University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2014 *Paper* 101

Set-based Tests for Genetic Association in Longitudinal Studies

Zihuai He[*] Min Zhang[†] Seunggeun Lee[‡]

Jennifer A. Smith[**] Xiuqing Guo[††] Walter Palmas[‡‡]

Sharon L.R. Kardia[§] Ana V. Diez Roux[¶] Bhramar Mukherjee[‖]

[*]University of Michigan School of Public Health, zihuai@umich.edu

[†]University of Michigan School of Public Health, mzhangst@umich.edu

[‡]University of Michigan School of Public Health

[**]University of Michigan School of Public Health

[††]Columbia University

[‡‡]Columbia University

[§]University of Michigan School of Public Health

[¶]University of Michigan School of Public Health

[‖]University of Michigan School of Public Health, bhramar@umich.edu

http://biostats.bepress.com/umichbiostat/paper101

Set-based Tests for Genetic Association in Longitudinal Studies

Zihuai He, Min Zhang, Seunggeun Lee, Jennifer A. Smith, Xiuqing Guo, Walter Palmas, Sharon L.R. Kardia, Ana V. Diez Roux, and Bhramar Mukherjee

Abstract

Genetic association studies with longitudinal markers of chronic diseases (e.g., blood pressure, body mass index) provide a valuable opportunity to explore how genetic variants affect traits over time by utilizing the full trajectory of longitudinal outcomes. Since these traits are likely influenced by the joint effect of multiple variants in a gene, a joint analysis of these variants considering linkage disequilibrium (LD) may help to explain additional phenotypic variation. In this article, we propose a longitudinal genetic random field model (LGRF), to test the association between a phenotype measured repeatedly during the course of an observational study and a set of genetic variants. Generalized score type tests are developed, which we show are robust to misspefication of within-subject correlation, a feature that is desirable for longitudinal analysis. In addition, a joint test incorporating gene-time interaction is further proposed. Computational advancement is made for scalable implementation of the proposed methods in large-scale genome-wide association studies (GWAS). The proposed methods are evaluated through extensive simulation studies and illustrated using data from the Multi-Ethnic Study of Atherosclerosis (MESA). Our simulation results indicate substantial gain in power using LGRF when compared with two commonly used existing alternatives: (i) single marker tests using longitudinal outcome and (ii) existing gene-based tests using the average value of repeated measurements as the outcome.

BIOMETRICS Manuscript, 1–??
March 2014
DOI: 000

Set-based Tests for Genetic Association in Longitudinal Studies

Zihuai He[1], Min Zhang[1,*], Seunggeun Lee[1], Jennifer A. Smith[2], Xiuqing Guo[3], Walter Palmas[4], Sharon L.R. Kardia[2], Ana V. Diez Roux[2], Bhramar Mukherjee[1,]**

[1]Department of Biostatistics, University of Michigan, Ann Arbor, U.S.A.

[2]Department of Epidemiology, University of Michigan, Ann Arbor, U.S.A.

[3]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, U.S.A.

[4]Department of Medicine, Columbia University, New York, U.S.A.

email: mzhangst@umich.edu

**email:* bhramar@umich.edu

SUMMARY: Genetic association studies with longitudinal markers of chronic diseases (e.g., blood pressure, body mass index) provide a valuable opportunity to explore how genetic variants affect traits over time by utilizing the full trajectory of longitudinal outcomes. Since these traits are likely influenced by the joint effect of multiple variants in a gene, a joint analysis of these variants considering linkage disequilibrium (LD) may help to explain additional phenotypic variation. In this article, we propose a longitudinal genetic random field model (LGRF), to test the association between a phenotype measured repeatedly during the course of an observational study and a set of genetic variants. Generalized score type tests are developed, which we show are robust to misspefication of within-subject correlation, a feature that is desirable for longitudinal analysis. In addition, a joint test incorporating gene-time interaction is further proposed. Computational advancement is made for scalable implementation of the proposed methods in large-scale genome-wide association studies (GWAS). The proposed methods are evaluated through extensive simulation studies and illustrated using data from the Multi-Ethnic Study of Atherosclerosis (MESA). Our simulation results indicate substantial gain in power using LGRF when compared with two commonly used existing alternatives: (i) single marker tests using longitudinal outcome and (ii) existing gene-based tests using the average value of repeated measurements as the outcome.

