Annals of human genetics WILEY

Multimarker omnibus tests by leveraging individual marker summary statistics from large biobanks

Angela M. Zigarelli[1] | Hanna M. Venera[2] | Brody A. Receveur[3] | Jack M. Wolf[4] |
Jason Westra[5] | Nathan L. Tintle[6]

[1]Department of Mathematics and Statistics, University of Massachusetts Amherst, Massachusetts, USA

[2]Division of Biostatistics, University of Michigan, Michigan, USA

[3]Department of Statistics, George Mason University, Virginia, USA

[4]Division of Biostatistics, University of Minnesota, Minnesota, USA

[5]Department of Math, Computer Science, and Statistics, Dordt University, Iowa, USA

[6]Department of Population Health Nursing Sciences, University of Illinois Chicago, Chicago, Illinois, USA

Correspondence
Nathan Tintle, Department of Population Health Nursing Sciences, University of Illinois Chicago, Chicago, IL, USA.
Email: ntintle@uic.edu

Funding information
National Institutes of Health, Grant/Award Number: R15HG006915; Dordt University

Abstract: As biobanks become increasingly popular, access to genotypic and phenotypic data continues to increase in the form of precomputed summary statistics (PCSS). Widespread accessibility of PCSS alleviates many issues related to biobank data, including that of data privacy and confidentiality, as well as high computational costs. However, questions remain about how to maximally leverage PCSS for downstream statistical analyses. Here we present a novel method for testing the association of an arbitrary number of single nucleotide variants (SNVs) on a linear combination of phenotypes after adjusting for covariates for common multimarker tests (e.g., SKAT, SKAT-O) without access to individual patient-level data (IPD). We validate exact formulas for each method, and demonstrate their accuracy through simulation studies and an application to fatty acid phenotypic data from the Framingham Heart Study.

KEYWORDS
genetic data banks, genetic markers, genetic privacy, genotype–phenotype associations, statistical

1 | INTRODUCTION

The availability and accessibility of data is an exciting prospect for advancements in science—especially in the area of biomedical research. However, the utilization of individual patient-level data (IPD) raises issues of data privacy and security, and in the biobank era, issues with computational cost and data processing time (Heatherly, 2016; Huppertz & Holzinger, 2014).

With the popularization of biobanks, there has been an explosion in the magnitude of available genotypic and phenotypic data, to facilitate the connection between genetics and human health. In an effort to make biobank data more accessible and usable, projects like GeneAtlas and PheWeb provide precomputed summary statistics (PCSS) eliminating the need for some researchers to access IPD (Neale, 2018; PheWeb, 2018). Most often these projects

The authors Hanna Venera and Angela Zigarelli contributed equally to this manuscript.

have computed simple linear regression results between many combinations of genotypes and phenotypes and then made summary statistics like β and SE(β) publicly available (Canela-Xandri et al., 2018; Sudlow et al., 2015). Though this process alleviates the issue of data privacy and confidentiality, as well as high computational costs, questions are raised about how useful PCSS can be given that the scope of PCSS is often merely results from simple linear regression models.

In an effort to better leverage PCSS in downstream analyses, several groups have published methods that utilize GWAS PCSS to perform meta-analyses, as well as multivariate methods to investigate relationships between phenotypes (Cichonska et al., 2016; Liu & Lin, 2018; Ray & Boehnke, 2018; Stephens, 2013; van der Sluis et al., 2013; Vuckovic et al., 2015). More recently, Gasdaska et al. (2019) developed a method utilizing PCSS from simple linear regression models to regress linear combinations of phenotypes against genotypes, with an extension by Wolf et al. (2021) for the multiplicative phenotype case. Subsequently, Wolf et al. (2020) expanded the approach to be able to use PCSS to adjust for covariates not included in the simple linear regression models on which the PCSS are based. Other recent developments in this area include Svishcheva et al. (2019) who presented an approach which leverages simple linear regression PCSS to perform gene-based (multimarker) tests on a single phenotype. Despite these advances, significant gaps in the literature remain. In particular, methods for conducting multimarker tests of association utilizing complex (e.g., linear/multiplicative combinations) phenotypes are lacking. Furthermore, post hoc covariate adjustment utilizing PCSS is important and, yet, is lacking for most methods utilizing PCSS to date. While metaSKAT (Lee et al., 2013) utilizes PCSS in a multimarker framework, this approach requires prespecification of phenotypes and covariates before PCSS are computed, limiting post hoc research exploration on different sets of phenotypes and/or covariates. Furthermore, multiSKAT (Dutta et al., 2019) tests for a general multivariate phenotype limiting power for research hypotheses involving prespecified, multivariate phenotypes.

To address these gaps, we present a method that calculates various multi-marker tests (e.g., SKAT; SKAT-O), for a linear combination of phenotypes using only PCSS from simple linear regression models of a single phenotype on a single genotype. Thus, our method allows phenotypes, genotypes, and covariates from different analyses to be combined using researcher specified linear combinations of phenotypes, genotype sets, and covariates. We provide an analytic framework for our methodology which is validated through simulation. Finally, we apply these methods on real data from the Framingham Heart Study.

2 | METHODS

The following sections outline first a method to calculate the F-test statistic, and second, a method to calculate rare variant tests. Both methods use only PCSS inputs, by which we mean only the estimates and standard errors from a simple linear regression model of a single phenotype on a single genetic marker (e.g., single nucleotide variant [SNV]).

2.1 | Notation and assumptions

Throughout this paper we use the column vectors $y_1, \ldots, y_m$ to represent a collection of vectors of $n \times 1$ measurements on m phenotypes such that $y_j = [y_{1j}, \ldots, y_{nj}]$ is a vector of n measurements on the jth phenotype. We define $y = \sum_{j=1}^{m} \tau_j y_j$ as n measurements across a weighted linear combination of m phenotypes where τ is an $m \times 1$ vector of weights corresponding to each phenotype. We use $X = [x_1, \ldots, x_p]$ to denote an $n \times p$ design matrix of n individuals on p variables including k SNVs for $p > 1$. We use $\bar{y}$ to represent the $m \times 1$ vector of means for each phenotype and $\bar{x}$ to represent the $p \times 1$ vector of means for the design matrix. We also define cov(X) as the empirical covariance matrix of X such that the i, jth entry in the matrix represents the covariance of x_i and x_j.

Moreover, we assume that $\sum_{j=1}^{m} \tau_j y_j = X\beta + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2 I_n)$ which is a standard assumption for both the F and rare variant tests used to derive the test statistics' sampling distributions. We also assume to have PCSS including the means for each phenotype, genotype, and covariate, as well as the full covariance matrix, that is, cov($y_1, \ldots, y_m, X$), which we will leverage to perform the omnibus tests of interest. These statistics can be aggregated from a variety of sources including single-marker GWAS results. For example, the sample covariance of a given SNV and a phenotype can be calculated by diving the GWAS simple linear regression slope coefficient by the sample variance of the SNV (Wolf et al., 2021). Figure 1 proposes one framework for collecting and compiling this information in practice.

