined multivariate and collapsing method (CMC) [9], and non-parametric weighted sum test (WST) [10], which imposes weights when collapsing rare variants. Several alternative burden methods are largely based on the same approach [11, 12, 13, 14]. Since all burden tests implicitly assume that all the rare variants in a region are causal and affect the phenotype in the same direction with similar magnitudes, they suffer from a substantial loss of power when these assumptions are violated [15, 16].

Kernel based test methods, such as the sequence kernel association test (SKAT) [17], are non-burden tests. Instead of aggregating variants, SKAT aggregating individual variant score test statistics with weights when SNP effects are modeled linearly. More generally, SKAT aggregates the associations between variants and the phenotype through a kernel matrix and can allow for SNP-SNP interac-

tions, i.e., epistatic effects. SKAT is especially powerful when a genetic region has both protective and deleterious variants or many non-causal variants. SKAT is derived as a variance component test in the induced mixed models when regression coefficients are assumed to be independent and follow a distribution with variance component. SKAT efficiently calculates the p-value analytically for large samples, and is hence computationally fast for analyzing genome-wide sequencing association studies. It has been shown that some non-burden tests[15, 18, 19] are a special case of SKAT [16, 17].

Although SKAT provides attractive power and makes few assumptions about the rare variant effects, it has several limitations. It can be less powerful than burden tests if a large proportion of the rare variants in a region are truly causal and influence the phenotype in the same direction[16, 17]. In addition, large sample based p-value calculations using SKAT can produce conservative type I errors for small sample case/control sequencing association studies, which could lead to power loss [17, 20]. This is particularly an issue in current exome sequencing studies, which are often of small sizes.

This paper aims to address the limitations of burden tests and SKAT, and has two objectives. First, we propose a unified test for rare variant effects by using the data to find the optimal linear combination of the burden test and SKAT to maximize the power. We show that this unified test belongs to an extended SKAT test family by allowing the regression coefficients of variants to be correlated [21]. We hence term this optimal unified test as SKAT-O, which is optimal in both scenarios. Specifically, using the data, it automatically behaves like the burden test in the situation when the burden test is more powerful than SKAT, and behaves like SKAT in the situation when the SKAT is more powerful than the burden test.

The second objective of this paper is to improve the performance of SKAT and SKAT-O in small sample case-control sequencing association studies. The original SKAT test has been found to be conservative for small samples [17, 20]. In this paper we develop an analytic adjustment method for SKAT and SKAT-O by estimating the small sample variance and kurtosis precisely. This allows us to precisly calculate the reference distribution for a small sample, thereby properly controlling the type I error. This is motivated by the fact that many of the current exome sequencing studies, such as those in the NHLBI Exome Sequencing Project, have small sample sizes, e.g., the Acute Lung Injury (ALI) exome sequencing data that are discussed in this paper have 88 subjects; the chronic Pseudomonas aerugi-

nosa infection exome sequencing data have 91 subjects [22]. The proposed small sample adjustment method is computationally fast and can be effectively applied to whole exome(genome) sequencing studies.

Using extensive simulations and analysis of the ALI exome sequencing data of the NHLBI Lung GO Exome Sequencing Project (ESP), we demonstrate that the small sample adjusted unified test (SKAT-O) has proper type I error rates for small sample sequencing association studies, has higher power in a wide range of settings and is more robust than SKAT and the burden tests.

2 Methods

For simplicity, we assume that we are interested in testing the association between rare variants in a region, e.g., a gene, and a complex trait. For whole exome sequencing studies (WES) and whole genome sequencing studies (WGS), the same method can be applied to one gene or one region at a time and then adjusted for multiple comparisons by the user's method of choice. For WGS, one can consider analysis of one window of the same size, e.g., 10kb, at a time using the moving window approach, or of different sizes, e.g., using haplotype blocks.

2.1 Sequence Kernel Association Test

Assume n subjects are sequenced in a region, e.g., a gene, that has m variants. For the i^{th} subject, let y_i denote a dichotomous phenotype, $\mathbf{G}_i = (g_{i1}, \ldots, g_{im})'$ the genotypes of the m variants ($g_{ij} = 0, 1, 2$), $\mathbf{X}_i = (x_{i1}, \ldots, x_{is})'$ the covariates. Without loss of generosity, we assume an additive genetic model and a binary trait. Results are similar for quantitative traits. To relate genotypes to a dichotomous phenotypes, we consider the logistic regression model

$$logit(\pi_i) = \gamma_0 + \mathbf{X}_i'\boldsymbol{\gamma}_1 + \mathbf{G}_i'\boldsymbol{\beta}, \tag{1}$$

where π_i is the disease probability, $\boldsymbol{\gamma}_1$ is an $s \times 1$ vector of regression coefficients of covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$ is an $m \times 1$ vector of regression coefficients of genetic variants. The standard m degrees of freedom (DF) test for no genetic association H_0: $\boldsymbol{\beta} = 0$ has little statistical power when m

is large. Several approaches have been proposed to reduce the DF and increase analysis power. Two classes of tests have been proposed: burden and non-burden tests.

Burden tests treat the β_j's to be the same up to a weight function, i.e., $\beta_j = w_j\beta_c$, where w_j is a weight function that may depend on properties of the j^{th} variant. For example, one can assume w_j to be a function of Minor Allele Frequency (MAF). Then (1) becomes

$$logit(\pi_i) = \boldsymbol{\gamma}_0 + \mathbf{X}_i'\boldsymbol{\gamma}_1 + \beta_c \left\{ \sum_{j=1}^{m} w_j g_{ij} \right\}, \tag{2}$$

and the association between the m genetic variants and a dichotomous trait can be tested using a one DF test for H_0: $\beta_c = 0$. Suppose $\widehat{\pi}_i$ is the estimated probability of y_i under the null hypothesis, i.e., $\widehat{\pi}_i$ is calculated by fitting the null model

$$logit(\pi_i) = \gamma_0 + \mathbf{X}_i'\boldsymbol{\gamma}_1. \tag{3}$$

Then the burden score statistic for testing $H_0 : \beta_c = 0$ is

$$Q_B = \left[\sum_{i=1}^{n} (y_i - \widehat{\pi}_i) \left(\sum_{j=1}^{m} w_j g_{ij} \right) \right]^2, \tag{4}$$

which asymptotically follows scaled χ_1^2 under the null hypothesis. This weighted burden test is equivalent to the weighted sum test of Madsen and Browning [10] and Han and Pan [13], where Madsen and Browning [10] assumed $w_j = 1/\sqrt{\widetilde{p}_j(1-\widetilde{p}_j)}$, where $\widetilde{p}_j$ is the estimated MAF for SNP j using controls. When all w_j are the same and analysis is restricted to rare variants, e.g., the variants with MAF$< 5\%$, Q_B is equivalent to the Morris and Zeggini (MZ) test [12]. The key limitation of the weighted burden test is that it assumes all rare variants in the region are causal and are associated with the trait in the same direction with the same magnitude after weighting, and thus the presence of both protective and deleterious variants or a large number of non-causal variants would substantially reduce its statistical power.

SKAT[17], which includes the C-alpha test[15] and the SSU test[18] as a special case, is a non-

burden test. SKAT assumes that the β_j in (1) are independent and follow an arbitrary distribution with mean 0 and variance $w_j^2\tau$. The null hypothesis $H_0 : \boldsymbol{\beta} = 0$ in model (1) is equivalent to the hypothesis $H_0 : \tau = 0$. Hence SKAT is a variance component test under the induced logistic mixed model [23]. Specifically, under the logistic model (1), the SKAT statistic can be written as

$$Q_S = (\mathbf{y} - \hat{\boldsymbol{\pi}})'\mathbf{K}(\mathbf{y} - \hat{\boldsymbol{\pi}}), \tag{5}$$

where $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \ldots, \hat{\pi}_n)'$ is a vector of the estimated probability of $\mathbf{y}$ under the null model (3), and $\mathbf{K} = \mathbf{GWWG}'$ is an $n \times n$ kernel matrix, where $\mathbf{G} = (\mathbf{G}_1, \ldots \mathbf{G}_n)'$ is an $n \times m$ genotype matrix, and $\mathbf{W} = diag\,(w_1, \ldots, w_m)$ is an $m \times m$ diagonal weight matrix. The SKAT statistic Q_S can be simplified as the weighted sum of the individual SNP score statistics as

$$Q_S = \sum_{j=1}^{m} w_j^2 S_j^2 = \sum_{j=1}^{m} w_j^2 \left\{ \sum_{i=1}^{n} g_{ij}(y_i - \hat{\pi}_i) \right\}^2, \tag{6}$$

where $S_j = \sum_{i=1}^{n} g_{ij}(y_i - \hat{\pi}_i)$ is the score statistic for testing $H_0 : \beta_j = 0$ in the individual SNP j only model

$$logit(\pi_i) = \gamma_0 + \mathbf{X}_i'\boldsymbol{\gamma} + g_{ij}\beta_j.$$

Note that the notation of the weights w_j here is slightly different from Wu *et al.*(2011) [17]. Our w_j^2 here was denoted as w_j in Wu *et al.*(2011) [17]. We modified the notation in this paper to allow for a simple notation for the burden test.

The weight w_j can be flexibly chosen using the observed data, such as a function of MAF, or external information, such as PolyPhen or SIFT score [24, 25]. For example, Beta density function of MAF can be used a weight function in which $w_j = Beta(p_j, a_1, a_2)$, where p_j is the estimated MAF for SNP j using all cases and controls, and the parameters a_1 and a_2 are pre-specified. The SKAT test statistic Q_S asymptotically follows a mixture of chi-square distributions [17]. For large samples, the p-value of SKAT can be quickly and accurately calculated by either matching the moments or inverting the characteristic function [26, 27, 28].

A comparison of the burden statistic Q_B in (4) and the SKAT statistic Q_S in (6) shows that the bur-

den test aggregates the variants first before performing regression, while SKAT aggregates individual variant test statistics. Hence SKAT is robust to the mixed signs of β's and a large fraction of non-causal variants.

2.2 Optimal Unified association test

The foregoing discussions suggest that burden tests are not powerful when the target region has many non-causal variants or causal variants have different directions of association, while SKAT is powerful in these situations [17]. However, if the target region has a high proportion of causal variants with the effects in the same direction, burden tests can be more powerful than SKAT. Because such prior biological knowledge is often unknown, and the underlying genetic mechanisms vary from one gene to another across the genome, it is of substantial interest to develop a test that is optimal for both scenarios in whole exome (genome) sequencing studies. Here we propose a unified test that includes burden tests and SKAT in one framework. In particular, the test statistic of the proposed unified test is

$$Q_\rho = \rho Q_B + (1-\rho)Q_S, \quad 0 \le \rho \le 1, \tag{7}$$

which is a weighted average of SKAT and burden test statistics. One can easily see that the unified test reduces to SKAT when $\rho = 0$ and to the burden test when $\rho = 1$, i.e, the class of tests Q_ρ ($0 \le \rho \le 1$) includes the burden test and SKAT as special cases. One can further show that the unified test (7) is equivalent to the generalized SKAT test [21] derived as the variance component score statistic assuming the regression coefficients β_j in (1) follow an arbitrary distribution with mean 0 and variance $w_j^2\tau$ and pairwise correlation ρ between different β_j's as

$$Q_\rho = (\mathbf{y} - \hat{\boldsymbol{\pi}})'\mathbf{K}_\rho(\mathbf{y} - \hat{\boldsymbol{\pi}}), \tag{8}$$

where $\mathbf{K}_\rho = \mathbf{G}\mathbf{W}\mathbf{R}_\rho\mathbf{W}\mathbf{G}'$ is an $n \times n$ kernel matrix, $\mathbf{R}_\rho = (1-\rho)\mathbf{I} + \rho\underline{1}\underline{1}'$ is an $m \times m$ compound symmetric matrix, and $\underline{1} = (1, \ldots 1)'$. This implies that the weight ρ in (7) can be interpreted as the correlation of the regression coefficients β_j's ($j = 1, \ldots, m$). If the regression coefficients β_j are perfectly correlated ($\rho = 1$), they will be all the same after weighting, and one should collapse the variants first

before running regression, i.e., using the burden test. If the regression coefficients are unrelated to each other, one should use SKAT.

In practice, the optimal weight ρ is unknown and needs to be estimated from the data to maximize the power. Lee *et al.*(2011) [21] proposed the optimal test procedure for the generalized SKAT, which selects the weight ρ to maximize the power. It follows that the optimal unified test can be calculated as

$$Q_{optimal} = \min_{0 \leq \rho \leq 1} p_\rho, \tag{9}$$

where p_ρ is the p-value computed based on a given ρ. The optimal unified test statistic can be easily obtained by the simple grid search: set a grid $0 = \rho_1 < \rho_2 < \ldots < \rho_b = 1$, then

$$Q_{optimal} = min\{p_{\rho_1}, \ldots, p_{\rho_b}\}.$$

For large samples, Lee *et al.*(2011) [21] showed that for given ρ, each test statistic Q_ρ can be decomposed into a mixture of two random variables, one asymptotically follows a chi-square distribution with one DF, and the other can be asymptotically approximated to a mixture of chi-square distributions with a proper adjustment. Hence the p$-$value of $Q_{optimal}$ can be quickly obtained analytically using a one-dimensional numerical integration. We term the optimal unified test as SKAT-O in view of the fact that it is an optimal test in the generalized SKAT family.

2.3 Small sample optimal unified test

One of the key strengths of SKAT and SKAT-O over the other competing methods is their ability to efficiently compute asymptotic p-values without the need for resampling, and it is easy to adjust for covariates. This is particularly advantageous in whole genome (exome) sequencing studies when a large number of tests is performed and one needs to control for multiple comparisons and account for population stratification. However, if the trait is binary and sample sizes are small, e.g., hundreds of subjects, the large sample based p-value calculations in Wu *et al.*(2011) [17] and Lee *et al.*(2011) [21] have been found to produce conservative results, which can lead to incorrect type I error control and power loss [17, 20, 21].

As most current whole exome sequencing studies, such as the those of NHLBI Exome Sequencing Project, have small sample sizes, there is a pressing need to develop a method that works well for small samples. We propose in this section small sample adjusted p-value calculations for SKAT and SKAT-O.

We first consider p-value calculations for SKAT when sample sizes are small. When variants are rare, and the genotype matrix $\mathbf{G}$ is sparse, the small sample variance of Q_S is much smaller than the asymptotic variance. Hence, we re-adjust the moments of the null distribution of Q_S. Suppose Q_S was obtained with known $\boldsymbol{\pi}$. Denote by $\mathbf{D} = diag\{\pi_i(1-\pi_i)\}$, where π_i is the probability of being a case under the null. Let $\widetilde{\mathbf{K}} = \mathbf{D}^{1/2}\mathbf{K}\mathbf{D}^{1/2}$, $\boldsymbol{\Lambda} = diag\{\lambda_1, \ldots, \lambda_q\}$ be a diagonal matrix of ordered non-zero eigenvalues, $\mathbf{U} = [\mathbf{u}_1, \ldots \mathbf{u}_q]$ be an $n \times q$ eigenvector matrix of $\widetilde{\mathbf{K}}$, and u_{ij} be the i^{th} element of $\mathbf{u}_j$. In the Appendix, we show that the small sample mean and variance of SKAT under the null hypothesis are

$$E\left[Q_S|\mathbf{U}, \boldsymbol{\Lambda}, \boldsymbol{\pi}\right] = \sum_{j=1}^{q} \lambda_j \quad \text{and} \quad Var\left[Q_S|\mathbf{U}, \boldsymbol{\Lambda}, \boldsymbol{\pi}\right] = \sum_{j=1,k=1}^{q} \lambda_j \lambda_k c_{jk}, \tag{10}$$

where

$$c_{jk} = \sum_{i=1}^{n} \frac{u_{ij}^2 u_{ik}^2 (3\pi_i^2 - 3\pi_i + 1)}{\pi_i(1-\pi_i)} + \sum_{i_1 \neq i_2}^{n} u_{i_1 j}^2 u_{i_2 k}^2 + 2 \sum_{i_1 \neq i_2}^{n} u_{i_1 j} u_{i_2 j} u_{i_1 k} u_{i_2 k} - 1.$$

A comparison of these results with those in Wu *et al.*(2011) [17] shows that the small sample mean of Q_s is the same as the asymptotic mean of Q_S but the small sample variance differs from the asymptotic variance. Using the estimated moments, the p-value can then be calculated as

$$1 - F((Q_S - \mu_Q)\sqrt{2df}/\sqrt{v_Q} + df|\chi_{df}^2), \tag{11}$$

where $F(\cdot|\chi_{df}^2)$ is the distribution function of χ_{df}^2, and

$$\mu_Q = \sum_{j=1}^{q} \lambda_j, \quad v_Q = \sum_{i,j=1}^{q} \lambda_i \lambda_j \widehat{c}_{ij}, \quad \text{and} \quad df = \frac{\sum_{j=1}^{q} \lambda_j^{*4}}{(\sum_{j=1}^{q} \lambda_j^{*2})^2},$$

and $\lambda_j^* = \lambda_j \widehat{c}_{jj}/\sqrt{2}$. $\widehat{c}_{jk}$ is an estimated c_{jk} with $\widehat{\boldsymbol{\pi}}$. We can apply the same approach to the optimal unified test SKAT-O; details are shown in the Appendix.