KEY WORDS: Genetic association; Generalized estimating equations; Generalized score test; Longitudinal study; Multi-marker test; Random field.

1. Introduction

Genome-wide association studies (GWAS) have been successful in identifying susceptibility loci for risk factors of chronic diseases. For genetic studies of cardiovascular disease risk factors, such as the Mulit-Ethnic Study of Atherosclerosis (MESA), observations at multiple time points are available for each individual (Bild, et al., 2002). The longitudinal nature of these studies results in more precise phenotypic characterization, enhancing the ability to associate genes or chromosomal regions with the phenotypes and assess gene-time interaction. However, current statistical methods for testing genetic association in longitudinal studies, in the presence of effect heterogeneity across time are limited, even for one single-nucleotide polymorphism (SNP) at a time analysis (Fan, et al., 2012; Furlotte, Eskin and Eyheramendy, 2012). Investigators often take a simple approach of collapsing the repeated measurements into a single value and hence the method is not able to harness the power of the complete information that is contained in the longitudinal trajectory. One can also apply the standard methods available for correlated outcome models to better utilize the longitudinal data, namely, random effects models (Fitzmaurice, Laird and Ware, 2011) and generalized estimating equations (GEE) (Zeger and Liang, 1986). These methods are primarily proposed for modeling and testing a limited number of SNPs, and cannot be directly applied to assess the joint association of a longitudinally varying outcome with an entire gene or a region with hundreds of SNPs without further modifications.

Recent studies showed the advantages of multi-marker tests over individual SNP analyses. First, the genetic markers in LD with the causal SNP(s) carry additional information and may enhance the power of identifying the true effect. Second, gene-based tests considerably reduce the burden of multiple comparisons. Third, Region-based methods are appealing for multi-ethnic cohorts due to differences in LD structure across ethnic groups and thus meta-analysis of a region-based statistic is likely to be more consistent than meta-analysis of single

marker tests across ethnicities. Last, gene-based tests enhance the power of identifying rare-variant association in next generation sequencing studies (Morris and Zeggini, 2010). Two notable existing approaches are the sequence kernel association tests (SKAT) (Wu, et al., 2011) and similarity regression (SIMreg) (Tzeng, et al., 2009). From a random field framework and borrowing ideas from spatial statistics, the genetic random field model (GenRF) was recently developed for modeling and testing joint associations (He, et al., 2014; Li, et al., 2014). So far, however, extensions of these methods are not available for longitudinal data.

It is desirable to have a multi-marker test for longitudinal studies that can incorporate the time-dependent variation in outcome, utilize all the variants in a gene or region and boost power in the presence of effect heterogeneity across time. Extending the GenRF method to the longitudinal setting, we propose a longitudinal genetic random field model (LGRF) and develop generalized score type tests to study the association between repeatedly measured phenotypes and a set of genetic variants in a gene or region. The methods are evaluated through extensive simulation studies and illustrated by analyzing the association between blood pressure and 29 candidate genomic regions across four ethnic groups in MESA.

2. Method

2.1 *Notation*

Consider a study population of m subjects, and the i-th subject has n_i repeated observations. Each subject is sequenced in a region of interest with q variants, and measured on p additional non-genetic covariates such as age, gender and other potential confounders. Let $Y_{i,l}$ be the phenotypic value for the l-th observation on the i-th subject, measured at time $t_{i,l}$; $\boldsymbol{G}_i = (G_{i,1}, G_{i,2}, \ldots, G_{i,q})^T$ be the genotypes for the q variants within the region where $G_{i,h} \in \{0, 1, 2\}$ for any $1 \leqslant h \leqslant q$, which does not change over time; $\boldsymbol{X}_{i,l} = (X_{i,l,1}, \ldots, X_{i,l,p})^T$ be the covariates corresponding to the l-th observation on the i-th subject, either time-varying or time-invariant. We denote $n = \sum_i n_i$, $\boldsymbol{Y}_{n\times 1} = (Y_{1,1}, \ldots, Y_{1,n_1}, Y_{2,1}, \ldots)^T$ and define $\boldsymbol{X}_{n\times p}$,

$\boldsymbol{G}_{n\times q}$ similarly for covariates and genotypes. We are interested in investigating the association between phenotype $\boldsymbol{Y}_{n\times 1}$ and variants $\boldsymbol{G}_{n\times q}$, adjusted for the effect of $\boldsymbol{X}_{n\times p}$.