Though we will assume to have a known covariance matrix of the phenotypes for our methods to produce calculations identical to IPD, it may be approximated using correlations of GWAS test statistics for each phenotype by techniques proposed by Kim et al. (2015) and Zhu et al. (2015).
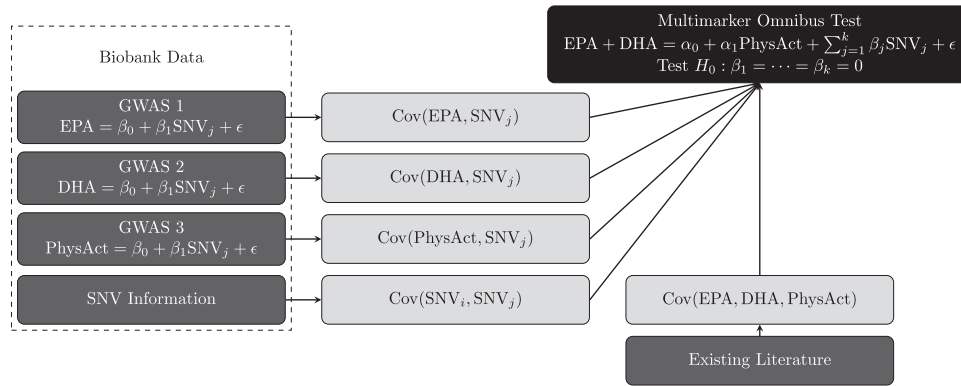
FIGURE 1 Workflow for an omnibus test using only precomputed summary statistics. Abbreviations: DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid; GWAS, genome-wide association study; MAF, minor allele frequency; PhysAct, physical activity; SNV, single-nucleotide variant

2.2 | F-tests for a linear model

The *F*-test can be used to evaluate the combined effect of several linear predictors on a response. Consider the design matrix $\tilde{X}$ that contains a subset of q selected attributes from the design matrix for $q < p$. For example, $\tilde{X}$ could be a collection of covariates that does not include the genetic markers of interest. Then, the test has null hypothesis $y = \tilde{X}\boldsymbol{\gamma} + \varepsilon$ for $\varepsilon \sim N(0, \sigma^2 I_n)$ and has test statistic:

$$F = \frac{(\text{SSR}_R - \text{SSR}_F)/(p - q)}{\text{SSR}_F/(n - p)} \quad (1)$$

which follows an $F_{p-q, n-p}$ distribution under the null hypothesis. We define SSR_F as the sum of squared residuals for the full model $\hat{y} = X\hat{\beta}$ with p predictors including the intercept, and SSR_R as the sum of squared residuals for the reduced model $\hat{y} = \tilde{X}\hat{\boldsymbol{\gamma}}$. The sum of squared residuals (SSR) can be expressed as follows (see Appendix for details):

$$\text{SSR} = y^T y - (X^T y)^T (X^T X)^{-1} X^T y \quad (2)$$

In order to express SSR in terms of PCSS, we use the following from Wolf et al. (2020):

$$y^T y = \sum_{h=1}^{m} \sum_{j=1}^{m} \tau_h \tau_j \left(\text{cov}(y_h, y_j)(n-1) + \bar{y}_h \bar{y}_j n \right) \quad (3)$$

$$X^T y = (n-1) \left[\text{cov}(x_1, y), \dots, \text{cov}(x_p, y) \right]^T + n\bar{x}\bar{y}^T \tau \quad (4)$$

$$X^T X = (n-1) \text{cov}(X) + n\bar{x}\bar{x}^T \quad (5)$$

where $\text{cov}(x_i, y) = \tau_1 \text{cov}(x_1, y_j) + \dots + \tau_m \text{cov}(x_p, y_j)$. We can then evaluate $y^T y, X^T y$, and $X^T X$ using the appropriate subset of X to calculate F from SSR and compare it to its null distribution to obtain a *p*-value under the null hypothesis.

2.3 | Rare variant tests

Rare variant tests including burden, SKAT, and SKAT-O also test the null hypothesis that a collection of features have no effect on a response, are useful in rarer genetic cases where the *F*-test may be less efficient (Lee et al., 2012; Li & SM, 2008; Wu et al., 2011). The following sections detail how to implement these tests using only PCSS inputs.

Consider the partition of our covariate space $X = [G, Z]$ where $G = [g_1, \dots, g_k]$ is an $n \times k$ genotype matrix, g_j is an $n \times 1$ column vector of minor allele counts at the *j*th variant, $Z = [z_1, \dots, z_c]$ is an $n \times c$ matrix of covariates, and z_l is an $n \times 1$ vector for the *l*th covariate. Similar to the *F*-test, the null hypothesis of the rare variant tests assumes that $y = Z\boldsymbol{\alpha} + \varepsilon$ and $\varepsilon \sim N(0, \sigma^2 I_n)$ The test statistics can be obtained by first calculating a site-specific score statistic for each SNV in G and then combining the score statistics into one statistic. We use $\bar{g}$ to represent the $k \times 1$ vector of means for each SNV and $\bar{z}$ to represent the $c \times 1$ vector of means for each covariate. For the null model $y = Z\boldsymbol{\alpha} + \varepsilon$, let $\hat{y} = Z\hat{\alpha}$ be the $n \times 1$ vector of the fitted values.

### 2.3.1 | Score statistic framework

A site score statistic that captures how much the minor alleles at variant *j* empirically contribute to increases in a

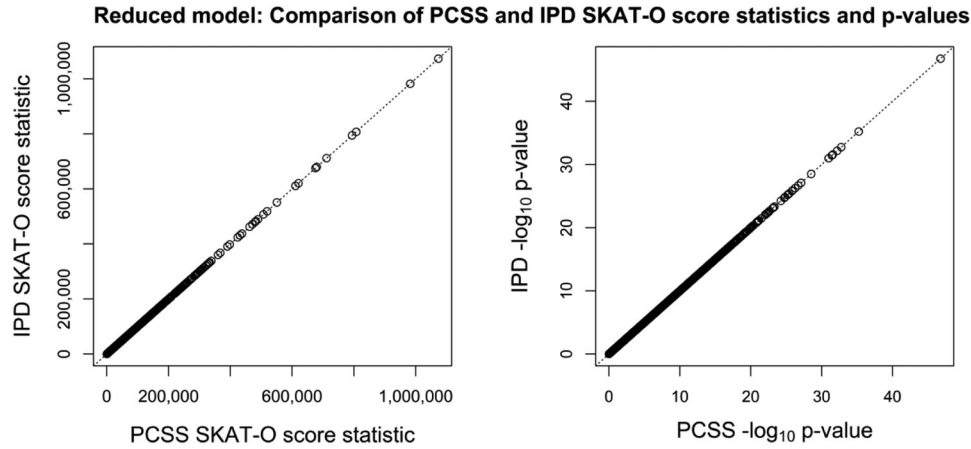**Example 1: Comparison of PCSS and IPD nested F-test statistics and p-values**
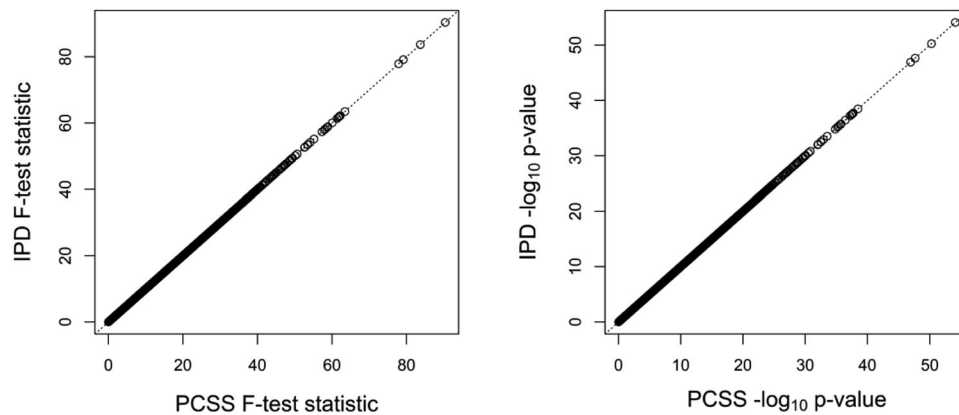
FIGURE 2 The plots depict the difference between score statistics (left) and log transformed *p*-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to a nested *F*-test adjusted for age and sex (example 1). These figures illustrate our method's high accuracy on simulated data