Note that the results here do not restrict the kernel matrix $\mathbf{K}$ to be the linear weighted kernel. This

proposed small sample adjustment procedure can be used for any types of kernel matrices such as IBS and IBS weighted kernels [17, 29]

2.4 Small sample SKAT and unified test with higher moments adjustments

In the previous section, we adjusted the asymptotic null distribution of Q_S and $Q_{optimal}$ using the small sample variance of Q_S and $Q_{optimal}$. If the sample size is very small, e.g., $n = 88$ in the ALI whole exome sequencing data, this approach may not be accurate enough to correct small sample type I error rates. We thus need to adjust higher moments, especially kurtosis. Unfortunately, deriving the analytical formula of the kurtosis of Q_S is a daunting task. Hence we propose a practical approach in which the kurtosis is estimated through a re-sampling method. When there is no covariate, the kurtosis of the null distribution of Q_S can be estimated from B permutation samples of phenotypes, and then the estimated kurtosis can be used to calculate the degrees of freedom parameter (df) in (11).

Specifically, suppose $Q^*_{s,b}$ ($b = 1, \ldots, B$) is the SKAT test statistic from the permutation sample $\mathbf{y}^*_b$. The sample kurtosis is

$$\hat{\gamma} = \frac{\hat{\mu}_4}{\hat{\sigma}^4} - 3,$$

where

$$\hat{\mu}_4 = \frac{1}{B}\sum_{b=1}^{B}(Q^*_{s,b} - \mu_Q)^4, \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{B}\sum_{b=1}^{B}(Q^*_{s,b} - \mu_Q)^2.$$

The degrees of freedom of the mixture of chi-square distribution (df) in (11) is modified as

$$df = 12/\hat{\gamma},$$

and the p-values can be calculated using the (11).

When there are covariates to adjust for, the simple permutation method cannot be used. Instead we propose to generate re-sampled phenotypes from the parametric bootstrap [30]. We first estimate π_i under the null model and use it to generate $\mathbf{y}^*_b$ with the same number of cases and controls.

It should be noted that our method has a computation time advantage over calculating p-values based on permutations or bootstrap samples that are obtained as a proportion of $Q^*_{s,b}$ larger than Q_S. For whole exome sequencing studies, one needs to calculate p-values at the $10^{-5} \sim 10^{-6}$ level to

account for multiple comparison adjustments for performing tests for 20,000 genes. This requires more than $10^7 \sim 10^8$ permutations or bootstraps for each gene. However, our approach requires sampling phenotypes under the null model only $10,000$ times to obtain stable estimates of the higher moments. Note that the null model is the same across different genes, and hence the same re-sampled bootstrap phenotypes under the null model can be used for all the genes across the genome. We hence can save a substantial amount of computation time.

2.5 Numerical Simulations

We conducted extensive simulation studies to evaluate the performance of the proposed methods for binary traits when sample sizes are small. We generated sequence data of European ancestry from 10,000 chromosomes over 1 Mb regions using the calibrated coalescent model [31]. We randomly selected regions with length 3Kb, and tested for associations in all simulation settings.

2.5.1 Type I Error Simulations

We first generated datasets under the null model to evaluate the type I error control of the proposed methods. Dichotomous phenotypes with 50% cases and 50% controls were generated from the null logistic regression model

$$logit(\pi_i) = \gamma_0 + 0.5X_{1i} + 0.5X_{2i},$$

where X_1 was a continuous covariate from $N(0,1)$, X_2 was a binary covariate from $Bernoulli(0.5)$, and γ_0 was chosen to create a trait prevalence of 0.01 under the null hypothesis. We applied six different methods to each of the randomly selected 3 Kb regions: 1) counting based burden test (N); 2) weighted burden test (W); 3) SKAT without small sample adjustment (SKAT); 4) unified test without small sample adjustment (SKAT-O); 5) small sample adjusted SKAT (adjusted SKAT); and 6) small sample adjusted unified test (adjusted SKAT-O).

For all methods except N, $Beta(1,25)$ weights were used to upweight rarer variants. For N, we used flat weights and restricted variants with observed MAF < 0.03. For both N and W, the likelihood ratio test was used to compute p-values. The p-values of the optimal unified tests were computed using the 11 points of equal-sized grids search of ρ from 0 to 1. For small sample adjusted SKAT and

adjusted SKAT-O, the sample kurtosis was estimated from 10,000 bootstrapped phenotype sets. Three different total sample sizes (n =200, 500, and 1,000) were considered, with 10,000 simulated data sets for each sample size.

To investigate type I error rates in the SKAT family tests when the α level is set at a level for exome-wide testing, we conducted simulations with slightly different settings. In order to reduce the computational burden, we first generated $20,000$ genotype sets of randomly selected regions, and then generated 500 phenotype sets for each genotype set. A total of 10^7 phenotypes were generated, and type I error rates were estimated by the proportion of p-values smaller than the given α level.

2.5.2 Power Simulations

To evaluate the power of the proposed unified tests and their small sample adjustments relative to the competing methods, we simulated datasets under the alternative model. As with the type I error simulations, we randomly selected $3kb$ regions from a broader 1Mb region, but we then randomly chose causal variants from the rare variants with true MAF < 0.03. The dichotomous phenotypes with 50% cases and 50% controls were simulated from

$$logit(\pi_i) = \gamma_0 + 0.5X_{i1} + 0.5X_{i2} + \beta_1 g_{i1} + \cdots + \beta_s g_{is},$$

where $(g_1, \cdots, g_s)$ were selected causal variants. Covariates X_1 and X_2 followed the same distribution as in the type I error simulation, and γ_0 was chosen to create a disease prevalence of disease 0.01 under the null hypothesis.

To study the effects of varying proportions of variants being causal variants, we considered three different settings in which 10%, 20%, and 50% of the rare variants were causal variants. For each setting, we considered three different sign configurations of the non-zero β's: all β_js were positive, 80% of β_js were positive, and 50% of β_js were positive. We used $|\beta_j| = c|\log_{10}(p_j)|/2$, where p_j was the MAF of the j^{th} variant. When 10% of the rare variants were causal, $c = \log(7)$, which gives an odds ratio equal to 7 for a variant with MAF = 0.01. When 20% and 50% of the rare variants were causal variants, $c = \log(5)$ and $\log(2.5)$, respectively, so the powers would not be too close to one and we can distinguish the powers of different methods. For each setting, 1000 datasets were generated and the

power was estimated as the proportion of p-values smaller than a given α level.

2.6 The NHLBI ALI Exome Sequencing Data

The Acute Lung Injury whole exome sequencing data were part of the LungGO of the NLBLI Exome Sequencing Project. We performed exome sequencing of 88 individuals with Acute Lung Injury (ALI)[32] selected from the extremes of the severity spectrum. Individuals with ALI and severe hypoxemia (Partial pressure of arterial oxygen/Fraction of inspired oxygen <200) were enrolled from the intensive care unit at the Massachusetts General Hospital. Those with very high or very low "ventilator-free days" (VFD), a composite variable measuring the degree of dependence on mechanical ventilation in the first 28 days of hospital admission[33], were selected for sequencing. Exome sequencing was completed on 88 subjects (n=43 high severity ALI (VFD<2), n=45 low severity ALI (VFD< 24)) at the Northwest Genomics Center at the University of Washington.

To call SNP variants, the GATK tool of the Broad Institute was used [34], and approximately 130,000 SNP variants on 17755 genes were identified. We subsequently filtered out variants with high missing rates (missing rate> 0.1) and low quality control scores using the GATK tool, i.e, keeping variants with Qual< 30, QD< 5, AB> 0.75 or SB> -0.10, % of missing $< 10\%$. This gave a total of 106,736 variants.

For SKAT and the unified test (SKAT-O), we used all the variants. For the weighted burden test (W) and the counting based burden test (N), due to the very small sample size, we used MAF< 0.05 as the criterion to define rare variants to be included in the analysis. Any genes with fewer than four rare variants with (MAF< 0.05) were excluded from the analysis, as these genes have little information about association with the phenotype given the small sample size. A total of 6,488 genes remained for analysis. All six methods discussed in the simulation study were applied to the data. The first two principal components calculated using EIGENSOFT[35] from all 106,736 variants were used as covariates to adjust possible population stratification.

3 Results

3.0.1 Type I Error Simulation Results

To investigate the type I error rates with exome-wide α levels, we generated 10^7 data sets. The results are given in Table 1. Three different $\alpha = 10^{-3}, 10^{-4}$ and 2.5×10^{-6} levels were considered. Note that $\alpha = 2.5 \times 10^{-6}$ is Bonferroni-adjusted level $\alpha = 0.05$ when simultaneously testing 20,000 genes. Table 1 clearly shows that the unadjusted SKAT and unified test (SKAT-O) had substantially deflated type I error rates for small sample sizes. The unified test (SKAT-O) was less conservative than SKAT, and had correct type I error control when the sample size was 1000. Both the proposed small sample adjusted SKAT and unified test (adjusted SKAT and adjusted SKAT-O) performed much better than their unadjusted counterparts in small samples. They controlled type I error rates accurately over all sample sizes and all significance levels. We also evaluated the type I error rates of the the burden tests and SKAT and SKAT-O tests at $\alpha = 0.05$ using 10,000 simulated datasets (Supplementary Table S1), and the results agreed with Table 1. Overall, our type I error simulation results confirm empirically that the proposed small sample adjustment methods accurately control type I error rates.

3.0.2 Power Simulation Results

We compared the powers for the burden tests, SKAT and the unified test (SKAT-O) and their small sample adjustments, i.e., all the six methods considered in the type I error simulations. The number of observed variants is given in Supplementary Table S2. On average, depending on sample sizes, 20 to 40 variants were observed in each region. We first considered the scenario that all causal variants were deleterious variants, i.e. the effects of the causal variants were all in the same direction. Figure 1 reports that by properly controlling the type I error, the small sample adjusted SKAT (adjusted SKAT) was more powerful than SKAT in every configuration, and the power gap was large when the sample size was small or when the significance level was small. Power for the unified test (SKAT-O) and its small sample adjustment (adjusted SKAT-O) showed a similar pattern. Between the two burden tests, W was more powerful than N for these simulation configurations, suggesting that proper weighting can increase power.

When only 10% of the rare variants were causal, small sample adjusted SKAT (adjusted SKAT) was the most powerful test. The burden tests had substantially lowest power, indicating that burden tests are not powerful in the presence of a large fraction of non-causal variants. When the proportion of causal rare variants increased, the burden tests performed better. When 50% of the rare variants were causal, the burden tests had a higher power than small sample adjusted SKAT.

The optimal unified tests (SKAT-O and adjusted SKAT-O) consistently performed very well in both settings above. They behave like SKAT when SKAT is more powerful than burden tests, and behave like burden tests when burden tests are more powerful than SKAT. Small sample adjusted optimal unified (adjusted SKAT-O) outperformed its unadjusted counterpart (SKAT-O), especially when sample sizes are small, e.g., n=200. When 20% of rare variants were causal, adjusted SKAT-O was the most powerful test.

We next performed simulations in which 20%/80% and 50%/50% of causal variants were protective/deleterious variants (Figure 2 and 3). The same odds ratio functions from above were used. Similar to the case when all causal variants were deleterious (Figure 1), small sample adjusted SKAT (adjusted SKAT) had higher power than SKAT, and small sample adjusted unified test (adjusted SKAT-O) had higher power than its unadjusted counterpart (SKAT-O). The presence of mixed protective and deleterious variants substantially reduced the powers of burden tests, since the effects of the causal variants canceled out. With 50%/50% of the causal variants being protective/deleterious, the powers of the burden tests were less than half those of SKAT and its small sample adjustment. The optimal unified test behaved similarly to SKAT but had better power than SKAT and the burden test when 50% of the rare variants were causal and 50%/50% of the causal variants being protective/deleterious. Small sample adjustment for both SKAT and the unified test improved the power. All tests had lower power relative to the situation in which all causal variants were deleterious (Figure 1). This is because for the given low prevalence, the presence of protective variants resulted in fewer causal variants selected into the case-control sample (Supplementary Table S3).

We present the optimal ρ values estimated by adjusted SKAT-O in Supplementary Figure S5. It shows that SKAT-O generally selects large ρ values when the percentage of causal variants is high and all causal variants are deleterious, and selects small ρs when either the percentage of causal variants is

low or some causal variants are protective. The estimated optimal ρ varies between different datasets as it accounts for sampling variation. We also conducted additional simulations for the extreme situation in which all rare variants in a region were causal and deleterious (Supplementary Figure S6). In this scenario, the theoretical optimal $\rho = 1$. As expected, W has the highest power. The adjusted SKAT-O has a slightly smaller power than W, since it assumes ρ is unknown and searches for the optimal ρ in [0,1]. However, the power gap between W and adjusted SKAT-O is small.

The power simulation results show that the optimal unified test (SKAT-O) is robust to the proportion of rare variants that are causal and to the directions of the causal variant effects (relative to the other tests); it performs very well in a wide range of situations; and it outperform SKAT and the burden tests. The proposed small sample adjustment increases the power by properly controlling for type I error rate, especially when the sample size is small, or α is very small.

3.1 Analysis of the NHLBI Acute Lung Injury (ALI) Exome Sequence Data

We applied the six methods used in simulation studies (burden tests, SKAT, the unified test (SKAT-O), and their small sample adjustments) to analysis of the NHLBI ALI exome sequencing data of 88 subjects to identify genes associated with ALI severity. We restricted our analysis to the genes with at least four variants with MAFs$<$0.05. A total of 6,488 genes were analyzed (see the Method Section).

Figure 4 gives the QQ plots of the p-values calculated using all the six methods. Given the small sample size, no p-value achieved the Bonferroni adjusted genome-wide significance at $\alpha = 7.7 \times 10^{-6}$. The QQ plots of unadjusted SKAT and unified test (SKAT-O) were skewed downward, suggesting these tests were conservative. Interestingly, the QQ plots of the burden tests had a slightly anti-conservative pattern. The QQ plots of small sample adjusted SKAT and unified test (adjusted SKAT-O) were close to the 45 degree line, suggesting that the proposed small sample adjustment methods worked well and properly controlled type I error rates. There were eight genes with p-values$< 10^{-3}$ by the adjusted SKAT-O. A total of 741 genes had the estimated optimal ρ values between 0.1 to 0.9.

We next restricted our analysis to the functional variants that are missense, nonsense, and splicing sites variants. Similar to the first analysis, we only considered genes that have at least four functional variants with MAF< 0.05. A total of 2,939 genes were used in analysis. The QQ plots of the six

methods are given in Supplementary Figure S1. The patterns of these QQ plots are similar to those in Figure 4. There were five genes with p-values $< 10^{-3}$ by the adjusted SKAT-O. Myosin light chain kinase (MYLK [MIM 600922]), a gene that was previously found to be associated with susceptibility to ALI [36, 37], was second most-significant in the adjusted SKAT analysis and 4th most significant in the adjusted SKAT-O analysis.

We compared the p-values obtained using small sample adjusted SKAT (adjusted SKAT) and the adjusted optimal unified test (adjusted SKAT-O) with those obtained using the burden test (W) (Supplementary Figure S2). These comparisons show that the p-values obtained using adjusted SKAT and W are quite different from each other, indicating that these two tests evaluate different aspects of association patterns. In contrast, the p-values obtained with adjusted SKAT-O were more highly correlated with those obtained with either adjusted SKAT or W as p-values declined, consistent with the expectation that the optimal unified test uses the data to adaptively choose an optimal test to maximize power.

4 Discussions

In this paper, we present a unified rare variant test framework that includes both burden tests and the non-burden test SKAT as special cases. The proposed optimal unified test (SKAT-O) procedure uses the data to adaptively select the best linear combination of the burden test and SKAT to maximize test power. Similar to SKAT, the proposed SKAT-O is computationally efficient and easily adjusts for covariates such as age, gender and principal components for population stratifications. We show in simulation studies that SKAT and burden tests can both lose power when underlying assumptions are violated. However, the optimal unified test SKAT-O is more robust in a wide range of circumstances we have considered. In the SKAT package, we also provide power and sample size calculations using SKAT, SKAT-O and their small sample adjustments to help investigators design sequencing association studies.

