Rare-variant association tests in longitudinal studies, with an application to the Multi-Ethnic Study of Atherosclerosis (MESA)

Running title: Rare variant association tests in longitudinal studies

Zihuai He[1], Seunggeun Lee[2], Min Zhang[2], Jennifer A. Smith[3], Xiuqing Guo[4], Walter Palmas[5], Sharon L.R. Kardia[3], Iuliana Ionita-Laza[1], Bhramar Mukherjee[2*]

[1]Department of Biostatistics, Columbia University, New York, NY 10032

[2]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109

[3]Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109

[4]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA 90509

[5]Department of Medicine, Columbia University, New York, NY 10032

[*]Correspondence:

bhramar@umich.edu, (734) 764-6544

Department of Biostatistics

School of Public Health

University of Michigan

1420 Washington Heights

Ann Arbor, MI 48109-2029

Abstract

Over the past few years, an increasing number of studies have identified rare variants that contribute to trait heritability. Due to the extreme rarity of some individual variants, gene-based association tests

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/gepi.22081.

have been proposed to aggregate the genetic variants within a gene, pathway or specific genomic region as opposed to a one-at-a-time single variant analysis. In addition, in longitudinal studies, statistical power to detect disease susceptibility rare variants can be improved through jointly testing repeatedly measured outcomes, which better describes the temporal development of the trait of interest. However, usual sandwich/model-based inference for sequencing studies with longitudinal outcomes and rare variants can produce deflated/inflated type I error rate without further corrections. In this paper, we develop a group of tests for rare-variant association based on outcomes with repeated measures. We propose new perturbation methods such that the type I error rate of the new tests is not only robust to misspecification of within-subject correlation, but also significantly improved for variants with extreme rarity in a study with small or moderate sample size. Through extensive simulation studies, we illustrate that substantially higher power can be achieved by utilizing longitudinal outcomes and our proposed finite sample adjustment. We illustrate our methods using data from the Multi-Ethnic Study of Atherosclerosis for exploring association of repeated measures of blood pressure with rare and common variants based on exome sequencing data on 6361 individuals.

Keywords: longitudinal studies; Multi-Ethnic Study of Atherosclerosis; sequence-based association tests.

Introduction

Although substantial progress has been made in the discovery of common variants associated with complex traits, much of the genetic heritability still remains unexplained. An increasing number of studies have now considered rare variants to explain additional heritability. Various gene-based association tests have been developed for cross-sectional data to aggregate the rare variants in a gene as opposed to a one-at-a-time single variant analysis (Lee, Abecasis, Boehnke and Lin, 2014). Among them, burden tests collapse multiple genetic variants into a single genetic score, then test the

association between the score and an outcome (Li and Leal, 2008; Madsen and Browning, 2009). They are especially powerful under the assumption that all variants in the set are associated with the outcome in the same direction, but violation of this assumption can lead to a loss of power. Variance component tests or dispersion tests test for the association by evaluating the variation of genetic effects for a group of variants (Neale, et al. 2011; Wu, et al. 2011; Li, et al. 2014). In contrast to burden tests, they are robust to genomic regions in which variants have both positive and negative effects. Since the underlying scenario is unknown in large-scale agnostic exploration of the genome, several methods have been proposed to combine these two methods, including the Fisher's combined probability test and the optimal unified sequence kernel association test (SKAT-O), which uses data to adaptively combine sequence kernel association test (SKAT) and burden test statistics (Derkach, Lawless and Sun, 2013; Sun, Zheng and Hsu, 2013; Lee, Wu and Lin, 2012).

To test genetic association in longitudinal studies, investigators often take a simple approach of collapsing the repeated measurements into a single value (average, baseline or last observation carried forward) and hence the method is not able to harness the power of the complete information that is contained in the longitudinal trajectory (He, et al. 2015; Ware, et al. 2016). For one-at-a-time single variant analysis, one can also apply the standard methods available for correlated outcome models to better utilize the longitudinal data, such as mixed effect models or generalized estimating equations (GEE) (Fan, et al. 2012; Furlotte, Eskin and Eyheramendy, 2012; Liang and Zeger, 1986). These methods are primarily proposed for modeling and testing a modest number of variants compared to the number of subjects. For gene-based analysis, several groups have recently extended the burden and dispersion tests to longitudinal studies through mixed effect models or generalized estimating equations (He, et al. 2015; Wang, Xu, Zhang, Wu and Wang, 2017). The mixed effect approaches are model-based, which can lead to inflated type I error rate when the within subject correlation is misspecified. Wang et al. (2017) proposed a practical strategy to reduce the inflation by combining multiple working correlation structures. Although it can work well for various scenarios, the type I error rate is not theoretically justified to be robust. The gene-based tests using GEE is robust to the misspecification of within-subject correlation, but the use of large-sample-based inference can

produce inaccurate type I errors rates when sample sizes are small or the minor allele frequencies are very low. So far, there is no extension of SKAT-O type tests to outcomes with repeated measures that can adaptively combine burden and dispersion tests. Development of such tests remains the central purpose of the current paper.

We propose a group of generalized score type tests for rare-variant association between a set of genetic variants and a phenotype measured repeatedly during the course of an observational study. The proposed tests include burden, dispersion, and an adaptively combined test of those two based on Fisher's and Minimum P-value approaches. They are GEE based tests that are robust to the misspecification of within-subject correlation. We also develop a perturbation method to address the difficulty of applying GEE based inference to rare variants to offer better small sample inference properties. The performance of the methods is evaluated through simulation studies and illustrated using repeated measures data on blood pressure measures on 6361 individuals from the Multi-Ethnic Study of Atherosclerosis (MESA) (Bild, et al. 2002).