Example 2: Comparison of PCSS and IPD nested F-test statistics and p-values
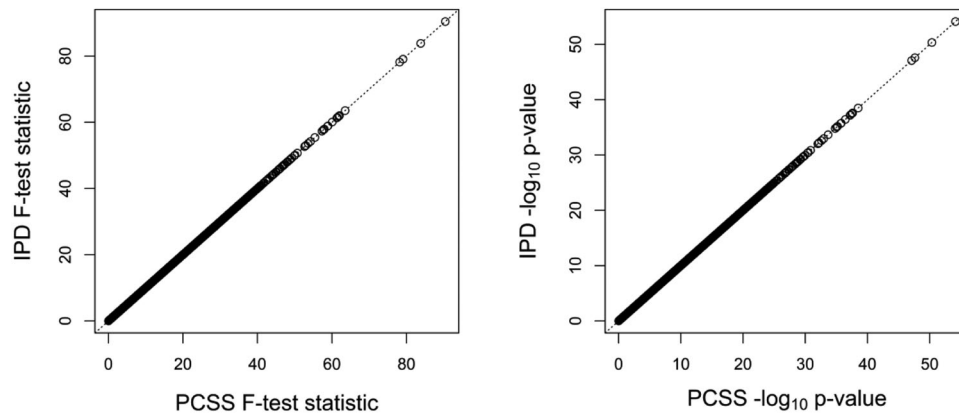
FIGURE 3 The plots depict the difference between score statistics (left) and log transformed *p*-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to a nested *F*-test adjusted for sex (example 2). These figures illustrate our method's high accuracy on simulated data

Full model: Comparison of PCSS and IPD SKAT-O score statistics and p-values
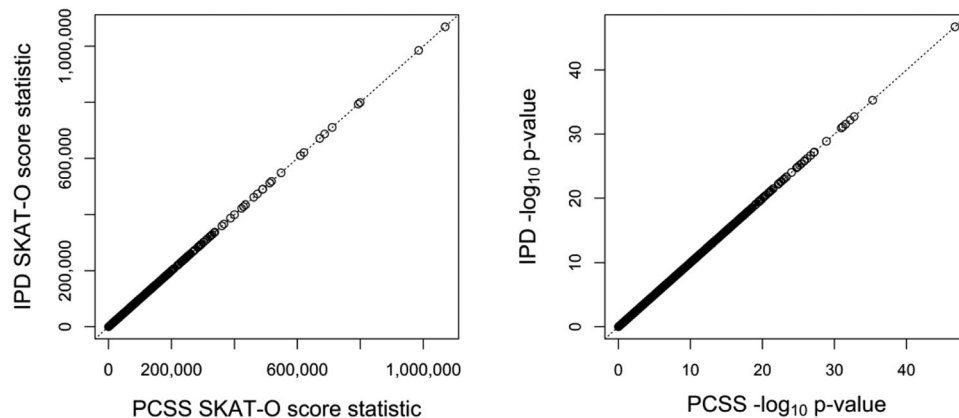
FIGURE 4 The plots depict the difference between score statistics (left) and log transformed *p*-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to the SKAT-O test for the full model, featuring three single nucleotide variants (SNVs), age, and sex. These figures illustrate our method's high accuracy on simulated data

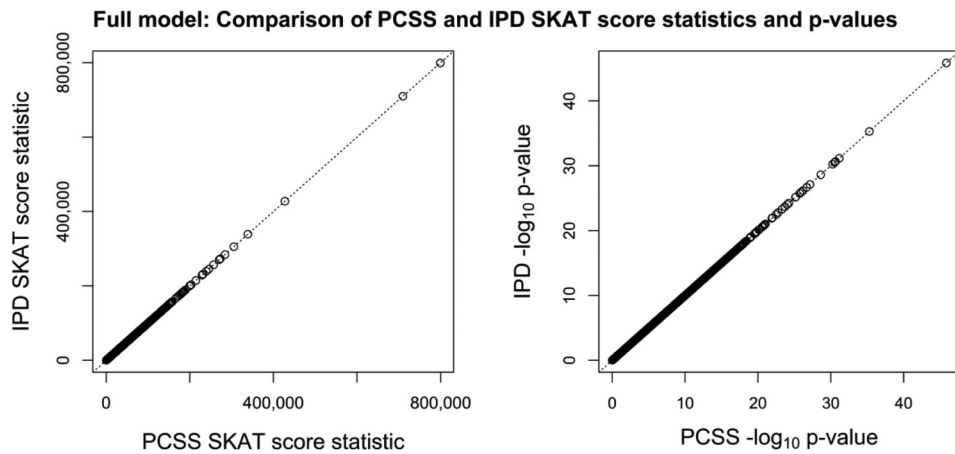**Reduced model: Comparison of PCSS and IPD SKAT-O score statistics and p-values**

FIGURE 5 The plots depict the difference between score statistics (left) and log transformed *p*-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to the SKAT-O test for the reduced model, featuring three single nucleotide variants (SNVs) and sex. These figures illustrate our method's high accuracy on simulated data

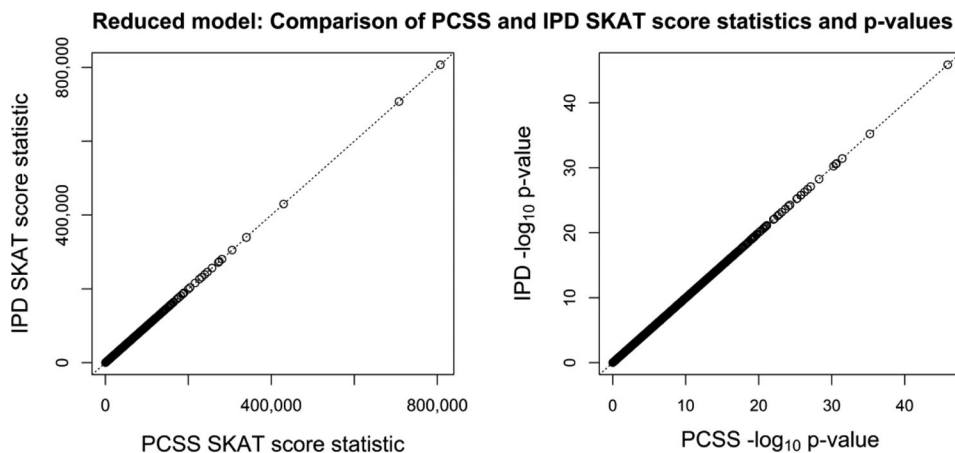**Full model: Comparison of PCSS and IPD SKAT score statistics and p-values**