In whole exome sequencing studies or whole genome sequencing studies, one would expect that some genes or genomic regions have a high proportion of causal variants with the same association direction, while other regions have many non-causal variants or causal variants with different associa-

tion directions. Applying only either a burden test or SKAT would decrease the chance of detecting all of those genes. However, the use of SKAT-O is more robust and will increase the chance of detecting these genes.

Although we have considered in this paper a wide range of simulation settings that are practical interest, we note that simulation results depend on simulation settings. Thus, our results of comparing different methods should be interpreted within the context of the range of simulation settings we have considered. It would be useful to examine the generality of the results in other simulation settings in the future.

Due to high sequencing costs, many of the existing whole exome sequencing studies have small sample sizes. As the second goal of this paper, we developed small sample adjustment methods to correct p-values for SKAT and SKAT-O to properly control the type I error rate and increase the power. Using extensive simulation studies and the NHLBI whole exomes from individuals who developed ALI, we demonstrated good performance of the proposed small sample adjustment methods both in terms of type I error control and power increase.

In this study, we only considered dichotomous traits. However, the application of SKAT-O to quantitative trait data is straightforward using equation (1) with a linear regression. Furthermore, we note that the small sample adjustment is not necessary for continuous traits when the normality assumption is true, because the small sample distributions of SKAT and SKAT-O are the same as their asymptotic distributions under normality.

We note that the proposed small sample adjustment methods are still computationally efficient even though we estimate the kurtosis using resampling. It only requires 10,000 bootstrap samples to accurately estimate the kurtosis, which is a substantially smaller computational burden compared to obtaining permutation or bootstrap p-values, which require 10^7 or 10^8 resampled phenotypes to accurately obtain p-values in the $10^{-5} \sim 10^{-6}$ ranges.

In simulation and real data analysis, we used a flexible beta weight to up-weight the influence of rarer variants. Similar results are obtained using logistic weight $w_j = exp((a_1 - p_j)a_2)/\{1 + exp((a_1 - p_j)a_2\}$ for the ALI exome sequencing data (see Supplementary Figures S3 and S4). In addition to using a function of MAF of variants as weights, functional information can also be used to choose variants

to be tested or to construct the weight. For example, only functional variants such as nonsense and missense variants can be used to test association, or functional information scores such as PolyPhen or SIFT scores [24, 25] can be used to construct a weight (an area under active study).

Recently several adaptive methods have been proposed to increase the power. For example, the VT test [11] tries to find the optimal MAF threshold of rare variants by varying the threshold, and EREC [20] estimates a regression coefficient of each variant and uses them as the weight. Those approaches could improve the power compared to the burden tests. However, the VT test makes similar assumptions to the burden tests, i.e, requires majority of rare variants under the optimal threshold to be causal and have effects in the same direction. The EREC method requires estimation of regression coefficients, which are difficult to be estimated stably for rare variants. Addition of a stabilizing constant in EREC can reduce the power relative to asymptotic calculations and make the test behave more like burden tests. Furthermore, these methods are computationally intensive when applied to large scale sequencing studies, e.g., whole exome(genome) sequencing studies, because they rely on a large number of permutation or bootstrap samples to compute p-values, and are difficult to control for covariates, such as population stratification. In contrast, SKAT-O and its small sample adjustment compute p-values efficiently and can be easily applied to whole exome(genome) sequencing studies.

With the rapid advance of bio-techonology, new biological knowledge will become available, and new sequencing technology and study designs will be devleoped. In the fast-moving next generation sequencing era, it is of significant importance to incorporate these new information to improve statistical and computational tools for detecting rare variants associated with complex diseases.

SUPPLEMENTAL DATA

Supplemental Data includes four figures one table can be found with this article online at *http://www.cell.com/AJHG*.

ACKNOWLEDGEMENTS

This work was supported by grants R37 CA076404 and PO1 CA134294 (S.L. and X.L.), RC2 HL101779 (D.C., S.L., X.L., M.W), R01HL060710 (D.C.), UC2HL102923 (M.B., K.B., M.E.,M.W). K.B. was supported in part by the Mary Beryl Patch Turnbull Scholar Program. The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute (NHLBI) and the contributions of the research institutions, study investigators, field staff and study participants in creating this resource for biomedical research. Funding for GO ESP was provided by NHLBI grants RC2 HL-103010 (HeartGO), RC2 HL-102923 (LungGO) and RC2 HL-102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL-102925 (BroadGO) and RC2 HL-102926 (SeattleGO).

Appendices

A. Mean and Variance of Q_S under the NULL hypothesis

Suppose $\widetilde{\mathbf{y}} = \mathbf{D}^{-1/2}(\mathbf{y} - \boldsymbol{\pi})$, where $\mathbf{D} = diag[\pi_1(1-\pi_1), \ldots, \pi_n, (1-\pi_n)]$. Then for all $j = 1, \ldots, q$, $E\left[(\widetilde{\mathbf{y}}'\mathbf{u}_j)^2 | \boldsymbol{\pi}, \mathbf{u}_j\right] = 1$, and

$$\begin{aligned} E\left[(\widetilde{\mathbf{y}}'\mathbf{u}_j)^4 | \boldsymbol{\pi}, \mathbf{u}_j\right] &= \sum_{i=1}^{n} u_{ij}^4 E(\widetilde{y}_i^4 | \boldsymbol{\pi}, \mathbf{u}_j) + 3\sum_{i \neq k}^{n} u_{ij}^2 u_{kj}^2 E(\widetilde{y}_i^2 | \boldsymbol{\pi}, \mathbf{u}_j) E(\widetilde{y}_k^2 | \boldsymbol{\pi}, \mathbf{u}_j) \\ &= \sum_{i=1}^{n} u_{ij}^4 \pi_i(1-\pi_i)(3\pi_i^2 - 3\pi_i + 1)/(\pi_i(1-\pi_i))^2 + 3\sum_{i \neq k}^{n} u_{ij}^2 u_{kj}^2, \end{aligned} \tag{A.1}$$

where $\widetilde{y}_i$ is the i^{th} element of $\widetilde{\mathbf{y}}$, and u_{ij} is the i^{th} element of $\mathbf{u}_j$. Therefore,

$$var\left[(\widetilde{\mathbf{y}}'\mathbf{u}_j)^2 | \boldsymbol{\pi}, \mathbf{u}_j\right] = \sum_{i=1}^{n} u_{ij}^4 \pi_i(1-\pi_i)(3\pi_i^2 - 3\pi_i + 1)/(\pi_i(1-\pi_i))^2 + 3\sum_{i \neq k}^{n} u_{ij}^2 u_{kj}^2 - 1.$$

Now we calculate the first two moments of Q given $\boldsymbol{\pi}$, $\mathbf{U}$ and $\boldsymbol{\Lambda}$.

$$E(Q_S | \boldsymbol{\pi}, \mathbf{U}, \boldsymbol{\Lambda}) = E(\sum_{j=1}^{q} \lambda_j \widetilde{\mathbf{y}}'\mathbf{u}_j\mathbf{u}_j'\widetilde{\mathbf{y}} | \boldsymbol{\pi}, \mathbf{U}, \boldsymbol{\Lambda}) = \sum_{j=1}^{q} \lambda_j, \quad \text{and}$$

$$
\begin{aligned}
E(Q_S^2|\boldsymbol{\pi},\mathbf{U},\boldsymbol{\Lambda}) &= E\left[\left(\sum_{j=1}^{q}\lambda_j\widetilde{\mathbf{y}}'\mathbf{u}_j\mathbf{u}_j'\widetilde{\mathbf{y}}\right)^2|\boldsymbol{\pi},\mathbf{U},\boldsymbol{\Lambda}\right] \\
&= \sum_{j=1}^{q}\lambda_j^2 E\left[(\widetilde{\mathbf{y}}'\mathbf{u}_j)^4|\boldsymbol{\pi},\mathbf{U},\boldsymbol{\Lambda}\right] \\
&\quad + \sum_{j\neq k}\lambda_j\lambda_k E\left[\left(\sum_{i_1,i_2}\widetilde{y}_{i_1}\widetilde{y}_{i_2}u_{i_1j}u_{i_2j}\right)\left(\sum_{l_1,l_2}\widetilde{y}_{l_1}\widetilde{y}_{l_2}u_{l_1k}u_{l_2k}\right)|\boldsymbol{\pi},\mathbf{U},\boldsymbol{\Lambda}\right] \qquad \text{(A.2)}
\end{aligned}
$$

Since $E(\widetilde{y}_i|\boldsymbol{\pi}) = 0$, the elements in the second term in (A.2) can contribute to the overall sum only when 1) $i_1 = i_2 = l_1 = l_2$, 2) $i_1 = i_2$ and $l_1 = l_2$, 3) $i_1 = l_1$ and 4) $i_2 = l_2$ or $i_1 = l_2$ and $i_2 = l_1$. Therefore

$$
\begin{aligned}
&E\left[\left(\sum_{i_1,i_2}\widetilde{y}_{i_1}\widetilde{y}_{i_2}u_{i_1j}u_{i_2j}\right)\left(\sum_{l_1,l_2}\widetilde{y}_{l_1}\widetilde{y}_{l_2}u_{l_1k}u_{l_2k}\right)|\boldsymbol{\pi},\mathbf{U},\boldsymbol{\Lambda}\right] \\
&= \sum_{i=1}^{n}u_{ij}^2u_{ik}^2E(\widetilde{y}_i^4|\boldsymbol{\pi}) + \sum_{i_1\neq i_2}^{n}u_{i_1j}^2u_{i_2k}^2E(a_{i_1}^2|\boldsymbol{\pi})E(a_{i_2}^2|\mathbf{p}) + 2\sum_{i_1\neq i_2}^{n}u_{i_1j}u_{i_2j}u_{i_1k}u_{i_2k}E(a_{i_1}^2|\boldsymbol{\pi})E(a_{i_2}^2|\boldsymbol{\pi}) \\
&= \sum_{i=1}^{n}u_{ij}^2u_{ik}^2\pi_i(1-\pi_i)(3\pi_i^2-3\pi_i+1)/(\pi_i(1-\pi_i))^2 + \sum_{i_1\neq i_2}^{n}u_{i_1j}^2u_{i_2k}^2 + 2\sum_{i_1\neq i_2}^{n}u_{i_1j}u_{i_2j}u_{i_1k}u_{i_2k}. \qquad \text{(A.3)}
\end{aligned}
$$

We can calculate the second moment of Q_S by combining (A.2) and (A.3).

B. Null distribution of Small Sample SKAT-O

Define $\mathbf{Z} = \mathbf{D}^{-1/2}\mathbf{G}\mathbf{W}$ and $\bar{\mathbf{z}} = (\bar{z}_1,\ldots,\bar{z}_n)'$, where $\bar{\mathbf{z}}_i = \sum_{j=1}^{m}z_{ij}/m$. We further let $\mathbf{M} = \bar{\mathbf{z}}(\bar{\mathbf{z}}'\bar{\mathbf{z}})^{-1}\bar{\mathbf{z}}'$ and

$$
\psi(\rho) = m^2\rho\bar{\mathbf{z}}'\bar{\mathbf{z}} + \frac{1-\rho}{\bar{\mathbf{z}}'\bar{\mathbf{z}}}\sum_{j=1}^{m}(\bar{\mathbf{z}}'\mathbf{z}_{.j})^2,
$$

where $\mathbf{z}_{.j}$ is the j^{th} column of $\mathbf{Z}$. Following the same argument in Lee *et al.*(2011) [21], it can be shown that Q_ρ is equivalent as

$$
(1-\rho)\kappa_1 + \psi(\rho)\kappa_2, \qquad \text{(A.4)}
$$

where

$$
\kappa_1 = (1-\rho)\widetilde{\mathbf{y}}'(\mathbf{I}-\mathbf{M})\mathbf{Z}\mathbf{Z}'(\mathbf{I}-\mathbf{M})\widetilde{\mathbf{y}} + 2(1-\rho)\widetilde{\mathbf{y}}'(\mathbf{I}-\mathbf{M})\mathbf{Z}\mathbf{Z}'\mathbf{M}\widetilde{\mathbf{y}}
$$

and

$$\kappa_2 = \widetilde{\mathbf{y}}'\bar{\mathbf{z}}\bar{\mathbf{z}}'\widetilde{\mathbf{y}}/\bar{\mathbf{z}}'\bar{\mathbf{z}}.$$

It can be shown that κ_2 asymptotically follows the χ^2_1 distribution, and κ_1 is asymptotically the same as

$$\sum_{k=1}^{q} \lambda_k \eta_k + \zeta,$$

where $\{\lambda_1, \ldots \lambda_q\}$ are non-zero eigenvalues of $\mathbf{Z}'(\mathbf{I}-\mathbf{M})\mathbf{Z}$, $\eta_k (k = 1, \ldots, q)$ are i.i.d χ^2_1 random variables, and ζ satisfies the following conditions:

$$E(\zeta) = 0, \quad Var(\zeta) = 4trace(\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{Z}'(\mathbf{I}-\mathbf{M})\mathbf{Z}),$$
$$Corr(\sum_{k=1}^{q} \lambda_k \eta_k, \zeta) = 0, \quad \text{and} \quad Corr(\kappa_2, \zeta) = 0.$$

We note that asymptotic p-values can be obtained by the one-dimensional integration. When the sample size is small, however, the asymptotic moments of κ_1 and κ_2 can be larger than small sample moments. Thus, we apply the same small sample adjustment procedure to null distributions of κ_1 and κ_2. We first compute small sample variance and kurtosis of κ_1 and κ_2, and apply the moment matching approximation to obtain their adjusted asymptotic distribution. To obtain a p-value, we apply the algorithm in Lee *et al.*(2011) with the adjusted null distribution κ_1 and κ_2.

WEB RESOURCES

An implementation of SKAT and SKAT-O and their small sample adjustments as well as power/sample size calculations in the R language can be found at *http://www.hsph.harvard.edu/∼xlin/software.html.*

More information on the NHLBI exome sequencing project can be found at *http://www.nhlbi.nih.gov/resources/exome.htm*.

References

[1] Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. Nature biotechnology *26*, 1135–1145.

[2] Mardis, E.R. (2008). Next-generation dna sequencing methods. Annu. Rev. Genomics Hum. Genet. *9*, 387–402.

[3] Bodmer, W. and Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. Nature genetics *40*, 695–701.

[4] Schork, N.J. and Murray, S.S. and Frazer, K.A. and Topol, E.J. (2009). Common vs. rare allele hypotheses for complex diseases. Current opinion in genetics & development *19*, 212–219.

[5] Cohen, J.C. and Kiss, R.S. and Pertsemlidis, A. and Marcel, Y.L. and McPherson, R. and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science *305*, 869–872.

[6] Ahituv, N. and Kavaslar, N. and Schackwitz, W. and Ustaszewska, A. and Martin, J. and Hébert, S. and Doelle, H. and Ersoy, B. and Kryukov, G. and Schmidt, S. and others. (2007). Medical sequencing at the extremes of human body mass. The American Journal of Human Genetics *80*, 779–791.

[7] Nejentsev, S. and Walker, N. and Riches, D. and Egholm, M. and Todd, J.A. (2009). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science *324*, 387–389.

[8] Morgenthaler, S. and 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). Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *615*, 28–56.

[9] Li, B. and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. The American Journal of Human Genetics *83*, 311–321.

[10] Madsen, B.E. and Browning, S.R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet *5*, e1000384.

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

[12] Morris, A.P. and Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genetic epidemiology *34*, 188–193.

[13] Han, F. and Pan, W. (2010). A data-adaptive sum test for disease association with multiple common or rare variants. Human Heredity *70*, 42–54.

[14] Zawistowski, M. and Gopalakrishnan, S. and Ding, J. and Li, Y. and Grimm, S. and Zollner, S. (2010). Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. The American Journal of Human Genetics *87*, 604–617.

[15] Neale, B.M. and Rivas, M.A. and Voight, B.F. and Altshuler, D. and Devlin, B. and Orho-Melander, M. and Kathiresan, S. and Purcell, S.M. and Roeder, K. and Daly, M.J. and others. (2011). Testing for an Unusual Distribution of Rare Variants. PLoS Genetics *7*, 161–165.

[16] Basu, S. and Pan, W. (2011). Comparison of Statistical Tests for Disease Association with Rare Variants. Genetic Epidemiology *35*, 606–619.

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

[18] Pan, W. (2009). Asymptotic tests of association with multiple snps in linkage disequilibrium. Genetic epidemiology *33*, 497–507.

[19] Tzeng, J.Y. and Zhang, D. (2007). Haplotype-based association analysis via variance-components score test. The American Journal of Human Genetics *81*, 927–938.

[20] Lin, D.Y. and Tang, Z.Z. (2011). A general framework for detecting disease associations with rare variants in sequencing studies. The American Journal of Human Genetics *89*, 354–367.