2.2 *Longitudinal Genetic Random Field Model*

The GenRF method (He, et al., 2014) is a gene-based association test motivated by the general idea that, if the genetic variants in a region are jointly associated with the phenotype, then subjects having similar genotypes in that region will have similar phenotype (Tzeng, et al., 2009). Motivated by development in spatial statistics (Cressie, 1993) and random field theory (Besag, 1974; Adler and Taylor, 2007), GenRF views phenotypic values as a random field on a genetic space where the vector of genotype sequences determines the location in the space; i.e., the phenotype at each location is a random variable and these random variables are possibly correlated depending on their spatial location, e.g., the closer the more similar. It directly regresses the phenotype of a given subject on that of all others, where the contribution of other subjects is weighted by their genotype similarity with the given subject. This leads to a conditional autoregressive model commonly used in spatial statistics to study spatial dependence.

With repeated measurements, one has to appropriately account for the within-subject correlation between outcomes to obtain valid inference and improve efficiency. Extending the GenRF model to the longitudinal setting, we propose a longitudinal GenRF (LGRF) model, where the conditional mean of each observation is modeled as a weighted sum of all other observations, including those from the same subject. In a longitudinal setting, one may expect that phenotypes from the same subject may be more similar due to reasons other than the shared genetic variants of interest. To capture this, we define a within-subject similarity, which depends on the time between two measurements on the same subject; for example, if two observations are measured closer in time, their within-subject similarity may be larger. Formally LGRF model is written as:

$$Y_{i,l}|\boldsymbol{Y}_{-(i,l)} = \boldsymbol{X}_{i,l}^T\boldsymbol{\beta} + \sum_{k\neq l} w(t_{i,k}, t_{i,l}; \boldsymbol{\eta})(Y_{i,k} - \boldsymbol{X}_{i,k}^T\boldsymbol{\beta}) + \gamma \sum_{(j,k)\neq(i,l)} s_{i,j}(Y_{j,k} - \boldsymbol{X}_{j,k}^T\boldsymbol{\beta}) + \varepsilon_{i,l}, \quad (1)$$

where $\boldsymbol{Y}_{-(i,l)}$ denotes all other phenotypic values except $Y_{i,l}$; $\boldsymbol{X}_{i,l}$ and $\boldsymbol{\beta}$ are, respectively, covariates and the corresponding regression coefficients, and thus $\boldsymbol{X}_{i,l}^T\boldsymbol{\beta}$ is the contribution to outcome mean from non-genetic covariates; $\varepsilon_{i,l} \sim$ i.i.d. $N(0, \sigma^2)$; $w(t_{i,k}, t_{i,l}; \boldsymbol{\eta})$ is the within-subject similarity between $Y_{i,k}$ and $Y_{i,l}$ with parameters $\boldsymbol{\eta}$ playing the role of introducing within-subject correlation between repeated measurements, similar to parameters in a correlation matrix in a GEE framework; $s_{i,j}$ is the genetic similarity between subjects i and j. Possible forms can be $s_{i,j} = \sum_{h=1}^{q}(G_{i,h} - 2p_h)(G_{j,h} - 2p_h)$ referred to as genetic relationship (GR) (Yang, et al., 2011) where p_h is the population allele frequency of h-th SNP in the region, and the identity-by-state (IBS) similarity: $s_{i,j} = \sum_{h=1}^{q}(2 - |G_{i,h} - G_{j,h}|)$. Parameter γ measures the magnitude of the joint association between genetic variants and the phenotype. If none of the genetic variants are associated with the phenotype, the phenotype of subject i will be irrelevant to the phenotypes of others regardless of their proximity in the genetic space, i.e., $\gamma = 0$. On the contrary, a large positive γ indicates a strong spatial dependence or equivalently genetic association. Thus, γ can be interpreted as the magnitude of the joint association between the q genetic variants and the phenotype. Briefly, the conditional autoregressive model relates each observation to others measured on the same subject by within-subject similarity $w(t_{i,k}, t_{i,l}; \boldsymbol{\eta})$, and all other observations (including other measurements on the same subject) in the study by genetic similarity $s_{i,j}$.