Methods

Notations and model:

Assume that we have a study population of m subjects and the i-th subject has n_i observations, $n = \sum_i n_i$. Let $Y_{i,j}$ be the quantitative outcome for the j-th observation of the i-th subject; $X_{i,j} = \left(X_{i,j}^1, \dots, X_{i,j}^p\right)^T$ be the p covariates which can include time (time-varying covariate), gender, body mass index (BMI) (baseline covariate), etc.; $G_i = \left(G_i^1, \dots, G_i^q\right)^T$ be the q time-invariant genetic variants sequenced in a region. We are interested in testing the association between $Y_{i,j}$ and G_i, adjusting for covariates $X_{i,j}$. The fixed effect model is given by

$$\mu_{i,j} = E\left(Y_{i,j}|X_{i,j}, G_i\right) = X_{i,j}^T\beta + G_i^T\gamma,$$

where β and γ are the coefficients for the covariates and genetic variants respectively. For simplicity, we define $Y_i = \left(Y_{i,1}, \dots, Y_{i,n_i}\right)^T$ as a vector of all observations on subject i; X_i, G_i are defined as the

matrix form of covariates and genetic variants, i.e. $X_i = \left(X_{i,1}, \ldots, X_{i,n_i}\right)^T$, $\tilde{G}_i = (G_i, \ldots, G_i)^T$. We note that G_i is repeated n_i times because genotype is time invariant. The matrix representation is given by

$$\mu_i = E(Y_i|X_i, G_i) = X_i\beta + \tilde{G}_i\gamma.$$

The above model gives a parameterization for testing the association between the genetic variants and response variable. When $\gamma^{q\times 1} = 0$, there is no joint association. Thus we consider the q dimensional hypothesis:

$$H_0: \gamma = 0 \; vs. \; H_1: \gamma \neq 0.$$

Generalized score type test:

To construct a simultaneous test for $H_0: \gamma = 0$, the classical approach is a q-degree of freedom likelihood ratio/Wald/score test. The power of such tests tends to diminish rapidly when the dimensionality q is large, which can be a common scenario when the sequenced region consists of hundreds of variants. Alternatively, we propose a score type test statistic by simply assembling the score statistics of the above fixed effect model. We consider the $q \times 1$ score vector with respect to γ,

$$S_\gamma(\beta, \zeta, \gamma) = \sum_{i=1}^{m} S_{\gamma,i}(\beta, \zeta, \gamma) = \sum_{i=1}^{m} \tilde{G}_i^T V_i^{-1}(\zeta)(Y_i - \mu_i)$$

where $V_i^{-1}(\zeta)$ is the working covariance matrix of subject i; ζ is a vector of parameters specifying the working covariance. Let $S_\gamma^k(\beta, \zeta, \gamma)$ be the k-th element of $S_\gamma(\beta, \zeta, \gamma)$. We define two test statistics as

$$Q_1 = \frac{1}{m}\sum_{k=1}^{q} w_k^2\left[S_\gamma^k(\hat{\beta}, \hat{\zeta}, 0)\right]^2, \qquad Q_2 = \frac{1}{m}\left[\sum_{k=1}^{q} w_k S_\gamma^k(\hat{\beta}, \hat{\zeta}, 0)\right]^2,$$

which are two different types of aggregation of the single variant score statistics; w_k is threshold indicator/weight for variant k. Specifically, we use Beta(1,25) distribution to up-weight variants with lower minor allele frequency, similar to SKAT; $\hat{\beta}$ and $\hat{\zeta}$ are estimated under H_0 by GEE. The form of

Q_1 is close to the dispersion tests, and Q_2 belongs to the class of burden tests.[5] Similar to SKAT-O, we can combine the two test statistics by

$$Q_\rho = (1-\rho)Q_1 + \rho Q_2, \qquad \rho \in [0,1].$$

Distribution of Q_ρ and perturbation method when ρ is fixed:

For a fixed ρ, we show in the supplementary materials that Q_ρ follows a weighted sum of chi-square distributions under H_0, where the mixture weights can be estimated by sandwich estimation as in GEE. However, the large-sample-based GEE inference can produce inaccurate type I errors rates when the sample size is small or the minor allele frequencies are very low. To address this, we use a perturbation method to approximate the distribution of Q_ρ (Wang, Lee, Zhu, Redline and Lin, 2013). We first generate a total of B samples of perturbed scores $\tilde{S}_b = \sum_{i=1}^m \tilde{G}_i^T V_i^{-1}(\hat{\zeta})(Y_i - \hat{\mu}_i)\, r_{b,i}$ and calculate the perturbed test statistic $\tilde{Q}_{\rho,b}$, where $b = 1, \dots, B$; r_i is a random variable sampled from the Rademacher distribution (a discrete distribution with equal chance of being -1 and 1). Then we calculate the sample mean $\hat{\mu}_{\rho,B}$, variance $\hat{\sigma}^2_{\rho,B}$ and kurtosis $\hat{\kappa}_{\rho,B} = \hat{\psi}_{\rho,B,4}/\left(\hat{\sigma}^2_{\rho,B}\right)^2 - 3$ of the perturbed test statistic, where $\hat{\psi}_{\rho,B,4}$ is the sample fourth central moments. To obtain cumulative distribution function of Q_ρ, we use the moment matching approximation with estimated ($\hat{\mu}_{\rho,B}$, $\hat{\sigma}^2_{\rho,B}$, $\hat{\kappa}_{\rho,B}$),

$$P_{H_0}(Q_\rho < x) = F\left((x - \hat{\mu}_{\rho,B})\sqrt{2df}/\hat{\sigma}_{\rho,B} + df \middle| \chi^2_{df}\right),$$

where $F(\cdot\,|\chi^2_{df})$ is the distribution function of χ^2_{df} and $df = 12/\hat{\kappa}_{\rho,B}$.

Adaptively combined test:

When $\rho = 1$, the test is more powerful under the assumption that all variants in the set are associated with the outcome with the same direction, but violation of this assumption can lead to a loss of power. When $\rho = 0$, the test is robust to genome regions in which variants have both positive and negative effects. Since both scenarios can arise and the optimal ρ is unknown, we adaptively combine Q_1 and Q_2. Let $\rho_1, \dots, \rho_L$ be L fixed values in the interval $[0,1]$, and $p_1, \dots, p_L$ be the p-values of tests based on $Q_{\rho_1}, \dots, Q_{\rho_L}$. We define two combined test statistic as

- Fisher's statistic: $T_{Fisher} = \sum_{k=1}^{L} -2\log p_k$

- MinP statistic: $T_{MinP} = \min(p_1, \dots, p_L)$.

Since the p-values $p_1, \dots, p_L$ are not independent, it poses a challenge to derive the distribution of T_{Fisher} and T_{MinP}. We propose a resampling method to calculate the p-value as follows.