[21] Lee, S. and Wu, M. and Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. Biostatistics. doi:10.1093/biostatistics/kxs014.

[22] Emond, M. and Louie, T. and Emerson, J.and Zhao, W. and Mathias, R. A.and Knowles, R.and Wright, F. A.and Rieder, M. J.and Tabor, H. K.and Nickerson, D. A.and Barnes, K. C.and NHLBI Go Exome Sequencing Project ESP-Lung GO and Bigson, R. L and Bamshad, M. I. (2012). Exome sequencing of extreme phenotypes identifies dctn4 as a modifier of chronic pseudomonas aeruginosa infection in cystic fibrosis. Nature Genetics.

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

[24] Ng, P.C. and Henikoff, S. (2003). Sift: Predicting amino acid changes that affect protein function. Nucleic acids research *31*, 3812–3814.

[25] Adzhubei, I.A. and Schmidt, S. and Peshkin, L. and Ramensky, V.E. and Gerasimova, A. and Bork, P. and Kondrashov, A.S. and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nature methods *7*, 248–249.

[26] Duchesne, P. and Lafaye De Micheaux, P. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. Computational Statistics & Data Analysis *54*, 858–862.

[27] Davies, R.B. (1980). Algorithm AS 155: The distribution of a linear combination of χ 2 random variables. Applied Statistics *29*, 323–333.

[28] Liu, H. and Tang, Y. and Zhang, H.H. (2009). A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. Computational Statistics & Data Analysis *53*, 853–856.

[29] Wessel, J. and Schork, N.J. (2006). Generalized genomic distance–based regression methodology for multilocus association analysis. The American Journal of Human Genetics *79*, 792–806.

[30] Davison, A.C. and Hinkley, D.V. and Canty, AJ. (1999). Bootstrap methods and their application. (Cambridge University Press).

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

[32] Artigas, A. and Bernard, G.R. and Carlet, J. and Dreyfuss, D. and Gattinoni, L. and Hudson, L. and Lamy, M. and Marini, J.J. and Matthay, M.A. and Pinsky, M.R. and others. (1998). The american-european consensus conference on ards, part 2. ventilatory, pharmacologic, supportive therapy, study design strategies, and issues related to recovery and remodeling. American journal of respiratory and critical care medicine *157*, 1332–1347.

[33] Schoenfeld, D.A. and Bernard, G.R. and others. (2002). Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. Critical care medicine *30*, 1772–1777.

[34] McKenna, A. and Hanna, M. and Banks, E. and Sivachenko, A. and Cibulskis, K. and Kernytsky, A. and Garimella, K. and Altshuler, D. and Gabriel, S. and Daly, M. and others. (2010). The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. Genome research *20*, 1297–1303.

[35] Patterson, N. and Price, A.L. and Reich, D. (2006). Population structure and eigenanalysis. PLoS genetics *2*, e190.

[36] Gao, L. and Grant, A. and Halder, I. and Brower, R. and Sevransky, J. and Maloney, J.P. and Moss, M. and Shanholtz, C. and Yates, C.R. and Meduri, G.U. and others. (2006). Novel polymorphisms in the myosin light chain kinase gene confer risk for acute lung injury. American journal of respiratory cell and molecular biology *34*, 487–495.

[37] Christie, J.D. and Ma S.F. and Aplenc, R. and Li, M. and Lanken, P.N. and Shaw, C.V. and Fuchs, B. and Albelda, S.M. and Flores, C. and Garcia, J.G. (2008). Variation in the myosin light chain kinase

gene is associated with development of acute lung injury after major trauma. Cricial Care Medicine *36*, 2794–2800.

FIGURES

Figure 1: **Power estimates for the six competing methods when all causal variants were deleterious.** Empirical power of the six methods for randomly selected 3kb regions and all causal variants were deleterious. From top to bottom, the plots consider the significance levels 0.01, 10^{-3}, and 2.5×10^{-6}, respectively. From left to right, the plots consider settings in which 10% of rare variants were causal, 20% of rare variants were causal, 50% of rare variants were causal. For causal variants, we assumed $|\beta_j| = c|\log_{10}(p_j)|/2$, where p_j was the MAF of the j^{th} variant. Different c was used for the three panels from left to right: $c = \log(7), \log(5), \log(2.5)$ for the percentage of causal variants being 10%, 20%, and 50% respectively. Hence the powers between the three panels from left to right are not comparable. Total sample sizes considered were 200, 500, and 1000, with half being cases in case-control studies.

Figure 2: **Power estimates for the six competing methods when $20\%/80\%$ of causal variants were protective/deterious.** Empirical power of the six methods for randomly selected 3kb regions and $20\%/80\%$ of causal variants were protective/deterious. From top to bottom, the plots consider the significance levels 0.01, 10^{-3}, and 2.5×10^{-6}, respectively. From left to right, the plots consider settings in which 10% of rare variants were causal, 20% of rare variants were causal, 50% of rare variants were causal. For causal variants, we assumed $|\beta_j| = c|\log_{10}(p_j)|/2$, where p_j was the MAF of the j^{th} variant. Different c was used for the three panels from the left to the right: $c = \log(7), \log(5), \log(2.5)$ for the percentage of causal variants being 10%, 20%, and 50% respectively. Hence the powers between the three panels from left to right are not comparable. Total sample sizes considered were 200, 500, and 1000, with half being cases in case-control studies.

Figure 3: **Power estimates for the six competing methods when $50\%/50\%$ of causal variants were protective/deterious.** Empirical power of the six methods for randomly selected 3kb regions and $50\%/50\%$ of causal variants were protective/deterious. From top to bottom, the plots consider the significance levels 0.01, 10^{-3}, and 2.5×10^{-6}, respectively. From left to right, the plots consider settings in which 10% of rare variants were causal, 20% of rare variants were causal, 50% of rare variants were causal. For causal variants, we assumed $|\beta_j| = c|\log_{10}(p_j)|/2$, where p_j was the MAF of the j^{th} variant. Different c was used for the three panels from left to right: $c = \log(7), \log(5), \log(2.5)$ for the percentage of causal variants being 10%, 20%, and 50%. Hence the powers between the three panels from left to right are not comparable. Total sample sizes considered were 200, 500, and 1000, with half being cases in case-control studies.

Figure 4: **Analysis of the ALI exome sequence data.** $-\log_{10}$ QQ plots of observed vs. expected p-values for the ALI exome sequence data for the six methods: burden tests (N,W), SKAT, SKAT-O, adjusted SKAT, adjusted SKAT-O. X-axis represents $-\log_{10}$ expected p-values, and Y-axis represents $-\log_{10}$ observed p-values. A total of 6,488 genes with at least four rare variants were tested for associations with ALI severity.

TABLES

Table 1: Simulation studies of type I error estimates of four different methods to test an association between randomly selected 3kb regions with dichotomous traits at stringent α levels $\alpha = 10^{-3}, 10^{-4}$ and 2.5×10^{-6}. Each entry represents type I error rate estimates as the proportion of p-values smaller than α under the null hypothesis based on 10^7 simulated phenotypes.

α	SKAT	SKAT-O	adjusted SKAT	adjusted SKAT-O
SampleSize = 200				
10^{-3}	1.84×10^{-4}	5.03×10^{-4}	1.13×10^{-3}	1.24×10^{-3}
10^{-4}	5.30×10^{-6}	3.20×10^{-5}	1.01×10^{-4}	1.04×10^{-4}
2.5×10^{-6}	1.00×10^{-7}	3.00×10^{-7}	3.20×10^{-6}	2.50×10^{-6}
SampleSize = 500				
10^{-3}	5.17×10^{-4}	8.14×10^{-4}	1.12×10^{-3}	1.16×10^{-3}
10^{-4}	2.95×10^{-5}	7.13×10^{-5}	1.14×10^{-4}	1.12×10^{-4}
2.5×10^{-6}	1.00×10^{-7}	1.00×10^{-6}	2.50×10^{-6}	2.30×10^{-6}
SampleSize = 1000				
10^{-3}	7.22×10^{-4}	1.00×10^{-3}	1.09×10^{-3}	1.12×10^{-3}
10^{-4}	5.59×10^{-5}	1.02×10^{-4}	1.22×10^{-4}	1.19×10^{-4}
2.5×10^{-6}	1.00×10^{-6}	2.80×10^{-6}	3.20×10^{-6}	3.10×10^{-6}

Figure1

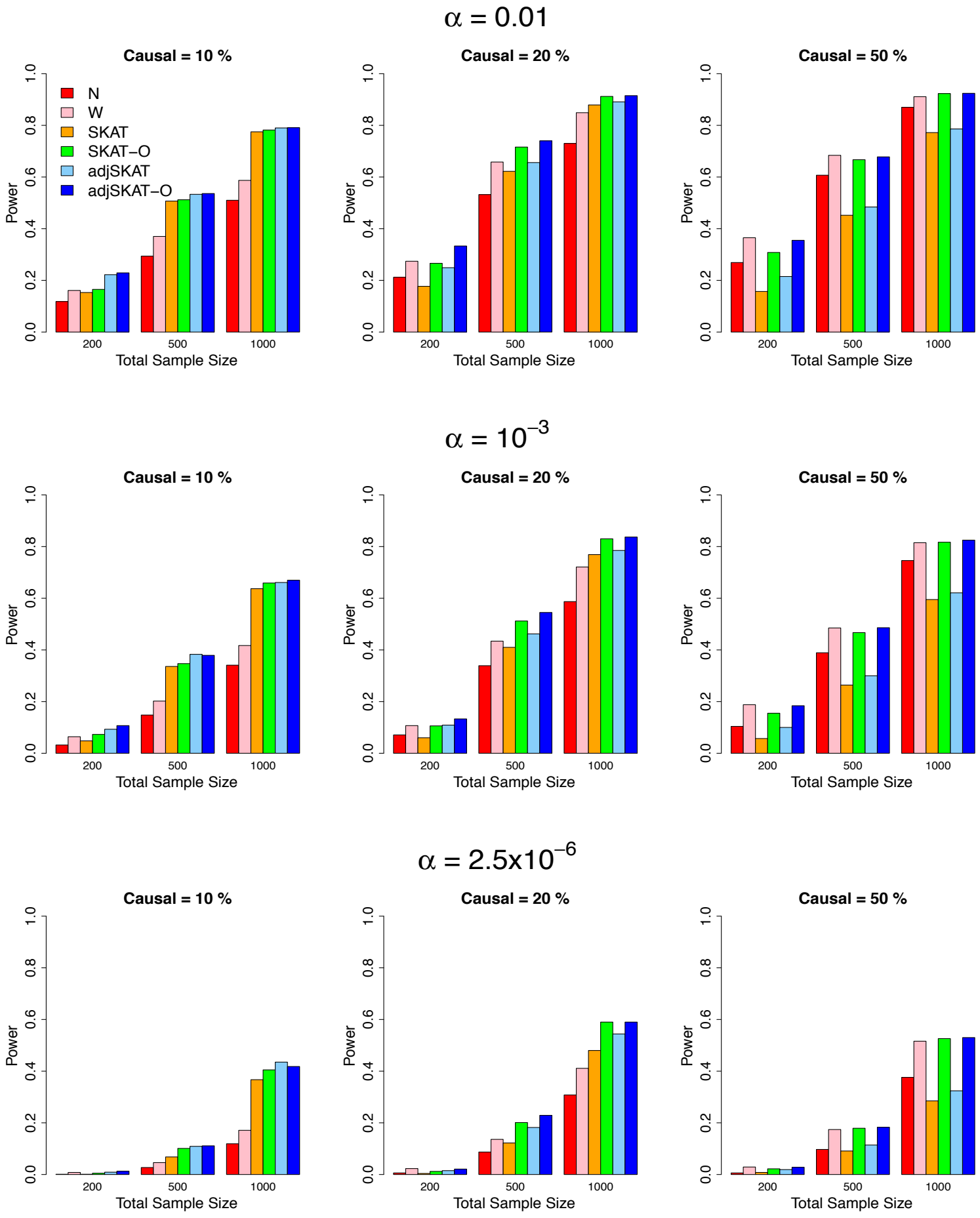

α = 0.01
Causal = 10 %
Causal = 20 %
Causal = 50 %
N
W
SKAT
SKAT−O
adjSKAT
adjSKAT−O
Power
Total Sample Size
200
500
1000
0.0
0.2
0.4
0.6
0.8
1.0
α = 10−3
α = 2.5x10−6

Figure2

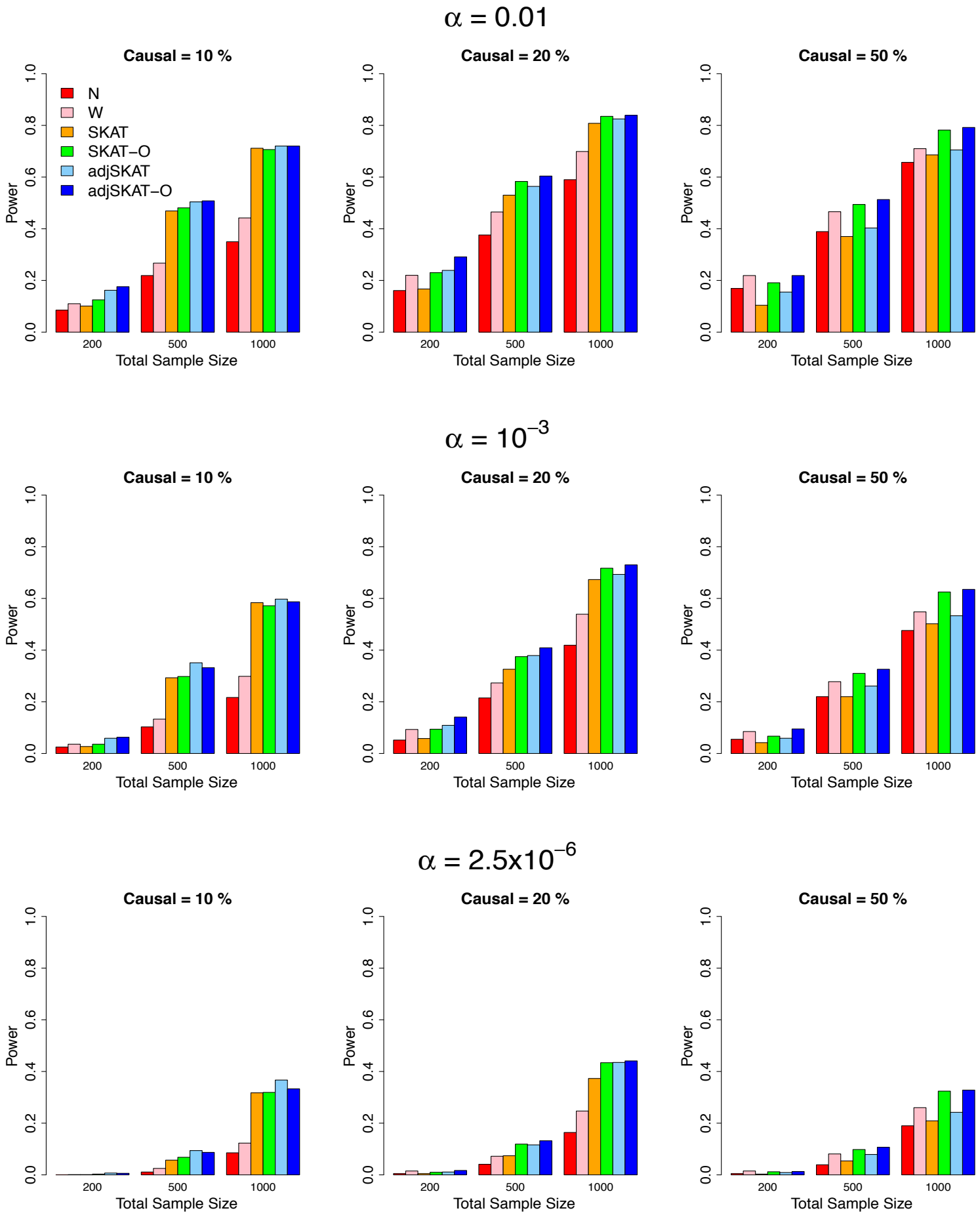

α = 0.01
Causal = 10 %
Causal = 20 %
Causal = 50 %
N
W
SKAT
SKAT−O
adjSKAT
adjSKAT−O
Power
Total Sample Size
200
500
1000
0.0
0.2
0.4
0.6
0.8
1.0
α = 10−3
α = 2.5x10−6