FIGURE 6 The plots depict the difference between score statistics (left) and log transformed *p*-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to the SKAT test for the full model, featuring three single nucleotide variants (SNVs), age, and sex. These figures illustrate our method's high accuracy on simulated data

Reduced model: Comparison of PCSS and IPD SKAT score statistics and p-values

FIGURE 7 The plots depict the difference between score statistics (left) and log transformed *p*-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to the SKAT test for the reduced model, featuring three single nucleotide variants (SNVs) and sex. These figures illustrate our method's high accuracy on simulated data

continuous phenotype as referenced in Wu et al. (2011) is

$$S_j = g_j^T (y - \hat{y}) / \hat{\sigma}^2 \qquad (6)$$

where $\hat{\sigma}^2$ is the unbiased estimator of σ^2 under the null model. This can be expressed using PCSS inputs as follows (see Appendix for derivation):

$$S_j = \frac{(n-1)}{\hat{\sigma}^2} \left(\text{cov}\left(g_j, y\right) - \sum_{l=1}^{c} \hat{\alpha}_l \, \text{cov}(g_j, z_l) \right) \qquad (7)$$

where

$$\hat{\alpha} = (Z^T Z)^{-1} Z^T y \qquad (8)$$

is the vector of slope coefficients under the covariate-only model, which can be computed by calculating $Z^T y$ and $Z^T Z$ using Equations 4 and 5 above. We note that $\hat{\sigma}^2 =$ SSR$/(n-q)$ using *SSR* from the covariate only model. If there is no covariate adjustment in the null model, the expression can be simplified to

$$S_j = (n-1)\,\text{cov}\left(g_j, y\right) / \hat{\sigma}^2$$

2.3.2 | Calculating the unified score test statistic

The site-specific score statistics can then be aggregated to assess the overall contribution of all k SNVs with $Q_{\text{SKAT}} = \sum_{j=1}^{k} w_j S_j^2$ and $Q_{\text{Burden}} = (\sum_{j=1}^{k} w_j S_j)^2$ where $\boldsymbol{w}$ is a vector of weights (Lee et al., 2012; Li & SM, 2008; Wu et al., 2011). A common set of weights $\boldsymbol{w}$ are the suggested weights as proposed in Wu et al. (2011) given as the square root of the density of a beta(2, 25) distribution evaluated at MAF$_j$ The unified score statistic merges Q_{SKAT} and Q_{Burden} to appeal to advantages of both test statistics, and may be written for the entire set of variants as referenced in Lee et al. (2012):

$$Q_\rho = \rho Q_{\text{Burden}} + (1 - \rho) Q_{\text{SKAT}} \qquad (9)$$

for some weight $\rho \in [0, 1]$. The optimized SKAT (SKAT-O) test statistic is a special case of the unified score Q_ρ such that the weight ρ yields the minimum p-value. The weight ρ^* can be calculated numerically or analytically as seen is Lee et al. (2012), and the SKAT-O test statistic is written as

$$Q_{\text{SKAT-O}} = \rho^* Q_{\text{Burden}} + (1 - \rho^*) Q_{\text{SKAT}} \qquad (10)$$

We note that burden and SKAT test statistics are also special cases of the unified score statistic Q_ρ that occur when ρ equals one and zero, respectively.

The test statistic's null distribution follows a mixture of independent $\chi^2(1)$ distribution such that under the null, $Q_\rho \sim \sum_{j=1}^{k} \lambda_j \chi_j^2(1)$ (Lee et al., 2013). We derive an expression for the weights for this null distribution Q_ρ in terms of PCSS.

Treating this as a meta-analysis for one cohort and following the framework proposed by Lee et al. (2013), we reference the following expression where $\boldsymbol{\Phi}$ is the $k \times k$ between-variant relationship matrix:

$$\Phi = (G^T G - (Z^T G)^T (Z^T Z)^{-1} Z^T G) / \hat{\sigma}^2 \qquad (11)$$

We find that $Z^T G$ can be expressed using PCSS as

$$Z^T G = (n-1)\,\text{cov}\,(Z, G) + n\bar{z}\bar{g}^T \qquad (12)$$

where cov(**Z**, **G**) represents the $c \times k$ covariance matrix of **Z** and **G** such that cov(**Z**, **G**)$_{(i,j)}$ = cov(z_i, g_j). Thus, **Φ** and $Z^T G$ can be expressed using only PCSS inputs, and we solve for the matrix Φ_ρ, which is defined as Lee et al. (2013):

$$\Phi_\rho = L_\rho^T W \Phi W L_\rho \qquad (13)$$

where L_ρ is the Cholesky decomposition of the $k \times k$ compound symmetric matrix $R_\rho = (1 - \rho) I + \rho 11^T$ and $\boldsymbol{W}$ is a diagonal matrix with weights $w_1, w_2, \ldots, w_k$ corresponding to the column order of $\boldsymbol{G}$ (Lee et al., 2013). Then the weights $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of Φ_ρ. To find the optimal ρ^*, we calculate $\mathbf{Q}_\rho$ for each ρ over a grid and use the corresponding weights to calculate the associated p-values using Davies method. Then, we choose ρ^* such that the p-value is minimized to find the SKAT-O score statistic Q_{SKAT-O} (Lee et al., 2013).

2.4 | Simulation

To evaluate the accuracy of our proposed method we performed a large simulation study. We simulated three correlated SNVs across 2000 subjects and generated minor allele frequency (MAF) using a beta(1, 4) (divided by four in order to simulate rare variants). The correlation matrix was sampled uniformly from the set of all possible 3×3 correlation matrices given these MAFs. Subjects' age and sex were generated independently from a Poisson distribution and a Bernoulli distribution, respectively. Each subject had three phenotypes: y_1, y_2, and y_3 which were generated from a multiple linear regression with coefficients age, sex, and the three SNVs. Error terms were generated from a multivariate normal distribution with mean 0 and covariance $\boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma}$ was randomly generated in each simulation using correlations for each pair of phenotypes drawn from a uniform(-0.5, 0.5) distribution, to

TABLE 1 The accuracy of our methods to estimate the nested F-test statistic for the reduced model adjusting for covariates using precomputed summary statistics (PCSS). Errors were minimal with low variance in all cases

Omnibus statistic	Full model	Reduced model	Mean error	Variance
F-test statistic	SNV1, SNV2, SNV3, Age, Sex	Age, sex	3.59×10^{-16}	9.48×10^{-27}
p-Value	SNV1, SNV2, SNV3, Age, Sex	Age, sex	1.03×10^{-17}	7.68×10^{-28}
F-test statistic	SNV1, SNV2, SNV3, Sex	Sex	-3.82×10^{-17}	4.13×10^{-27}
p-Value	SNV1, SNV2, SNV3, Sex	Sex	3.48×10^{-17}	4.15×10^{-28}

TABLE 2 The accuracy of our methods to estimate the rare variant test statistics for SKAT-O, SKAT, and burden tests using precomputed summary statistics (PCSS) are depicted below