- Fisher's statistic: we calculate the p-values $(p_1^b, p_2^b, \dots, p_L^b)$ using the aforementioned perturbation method with respect to $\rho_1, \dots, \rho_L$, and calculate the unified test statistic

$$T_{Fisher,b} = \sum_{k=1}^{L} -2\log p_k^b, b = 1, \dots, B.$$

Note that each $-2\log p_k^b$ follows a chi-square distribution with degree of freedom one. We approximate the distribution of T_{Fisher} by using the moment matching approximation. We estimate the moments of T_{Fisher} using the sample mean $\hat{\mu}_{Fisher,B}$, variance $\hat{\sigma}^2_{Fisher,B}$ and kurtosis $\hat{\kappa}_{Fisher,B} = \hat{\psi}_{Fisher,B,4}/\left(\hat{\sigma}^2_{Fisher,B}\right)^2 - 3$ of the resampling based statistic, where $\hat{\psi}_{B,4}$ is the sample fourth central moments. The cumulative distribution function of T_{Fisher} is

$$P_{H_0}(T_{Fisher} < x) = F\left(\left(x - \hat{\mu}_{Fisher,B}\right)\sqrt{2df}/\hat{\sigma}_{Fisher,B} + df \middle| \chi^2_{df}\right),$$

where $F(\cdot\,|\chi^2_{df})$ is the distribution function of χ^2_{df} and $df = 12/\hat{\kappa}_{Fisher,B}$.[11]

- MinP statistic: we define $\delta = \left(\Phi^{-1}(p_1), \dots, \Phi^{-1}(p_L)\right)^T$. The marginal distribution of $\Phi^{-1}(p_k)$ follows a normal distribution with mean 0 and variance 1 under H_0. We approximate their joint distribution by a multivariate normal distribution, i.e. $\delta \sim N(0, D)$. To estimate D, we calculate the p-values $(p_1^b, p_2^b, \dots, p_L^b)$ using the aforementioned perturbation method with respect to $\rho_1, \dots, \rho_L$, and define

$$\delta_b = \left(\Phi^{-1}\left(p_1^b\right), \dots, \Phi^{-1}\left(p_L^b\right)\right)^T, b = 1, \dots, B.$$

Then we estimate D by $\widehat{D} = \frac{1}{B}\sum_{b=1}^{B} \delta_b \delta_b^T$. The calibrated p-value can be calculated by

$$P_{H_0}(T_{MinP} < x) = 1 - P(\delta > [\Phi^{-1}(x), \dots, \Phi^{-1}(x)]^T)$$

It is worth noting that these tests use a similar strategy as SKAT-O. Namely, the MinP test defines the test statistic same as SKAT-O, but use an alternative procedure for a robust inference in longitudinal studies. The Fisher's statistic is an alternative strategy to combine p-values, which is nearly comparable to the MinP statistic but slightly more powerful when the significance is homogeneous for multiple $p_1, \ldots, p_L \in [0,1]$. Since the focus of this paper is on utilizing longitudinal outcomes, we restrict $\rho = 0,1$. We note that the power difference between this simplified test and a full spectrum of ρ is often negligible.

The theoretical aspects of these score based tests for rare-variants have been discussed in many existing papers for cross-sectional data (Wu, et al. 2011; Lee, et al. 2012; Derkach, et al. 2013; Lee, et al. 2014). The novelty of our proposed tests lies in their robustness to within-subject correlation and the small sample adjustment. The proposed perturbation procedure to calculate the analytical p-values of MinP and Fisher's test statistics is also new. The algorithm efficiently estimates moments of the test statistics, such that extreme p-values at genome-wide level can be computed. Traditional resampling procedure usually does not guarantee the robustness to within-subject correlation, and requires large number of replicates to achieve the correct p-values at genome-wide level.

Numerical Simulations

Since there is no adaptively combined test developed for rare-variant association in longitudinal studies, we mainly compared the performance of the proposed methods using longitudinal data with SKAT-O using the average/baseline value of the repeated measures. We also considered alternative methods for longitudinal studies, such as dispersion/burden test using sandwich/model-based inference. The tests using model-based inference assume a compound-symmetry/auto-regressive within-subject correlation structure in a mixed model, but violation of this assumption can lead to inflated type I error rate. We note that this model-based inference is similar to the longitudinal sequence kernel association test (LSKAT) and longitudinal burden test (LBT) proposed by Wang et al. (2017), where they assume the within-subject correlation to be a mixture of compound-symmetry

and auto-regressive structure. Their method practically reduces the type I error inflation and have equivalent power as model-based inference when the within-subject correlation is correctly specified, although the type I error rate is not theoretically justified to be robust.

Sequencing data were generated from 10,000 haplotypes over 200k regions (3845 genetic variants) using the calibration coalescent model (COSI), with mimicking the linkage disequilibrium (LD) structure of European ancestry samples (Schaffner, et al. 2005). The simulation studies focus on the variants with minor allele frequency (MAF) <0.05. We randomly selected 3kb regions (38.3 MAF<0.05 variants on average) and form a sample with 500, 1000, 2000 and 5000 individuals for each replicate. We first simulated the complete data with four repeated measurements, and then applied a missingness indicator with 4% fixed drop-out rate at each exam assuming data missing completely at random.

Type I error simulations

To examine the type I error rate of the proposed methods, we simulated continuous phenotypes from the following model

$$Y_{ij} = b_i + \beta_{time} t_{ij} + 0.5X_{1,i} + 0.5X_{2,ij} + \epsilon_{ij},$$

where $t_{ij} = 2 \times (j-1)$ (0, 2, 4, 6 standing for years since the initiation of the study), $\beta_{time} = 3$; $X_{1,i}$ and $X_{2,ij}$ are time invariant and time-varying covariates respectively; $b_i \sim N(0,1)$, $\epsilon_{ij} \sim N(0,1)$ and they are all independent (estimated within-subject correlation ~ 0.46); $j = 1,2,3,4$. The simulation setting is similar to Lee et al. (2012). We simulated 10^6 replicates to examine the type I error rate at $\alpha = 0.01$, 0.001 and 0.0001 as the sample size varies from 500 to 5000. We also examine the type I error rate when the within-subject correlation follows an auto-regressive model of order 1. Results are presented in Table 1 and 2.