Figure3

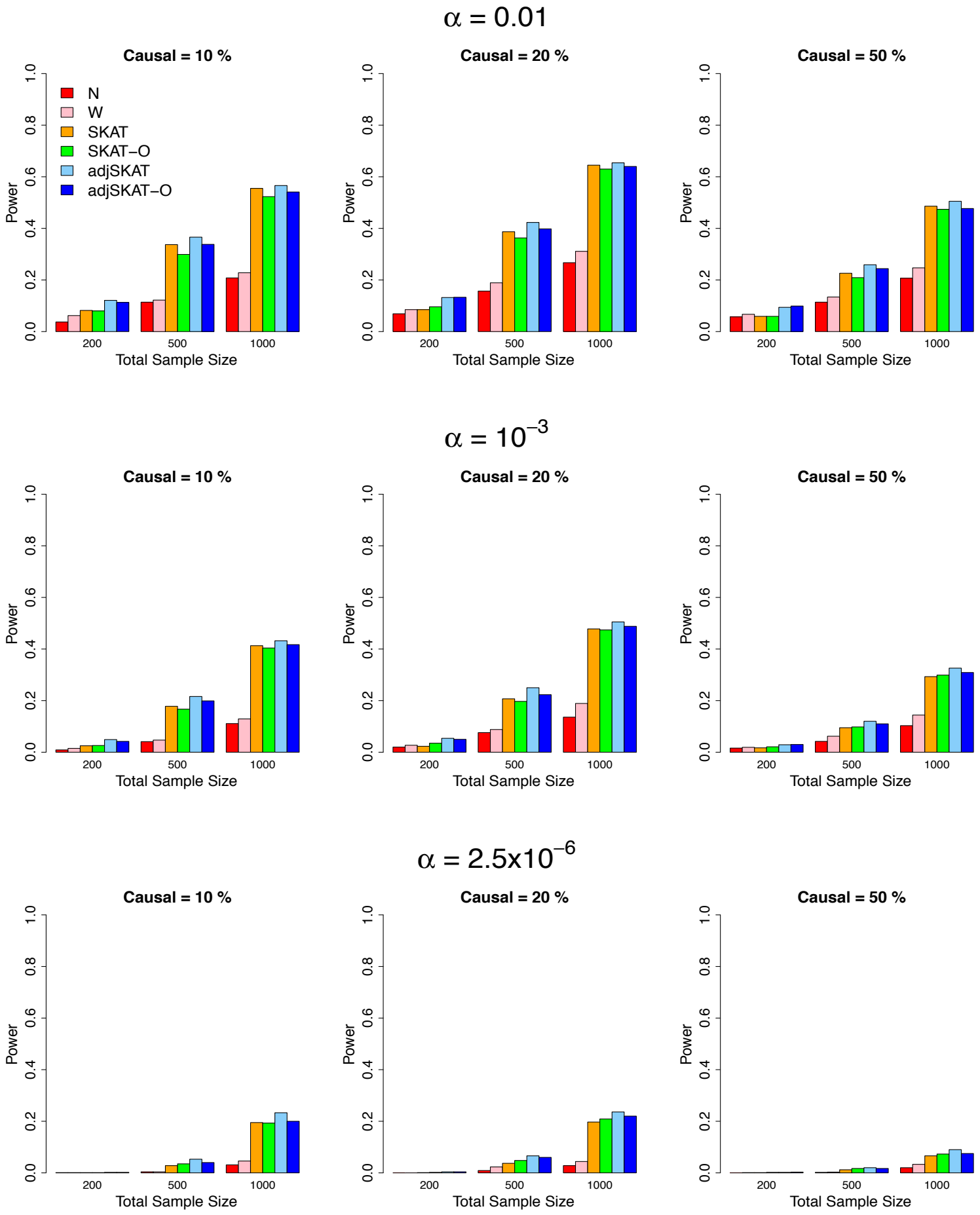
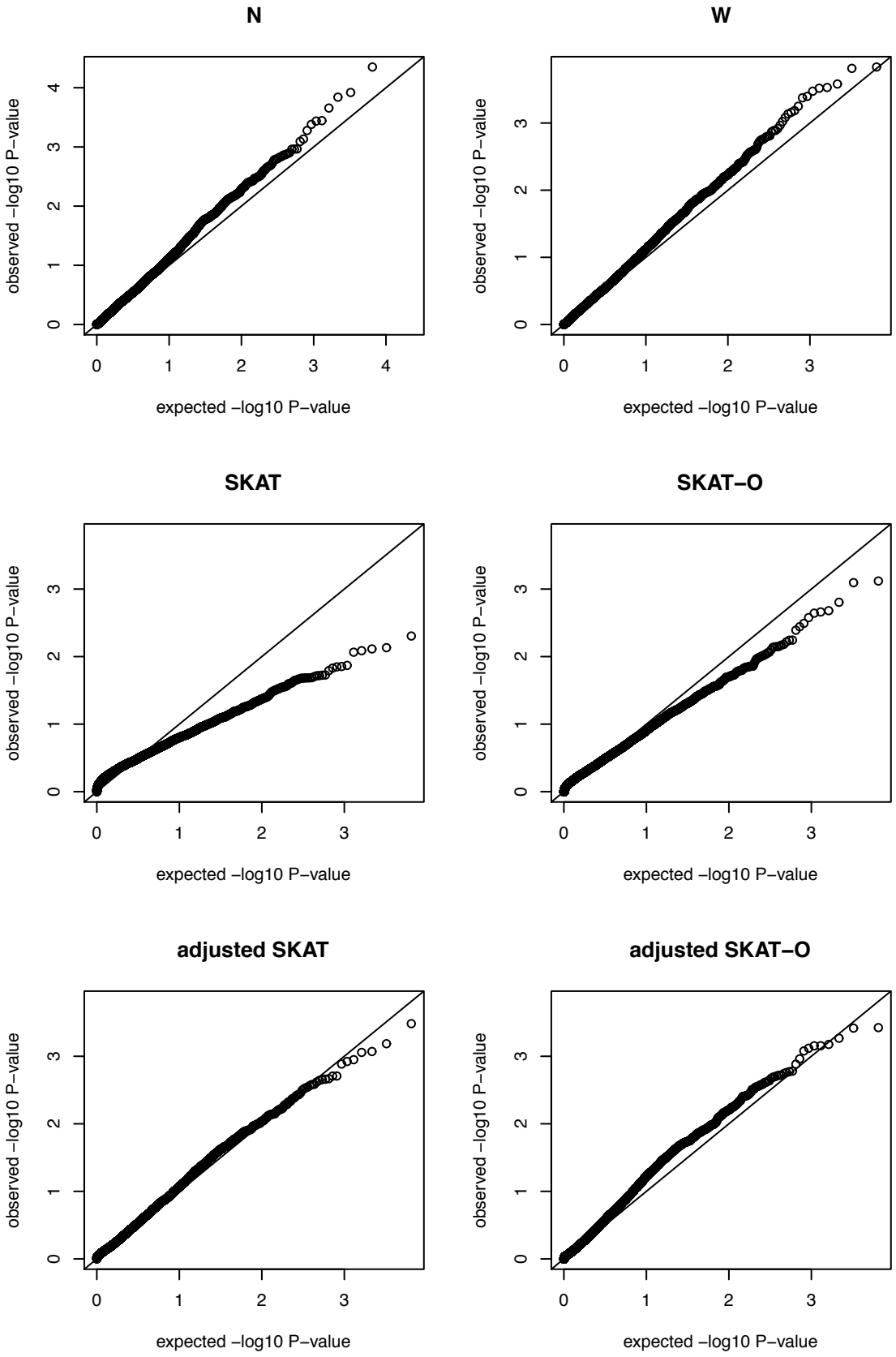

α = 0.01
Causal = 10 %
Causal = 20 %
Causal = 50 %
N
W
SKAT
SKAT−O
adjSKAT
adjSKAT−O
Power
0.0
0.2
0.4
0.6
0.8
1.0
200
500
1000
Total Sample Size
α = 10^−3
Causal = 10 %
Causal = 20 %
Causal = 50 %
Power
Total Sample Size
α = 2.5x10^−6
Causal = 10 %
Causal = 20 %
Causal = 50 %
Power
Total Sample Size

Figure4


N
W
SKAT
SKAT−O
adjusted SKAT
adjusted SKAT−O
observed −log10 P−value
expected −log10 P−value

Supplemental Data

Optimal unified approach for rare variant association testing with application to small sample case-control whole-exome sequencing studies

Seunggeun Lee, Mary Emond, Michael Bamshad, Kathleen Barnes, Mark Rieder, Deborah Nickerson, NHLBI GO Exome Sequencing Project -- ESP Lung Project Team, David C. Christiani, Mark M. Wurfel, and Xihong Lin

Figure S1: Analysis of functional variants of the ALI exome sequence data.
− $\log_{10}$ QQ plots of observed vs. expected p-values for the ALI whole exome sequence data using the six methods: burden tests (W,N), SKAT, SKAT-O, adjusted SKAT, adjusted SKAT-O. X-axis represents − $\log_{10}$ expected p-values, and Y-axis represents − $\log_{10}$ observed p-values. Total 2,939 genes with at least four rare functional variants were tested for associations with ALI severity.

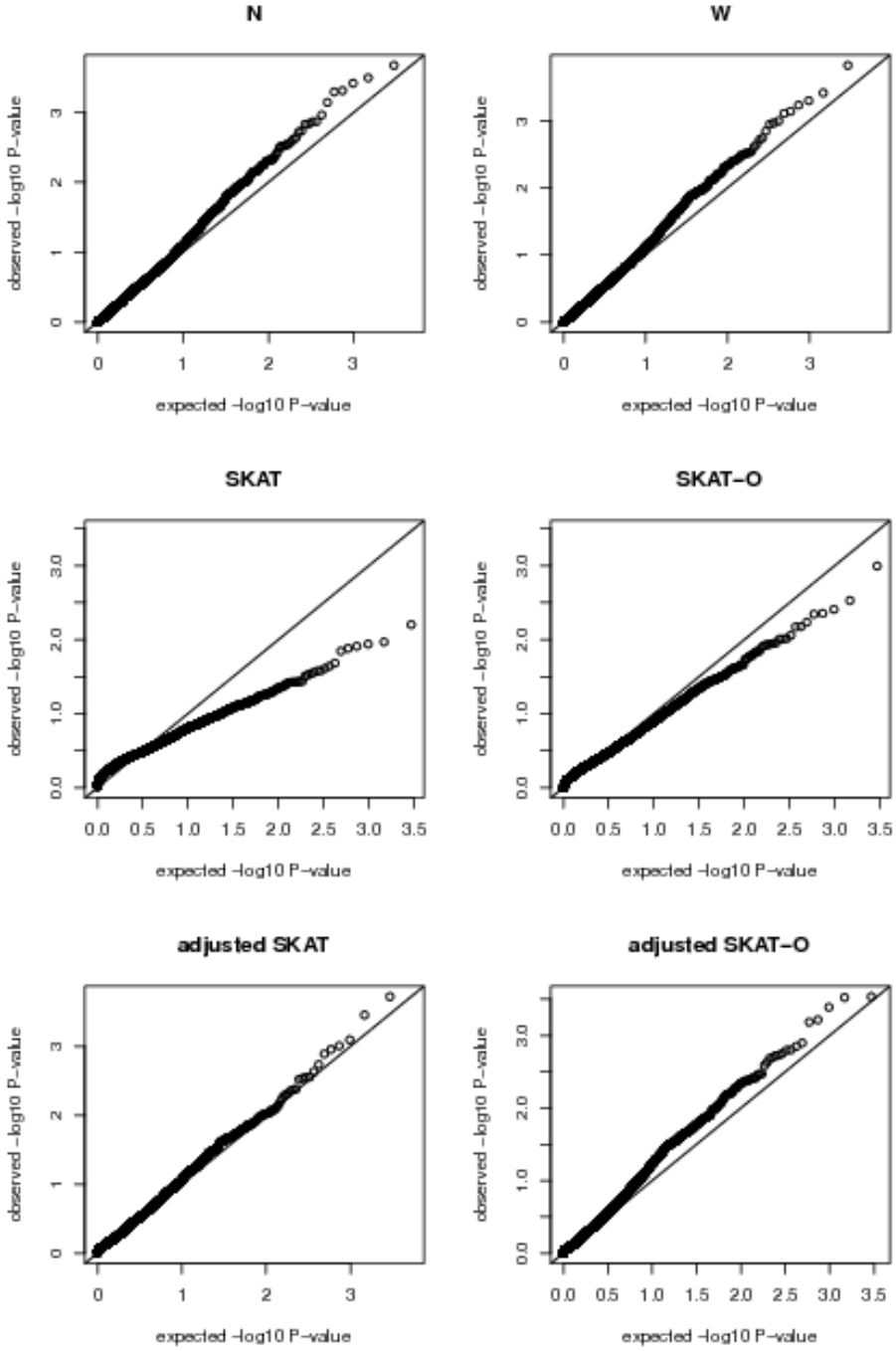


Figure S2: Comparison of burden test (W), adjusted SKAT, and adjusted SKAT-O.
Scatter plots of $-\log_{10}$ p-values to compare burden test (W), adjusted SKAT, adjusted SKAT-O. The top panel considers testing all variants, and bottom panel considers testing functional variants.

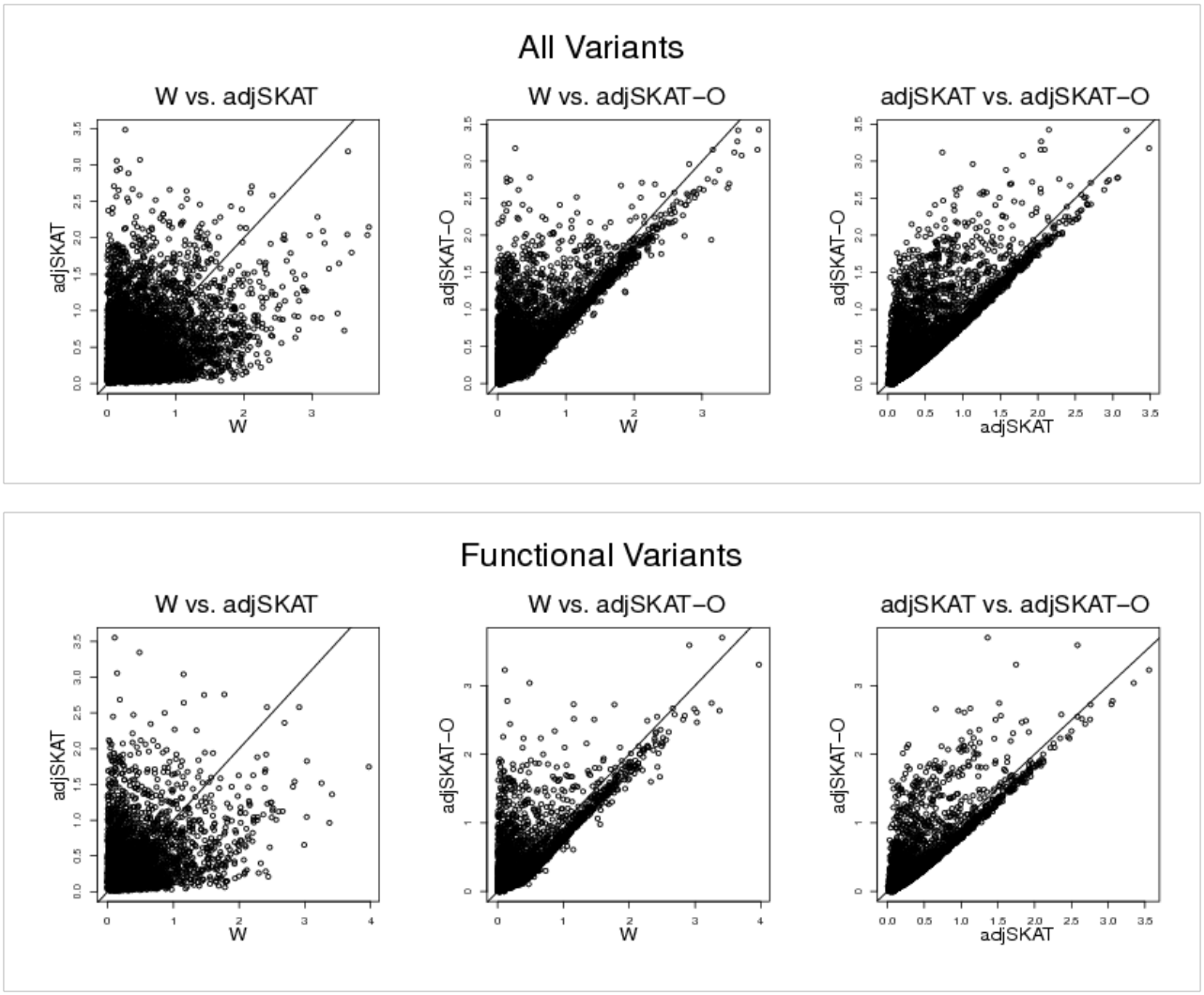


Figure S3: Analysis of the ALI exome sequence data with logistic weight.
$-\log_{10}$ QQ plots of observed vs. expected p-values for the ALI whole exome sequence data with logistic weight ($w_j = exp((a1 - pj)a2)/\{1+exp((a1 - pj)a2\}$) with $a_1 = 0.07$ and $a_2 = 150$. X-axis represents $-\log_{10}$ expected p-values, and Y-axis represents $-\log_{10}$ observed p-values. Total 6,488 genes with at least four rare variants were tested for associations with ALI severity.