Omnibus Statistic	Full model (Age)		Reduced model (No age)	
	Mean error	Variance	Mean error	Variance
SKAT-O Score statistic	7.84×10^{-14}	4.40×10^{-22}	-1.62×10^{-14}	4.05×10^{-22}
SKAT-O p-Value	3.23×10^{-18}	2.87×10^{-31}	3.01×10^{-18}	1.96×10^{-31}
SKAT Score statistic	6.23×10^{-14}	2.08×10^{-22}	4.33×10^{-14}	1.40×10^{-22}
SKAT p-value	3.56×10^{-18}	2.44×10^{-31}	8.46×10^{-19}	1.90×10^{-31}
Burden score statistic	1.09×10^{-13}	5.63×10^{-22}	-5.11×10^{-14}	6.78×10^{-22}
Burden p-value	-7.02×10^{-19}	7.70×10^{-31}	1.60×10^{-18}	5.49×10^{-31}

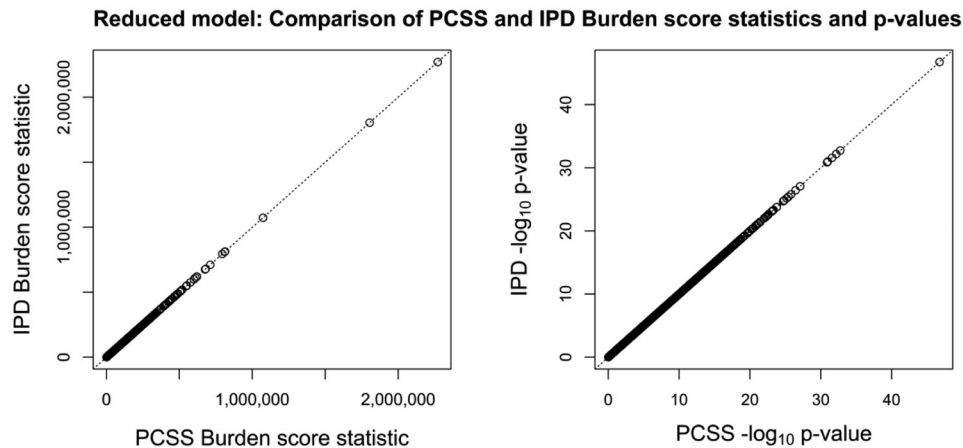*Note*: Errors were minimal with low variance in all cases.
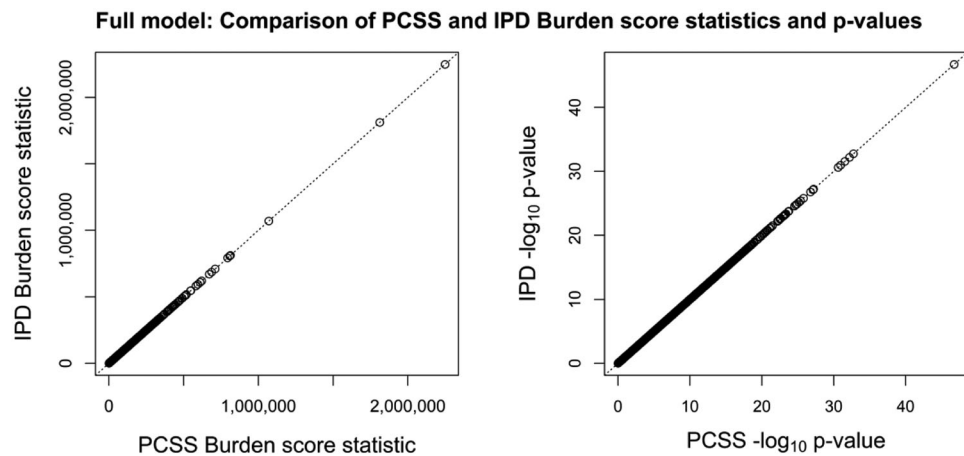
FIGURE 8 The plots depict the difference between score statistics (left) and log transformed p-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to the Burden test for the full model, featuring three single nucleotide variants (SNVs), age, and sex. These figures illustrate our method's high accuracy on simulated data

simulate positive and negative relationships between phenotypes, and error variances held fixed at one for each phenotype. Coefficients for the SNVs and covariates were generated to simulate both negative and positive linear relationships with the phenotypes. We held the intercept at a constant value of zero across all simulations to represent an arbitrary intercept. We then considered the linear combination $\tau_1 y_1 + \tau_2 y_2 + \tau_3 y_3$ as the primary response for all of our analyses where arbitrary weights τ_j were independently generated from a uniform(0.5, 5) distribution.

2.4.1 | F-test simulation

To assess our F-test method's accuracy compared to an IPD-computed F-test, we calculated SSR using the simulated IPD and PCSS for the full model using all predictors, and a reduced model excluding both SNV1 and SNV2. We then performed a nested F-test using only PCSS, calculated the F-test statistic and the corresponding p-value for the linear combination of phenotypes, and evaluated bias in our methods by comparing our result to the F-test statistic using IPD.

FIGURE 9 The plots depict the difference between score statistics (left) and log transformed *p*-values (right) on our method using precomputed summary statistics (PCSS) compared to individual patient-level data (IPD) as applied to the Burden test for the reduced model, featuring three single nucleotide variants (SNVs) and sex. These figures illustrate our method's high accuracy on simulated data

2.4.2 | Rare variant test simulation

To assess our rare variant tests, we calculated the test statistic $Q_{\rho*}$ that combined all S_j across the region using the simulated IPD and PCSS using the suggested weights as proposed in Wu et al. (2011). To assess the potential effect of covariates in the model, we also calculated a *p*-value using a standard SKAT-O test on the IPD. We note that by calculating the SKAT-O score statistic, SKAT and burden are trivial extensions.

2.5 | Real data example

Several groups have investigated the relationship between SNPs and fatty acid levels in the blood using data from the Framingham Heart Study (Kalsbeek et al., 2018; Tintle et al., 2015; Veenstra et al., 2017; Wolf et al., 2020). We applied our method to an unrelated subset of Generation 3 and Offspring cohorts from the Framingham Heart Study using the sequence data on 1212 subjects across 20 genes with common variants associated with omega 3 fatty acids as indicated by the GWAS catalog. We tested if the rare variants in these genes had an effect on the omega 3 index (EPA + DHA) adjusting for age and sex. We used dbGene (NLM, n.d.) for the start and end points of each gene referencing human genome build 38, and all variants in the minor allele frequency window between 0.01 and 0.05. When examining the nested F-test, we used all variants with minor allele frequency greater than 0.05 for the SKAT-O test, and an absolute correlation less than 0.9 compared to all other SNPs in the subset.

After calculating basic summary statistics for our phenotypes (EPA, DHA, age, and sex) and all genotypes, we used the summary statistics to calculate the SKAT-O score

statistics and *p*-value. We then compared our results to those found using IPD and the meta-analysis null distribution. For the nested *F*-test, we compared a model with all SNPs, age, and sex to a model with only age and sex.

3 | RESULTS

The following sections detail our results from our simulation and real data application to the Framingham Heart Study.

3.1 | Simulation *F*-test statistic results

We performed two nested *F*-tests with 3 SNVs: one full model adjusting for age and sex, and one full model adjusting only for sex. Table 1 shows the accuracy of our results for both nested *F*-test simulations averaged across 100,000 iterations. We show the precision of our method with respect to calculated test statistics and the associated *p*-values in Figures 2 and 3. In short, our method to describe covariate-adjusted models with multiple SNVs on a linear combination of phenotypes with a nested *F*-test proved to be exact to rounding errors. As expected, due to the exactness of the method, simulation shows maintenance of the type I error rate and power as compared to the IPD method (detailed results not shown).