Empirical power simulations

To evaluate the power, the continuous phenotype was simulated from

$$Y_{ij} = b_i + \beta_{time} t_{ij} + 0.5X_{1,i} + 0.5X_{2,ij} + \beta_1 g_1 + \cdots + \beta_s g_s + \epsilon_{ij},$$

where $(g_1, \ldots, g_s)$ were selected causal variants; b_i, t_{ij}, β_{time}, $X_{1,i}$ and $X_{2,ij}$ are defined same as the type I error simulation. Similar to Lee et al. (2012), we considered simulations in which 10%, 20% or 50% of variants were causal, and set $\beta_k = c|\log_{10} m_k|$, where m_j is the MAF if the j-th variant. We set $c = 0.8$, 0.4 and 0.2 when 10%, 20% and 50% of the rare variants were causal to compensate for the increased number of causal variants. We allow the sample size to vary as $m = 500, 1000$, 2000 and 5000. The power was estimated as the proportion of p-values less than $\alpha = 0.001$. Results are presented in Table 3. We additionally present results when 20%/50% of causal variances have negative βs in Table S1 and S2.

To evaluate the use of longitudinal information, we evaluate a spectrum of β_{time} from 0 to 3, reflecting scenarios from no time effect to a strong time effect; $b_i \sim N(0, \sigma^2)$ where σ^2 =0.25, 1, (estimated within-subject correlation ~ 0.2, 0.5 respectively); 20% variants were causal, and set $\beta_k = 0.4|\log_{10} m_k|$, where m_j is the MAF if the j-th variant. The power was estimated as the proportion of p-values less than $\alpha = 0.001$. Results are presented in Table 4.

Results

Simulation of Type I Error Rate

The empirical type I error rates are presented in Table 1 and 2 for $\alpha = 0.01, 0.001$ and 0.0001 and sample sizes 500, 1000, 2000 and 5000. The results show that the tests using sandwich-based inference as in usual GEE suffer from significantly conservative type I error rate, especially for small sample size (sandwich dispersion: 0.0001, sandwich burden: 0.0005 at $\alpha = 0.001$ when $m = 500$ and the correlation structure is compound symmetry). In addition, the tests using model-based inference suffer from significantly inflated type I error rate when the working correlation structure is misspecified (sandwich dispersion: 0.006, sandwich burden: 0.0033 at $\alpha = 0.001$ when $m = 500$, where the working correlation is compound symmetry but the underlying within-subject correlation is auto-regressive). We note that directly applying SKAT-O to the average of repeated measurements

leads to slightly inflated type I error rate because both the mean model and homogenous assumption are not valid due to missing data.

While sandwich/model-based inference can lead to either deflated/inflated type I error rates, the type I error rates of the proposed tests based on the perturbation approach are preserved for all α levels and sample sizes. The type I error rates are well controlled even if the working correlation is misspecified (Table 2, the working correlation is compound symmetry but the underlying within-subject correlation is auto-regressive). In the genome-wide analysis of the MESA blood pressure measures in association with the exome-chip data, the QQ-plots show that the distribution of p-values generally follows a global null (Figures S1-S4). The results illustrate that the proposed tests are valid methods for a genome-wide analysis of rare variants in longitudinal studies.

Power Gain from Utilizing Longitudinal Information

We compare the proposed tests with SKAT-O using the average/baseline value of the repeated measures (Table 3). The proposed adaptively combined tests using longitudinal data have higher power than SKAT-O using average/baseline of the repeated measurements (e.g. Fisher: 0.61; MinP: 0.59; SKAT-O-average: 0.19; SKAT-O-baseline: 0.50 when $m = 2000$ and causal proportion is 0.2). We also observed that the adaptively combined test generally achieve the maximum power of dispersion and burden tests, which is a desired property as the underlying causal scenario (causal proportion, direction of effects) is usually unknown. Additional simulation studies with bi-direction effects are included in Table S1 and S2. The results show similar pattern.

To further investigate the improved power due to using longitudinal outcomes, we evaluated the methods over a spectrum of time effect, from no effect to a strong effect. The results are summarized in Table 4. We observed that power of SKAT-O using average of repeated measurements substantially decreases as the time effect increases. We additionally evaluated the methods when complete data is simulated and observed there is no such power loss (data not shown). This shows that the misspecified mean model and heterozygosity in variance due to missing data not only cause inflated type I error rate (Table 1), but also reduce power. Since missing data commonly exists in

longitudinal studies over a period of time, directly applying methods for cross-sectional study to the average of longitudinal outcomes is less than optimal.

Application to The Multi-Ethnic Study of Atherosclerosis (MESA)

We illustrate the use of the method by applying it to exome-chip data and blood pressure measures in MESA. MESA is a collaborative longitudinal study initiated in July 2000 to investigate the prevalence, correlates, and progression of subclinical cardiovascular disease.[16] From 2000 to 2007, four examinations of blood pressure were conducted over 18- to 24- month periods. A total of 6361 subjects consisting of 2526 European Americans (EUR), 1611 African Americans (AFA), 1449 Hispanics (HIS) and 775 Asian of Chinese descent (CHN) with genome-wide genotype data, systolic blood pressure (sBP) and diastolic blood pressure (dBP) outcomes were considered in the current analysis. We adjusted the actual blood pressures for participants taking antihypertensive medications using the standard procedure of adding 10 mmHg to systolic blood pressure and 5 mmHg to diastolic blood pressure (Cui, Hopper and Harrap, 2003). Genetic variants were genotyped using the Illumina HumanExome BeadChip 12-v1. We annotated variants to genes using Annovar (Wang, Li and Hakonarson, 2010). We conducted ethnicity-specific analysis of the association between systolic and diastolic blood pressures and genetic variants adjusting for age, gender, BMI and the leading four ethnicity-specific genetic principal components (PCs). Ethnicity-specific PCs are estimated using genome-wide genotyping data from the Affymetrix HumanGenome SNP Array 6.0. Then the ethnicity-specific p-values were combined using Fisher's method for a meta-analysis (Fisher, 1925). We present the top three genes for systolic and diastolic blood pressures in Table S3. We also present the results for 18 genes around index SNPs that were significant (p-value $< 10^{-9}$) in the International Consortium for Blood Pressure genome-wide association studies in Table S4 (International Consortium for Blood Pressure Genome-Wide Association Studies, 2011).