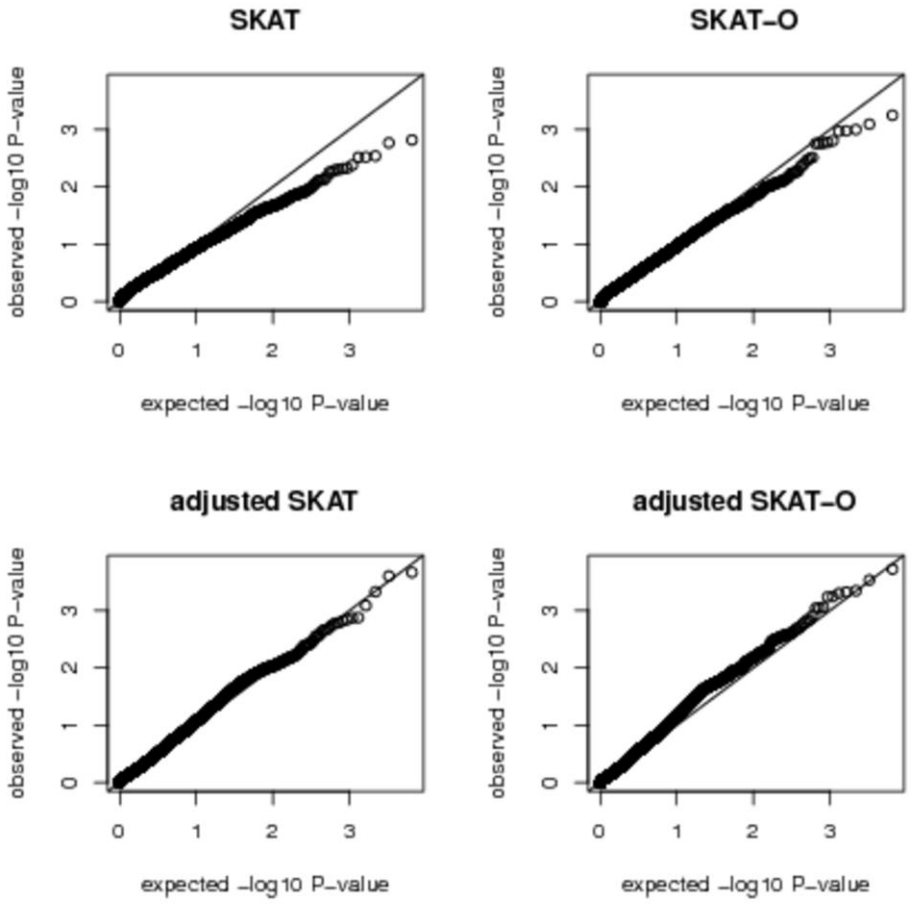


Figure S4: Analysis of the functional variants of the ALI exome sequence data with logistic weight.
$-\log_{10}$ QQ plots of observed vs. expected p-values for the ALI whole exome sequence data with logistic weight ($w_j = exp((a1 - pj)a2)/\{1+exp((a1 - pj)a2\}$) with $a_1 = 0.07$ and $a_2 = 150$. X-axis represents $-\log_{10}$ expected p-values, and Y-axis represents $-\log_{10}$ observed p-values. Total 2,939 genes with at least four rare variants were tested for associations with ALI severity.

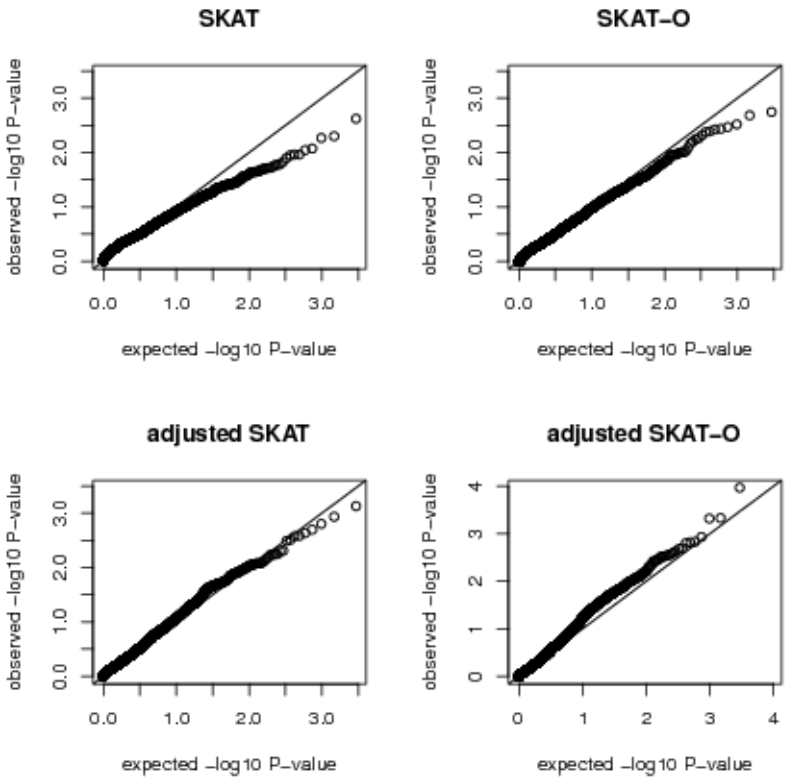


Figure S5: Estimated optimal *ρ* of the power simulation.
Box plots of the estimated optimal *ρ* in the power simulation studies. From top to bottom, the plots consider the setting in which percentage of protective/deleterious causal variants= 0/100, = 20/80 and = 50/50, respectively. From left to right, the plots consider the settings in which 10%, 20% and 50% of the rare variants were causal, respectively.

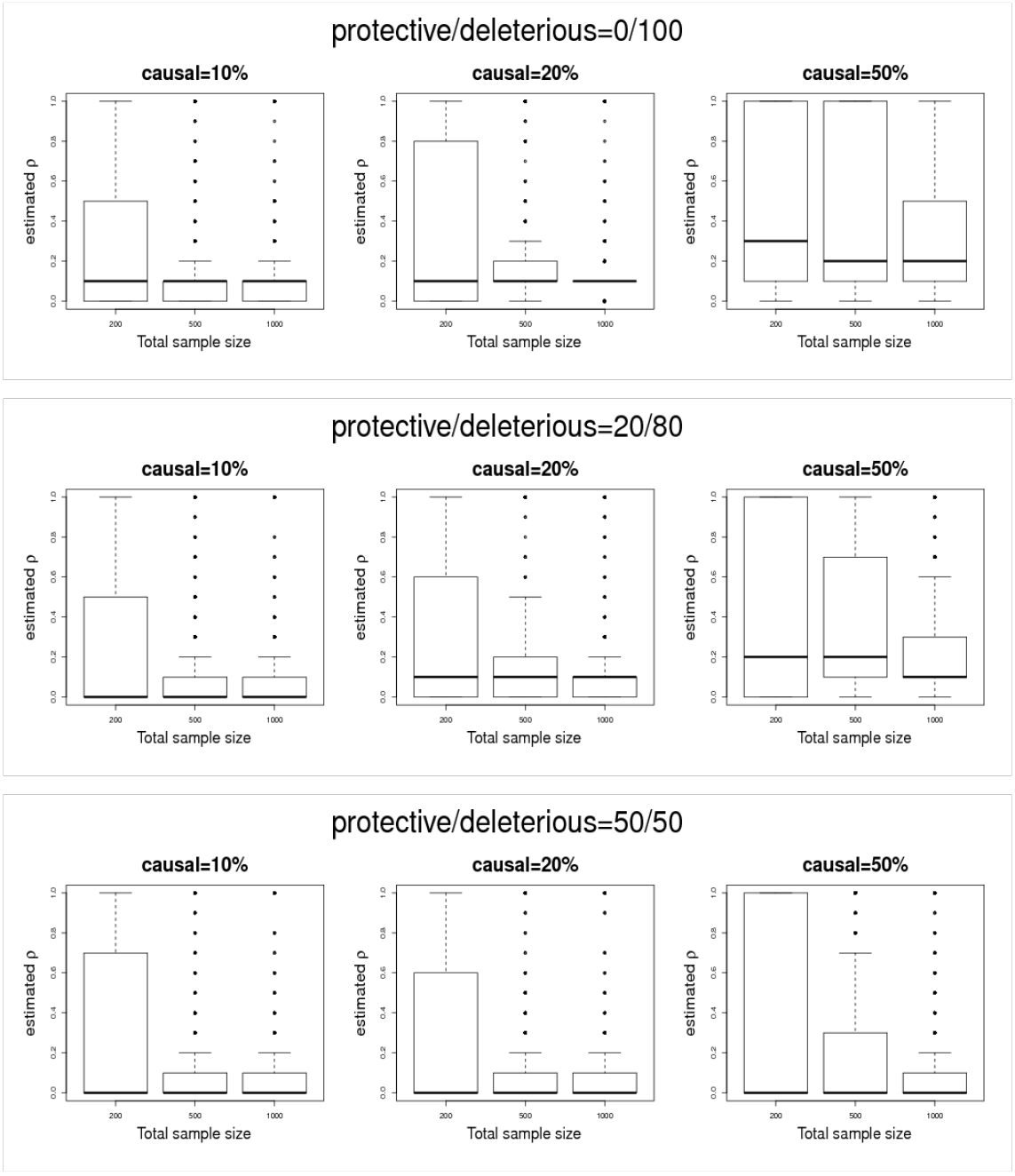


Figure S6: Power comparison for SKAT, omnibus and burden tests when all rare variants are deleterious causal variants.
Empirical power of the four methods for randomly selected 3kb regions with all the rare variants being deleterious causal variants, i.e., 100% causal. “Ominbus” represents the simple omnibus test that uses the smallest p-value of adjusted SKAT (adjSKAT) and W as the test statistics. Since it used the minimum p-value of two different tests, the multiple tests was corrected by the Bonferroni correction. From left to right, the plots consider the significance levels 0.01, 10^{-3}, and 2.5×10^{-6}, respectively. For causal variants, we assumed $|\beta_j| = c|\log10(p_j)|/2$, where p_j was the MAF of the j_{th} variant, and c = log(1.5). Total sample sizes considered were 200, 500, and 1000, with half being cases in case-control studies.

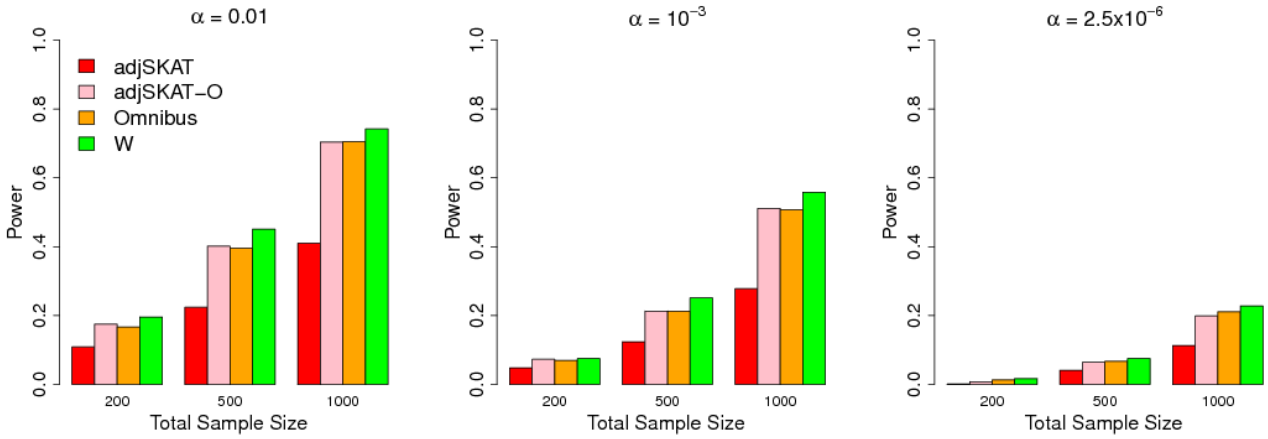


Table S1: Type I error rates of the burden tests and the SKAT family methods.

α	N	W	SKAT	SKAT-O	adjusted SKAT	adjusted SKAT-O
SampleSize = 200						
0.05	5.72×10^{-2}	5.55×10^{-2}	3.36×10^{-2}	4.26×10^{-2}	5.34×10^{-2}	5.96×10^{-2}
0.01	1.28×10^{-2}	1.32×10^{-2}	2.90×10^{-3}	5.80×10^{-3}	9.75×10^{-3}	1.14×10^{-2}
SampleSize = 500						
0.05	5.53×10^{-2}	5.29×10^{-2}	4.59×10^{-2}	4.76×10^{-2}	5.26×10^{-2}	5.39×10^{-2}
0.01	1.09×10^{-2}	1.06×10^{-2}	7.95×10^{-3}	9.00×10^{-3}	1.13×10^{-2}	1.12×10^{-2}
SampleSize = 1000						
0.05	5.05×10^{-2}	5.14×10^{-2}	4.45×10^{-2}	4.71×10^{-2}	4.73×10^{-2}	4.91×10^{-2}
0.01	1.05×10^{-2}	1.14×10^{-2}	8.30×10^{-3}	1.00×10^{-2}	9.70×10^{-3}	1.11×10^{-2}

Simulation Studies of type I error estimates of six different methods to test an association between randomly selected 3kb regions with dichotomous traits at $\alpha = 0.01$ and 0.05. Each entry represents type I error rate estimates as the proportion of p-values smaller than α under the null hypothesis based on 10,000 simulated datasets.

Table S2: Observed number of variants within randomly selected 3kb regions in the power simulation.

Total sample size	% of causal variants	% of protective variants	All	case	control
200	10%	0%	20.69	17.03	14.33
		20%	20.05	16.38	14.14
		50%	19.23	15.41	14.11
	20%	0%	21.71	18.37	14.11
		20%	21.03	17.44	14.11
		50%	19.91	15.90	14.29
	50%	0%	22.36	19.42	14.13
		20%	21.15	17.87	14.08
		50%	19.53	15.68	14.19
500	10%	0%	28.04	22.64	19.52
		20%	27.40	21.95	19.39
		50%	26.83	20.95	19.72
	20%	0%	29.58	24.71	19.35
		20%	28.98	23.80	19.31
		50%	27.17	21.40	19.50
	50%	0%	31.18	27.03	19.41
		20%	29.78	24.98	19.56
		50%	27.41	21.67	19.59
1000	10%	0%	34.97	28.16	25.24
		20%	34.73	27.62	25.43
		50%	33.63	26.17	25.23
	20%	0%	36.74	30.56	25.06
		20%	35.84	29.40	25.03
		50%	34.58	26.99	25.41
	50%	0%	39.16	34.19	25.14
		20%	37.88	32.01	25.50
		50%	34.46	27.10	25.21

Each entry represents the average number of observed variants in the simulated datasets. “all” represents the number of variants observed among all samples. “case” and “control” represent the number of observed variants among cases and controls, respectively.

Table S3: Observed number of causal variants within randomly selected 3kb regions in the power simulation.