3.2 | Simulation rare variant test statistic results

We tested the full model with three SNVs adjusting for age and sex, and the reduced model removing age. Table 2 shows the flexibility of our results for the SKAT-O

simulations averaged across 100,000 iterations, as well as results for SKAT and burden test for completeness, noting that these results are trivial after showing the accuracy of the SKAT-O test. We show the precision of our method with respect to calculated test statistics and the associated *p*-values for each of the three tests in Figures 4 and 5 for the SKAT-O test. Similarly, Figures 6 and 7 for SKAT and Figures 8 and 9 for burden tests are included in the Appendix, as they are trivial extensions. Again, our method to describe covariate adjusted models with multiple SNVs with a SKAT-O test statistic proved to be exact to rounding errors. As expected, due to the exactness of the method, simulation shows maintenance of the type I error rate and power as compared to the IPD method (detailed results not shown).

3.3 | Real data results

As indicated in Appendix Tables C1 and C2, the results from our PCSS methods using the Generation 3 and Offspring cohorts from the Framingham Heart Study are equal to IPD methods to rounding errors. No gene *p*-value was less than the Bonferroni corrected cut off of 0.0025. Small bias is due to rounding error and missing data and the bias is similar across a varied number of different SNVs.

4 | CONCLUSIONS

We have developed and demonstrated exact methods for *F*-tests and common rare variant tests on a linear combination of continuous phenotypes regressed on an arbitrary number of SNVs and covariates using only PCSS. We provided the mathematical framework behind these methods, and validated them through simulation and a real data application using the Framingham heart study.

Since the method we have developed is mathematically exact, limitations due to methodological approximations are limited. There are, however, some limitations worth noting. For example, the rare variant tests we have derived here are only robust for continuous phenotypes. Extensions of the method as shown to binary or other categorical variables may not be robust and is an area of future research. We note that while we simulated the variance of the SNVs under the assumption of Hardy Weinberg equilibrium using the minor allele frequency. In practice, if the variance of the SNVs is known, the HWE assumption becomes unnecessary.

We have applied the method to a well-studied genome wide dataset (Framingham) and shown that the method provides exact solutions in this context as well. However, future research is needed in order to more fully vet the method on real data to explore its robustness to larger amounts of missing data, skewness and outliers, and situations that arise when some PCSS are not available for the cohort of interest.

Finally, we note that estimating the covariance matrix of the phenotypes using the correlation matrix of slope coefficients and each individual genotype has been shown by Kim et al. (2015) and Zhu et al. (2015) and can be utilized if needed. Earlier work on combined phenotypes (Gasdaska et al., 2019) using this approach showed unbiased behavior with some power loss due to an increase in variability. In the case where correlations or values are estimated and investigators wish to evaluate the sensitivity of their findings to these estimated quantities, we recommend rerunning the method using different estimates of these quantities to qualitatively understand the sensitivity of their findings to the estimated values. Future simulation studies should continue to explore the sensitivity of this approach and other PCSS-based methods to estimated and incorrect PCSS inputs.

CONFLICT OF INTEREST
The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in dbGaP at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v32.p13 [phs000007].

ORCID
Jack M. Wolf https://orcid.org/0000-0002-8919-8740

REFERENCES
Canela-Xandri, O., Rawlik, K., & Tenesa, A. (2018). An atlas of genetic associations in UK biobank. *Nature Genetics*, *50*, 1593–1599.

Cichonska, A., Rousu, J., Marttinen, P., Kangas, A., Soininen, P., Lehtimäki, T., Raitakari, O. T., Järvelin, M. R., Salomaa, V., Ala-Korpela, M., Ripatti, S., & Pirinen, M. (2016). metacca: Summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics*, *32*(13), 1981–1989.

Dutta, D., Scott, L., Boehnke, M., & Lee, S. (2019). Multi-skat: General framework to test for rare-variant association with multiple phenotypes. *Genetic Epidemiology*, *43*(1), 4–23.

Gasdaska, A., Friend, D., Chen, R., Westra, J., Zawitowski, M., Lindsey, W., & Tintle, N. (2019). Leveraging summary statistics to make inferences about complex phenotypes in large biobanks. *Pacific Symposium on Biocomputing*, *24*, 391–402.

Heatherly, R. (2016). Privacy and security within biobanking: The role of information technology. *Journal of Law, Medicine Ethics*, *44*(1), 156–160.

Huppertz, B., & Holzinger, A. (2014). Biobanks – a source of large biological data sets: Open problems and future challenges. In A. Holzinger, & I. Jurisica (Eds.), *Interactive knowledge discovery and data mining in biomedical informatics*. Springer.

Kalsbeek, A., Veenstra, J., Westra, J., Disselkoen, C., Koch, K., McKenzie, K. A., O'Bott, J., Vander Woude, J., Fischer, K., Shearer, G. C., Harris, W. S., & Tintle, N. L. (2018). A genome-wide association study of red-blood cell fatty acids and ratios incorporating dietary covariates: Framingham heart study offspring cohort. *PLoS ONE*, *13*(4), e0194882.

Kim, J., Bai, Y., & Pan, W. (2015). An adaptive association test for multiple phenotypes with gwas summary statistics. *Genetic Epidemiology*, *39*(8), 651–663.

Lee, S., Emond, M., Bamshad, M., Barnes, K., Rieder, M., & Nickerson, D. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American Journal of Human Genetics*, *91*(2), 224–237.

Lee, S., Teslovich, T. M., Boehnke, M., & Lin, X. (2013). General framework for meta-analysis of rare variants in sequencing association studies. *American Journal of Human Genetics*, *93*(1), 42–53.

Li, B., & SM, L. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *American Journal of Human Genetics*, *83*(3), 311–321.

Liu, Z., & Lin, X. (2018). Multiple phenotype association tests using summary statistics in genome-wide association studies. *Biometrics*, *74*(1), 165–175.

Neale, B. M. (2018). *Biobank gwas*. Retrieved from http://www.nealelab.is/uk-biobank/

NLM. (n.d.). Dbgene. https://www.ncbi.nlm.nih.gov/gene

Pheweb. (2018). Retrieved from https://pheweb.sph.umich.edu/

Ray, D., & Boehnke, M. (2018). Methods for meta-analysis of multiple traits using gwas summary statistics. *Genetic Epidemiology*, *42*(2), 134–145.

Stephens, M. (2013). A unified framework for association analysis with multiple related phenotypes. *PLoS ONE*, *14*(3), e0213951.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, *12*(3), e1001779.

Svishcheva, G. R., Belonogova, N. M., Zorkoltseva, I. V., Kirichenko, A. V., & Axenovich, T. I (2019). Gene-based association tests using gwas summary statistics. *Bioinformatics*, *35*(19), 3701–3708.

Tintle, N. L., Pottala, J. V., Lacey, S., Ramachandrane, V., Westra, J., Rogers, A., Clark, J., Olthoff, B., Larson, M., Harris, W., & Sheareri, G. C. (2015). A genome-wide association study of saturated, mono- and polyunsaturated red blood cell fatty acids in the framingham heart offspring study. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, *94*, 65–72.

van der Sluis, S., Posthuma, D., & Dolan, C. (2013). Tates: Efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLOS Genetics*, *9*, e1003235.