Utilizing the longitudinal trajectory, we identified a protein-coding gene, *ZNF473* that exhibits suggestive association with systolic blood pressure in Hispanics (p-value = 2.4×10^{-6} for Fisher's test, 1.8×10^{-6} for MinP test). The corresponding p-values for diastolic blood pressure are

also suggestive (p-value = 0.0016 for Fisher's test, 0.0013 for MinP test). The significance is more pronounced for the burden test than the dispersion test (9.2×10^{-7} vs. 8.1×10^{-6} for systolic blood pressure, 7.1×10^{-4} vs. 0.0045 for diastolic blood pressure). We present the detailed results in Table 5. The identified gene *ZNF473* encodes a member of the Krueppel C2H2-type zinc-finger family of proteins, a component of the U7 snRNP complex. The encoded protein plays a role in histone 3'-end pre-mRNA processing and may be required for cell cycle progression to S phase. Bone mineral density might be correlated with the expression level and methylation status of this gene (O'Leary, et al. 2015). We additionally perform the single SNP analysis in gene *ZNF473*. We present the results in Table S5. We observed that *exm1493401* at position 50545025 (hg19) is the SNP that exhibits the smallest p-value (p-value = 2.6×10^{-5}) associated with systolic blood pressure, and there are multiple SNPs highly correlated to *exm1493401*. Another suggestive SNP is *exm1493479* at 50549462 (p-value = 0.0013). We also present the genome-wide meta-analysis p-values and ethnicity specific p-values in Hispanics in Figures S1-S4. Although the QQ-plots do not show inflation due to population stratification, we note that the sample size of Hispanics (1449 subjects with several repeated measurements per subject) is relatively small for the identification of rare-variant association and the association was not observed in MESA Europeans, African-Americans or Chinese. Therefore, future replication studies with a larger sample size will be needed to verify this association.

MESA samples are collected from six study sites (Table S6). Since the association presented in this paper is identified in Hispanics, we characterize the amount of admixture due to European, African and Native American (NA) in MESA Hispanic samples. The MESA Hispanic samples consist of individuals from Central America, Cuba, Dominican Republic, Mexico, Puerto Rico and South America. The amount of admixture in each group has been extensively evaluated by Manichaikul et al., 2012. Since the focus of this paper is on the development of new association tests for longitudinal study, we directly cite the existing results in Table S7. We also calculate the amount of admixture within each study site, and present the results in Table S8. We present a plot of the PCs versus the self-reported Hispanic origins in Figure S5. Then we identify the two subpopulations (Mexicans and Caribbean). The classification is mainly based on self-reported origin, with reclassification of some

individuals based on the leading four PCs. We present the results in Figure S6 and Table S9. We observe that the resulting clusters of ancestry showed good agreement with self-reported country/region of origin. In addition to adjusting for the top four ethnicity-specific PCs, we further conducted sensitivity analysis of *ZNF473* to evaluate how different adjustments of population stratification affect the results: a. Since Hispanics are an admixed population with European, African and Native American ancestries, we applied LAMP to the variants that can be matched with reference samples from HapMap3 and HGDP in a 10Mb window around gene ZNF473. African (AFR) and European (EUR) samples are from HapMap3 (African ancestry: YRI, ASE and LWK, European ancestry: CEU and TSI); Native American samples are from HGDP (Colombian, Karitiana, Maya, Pima, and Surui). We present the LAMP results in Table S10. We observed that overall African, European and Native American ancestries account for 29.7%, 52.5%, 17.8% of the local ancestry in this region, respectively. These local ancestry effects are then included as covariates in the sensitivity analysis; b. We included self-reported Hispanic origins as covariates (Figure S5); c. We included classification of Hispanics (Mexican vs Caribbean) as a covariate. The classification is mainly based on self reported origin, and we reclassified some individuals based on the leading four PCs (Figure S6, Figure 2 in this response letter). d. We conducted stratified analysis within Mexicans/Caribbean. We present the results in Table S11. The results show that the significance remains similarly.

Discussion

In this paper, we propose a group of rare-variant association tests that can utilize the longitudinal trajectory of outcomes. The new tests include burden, dispersion, and an adaptively combined test of those two based on Fisher's and Minimum P-value approaches. The tests can incorporate time varying covariates as fixed effects and are robust to misspecification of the within subject correlation structure. Using extensive simulation studies, we illustrate that substantial power gain can be achieved by jointly modeling the repeated measurements and using the complete information, compared to simple approaches of collapsing the repeated measurements into a single average/baseline value. The analysis of blood pressure measures of 6361 individuals in MESA in association with exome

sequencing data further illustrates the use of the methods and identified a protein-coding gene, *ZNF473*, suggestively associated with systolic blood pressure in Hispanics.

One attractive feature of the proposed tests is that they are theoretically robust to misspecification of within-subject correlation by using a GEE based inference with a novel perturbation method. Unlike model-based inference that can lead to inflated/deflated type I error rate when the working correlation structure is misspecified, the proposed tests have much improved type I error control. We also developed a novel perturbation method to address the difficulty of applying robust variance inference to rare variants, especially when the sample size is relatively small.

The ability to adaptively combine dispersion and burden tests in longitudinal studies, and obtain an analytical p-value is another attractive feature. Unlike usual permutation/perturbation based methods to combine multiple p-values or statistics, the proposed method only uses the resampling technic to estimate moments of the test statistics so that the p-value is still calculated analytically, which enables a direct calculation of p-value at genome-wide significance level (2.5×10^{-6}). This feature drastically reduced the required number of resampling replicates. In addition, we only need to sample those resampling replicates once for a genome-wide analysis of approximately 20,000 genes. These features make the proposed methods more suitable to large-scale genome-wide analysis.

We carefully evaluated various factors that may influence the power of gene-based tests in longitudinal studies, namely, magnitude of time effect on the outcome variables, percentage of missing data and strength of within-subject correlation. We observed that association tests using longitudinal trajectory have more pronounced power improvement over tests using average/baseline value of repeated measurements in the presence of larger time effect and missing data. In a longitudinal study like MESA, not only the longitudinal outcomes precisely describe the phenotype progression, the rich information on time varying exposures and their interactions with genotype may also improve the discovery process. However, an analysis using the average of repeated measurements of an exposure will reduce the variation in the exposure and substantially reduce the power. We expect that a potential future extension of the proposed methods towards separately testing

gene-time or gene-environment interaction in longitudinal studies with time dependent covariates may enhance the discovery process.

Supplemental Data Description

Supplemental Data include 6 figures and 11 tables.

Web Resources

SKAT: https://www.hsph.harvard.edu/skat/

Annovar: http://annovar.openbioinformatics.org/en/latest/

LAMP: http://lamp.icsi.berkeley.edu/lamp/

Software Package: the methods are implemented in R package LGEWIS to facilitate the analysis using longitudinal outcomes. The package is available on CRAN: https://cran.r-project.org/web/packages/LGEWIS/index.html.