According to the factorization theorem of Besag (1974), the conditional model (1) uniquely determines a joint distribution of $\boldsymbol{Y}$:

$$\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}, \boldsymbol{v} \sim N(0, \sigma^2\{\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{\eta}) - \gamma\boldsymbol{S}\}^{-1}), \quad (2)$$

where $\boldsymbol{I}$ is an $n \times n$ identity matrix; $\boldsymbol{W}(\boldsymbol{\eta})$ and $\boldsymbol{S}$ are matrices $(n \times n)$ composed of $w(t_{i,k}, t_{i,l}; \boldsymbol{\eta})$ and $s_{i,j}$, respectively. Specifically, the within-subject similarity matrix $\boldsymbol{W}(\boldsymbol{\eta})$ is block diagonal with block i $(n_i \times n_i)$ corresponding to subject i and the (k, l)-th element of

block i is $w(t_{i,k}, t_{i,l}; \boldsymbol{\eta})$ except for diagonal elements of $\boldsymbol{W}(\boldsymbol{\eta})$. The genetic similarity matrix $\boldsymbol{S}$ is composed of $m \times m$ block matrices with dimension $n_i \times n_j, i, j = 1, \ldots, m$, and all elements in the (i, j)-th block are $s_{i,j}$ except for the diagonal elements of $\boldsymbol{S}$. The diagonal elements of $\boldsymbol{W}(\boldsymbol{\eta})$ and $\boldsymbol{S}$ are 0 as in model (1) observations are not compared with themselves. To evaluate the joint association of multiple genetic variants with the phenotype we can test the null hypothesis $H_0 : \gamma = 0$ involving a single parameter in the precision matrix (or equivalently in the variance matrix).

With respect to the within-subject similarity, the random field model focuses on how the observations are related, regardless of the direction (past or future) as opposed to transition models which condition each observation only on the past observations. However, they can result in very similar marginal correlation structures such as the first-order auto-regressive (AR1) correlation. Examples of plausible $\boldsymbol{W}(\boldsymbol{\eta})$ are given below.

Example 1: One might assume observations from the same subject to be equally similar and sets $w(t_{i,k}, t_{i,l}; \boldsymbol{\eta}) = \eta$ for $\forall i, k, l$, and in matrix notation, $\boldsymbol{W}(\boldsymbol{\eta}) = \eta \boldsymbol{T}$, where $\boldsymbol{T}$ is a block diagonal matrix with block i, $i = 1, \ldots, m$, an $n_i \times n_i$ matrix with 0's in the diagonal and 1's off-diagonal. Under $H_0 : \gamma = 0$, the corresponding covariance matrix is $\sigma^2(\boldsymbol{I} - \eta \boldsymbol{T})^{-1}$. This specification is equivalent to the usual compound symmetric correlation.

Example 2: One might assume each observation conditionally depends on only the nearest observations before and after it (Markov property): $w(t_{i,k}, t_{i,l}; \boldsymbol{\eta}) = \eta$ if $|k - l| = 1$, and 0 otherwise. This is an approximation of the usual AR1 correlation by ignoring the edge effect (Qu, et al., 2000). Again $\boldsymbol{W}(\boldsymbol{\eta}) = \eta \boldsymbol{T}$ for a block diagonal matrix $\boldsymbol{T}$, where the (k, l)-th element of the i-th block is 1 if $|k - l| = 1$ and 0 otherwise.