Veenstra, J., Kalsbeek, A., Westra, J., Disselkoen, C., Smith, C. E., & Tintle, N. (2017). Genome-wide interaction study of omega-3 pufas and other fatty acids on inflammatory biomarkers of cardiovascular health in the framingham heart study. *Nutrients*, *9*(8), 900.

Vuckovic, D., Gasparini, P., Soranzo, N., & Iotchkova, V. (2015). Multimeta: An r package for meta-analyzing multi-phenotype genome-wide association studies. *Bioinformatics*, *31*(16), 2754–2756.

Wolf, J., Barnard, M., Xia, X., Ryder, N., Westra, J., & Tintle, N. (2020). Computationally efficient, exact, covariate-adjusted genetic principal component analysis by leveraging individual marker summary statistics from large biobanks. *Pacific Symposium on Biocomputing*, *25*, 719–730.

Wolf, J., Westra, J., & Tintle, N. (2021). Using summary statistics to model multiplicative combinations of initially analyzed phenotypes with a flexible choice of covariates. *Frontiers in Genetics*, *12*, https://doi.org/10.3389/fgene.2021.74590

Wu, M., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, *89*(1), 82–93.

Zhu, X., Feng, T., Tayo, B., Liang, J., Young, J., Franceschini, N., Smith, J. A., Yanek, L. R., Sun, Y. V., Edwards, T. L., Chen, W., Nalls, M., Fox, E., Sale, M., Bottinger, E., Rotimi, C., Liu, Y., McKnight, B., Liu, K., … Redline, S., COGENT BP Consortium. (2015). Meta-analysis of correlated traits via summary statistics from gwass with an application in hypertension. *American Journal of Human Genetics*, *96*(1), 21–36.

APPENDIX 1

A SSR DERIVATION

For a linear model of the form $\hat{y} = X\hat{\beta}$ fit via least squares, we can simplify the sum of squared residuals such that it can be computed using only summary statistics by the following:

$$
\begin{aligned}
SSR &= (y - X\hat{\beta})^T \left(y - X\hat{\beta} \right) \\
&= y^T y - 2y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta} \\
&= y^T y - 2y^T X(X^T X)^{-1} X^T y + y^T X(X^T X)^{-1} X^T X(X^T X)^{-1} X^T y \\
&= y^T y - 2y^T X(X^T X)^{-1} X^T y + y^T X(X^T X)^{-1} X^T X^T y \\
&= y^T y - y^T X(X^T X)^{-1} X^T y \\
&= \left(y^T y \right) - (X^T y)^T (X^T X)^{-1} X^T y
\end{aligned}
$$

TABLE C1 The accuracy of our methods to perform nested F-tests using PCSS from the Framingham Heart Study

Gene	Num SNVs	F Stat	p-Value	Bias F‡	Bias p-value‡	F Stat	p-Value	Bias F‡	Bias p-value ‡
			With covariate adjustment (age)					Without covariate adjustment (no age)	
ADRA1D	51	0.664	0.967	$-2.08 \times 10{-}10$	$1.01 \times 10{-}10$	0.638	0.978	$1.62 \times 10{-}10$	$-5.75 \times 10{-}11$
AHI1	74	1.364	0.025	$-1.24 \times 10{-}11$	$3.41 \times 10{-}12$	1.501	0.005	$1.28 \times 10{-}11$	$-7.97 \times 10{-}13$
AMIGO2	4	0.987	0.413	$-2.34 \times 10{-}11$	$1.28 \times 10{-}11$	1.083	0.363	$-2.50 \times 10{-}11$	$1.24 \times 10{-}11$
CCDC141	219	1.045	0.329	$5.86 \times 10{-}12$	$-1.94 \times 10{-}11$	1.058	0.287	$2.13 \times 10{-}10$	$-6.56 \times 10{-}10$
CD96	70	1.007	0.464	$-2.06 \times 10{-}11$	$4.67 \times 10{-}11$	1.164	0.174	$-1.88 \times 10{-}11$	$2.49 \times 10{-}11$
CR1L	77	0.853	0.811	$-6.31 \times 10{-}12$	$1.15 \times 10{-}11$	0.878	0.763	$-1.51 \times 10{-}11$	$3.07 \times 10{-}11$
DSPP	68	1.090	0.293	$-2.83 \times 10{-}11$	$5.18 \times 10{-}11$	1.088	0.297	$1.02 \times 10{-}11$	$-1.88 \times 10{-}11$
ELOVL1	3	0.474	0.700	$1.55 \times 10{-}11$	$-1.09 \times 10{-}11$	0.626	0.598	$-2.92 \times 10{-}12$	$1.87 \times 10{-}12$
ELOVL2	27	1.569	0.032	$1.58 \times 10{-}11$	$-3.07 \times 10{-}12$	1.621	0.024	$-5.00 \times 10{-}12$	$7.31 \times 10{-}13$
EPHA2	37	0.972	0.518	$-7.15 \times 10{-}12$	$1.23 \times 10{-}11$	1.044	0.398	$1.77 \times 10{-}12$	$-2.79 \times 10{-}12$
FADS1	3	4.292	0.005	$-3.59 \times 10{-}11$	$2.51 \times 10{-}13$	4.550	0.004	$-1.94 \times 10{-}12$	$9.50 \times 10{-}15$
FADS2	63	1.223	0.117	$-1.44 \times 10{-}11$	$1.35 \times 10{-}11$	1.261	0.086	$-6.46 \times 10{-}12$	$4.72 \times 10{-}12$
FADS3	21	1.479	0.075	$-1.11 \times 10{-}11$	$3.88 \times 10{-}12$	1.646	0.033	$6.41 \times 10{-}12$	$-1.09 \times 10{-}12$
FFAR4	24	0.619	0.924	$8.86 \times 10{-}12$	$-6.05 \times 10{-}12$	0.703	0.852	$9.59 \times 10{-}13$	$-9.65 \times 10{-}13$
FILNC1	81	1.159	0.165	$3.54 \times 10{-}11$	$-4.85 \times 10{-}11$	1.202	0.114	$8.67 \times 10{-}12$	$-9.01 \times 10{-}12$
LRRC3B	92	1.208	0.096	$1.35 \times 10{-}11$	$-1.31 \times 10{-}11$	1.257	0.057	$-1.02 \times 10{-}11$	$6.43 \times 10{-}12$
ME1	111	1.023	0.421	$3.20 \times 10{-}11$	$-8.74 \times 10{-}11$	1.021	0.427	$-9.26 \times 10{-}12$	$2.54 \times 10{-}11$
POLR1D	47	1.145	0.236	$-5.45 \times 10{-}11$	$7.26 \times 10{-}11$	1.141	0.241	$-1.20 \times 10{-}11$	$1.62 \times 10{-}11$
PTGS2	7	1.697	0.106	$2.57 \times 10{-}11$	$-6.14 \times 10{-}12$	1.756	0.093	$1.99 \times 10{-}12$	$-4.23 \times 10{-}13$
SCFD1	63	1.104	0.273	$-1.03 \times 10{-}11$	$1.75 \times 10{-}11$	1.077	0.322	$1.09 \times 10{-}12$	$-2.02 \times 10{-}12$
TMEM258	2	2.239	0.107	$-1.21 \times 10{-}10$	$1.29 \times 10{-}11$	2.657	0.071	$3.67 \times 10{-}12$	$-2.58 \times 10{-}13$
WDR70	343	0.901	0.872	$-1.13 \times 10{-}10$	$2.88 \times 10{-}10$	0.880	0.917	$-6.04 \times 10{-}11$	$1.14 \times 10{-}10$

‡ Bias was calculated as the difference of the PCSS estimate and the individual patient-level data (IPD) estimate.