Acknowledgements

We gratefully acknowledge support from NSF DMS 1406712 and NIH/NIEHS grant ES020811, NIH/NHGRI grant R01HG008773, NIH/NHLBI HL101161, and NIMHHD Grant 2P60MD002249 Center for Integrative Approaches to health Disparities (CIAHD). MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Provision of exome chip genotyping was provided in part by support of NHLBI contract N02-HL-64278 and UCLA CTSI UL1-TR001881, and the S.Calif DRC DK063491. The authors have no conflict of interest to declare.

References

Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Roux, A.V.D., Folsom, A.R., Greenland, P., JacobsJr, D.R., Kronmal, R., Liu, K. and Nelson, J.C. (2002). Multi-ethnic study of atherosclerosis: objectives and design. American Journal of Epidemiology, 156, 871-881.

Cui, J.S., Hopper, J.L., and Harrap, S.B. (2003) Anti-hypertensive treatments obscure familial constributions to blood pressure variation. Hypertension, 41, 207-210.

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

Fan, R., Zhang, Y., Albert, P., Liu, A., Wang, Y., and Xiong, M. (2012). Longitudinal association analysis of quantitative traits. Genetic Epidemiology, doi 10.1002/gepi. 21673.

Fisher, R.A. (1925). Statistical methods for research workers. Genesis Publishing Pvt Ltd.

Furlotte, N., Eskin, E., and Eyheramendy, S. (2012). Genome-wide association mapping with longitudinal data. Genetic Epidemiology, 36, 463–471.

He, Z., Zhang, M., Lee, S., Smith, J.A., Guo, X., Palmas, W., Kardia, S.L., Roux, A.V.D. and Mukherjee, B. (2015). Set-based tests for genetic association in longitudinal studies. Biometrics, 71, 606-615.

International Consortium for Blood Pressure Genome-Wide Association Studies (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature, 478, 103-109.

Lee, S., Abecasis, G. R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. The American Journal of Human Genetics, 95, 5-23.

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

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.

Li, M., He, Z., Zhang, M., Zhan, X., Wei, C., Elston, R. C., & Lu, Q. (2014). A generalized genetic random field method for the genetic association analysis of sequencing data. Genetic Epidemiology, 38, 242-253.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 13-22.

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

Manichaikul, A., Palmas, W., Rodriguez, C.J., Peralta, C.A., Divers, J., Guo, X., Chen, W.M., Wong, Q., Williams, K., Kerr, K.F. and Taylor, K.D., 2012. Population structure of Hispanics in the United States: the multi-ethnic study of atherosclerosis. PLoS Genetics, 8, p.e1002640.

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M. et al. (2011). Testing for an unusual distribution of rare variants. PLoS Genetics, 7, e1001322.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. and Astashyn, A. (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research, p.gkv1189.

Sankararaman, S., Sridhar, S., Kimmel, G. and Halperin, E. (2008). Estimating local ancestry in admixed populations. The American Journal of Human Genetics, 82, 290-303.

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

Sun, J., Zheng, Y., and Hsu, L. (2013). A unified mixed-effects model for rare-variant association in sequencing studies. Genetic Epidemiology, 37, 334-344.

Wang, K., Li, M. and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research, 38, e164-e164.

Wang, X., Lee, S., Zhu, X., Redline, S., and Lin, X. (2013). GEE-based SNP set association test for continuous and discrete traits in family-based association studies. Genetic Epidemiology, 37, 778-786.

Wang, Z., Xu, K., Zhang, X., Wu, X. and Wang, Z. (2017). Longitudinal SNP- set association analysis of quantitative phenotypes. Genetic Epidemiology, 41, pp.81-93.

Ware, E.B., Smith, J.A., Mukherjee, B., Lee, S., Kardia, S.L. and Diez-Roux, A.V. (2016). Applying novel methods for assessing individual-and neighborhood-level social and psychosocial environment interactions with genetic factors in the prediction of depressive symptoms in the multi-ethnic study of atherosclerosis. Behavior Genetics, 46, pp.89-99.

Wu, M. C., Lee, S., Cai, T., Li, Y., 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.

Tables

Table 1. Type I error estimates of the proposed tests based on 1,000,000 replicates. The within-subject correlation structure is compound symmetry. S/M/P-Dispersion/Burden: Dispersion/Burden test using Sandwich/Model-based/Perturbation inference. SKAT-O-average: SKAT-O using the average value of the repeated measures.

Sample Size	Level α	S-Dispersion	S-Burden	M-Dispersion	M-Burden	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average
500	0.01	0.0029	0.0080	0.0098	0.0099	0.0110	0.0107	0.0105	0.0105	0.0170
	0.001	0.0001	0.0005	0.0008	0.0009	0.0010	0.0010	0.0011	0.0010	0.0036
	0.0001	1.31E-06	2.89E-05	6.96E-05	1.04E-04	8.60E-05	9.70E-05	1.19E-04	9.30E-05	8.62E-04
1000	0.01	0.0055	0.0091	0.0097	0.0100	0.0106	0.0107	0.0104	0.0103	0.0144
	0.001	0.0003	0.0007	0.0009	0.0010	0.0011	0.0011	0.0012	0.0010	0.0025
	0.0001	1.46E-05	5.47E-05	6.92E-05	9.84E-05	1.11E-04	9.00E-05	1.23E-04	9.40E-05	5.20E-04
2000	0.01	0.0074	0.0095	0.0096	0.0101	0.0101	0.0104	0.0102	0.0099	0.0127
	0.001	0.0005	0.0008	0.0009	0.0010	0.0011	0.0011	0.0011	0.0011	0.0018
	0.0001	3.92E-05	7.29E-05	7.29E-05	1.06E-04	1.30E-04	9.07E-05	1.27E-04	1.08E-04	3.22E-04
5000	0.01	0.0086	0.0101	0.0096	0.0102	0.0100	0.0102	0.0100	0.0097	0.0118
	0.001	0.0008	0.0013	0.0009	0.0013	0.0011	0.0010	0.0011	0.0011	0.0014
	0.0001	2.12E-05	1.27E-04	2.12E-05	1.27E-04	1.13E-04	1.23E-04	1.20E-04	1.23E-04	2.00E-04

Table 2. Type I error estimates of the proposed tests based on 1,000,000 replicates. The within-subject correlation structure follows an auto-regressive model of order 1. S/M/P-Dispersion/Burden: Dispersion/Burden test using Sandwich/Model-based/Perturbation inference. SKAT-O-average: SKAT-O using the average value of the repeated measures.