In addition, multiple within-subject similarities can be combined for a better working precision matrix, adaptively approximating the underlying structure. Taking $\boldsymbol{W}(\boldsymbol{\eta})$ to be linear in $\boldsymbol{\eta}$, e.g., the two examples given above and their linear combinations, can lead to a rich

class to accommodate many commonly used working correlation structures. A similar idea has been studied by Qu, et al. (2000) to improve efficiency of estimation over GEE method.

As in the GEE framework, the within-subject similarity matrix $\boldsymbol{W}(\boldsymbol{\eta})$, or equivalently the correlation matrix $\{\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{\eta})\}^{-1}$ under the null, is only a working assumption that is not required to be correct for valid inference. Thus we present our test using a working within-subject similarity matrix that is of the form $\eta\boldsymbol{T}$, as in the two examples, and note the method applies to more general $\boldsymbol{W}(\boldsymbol{\eta})$. For simplicity, the matrix representation of the LGRF model is given by:

$$\boldsymbol{Y}|\boldsymbol{Y}_{-} = \boldsymbol{X}\boldsymbol{\beta} + (\eta\boldsymbol{T} + \gamma\boldsymbol{S})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \tag{3}$$

where $\boldsymbol{Y}$ is the n dimensional vector of all observations; $\boldsymbol{Y}|\boldsymbol{Y}_{-}$ stands for that each observation $Y_{(i,l)}$ is conditional on all other observations $\boldsymbol{Y}_{-(i,l)}$; Matrices $\boldsymbol{T}$ and $\boldsymbol{S}$ have diagonal elements equal to zero, to reflect that the mean of each element of $\boldsymbol{Y}$ only depends on other elements but not on itself; $\boldsymbol{\varepsilon} = (\varepsilon_{1,1}, \ldots, \varepsilon_{1,n_1}, \varepsilon_{2,1}, \ldots)^T$ is the residual vector. Since the genetic similarities are compared across all observations, the model does not have the Markov property, i.e., each observation has finite neighbors, typically assumed in a conditional auto-regressive model in spatial statistics. Thus the regular likelihood-ratio test or score test used in spatial statistics for testing spatial auto-correlation cannot be applied directly. Also, because of the within-subject similarly, the pseudo-likelihood approach developed by He, et al. (2014) does not apply. Instead, we propose a set of generalized score type tests.

2.3 *Association Test under the Longitudinal Genetic Random Field Model*

In this subsection we focus on developing a generalized score type test for testing $H_0 : \gamma = 0$ under model (3). The inference procedure is developed by treating the within-subject correlation as a working model, leading to a test that is robust to misspecification of the correlation structure. Model (3) states, given all other observations, the conditional mean of each observation is linearly related to others, i.e., $E(\boldsymbol{Y}|\boldsymbol{Y}_{-}) = \boldsymbol{X}\boldsymbol{\beta} + (\eta\boldsymbol{T} + \gamma\boldsymbol{S})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$.

Adopting the similar argument for the usual GEE method (Zeger and Liang, 1986) to our conditional auto-regressive model, we construct the following generalized estimating function:

$$U_\gamma(\boldsymbol{\beta}, \eta, \gamma) = \frac{\partial E(\boldsymbol{Y}|\boldsymbol{Y}_-)}{\partial \gamma}^T \{\boldsymbol{Y} - E(\boldsymbol{Y}|\boldsymbol{Y}_-)\} = (\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}(\boldsymbol{I} - \eta \boldsymbol{T} - \gamma \boldsymbol{S})(\boldsymbol{Y} - \boldsymbol{\mu}), \quad (4)$$

where $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$. The estimating equation is quadratic in $\boldsymbol{Y}$ because γ is a coefficient in an auto-regressive model and corresponds to a parameter in the marginal variance as in (2). In the Supplementary Materials section 1.1, we show that the above estimating function is unbiased in the sense that its expectation is zero under the truth. Therefore, following Boos (1992), we refer to it as a "generalized" score and the score evaluated at $\gamma = 0$, i.e., $U_\gamma(\boldsymbol{\beta}, \eta, 0) = (\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}(\boldsymbol{I} - \eta \boldsymbol{T})(\boldsymbol{Y} - \boldsymbol{\mu})$, can be used to construct a generalized score type test. Due to the unbiasedness, we show that $U_\gamma(\boldsymbol{\beta}, \eta, 0)$ has mean 0 under H_0 and positive mean $\gamma E\{(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}^2 (\boldsymbol{Y} - \boldsymbol{\mu})\}$ under $H_1 : \gamma > 0$. This rationale leads to constructing a generalized score statistic