Total sample size	% of protective variants	observed causal			observed harmful			observed protective		
		all	Case	control	all	case	control	all	case	control
		10% variants were causal (average number of causal variants = 4.9)								
200	0%	3.38	3.38	0.00	3.33	3.33	0.00	0.70	0.70	0.00
	20%	3.02	2.90	0.12	2.89	2.87	0.02	0.72	0.60	0.12
	50%	2.31	1.94	0.37	2.00	1.91	0.09	0.78	0.43	0.36
500	0%	4.39	4.39	0.00	4.37	4.37	0.00	1.16	1.16	0.00
	20%	3.95	3.78	0.17	3.81	3.77	0.05	1.12	0.95	0.17
	50%	3.02	2.43	0.59	2.58	2.42	0.16	1.20	0.63	0.57
1000	0%	4.74	4.74	0.00	4.74	4.74	0.00	1.56	1.56	0.00
	20%	4.33	4.07	0.27	4.12	4.05	0.06	1.62	1.36	0.26
	50%	3.43	2.62	0.81	2.82	2.62	0.20	1.66	0.87	0.80
		20% variants were causal (average number of causal variants = 10.2)								
200	0%	5.85	5.85	0.00	5.68	5.68	0.00	1.49	1.49	0.00
	20%	5.33	5.02	0.31	4.94	4.85	0.09	1.62	1.33	0.29
	50%	4.05	3.20	0.84	3.34	3.11	0.23	1.59	0.79	0.80
500	0%	8.15	8.15	0.00	8.01	8.01	0.00	2.49	2.49	0.00
	20%	7.42	6.93	0.48	6.96	6.82	0.14	2.51	2.05	0.46
	50%	5.67	4.34	1.33	4.70	4.27	0.42	2.60	1.32	1.28
1000	0%	9.41	9.41	0.00	9.33	9.33	0.00	3.47	3.47	0.00
	20%	8.55	7.88	0.67	8.03	7.81	0.22	3.58	2.94	0.64
	50%	6.96	5.03	1.94	5.56	4.98	0.58	3.70	1.84	1.86
		50% variants were causal (average number of causal variants = 26.3)								
200	0%	10.14	10.14	0.00	9.08	9.08	0.00	4.10	4.10	0.00
	20%	9.23	8.31	0.92	7.84	7.39	0.45	4.23	3.44	0.79
	50%	7.47	5.17	2.30	5.64	4.61	1.04	4.08	2.11	1.97
500	0%	15.45	15.45	0.00	14.27	14.27	0.00	6.75	6.75	0.00
	20%	13.97	12.56	1.41	12.20	11.56	0.64	6.75	5.48	1.26
	50%	11.67	7.82	3.86	8.94	7.20	1.74	6.86	3.39	3.47
1000	0%	19.80	19.80	0.00	18.63	18.63	0.00	9.53	9.53	0.00
	20%	18.23	16.20	2.03	16.13	15.20	0.93	9.60	7.78	1.82
	50%	15.14	10.00	5.14	11.64	9.43	2.21	9.54	4.80	4.73

Each entry represents the average number of observed causal variants (harmful + protective), observed harmful variants, and observed protective variants in the simulated datasets. “all” represents the number of causal variants observed among all samples. “case” and “control” represent the number of observed causal variants among cases and controls, respectively. Harmful variants increase the chance to be a case ($\beta > 0$) and protective variants reduce the chance to be a case ($\beta < 0$).

Further acknowledgements

HeartGO:

Atherosclerosis Risk in Communities (ARIC): NHLBI (N01 HC-55015, N01 HC-55016, N01HC-55017, N01 HC-55018, N01 HC-55019, N01 HC-55020, N01 HC-55021); **Cardiovascular Health Study (CHS)**: NHLBI (N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, and grant HL080295), with additional support from NINDS and from NIA (AG-023629, AG-15928, AG-20098, and AG-027058); **Coronary Artery Risk Development in Young Adults (CARDIA)**: NHLBI (N01-HC95095 & N01-HC48047, N01-HC48048, N01-HC48049, and N01-HC48050); **Framingham Heart Study (FHS)**: NHLBI (N01-HC-25195 and grant R01 NS17950) with additional support from NIA (AG08122 and AG033193); **Jackson Heart Study (JHS)**: NHLBI and the National Institute on Minority Health and Health Disparities (N01 HC-95170, N01 HC-95171 and N01 HC-95172); **Multi-Ethnic Study of Atherosclerosis (MESA)**: NHLBI (N01-HC-95159 through N01-HC-95169 and RR-024156).

Lung GO:

Cystic Fibrosis (CF): Cystic Fibrosis Foundation (GIBSON07K0, KNOWLE00A0, OBSERV04K0, RDP R026), the NHLBI (R01 HL-068890, R02 HL-095396), NIH National Center for Research Resources (UL1 RR-025014), and the National Human Genome Research Institute (NHGRI) (5R00 HG-004316). **Chronic Obstructive Pulmonary Disease (COPDGene):** NHLBI (U01 HL-089897, U01 HL-089856), and the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, and Sunovian. The COPDGene clinical centers and investigators are available at www.copdgene.org. **Acute Lung Injury (ALI)**: NHLBI (RC2 HL-101779). **Lung Health Study (LHS)**: NHLBI (RC2 HL-066583), the NHGRI (HG-004738), and the NHLBI Division of Lung Diseases (HR-46002). **Pulmonary Arterial Hypertension (PAH):** NIH (P50 HL-084946, K23 AR-52742), and the NHLBI (F32 HL-083714). **Asthma:** NHLBI (RC2 HL-101651), and the NIH (HL-077916, HL-69197, HL-76285, M01 RR-07122).

SWISS and ISGS:

Siblings with Ischemic Stroke Study (SWISS): National Institute of Neurological Disorders and Stroke (NINDS) (R01 NS039987); Ischemic Stroke Genetics Study (ISGS): NINDS (R01 NS042733)

WHISP:

Women's Health Initiative (WHI): The WHI Sequencing Project is funded by the NHLBI (HL-102924) as well as the National Institutes of Health (NIH), U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 44221. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf

NHLBI GO Exome Sequencing Project

BroadGO

Stacey B. Gabriel (Broad Institute)[4, 5, 11, 16, 17], David M. Altshuler (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[1, 5, 7, 17], Gonçalo R. Abecasis (University of Michigan)[3, 5, 9, 13, 15, 17], Hooman Allayee (University of Southern California)[5], Sharon Cresci (Washington University School of Medicine)[5], Mark J. Daly (Broad Institute, Massachusetts General Hospital), Paul I. W. de Bakker (Broad Institute, Harvard Medical School, University Medical Center Utrecht)[3, 15], Mark A. DePristo (Broad Institute)[4, 13, 15, 16], Ron Do (Broad Institute)[5, 9, 13, 15], Peter Donnelly (University of Oxford)[5], Deborah N. Farlow (Broad Institute)[3, 4, 5, 12, 14, 16, 17], Tim Fennell (Broad Institute), Kiran Garimella (University of Oxford)[4, 16], Stanley L. Hazen (Cleveland Clinic)[5], Youna Hu (University of Michigan)[3, 9, 15], Daniel M. Jordan (Harvard Medical School, Harvard University)[13], Goo Jun (University of Michigan)[13], Sekar Kathiresan (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[5, 8, 9, 12, 14, 15, 17, 20], Hyun Min Kang (University of Michigan)[9, 13, 16], Adam Kiezun (Broad Institute)[5, 13, 15], Guillaume Lettre (Broad Institute, Montreal Heart Institute, Université de Montréal)[1, 2, 13, 15], Bingshan Li (University of Michigan)[3], Mingyao Li (University of Pennsylvania)[5], Christopher H. Newton-Cheh (Broad Institute, Massachusetts General Hospital, Harvard Medical School)[3, 8, 15], Sandosh Padmanabhan (University of Glasgow School of Medicine)[3, 12, 15], Gina Peloso (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[5], Sara Pulit (Broad Institute)[3, 15], Daniel J. Rader (University of Pennsylvania)[5], David Reich (Broad Institute, Harvard Medical School)[15], Muredach P. Reilly (University of Pennsylvania)[5], Manuel A. Rivas (Broad Institute, Massachusetts General Hospital)[5], Steve Schwartz (Fred Hutchinson Cancer Research Center)[5, 12], Laura Scott

(University of Michigan)[1], David S. Siscovick (University of Washington)[5, 1, 25], John A. Spertus (University of Missouri Kansas City)[5], Nathaniel O. Stitziel (Brigham and Women's Hospital)[5, 15], Nina Stoletzki (Brigham and Women's Hospital, Broad Institute, Harvard Medical School)[13], Shamil R. Sunyaev (Brigham and Women's Hospital, Broad Institute, Harvard Medical School)[1, 3, 5, 13, 15], Benjamin F. Voight (Broad Institute, Massachusetts General Hospital), Cristen J. Willer (University of Michigan)[1, 9, 13, 15]

HeartGO

Stephen S. Rich (University of Virginia)[2, 4, 7, 8, 9, 11,14, 15, 17, 18, 31], Ermeg Akylbekova (Jackson State University, University of Mississippi Medical Center)[29], Larry D. Atwood (Boston University)[1, 11, 28], Christie M. Ballantyne (Baylor College of Medicine, Methodist DeBakey Heart Center)[9, 22], Maja Barbalic (University of Texas Health Science Center Houston)[9, 14, 15, 17, 22], R. Graham Barr (Columbia University Medical Center)[10, 31], Emelia J. Benjamin (Boston University)[14, 20, 28], Joshua Bis (University of Washington)[15, 23], Eric Boerwinkle (University of Texas Health Science Center Houston)[3, 5, 9, 13, 15, 17, 22], Donald W. Bowden (Wake Forest University)[1, 31], Jennifer Brody (University of Washington)[3, 5, 15, 23], Matthew Budoff (Harbor-UCLA Medical Center)[31], Greg Burke (Wake Forest University)[5, 31], Sarah Buxbaum (Jackson State University)[3, 13, 15, 29], Jeff Carr (Wake Forest University)[25, 29, 31], Donna T. Chen (University of Virginia)[6, 11], Ida Y. Chen (Cedars-Sinai Medical Center)[1, 31], Wei-Min Chen (University of Virginia)[13, 15, 18], Pat Concannon (University of Virginia)[11], Jacy Crosby (University of Texas Health Science Center Houston)[22], L. Adrienne Cupples (Boston University)[1, 3, 5, 9, 13, 15, 18, 28], Ralph D'Agostino (Boston University)[28], Anita L. DeStefano (Boston University)[13, 18, 28], Albert Dreisbach (University of Mississippi Medical Center)[3, 29], Josée Dupuis (Boston University)[1, 28], J. Peter Durda (University of Vermont)[15, 23], Jaclyn Ellis (University of North Carolina Chapel Hill)[1], Aaron R. Folsom (University of Minnesota)[5, 22], Myriam Fornage (University of Texas Health Science Center Houston)[3, 18, 25], Caroline S. Fox (National Heart, Lung, and Blood Institute)[1, 28], Ervin Fox (University of Mississippi Medical Center)[3, 9, 29], Vincent Funari (Cedars-Sinai Medical Center)[1, 11, 31], Santhi K. Ganesh (University of Michigan)[2, 22], Julius Gardin (Hackensack University Medical Center)[25], David Goff (Wake Forest University)[25], Ora Gordon (Cedars-Sinai Medical Center)[11, 31], Wayne Grody (University of California Los Angeles)[11, 31], Myron Gross (University of Minnesota)[1, 5, 14, 25], Xiuqing Guo (Cedars-Sinai Medical Center)[3, 15, 31], Ira M. Hall (University of Virginia), Nancy L. Heard-Costa (Boston University)[1, 11, 28], Susan R. Heckbert (University of Washington)[10, 14, 20, 23], Nicholas Heintz (University of Vermont), David M. Herrington (Wake Forest University)[5, 31], DeMarc Hickson (Jackson State University, University of Mississippi Medical Center)[29], Jie Huang (National Heart, Lung, and Blood Institute)[5, 28], Shih-Jen Hwang (Boston University, National Heart, Lung, and Blood Institute)[3, 28], David R. Jacobs (University of Minnesota)[25], Nancy S. Jenny (University of Vermont)[1, 2, 23], Andrew D. Johnson (National Heart, Lung, and Blood Institute)[2, 5, 11, 28], Craig W. Johnson (University of Washington)[15, 31], Steven Kawut (University of Pennsylvania)[10,31], Richard Kronmal (University of Washington)[31], Raluca Kurz (Cedars-Sinai Medical Center)[11, 31], Ethan M. Lange (University of North Carolina Chapel Hill)[3, 5, 9, 13, 34], Leslie A. Lange (University of North Carolina Chapel Hill)[1, 2, 3, 5, 9, 12, 13, 15, 17, 18, 20, 25, 34], Martin G. Larson (Boston University)[3, 15, 28], Mark Lawson (University of Virginia), Cora E. Lewis (University of Alabama at Birmingham)[25,34], Daniel Levy (National Heart, Lung, and Blood Institute)[3, 15, 17, 28], Dalin Li (Cedars-Sinai Medical Center)[11, 15, 31], Honghuang Lin (Boston University)[20, 28], Chunyu Liu (National Heart, Lung, and

Blood Institute)[3, 28], Jiankang Liu (University of Mississippi Medical Center)[1, 29], Kiang Liu (Northwestern University)[25], Xiaoming Liu (University of Texas Health Science Center Houston)[15, 22], Yongmei Liu (Wake Forest University)[2, 5, 31], William T. Longstreth (University of Washington)[18, 23], Cay Loria (National Heart, Lung, and Blood Institute)[25], Thomas Lumley (University of Auckland)[9, 23], Kathryn Lunetta (Boston University)[28], Aaron J. Mackey (University of Virginia)[16, 18], Rachel Mackey (University of Pittsburgh)[1, 23, 31], Ani Manichaikul (University of Virginia)[8, 15, 18, 31], Taylor Maxwell (University of Texas Health Science Center Houston)[22], Barbara McKnight (University of Washington)[15, 23], James B. Meigs (Brigham and Women's Hospital, Harvard Medical School, Massachusetts General Hospital)[1, 28], Alanna C. Morrison (University of Texas Health Science Center Houston)[3, 15, 17], Solomon K. Musani (University of Mississippi Medical Center)[3, 29], Josyf C. Mychaleckyj (University of Virginia)[13, 15, 31], Jennifer A. Nettleton (University of Texas Health Science Center Houston)[9, 22], Kari North (University of North Carolina Chapel Hill)[1, 3, 9, 10, 13, 15, 17, 34], Christopher J. O'Donnell (Massachusetts General Hospital, National Heart, Lung, and Blood Institute)[2, 5, 9, 11, 12, 14, 15, 17, 20, 28], Daniel O'Leary (Tufts University School of Medicine)[25, 31], Frank Ong (Cedars-Sinai Medical Center)[3, 11, 31], Walter Palmas (Columbia University)[3, 15, 31], James S. Pankow (University of Minnesota)[1, 22], Nathan D. Pankratz (Indiana University School of Medicine)[15, 25], Shom Paul (University of Virginia), Marco Perez (Stanford University School of Medicine), Sharina D. Person (University of Alabama at Birmingham, University of Alabama at Tuscaloosa)[25], Joseph Polak (Tufts University School of Medicine)[31], Wendy S. Post (Johns Hopkins University)[3, 9, 11, 14, 20, 31], Bruce M. Psaty (Group Health Research Institute, University of Washington)[3, 5, 9, 11, 14, 15, 23], Aaron R. Quinlan (University of Virginia)[18, 19], Leslie J. Raffel (Cedars-Sinai Medical Center)[6, 11, 31], Vasan S. Ramachandran (Boston University)[3, 28], Alexander P. Reiner (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 9, 11, 12, 13, 14, 15, 20, 25, 34], Kenneth Rice (University of Washington)[15, 23], Jerome I. Rotter (Cedars-Sinai Medical Center)[1, 3, 6, 8, 11, 15, 31], Jill P. Sanders (University of Vermont)[23], Pamela Schreiner (University of Minnesota)[25], Sudha Seshadri (Boston University)[18, 28], Steve Shea (Brigham and Women's Hospital, Harvard University)[28], Stephen Sidney (Kaiser Permanente Division of Research, Oakland, CA)[25], Kevin Silverstein (University of Minnesota)[25], David S. Siscovick (University of Washington)[5, 1, 25], Nicholas L. Smith (University of Washington)[2, 15, 20, 23], Nona Sotoodehnia (University of Washington)[3, 15, 23], Asoke Srinivasan (Tougaloo College)[29], Herman A. Taylor (Jackson State University, Tougaloo College, University of Mississippi Medical Center)[5, 29], Kent Taylor (Cedars-Sinai Medical Center)[31], Fridtjof Thomas (University of Texas Health Science Center Houston)[3, 22], Russell P. Tracy (University of Vermont)[5, 9, 11, 12, 14, 15, 17, 20, 23], Michael Y. Tsai (University of Minnesota)[9, 31], Kelly A. Volcik (University of Texas Health Science Center Houston)[22], Chrstina L Wassel (University of California San Diego)[9, 15, 31], Karol Watson (University of California Los Angeles)[31], Gina Wei (National Heart, Lung, and Blood Institute)[25], Wendy White (Tougaloo College)[29], Kerri L. Wiggins (University of Vermont)[23], Jemma B. Wilk (Boston University)[10, 28], O. Dale Williams (Florida International University)[25], Gregory Wilson (Jackson State University)[29], James G. Wilson (University of Mississippi Medical Center)[1, 2, 5, 8, 9, 11, 12, 14, 17, 20, 29], Phillip Wolf (Boston University)[28], Neil A. Zakai (University of Vermont)[2, 23]

ISGS and SWISS

John Hardy (Reta Lila Weston Research Laboratories, Institute of Neurology, University College London)[18], James F. Meschia (Mayo Clinic)[18], Michael Nalls (National Institute on Aging)[2, 18], Stephen

S. Rich (University of Virginia)[2, 4, 7, 8, 9, 11, 14, 15, 17, 18, 31], Andrew Singleton (National Institute on Aging)[18], Brad Worrall (University of Virginia)[18]