TABLE C2 The accuracy of our methods to estimate SKAT-O test statistics using PCSS from the Framingham Heart Study. SKAT-O test statistics have been scaled by 108 for readability

| Gene | Num SNVs | With covariate adjustment (age) | | | | Without covariate adjustment (no age) | | | |
		F Stat	*p*-Value	Bias F[‡]	Bias *p*-value[‡]	*F* Stat	*p*-Value	Bias *F*[‡]	Bias *p*-Value[‡]
ADRA1D	58	64.49	0.196	$3.89 \times 10{-04}$	$-6.99 \times 10{-15}$	69.04	0.172	$-1.70 \times 10{-04}$	$3.00 \times 10{-15}$
AHI1	419	95.40	0.233	$5.89 \times 10{-04}$	$-2.00 \times 10{-14}$	100.88	0.176	$-2.00 \times 10{-04}$	$6.00 \times 10{-15}$
AMIGO2	3	0.43	0.368	$2.40 \times 10{-06}$	$-6.00 \times 10{-15}$	0.40	0.375	$-7.97 \times 10{-07}$	$2.00 \times 10{-15}$
CCDC141	328	57.80	0.437	$3.40 \times 10{-04}$	$-4.80 \times 10{-14}$	52.61	0.527	$-1.10 \times 10{-04}$	$1.70 \times 10{-14}$
CD96	110	3.22	0.891	$5.30 \times 10{-05}$	$-7.99 \times 10{-15}$	58.61	0.553	$-1.39 \times 10{-04}$	$3.00 \times 10{-15}$
CR1L	182	24.84	0.477	$1.50 \times 10{-04}$	$-2.30 \times 10{-14}$	26.28	0.407	$-5.01 \times 10{-05}$	$6.00 \times 10{-15}$
DSPP	55	0.84	0.837	$-3.40 \times 10{-06}$	$6.00 \times 10{-15}$	0.38	0.889	$-6.03 \times 10{-07}$	$9.99 \times 10{-16}$
ELOVL1	5	0.57	0.417	$2.60 \times 10{-06}$	$-3.00 \times 10{-15}$	0.89	0.429	$-1.80 \times 10{-06}$	$6.99 \times 10{-15}$
ELOVL2	67	14.92	0.203	$9.01 \times 10{-05}$	$-3.90 \times 10{-14}$	13.84	0.251	$-3.00 \times 10{-05}$	$1.30 \times 10{-14}$
ELOVL3	3	0.40	0.521	$2.50 \times 10{-06}$	$-1.40 \times 10{-14}$	0.52	0.390	$-1.30 \times 10{-06}$	$6.99 \times 10{-15}$
EPHA2	51	13.91	0.056	$8.01 \times 10{-05}$	$-1.13 \times 10{-14}$	15.07	0.041	$-3.00 \times 10{-05}$	$2.60 \times 10{-15}$
FADS1	12	1.85	0.569	$1.10 \times 10{-05}$	$-2.20 \times 10{-14}$	1.05	0.525	$-2.00 \times 10{-06}$	$9.99 \times 10{-16}$
FADS2	89	16.35	0.276	$1.00 \times 10{-04}$	$-2.80 \times 10{-14}$	18.22	0.171	$-4.01 \times 10{-05}$	$6.99 \times 10{-15}$
FADS3	10	3.42	0.116	$2.10 \times 10{-05}$	$-9.99 \times 10{-15}$	3.04	0.073	$-6.97 \times 10{-06}$	$3.79 \times 10{-15}$
FFAR4	41	16.73	0.511	$7.01 \times 10{-05}$	$-2.00 \times 10{-15}$	6.00	0.562	$-1.20 \times 10{-05}$	$7.99 \times 10{-15}$
FILNC1	98	21.16	0.173	$1.30 \times 10{-04}$	$-2.50 \times 10{-14}$	20.54	0.165	$-5.01 \times 10{-05}$	$7.99 \times 10{-15}$
LRRC3B	176	966.21	0.119	$6.90 \times 10{-03}$	$-6.99 \times 10{-15}$	1270.36	0.068	$-2.99 \times 10{-03}$	$1.80 \times 10{-15}$
ME1	316	58.06	0.450	$3.50 \times 10{-04}$	$-3.50 \times 10{-14}$	54.61	0.477	$-1.11 \times 10{-04}$	$1.40 \times 10{-14}$
POLR1D	54	26.39	0.062	$1.40 \times 10{-04}$	$-4.91 \times 10{-15}$	59.55	0.091	$-1.20 \times 10{-04}$	$2.19 \times 10{-15}$
PTGS2	10	40.39	0.023	$2.40 \times 10{-04}$	$-2.30 \times 10{-15}$	44.28	0.016	$-8.96 \times 10{-05}$	$5.00 \times 10{-16}$
SCFD1	270	90.74	0.061	$5.19 \times 10{-04}$	$-7.00 \times 10{-15}$	75.25	0.106	$-1.50 \times 10{-04}$	$4.00 \times 10{-15}$
TMEM258	3	0.19	0.624	$1.70 \times 10{-06}$	$-9.99 \times 10{-15}$	0.70	0.339	$-1.19 \times 10{-06}$	$9.99 \times 10{-16}$
WDR70	455	18.01	0.838	$1.30 \times 10{-04}$	$-3.00 \times 10{-15}$	70.15	0.720	$-1.40 \times 10{-04}$	$2.70 \times 10{-14}$

[‡] Bias was calculated as the difference of the PCSS estimate and the individual patient-level data (IPD) estimate.

B S_J DERIVATION

Given a null model $y = Z\alpha + \varepsilon$ and least squares estimate $\hat{y} = Z\hat{\alpha}$, the contribution to the score statistic from any one SNV,

g_j, can be expressed as (assuming that tshe null model includes an intercept term so the residuals, $y_i - \hat{y}_i$, sum to zero):

$$S_j = g_j^T (y - \hat{y}) / \hat{\sigma}$$
$$= \sum_{i=1}^{n} g_{ij} (y_i - \hat{y}) / \hat{\sigma}$$
$$= \left((n-1) \operatorname{cov}(g_j, y - \hat{y}) + \bar{g}_j \sum_{i=1}^{n} y_i - \hat{y}_i \right) / \hat{\sigma}$$
$$= (n-1) \operatorname{cov}(g_j, y - \hat{y}) / \hat{\sigma}$$
$$= (n-1) \left(\operatorname{cov}(g_j, y) - \operatorname{cov}(g_j, \hat{y}) \right) / \hat{\sigma}$$
$$= (n-1) \left(\operatorname{cov}(g_j, y) - \sum_{l=1}^{c} \hat{\alpha}_l \operatorname{cov}(g_j, z_l) \right) / \hat{\sigma}$$

C REAL DATA RESULTS