Sample Size	Level α	S-Dispersion	S-Burden	M-Dispersion	M-Burden	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average
500	0.01	0.0029	0.0081	0.0373	0.0218	0.0110	0.0110	0.0105	0.0104	0.0174
	0.001	0.0001	0.0006	0.0059	0.0034	0.0010	0.0011	0.0011	0.0010	0.0037
	0.0001	1.02E-06	3.58E-05	8.03E-04	5.60E-04	7.00E-05	1.06E-04	1.10E-04	8.30E-05	9.08E-04
1000	0.01	0.0052	0.0093	0.0376	0.0225	0.0106	0.0107	0.0104	0.0103	0.0144
	0.001	0.0003	0.0007	0.0059	0.0036	0.0011	0.0011	0.0012	0.0010	0.0026
	0.0001	2.05E-05	7.06E-05	8.40E-04	5.83E-04	1.07E-04	1.09E-04	1.32E-04	1.06E-04	5.15E-04
2000	0.01	0.0074	0.0091	0.0376	0.0215	0.0103	0.0105	0.0102	0.0101	0.0127
	0.001	0.0006	0.0009	0.0060	0.0033	0.0012	0.0011	0.0012	0.0011	0.0018
	0.0001	4.44E-05	1.48E-04	9.39E-04	5.84E-04	1.34E-04	9.37E-05	1.35E-04	1.12E-04	3.02E-04
5000	0.01	0.0089	0.0090	0.0373	0.0210	0.0099	0.0105	0.0102	0.0099	0.0116
	0.001	0.0007	0.0009	0.0061	0.0034	0.0011	0.0011	0.0012	0.0011	0.0015
	0.0001	4.96E-05	1.49E-04	7.94E-04	4.96E-04	1.30E-04	1.32E-04	1.37E-04	1.24E-04	1.93E-04

Table 3. Power evaluation when all causal variants have positive effects at $\boldsymbol{\alpha = 0.001}$ based on 1000 replicates. m: sample size. P-Dispersion/Burden: Dispersion/Burden test using the proposed perturbation method. SKAT-O average/baseline: SKAT-O using the average/baseline value of the repeated measures.

Causal Proportion	m	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average	SKAT-O Baseline
0.1	500	0.29	0.15	0.29	0.28	0.07	0.26
0.1	1000	0.42	0.26	0.43	0.42	0.21	0.44
0.1	2000	0.50	0.37	0.49	0.48	0.34	0.47
0.1	5000	0.65	0.59	0.67	0.67	0.49	0.64
0.2	500	0.15	0.13	0.17	0.15	0.03	0.12
0.2	1000	0.36	0.28	0.38	0.36	0.09	0.29
0.2	2000	0.58	0.48	0.61	0.59	0.19	0.50
0.2	5000	0.74	0.70	0.77	0.76	0.48	0.72
0.5	500	0.11	0.17	0.18	0.15	0.03	0.12
0.5	1000	0.26	0.37	0.40	0.36	0.07	0.28
0.5	2000	0.52	0.69	0.71	0.68	0.20	0.53
0.5	5000	0.87	0.93	0.94	0.94	0.50	0.88

Table 4. Power evaluation for the use of longitudinal information at $\boldsymbol{\alpha = 0.001}$ based on 1000 replicates. The sample size is 5000 and 20% genetic variants are causal with all positive effects. There is 4% missing data at each exam. SKAT-O-average/baseline: SKAT-O using the average/baseline value of the repeated measures.

Correlation	Time Effect	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average	SKAT-O Baseline
0.2	0	0.84	0.84	0.88	0.88	0.91	0.79
	1	0.84	0.84	0.89	0.88	0.83	0.79
	2	0.84	0.84	0.88	0.88	0.69	0.79
	3	0.85	0.85	0.89	0.88	0.52	0.79
0.5	0	0.74	0.71	0.78	0.77	0.79	0.72
	1	0.74	0.71	0.77	0.76	0.75	0.72
	2	0.74	0.71	0.78	0.77	0.63	0.72
	3	0.74	0.70	0.77	0.77	0.48	0.72

Table 5. Analysis of gene *ZNF473*, the most significant gene in the genome-wide longitudinal data analysis of MESA exome-chip data. Each cell presents the p-value. P-values below the gene-based genome-wide significance level $\boldsymbol{2.5 \times 10^{-6}}$ are bolded.

		Systolic Blood Pressure				Diastolic Blood Pressure			
	# Variants	Dispersion	Burden	Fisher	MinP	Dispersion	Burden	Fisher	MinP
Longitudinal Measures									
EUR	19	0.6828	0.8087	0.7978	0.8225	0.8281	0.3476	0.5852	0.5135
CHN	11	0.9419	0.5907	0.8855	0.7884	0.7143	0.7122	0.7657	0.8715
AFA	24	0.6313	0.1823	0.3360	0.3082	0.3784	0.3812	0.3883	0.5388
HIS	16	8.1E-06	**9.2E-07**	**2.4E-06**	**1.8E-06**	0.0045	7.1E-04	0.0016	0.0013
Meta	-	0.0014	7.0E-05	3.5E-04	2.4E-04	0.0873	0.0137	0.0378	0.0395
Baseline Measure									
EUR	19	0.3747	0.7303	0.5710	0.5431	0.6836	0.3310	0.5129	0.5101
CHN	11	1.0000	0.6355	1.0000	0.8257	0.6786	0.8258	0.8216	0.8438
AFA	24	0.4689	1.0000	0.7671	0.6704	0.5764	0.5560	0.6229	0.7459
HIS	16	4.2E-04	1.0E-04	1.8E-04	1.9E-04	0.0031	7.3E-04	0.0014	0.0013
Meta	-	0.3747	0.7303	0.5710	0.5431	0.6836	0.3310	0.5129	0.5101