LungGO

Michael J. Bamshad (Seattle Children's Hospital, University of Washington)[4, 6, 7, 8, 10, 11, 13, 15, 17, 27], Kathleen C. Barnes (Johns Hopkins University)[2, 10, 12, 14, 15, 17, 20, 24, 30, 32], Ibrahim Abdulhamid (Children's Hospital of Michigan)[27], Frank Accurso (University of Colorado)[27], Ran Anbar (Upstate Medical University)[27], Terri Beaty (Johns Hopkins University)[24, 30], Abigail Bigham (University of Washington)[13, 15, 27], Phillip Black (Children's Mercy Hospital)[27], Eugene Bleecker (Wake Forest University)[33], Kati Buckingham (University of Washington)[27], Anne Marie Cairns (Maine Medical Center)[27], Wei-Min Chen (University of Virginia)[13, 15, 18], Daniel Caplan (Emory University)[27], Barbara Chatfield (University of Utah)[27], Aaron Chidekel (A.I. Dupont Institute Medical Center)[27], Michael Cho (Brigham and Women's Hospital, Harvard Medical School)[13, 15, 24], David C. Christiani (Massachusetts General Hospital)[21], James D. Crapo (National Jewish Health)[24, 30], Julia Crouch (Seattle Children's Hospital)6, Denise Daley (University of British Columbia)[30], Anthony Dang (University of North Carolina Chapel Hill)[26], Hong Dang (University of North Carolina Chapel Hill)[26], Alicia De Paula (Ochsner Health System)[27], Joan DeCelie-Germana (Schneider Children's Hospital)[27], Allen Dozor (New York Medical College, Westchester Medical Center)[27,] Mitch Drumm (University of North Carolina Chapel Hill)[26], Maynard Dyson (Cook Children's Med. Center)[27], Julia Emerson (Seattle Children's Hospital, University of Washington)[27], Mary J. Emond (University of Washington)[10, 13, 15, 17, 27], Thomas Ferkol (St. Louis Children's Hospital, Washington University School of Medicine)[27], Robert Fink (Children's Medical Center of Dayton)[27], Cassandra Foster (Johns Hopkins University)[30], Deborah Froh (University of Virginia)[27], Li Gao (Johns Hopkins University)[24, 30, 32], William Gershan (Children's Hospital of Wisconsin)[27], Ronald L. Gibson (Seattle Children's Hospital, University of Washington)[10, 27], Elizabeth Godwin (University of North Carolina Chapel Hill)[26], Magdalen Gondor (All Children's Hospital Cystic Fibrosis Center)[27], Hector Gutierrez (University of Alabama at Birmingham)[27], Nadia N. Hansel (Johns Hopkins University, Johns Hopkins University School of Public Health)[10, 15, 30], Paul M. Hassoun (Johns Hopkins University)[10, 14, 32], Peter Hiatt (Texas Children's Hospital)[27], John E. Hokanson (University of Colorado)[24], Michelle Howenstine (Indiana University, Riley Hospital for Children)[27], Laura K. Hummer (Johns Hopkins University)[32], Jamshed Kanga (University of Kentucky)[27], Yoonhee Kim (National Human Genome Research Institute)[24, 32], Michael R. Knowles (University of North Carolina Chapel Hill)[10, 26], Michael Konstan (Rainbow Babies & Children's Hospital)[27], Thomas Lahiri (Vermont Children's Hospital at Fletcher Allen Health Care)[27], Nan Laird (Harvard School of Public Health)[24], Christoph Lange (Harvard School of Public Health)[24], Lin Lin (Harvard Medical School)[21], Xihong Lin (Harvard School of Public Health)[21], Tin L. Louie (University of Washington)[13, 15, 27], David Lynch (National Jewish Health)[24], Barry Make (National Jewish Health)[24], Thomas R. Martin (University of Washington, VA Puget Sound Medical Center)[10, 21], Steve C. Mathai (Johns Hopkins University)[32], Rasika A. Mathias (Johns Hopkins University)[10, 13, 15, 30, 32], John McNamara (Children's Hospitals and Clinics of Minnesota)[27], Sharon McNamara (Seattle Children's Hospital)[27], Deborah Meyers (Wake Forest University)[33], Susan Millard (DeVos Children's Butterworth Hospital, Spectrum Health Systems)[27], Peter Mogayzel (Johns Hopkins University)[27], Richard Moss (Stanford University)[27], Tanda Murray (Johns Hopkins University)[30], Dennis Nielson (University of California at San Francisco)[27],

Blakeslee Noyes (Cardinal Glennon Children's Hospital)[27], Wanda O'Neal (University of North Carolina Chapel Hill)[26], David Orenstein (Children's Hospital of Pittsburgh)[27], Brian O'Sullivan (University of Massachusetts Memorial Health Care)[27], Rhonda Pace (University of North Carolina Chapel Hill)[26], Peter Pare (St. Paul's Hospital)[30], H. Worth Parker (Dartmouth-Hitchcock Medical Center, New Hampshire Cystic Fibrosis Center)[27], Mary Ann Passero (Rhode Island Hospital)[27], Elizabeth Perkett (Vanderbilt University)[27,] Adrienne Prestridge (Children's Memorial Hospital)[27], Nicholas M. Rafaels (Johns Hopkins University)[30], Bonnie Ramsey (Seattle Children's Hospital, University of Washington)[27], Elizabeth Regan (National Jewish Health)[24], Clement Ren (University of Rochester)[27], George Retsch-Bogart (University of North Carolina Chapel Hill)[27], Michael Rock (University of Wisconsin Hospital and Clinics)[27], Antony Rosen (Johns Hopkins University)[32], Margaret Rosenfeld (Seattle Children's Hospital, University of Washington)[27], Ingo Ruczinski (Johns Hopkins University School of Public Health)[13, 15, 30], Andrew Sanford (University of British Columbia)[30], David Schaeffer (Nemours Children's Clinic)[27], Cindy Sell (University of North Carolina Chapel Hill)[26], Daniel Sheehan (Children's Hospital of Buffalo)[27], Edwin K. Silverman (Brigham and Women's Hospital, Harvard Medical School)[24, 30], Don Sin (Children's Medical Center of Dayton)[30], Terry Spencer (Elliot Health System)[27], Jackie Stonebraker (University of North Carolina Chapel Hill)[26], Holly K. Tabor (Seattle Children's Hospital, University of Washington)[6, 10, 11, 17, 27], Laurie Varlotta (St. Christopher's Hospital for Children)[27], Candelaria I. Vergara (Johns Hopkins University)[30], Robert Weiss [30], Fred Wigley (Johns Hopkins University)[32], Robert A. Wise (Johns Hopkins University)[30], Fred A. Wright (University of North Carolina Chapel Hill)[26], Mark M. Wurfel (University of Washington)[10, 14, 21], Robert Zanni (Monmouth Medical Center)[27], Fei Zou (University of North Carolina Chapel Hill)[26]

SeattleGO

Deborah A. Nickerson (University of Washington)[3, 4, 5, 7, 8, 9, 11, 15, 17, 18, 19], Mark J. Rieder (University of Washington)[4, 11, 13, 15, 16, 17, 19], Phil Green (University of Washington), Jay Shendure (University of Washington)[1, 8, 14, 16, 17], Joshua M. Akey (University of Washington)[13, 14, 15], Michael J. Bamshad (Seattle Children's Hospital, University of Washington)[4, 6, 7, 8, 10, 11, 13, 15, 17, 27], Carlos D. Bustamante (Stanford University School of Medicine)[3, 13, 15], David R. Crosslin (University of Washington)[2, 9], Evan E. Eichler (University of Washington)[19], P. Keolu Fox[2], Wenqing Fu (University of Washington)[13], Adam Gordon (University of Washington)[11], Simon Gravel (Stanford University School of Medicine)[13, 15], Gail P. Jarvik (University of Washington)[9, 15], Jill M. Johnsen (Puget Sound Blood Center, University of Washington)[2], Mengyuan Kan (Baylor College of Medicine)[13], Eimear E. Kenny (Stanford University School of Medicine)[3, 13, 15], Jeffrey M. Kidd (Stanford University School of Medicine)[13, 15], Fremiet Lara-Garduno (Baylor College of Medicine)[15], Suzanne M. Leal (Baylor College of Medicine)[1, 13, 15, 16, 17, 19, 20], Dajiang J. Liu (Baylor College of Medicine)[13, 15], Sean McGee (University of Washington)[13, 15, 19], Timothy D. O'Connor (University of Washington)[13], Bryan Paeper (University of Washington)[16], Peggy D. Robertson (University of Washington)[4], Joshua D. Smith (University of Washington)[4, 16, 19], Jeffrey C. Staples (University of Washington), Jacob A. Tennessen (University of Washington)[13], Emily H. Turner (University of Washington)[4, 16], Gao Wang (Baylor College of Medicine)[1,13,20], Qian Yi (University of Washington)[4]

WHISP

Rebecca Jackson (Ohio State University)[1, 2, 4, 5, 8, 12, 14, 15, 17, 18, 20, 34], Kari North (University of North Carolina Chapel Hill)[1, 3, 9, 10, 13, 15, 17, 34,] Ulrike Peters (Fred Hutchinson Cancer Research Center)[1, 3, 11, 12, 13, 15, 17, 18, 34], Christopher S. Carlson (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 12, 13, 14, 15, 16, 17, 18, 19, 34], Garnet Anderson (Fred Hutchinson Cancer Research Center)[34], Hoda Anton-Culver (University of California at Irvine)[34], Themistocles L. Assimes (Stanford University School of Medicine)[5, 9, 11, 34], Paul L. Auer (Fred Hutchinson Cancer Research Center)[1, 2, 3, 5, 11, 12, 13, 15, 16, 18, 34], Shirley Beresford (Fred Hutchinson Cancer Research Center)[34], Chris Bizon (University of North Carolina Chapel Hill)[3, 9, 13, 15, 34], Henry Black (Rush Medical Center)[34], Robert Brunner (University of Nevada)[34], Robert Brzyski (University of Texas Health Science Center San Antonio)[34], Dale Burwen (National Heart, Lung, and Blood Institute WHI Project Office)[34], Bette Caan (Kaiser Permanente Division of Research, Oakland, CA)[34], Cara L. Carty (Fred Hutchinson Cancer Research Center)[18, 34], Rowan Chlebowski (Los Angeles Biomedical Research Institute)[34], Steven Cummings (University of California at San Francisco)[34], J. David Curb* (University of Hawaii)[9, 18, 34], Charles B. Eaton (Brown University, Memorial Hospital of Rhode Island)[12, 34], Leslie Ford (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[34], Nora Franceschini (University of North Carolina Chapel Hill)[2, 3, 9, 10, 15, 34], Stephanie M. Fullerton (University of Washington)[6, 11, 34], Margery Gass (University of Cincinnati)[34], Nancy Geller (National Heart, Lung, and Blood Institute WHI Project Office)[34], Gerardo Heiss (University of North Carolina Chapel Hill)[5, 34], Barbara V. Howard (Howard University, MedStar Research Institute)34, Li Hsu (Fred Hutchinson Cancer Research Center)[1, 13, 15, 18, 34], Carolyn M. Hutter (Fred Hutchinson Cancer Research Center)[13, 15, 18, 34], John Ioannidis (Stanford University School of Medicine)[11, 34], Shuo Jiao (Fred Hutchinson Cancer Research Center)[34], Karen C. Johnson (University of Tennessee Health Science Center)[3, 34], Charles Kooperberg (Fred Hutchinson Cancer Research Center)[1, 5, 9, 13, 14, 15, 17, 18, 34], Lewis Kuller (University of Pittsburgh)[34], Andrea LaCroix (Fred Hutchinson Cancer Research Center)[34], Kamakshi Lakshminarayan (University of Minnesota)[18, 34], Dorothy Lane (State University of New York at Stony Brook)[34], Ethan M. Lange (University of North Carolina Chapel Hill)[3, 5, 9, 13, 34], Leslie A. Lange (University of North Carolina Chapel Hill)[1, 2, 3, 5, 9, 12, 13, 15, 17, 18, 20, 25, 34], Norman Lasser (University of Medicine and Dentistry of New Jersey)[34], Erin LeBlanc (Kaiser Permanente Center for Health Research, Portland, OR)[34], Cora E. Lewis (University of Alabama at Birmingham)[25,34], Kuo-Ping Li (University of North Carolina Chapel Hill)[9, 34], Marian Limacher (University of Florida)[34], Dan-Yu Lin (University of North Carolina Chapel Hill)[1, 3, 9, 13, 15, 34], Benjamin A. Logsdon (Fred Hutchinson Cancer Research Center)[2, 34], Shari Ludlam (National Heart, Lung, and Blood Institute WHI Project Office)[34], JoAnn E. Manson (Brigham and Women's Hospital, Harvard School of Public Health)[34], Karen Margolis (University of Minnesota)[34], Lisa Martin (George Washington University Medical Center)[9, 34], Joan McGowan (National Heart, Lung, and Blood Institute WHI Project Office)[34], Keri L. Monda (Amgen, Inc.)[1, 15, 34], Jane Morley Kotchen (Medical College of Wisconsin)[34], Lauren Nathan (University of California Los Angeles)[34], Judith Ockene (Fallon Clinic, University of Massachusetts)[34], Mary Jo O'Sullivan (University of Miami)[34], Lawrence S. Phillips (Emory University)[34], Ross L. Prentice (Fred Hutchinson Cancer Research Center)[34], Alexander P. Reiner (Fred Hutchinson Cancer Research

Center, University of Washington)[1, 2, 3, 5, 9, 11, 12, 13, 14, 15, 20, 25, 34], John Robbins (University of California at Davis)[34], Jennifer G. Robinson (University of Iowa)[9, 11, 18, 34], Jacques E. Rossouw (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[5, 14, 17, 20, 34], Haleh Sangi-Haghpeykar (Baylor College of Medicine)[34], Gloria E. Sarto (University of Wisconsin)[34], Sally Shumaker (Wake Forest University)[34], Michael S. Simon (Wayne State University)[34], Marcia L. Stefanick (Stanford University School of Medicine)[34], Evan Stein (Medical Research Labs)[34], Hua Tang (Stanford University)[2, 34], Kira C. Taylor (University of Louisville)[1, 3, 13, 15, 20, 34], Cynthia A. Thomson (University of Arizona)[34], Timothy A. Thornton (University of Washington)[13, 15, 18, 34], Linda Van Horn (Northwestern University)[34], Mara Vitolins (Wake Forest University)[34], Jean Wactawski-Wende (University of Buffalo)[34], Robert Wallace (University of Iowa)[2, 34], Sylvia Wassertheil-Smoller (Boston University)[18, 34], Donglin Zeng (University of North Carolina Chapel Hill)[9, 34]

NHLBI GO ESP Project Team

Deborah Applebaum-Bowden (National Heart, Lung, and Blood Institute)[4, 7, 12, 17], Michael Feolo (National Center for Biotechnology Information)[12], Weiniu Gan (National Heart, Lung, and Blood Institute)[7, 8, 16, 17], Dina N. Paltoo (National Heart, Lung, and Blood Institute)[4, 6, 11, 17], Jacques E. Rossouw (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[5, 14, 17, 20, 34], Phyliss Sholinsky (National Heart, Lung, and Blood Institute)[4, 12, 17], Anne Sturcke (National Center for Biotechnology Information)[12]

*deceased

ESP Groups

[1]Anthropometry Project Team, [2]Blood Count/Hematology Project Team, [3]Blood Pressure Project Team, [4]Data Flow Working Group, [5]Early MI Project Team, [6]ELSI Working Group, [7]Executive Committee, [8]Family Study Project Team, [9]Lipids Project Team, [10]Lung Project Team, [11]Personal Genomics Project Team, [12]Phenotype and Harmonization Working Group, [13]Population Genetics and Statistical Analysis Working Group, [14]Publications and Presentations Working Group, [15]Quantitative Analysis Ad Hoc Task Group, [16]Sequencing and Genotyping Working Group, [17]Steering Committee, [18]Stroke Project Team, [19]Structural Variation Working Group, [20]Subclinical/Quantitative Project Team

ESP Cohorts

[21]Acute Lung Injury (ALI), [22]Atherosclerosis Risk in Communities (ARIC), [23]Cardiovascular Health Study (CHS), [24]Chronic Obstructive Pulmonary Disease (COPDGene), [25]Coronary Artery Risk Development in Young Adults (CARDIA), [26]Cystic Fibrosis (CF), [27]Early Pseudomonas Infection Control (EPIC), [28]Framingham Heart Study (FHS), [29]Jackson Heart Study (JHS), [30]Lung Health Study (LHS), [31]Multi-

Ethnic Study of Atherosclerosis (MESA), [32]Pulmonary Arterial Hypertension (PAH), [33]Severe Asthma Research Program (SARP), [34]Women's Health Initiative (WHI)