Sample Size	Level α	S-Dispersion	S-Burden	M-Dispersion	M-Burden	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average
500	0.01	0.0029	0.0080	0.0098	0.0099	0.0110	0.0107	0.0105	0.0105	0.0170
	0.001	0.0001	0.0005	0.0008	0.0009	0.0010	0.0010	0.0011	0.0010	0.0036
	0.0001	1.31E-06	2.89E-05	6.96E-05	1.04E-04	8.60E-05	9.70E-05	1.19E-04	9.30E-05	8.62E-04
1000	0.01	0.0055	0.0091	0.0097	0.0100	0.0106	0.0107	0.0104	0.0103	0.0144
	0.001	0.0003	0.0007	0.0009	0.0010	0.0011	0.0011	0.0012	0.0010	0.0025
	0.0001	1.46E-05	5.47E-05	6.92E-05	9.84E-05	1.11E-04	9.00E-05	1.23E-04	9.40E-05	5.20E-04
2000	0.01	0.0074	0.0095	0.0096	0.0101	0.0101	0.0104	0.0102	0.0099	0.0127
	0.001	0.0005	0.0008	0.0009	0.0010	0.0011	0.0011	0.0011	0.0011	0.0018
	0.0001	3.92E-05	7.29E-05	7.29E-05	1.06E-04	1.30E-04	9.07E-05	1.27E-04	1.08E-04	3.22E-04
5000	0.01	0.0086	0.0101	0.0096	0.0102	0.0100	0.0102	0.0100	0.0097	0.0118
	0.001	0.0008	0.0013	0.0009	0.0013	0.0011	0.0010	0.0011	0.0011	0.0014
	0.0001	2.12E-05	1.27E-04	2.12E-05	1.27E-04	1.13E-04	1.23E-04	1.20E-04	1.23E-04	2.00E-04

Sample Size	Level α	S-Dispersion	S-Burden	M-Dispersion	M-Burden	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average
500	0.01	0.0029	0.0081	0.0373	0.0218	0.0110	0.0110	0.0105	0.0104	0.0174
	0.001	0.0001	0.0006	0.0059	0.0034	0.0010	0.0011	0.0011	0.0010	0.0037
	0.0001	1.02E-06	3.58E-05	8.03E-04	5.60E-04	7.00E-05	1.06E-04	1.10E-04	8.30E-05	9.08E-04
1000	0.01	0.0052	0.0093	0.0376	0.0225	0.0106	0.0107	0.0104	0.0103	0.0144
	0.001	0.0003	0.0007	0.0059	0.0036	0.0011	0.0011	0.0012	0.0010	0.0026
	0.0001	2.05E-05	7.06E-05	8.40E-04	5.83E-04	1.07E-04	1.09E-04	1.32E-04	1.06E-04	5.15E-04
2000	0.01	0.0074	0.0091	0.0376	0.0215	0.0103	0.0105	0.0102	0.0101	0.0127
	0.001	0.0006	0.0009	0.0060	0.0033	0.0012	0.0011	0.0012	0.0011	0.0018
	0.0001	4.44E-05	1.48E-04	9.39E-04	5.84E-04	1.34E-04	9.37E-05	1.35E-04	1.12E-04	3.02E-04
5000	0.01	0.0089	0.0090	0.0373	0.0210	0.0099	0.0105	0.0102	0.0099	0.0116
	0.001	0.0007	0.0009	0.0061	0.0034	0.0011	0.0011	0.0012	0.0011	0.0015
	0.0001	4.96E-05	1.49E-04	7.94E-04	4.96E-04	1.30E-04	1.32E-04	1.37E-04	1.24E-04	1.93E-04

Causal Proportion	m	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average	SKAT-O Baseline
0.1	500	0.29	0.15	0.29	0.28	0.07	0.26
0.1	1000	0.42	0.26	0.43	0.42	0.21	0.44
0.1	2000	0.50	0.37	0.49	0.48	0.34	0.47
0.1	5000	0.65	0.59	0.67	0.67	0.49	0.64

0.2	500	0.15	0.13	0.17	0.15	0.03	0.12
0.2	1000	0.36	0.28	0.38	0.36	0.09	0.29
0.2	2000	0.58	0.48	0.61	0.59	0.19	0.50
0.2	5000	0.74	0.70	0.77	0.76	0.48	0.72
0.5	500	0.11	0.17	0.18	0.15	0.03	0.12
0.5	1000	0.26	0.37	0.40	0.36	0.07	0.28
0.5	2000	0.52	0.69	0.71	0.68	0.20	0.53
0.5	5000	0.87	0.93	0.94	0.94	0.50	0.88

Correlation	Time Effect	P-Dispersion	P-Burden	P-Fisher	P-MinP	SKAT-O Average	SKAT-O Baseline
0.2	0	0.84	0.84	0.88	0.88	0.91	0.79
	1	0.84	0.84	0.89	0.88	0.83	0.79
	2	0.84	0.84	0.88	0.88	0.69	0.79
	3	0.85	0.85	0.89	0.88	0.52	0.79
0.5	0	0.74	0.71	0.78	0.77	0.79	0.72
	1	0.74	0.71	0.77	0.76	0.75	0.72
	2	0.74	0.71	0.78	0.77	0.63	0.72
	3	0.74	0.70	0.77	0.77	0.48	0.72

		Systolic Blood Pressure				Diastolic Blood Pressure			
	# Variants	Dispersion	Burden	Fisher	MinP	Dispersion	Burden	Fisher	MinP
Longitudinal Measures									
EUR	19	0.6828	0.8087	0.7978	0.8225	0.8281	0.3476	0.5852	0.5135
CHN	11	0.9419	0.5907	0.8855	0.7884	0.7143	0.7122	0.7657	0.8715
AFA	24	0.6313	0.1823	0.3360	0.3082	0.3784	0.3812	0.3883	0.5388
HIS	16	8.1E-06	**9.2E-07**	**2.4E-06**	**1.8E-06**	0.0045	7.1E-04	0.0016	0.0013
Meta	-	0.0014	7.0E-05	3.5E-04	2.4E-04	0.0873	0.0137	0.0378	0.0395
Baseline Measure									
EUR	19	0.3747	0.7303	0.5710	0.5431	0.6836	0.3310	0.5129	0.5101
CHN	11	1.0000	0.6355	1.0000	0.8257	0.6786	0.8258	0.8216	0.8438
AFA	24	0.4689	1.0000	0.7671	0.6704	0.5764	0.5560	0.6229	0.7459
HIS	16	4.2E-04	1.0E-04	1.8E-04	1.9E-04	0.0031	7.3E-04	0.0014	0.0013
Meta	-	0.3747	0.7303	0.5710	0.5431	0.6836	0.3310	0.5129	0.5101