$$Q_G = \frac{U_\gamma(\widehat{\boldsymbol{\beta}}, \widehat{\eta}, 0)}{m} = \frac{(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})^T \boldsymbol{S}(\boldsymbol{I} - \widehat{\eta}\boldsymbol{T})(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})}{m} \quad (5)$$

and rejecting H_0 when it is sufficiently large. In (5), $\widehat{\boldsymbol{\mu}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta}$ are estimates under the null hypothesis that $\gamma = 0$. Specifically, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta}$ are the solution to the following estimating equations:

$$\begin{cases} U_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \eta, 0) = \boldsymbol{X}^T(\boldsymbol{I} - \eta \boldsymbol{T})(\boldsymbol{Y} - \boldsymbol{\mu}) = 0 \\ U_\eta(\boldsymbol{\beta}, \eta, 0) = (\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{T}(\boldsymbol{I} - \eta \boldsymbol{T})(\boldsymbol{Y} - \boldsymbol{\mu}) = 0. \end{cases}$$

The first equation is the usual estimating equation for estimating $\boldsymbol{\beta}$ in GEE based on the the joint distribution (2) as $\boldsymbol{I} - \eta \boldsymbol{T}$ is proportional to the inverse of a working correlation matrix under H_0. The second equation is derived by considering the estimating function $\frac{\partial E(\boldsymbol{Y}|\boldsymbol{Y}_-)}{\partial \eta}^T \{\boldsymbol{Y} - E(\boldsymbol{Y}|\boldsymbol{Y}_-)\}$ under H_0. It is worth noting that the second estimating equation is linear in η. Thus the estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta}$ can be calculated by iteratively solving linear equations. This property remains when we linearly combine multiple within-subject similarities, leading to an efficient way to estimate the correlation structure.

We derive an asymptotically equivalent representation of Q_G under H_0 and show that this representation allows us to achieve theoretical protection against the misspecification of within-subject correlation as well as facilitating computationally efficient implementation suitable for large-scale studies. Specifically, we show in Supplementary Materials sections 1.2 and 1.3 that for all the genetic similarity metrics introduced previously, under H_0, Q_G can be represented as

$$Q_G = \frac{1}{2m}\boldsymbol{R}_1(\eta_0, \widehat{\boldsymbol{\beta}})^T \begin{pmatrix} \mathbf{0}_{dq} & \boldsymbol{I}_{dq} \\ \boldsymbol{I}_{dq} & \mathbf{0}_{dq} \end{pmatrix} \boldsymbol{R}_1(\eta_0, \widehat{\boldsymbol{\beta}}) + c + o_p(1),$$

where η_0 is the true parameter under H_0; $\boldsymbol{I}_{dq}$ is a $dq \times dq$ identity matrix and $\mathbf{0}_{dq}$ is a zero matrix; $\boldsymbol{R}_1(\eta, \boldsymbol{\beta}) = \widetilde{\boldsymbol{Z}}(\eta)^T(\boldsymbol{Y} - \boldsymbol{\mu})$ and $\widetilde{\boldsymbol{Z}}(\eta) = \{(\boldsymbol{I} - \eta\boldsymbol{T})\boldsymbol{Z}, \boldsymbol{Z}\}$; $\boldsymbol{Z}$ is an $n \times dq$ matrix for some integer d, and c is a constant. The exact form or value of $\boldsymbol{Z}$, d and c depend on the chosen genetic similarity and the details are given in the Supplementary Materials sections 1.2 and 1.4. For example, for GR similarity, $\boldsymbol{Z}, (n \times q)$, is the centered genotype matrix, i.e., each column of the genotype matrix $\boldsymbol{G}$ is now centered by the genotype population mean $2p_h$. Note that $\boldsymbol{R}_1(\eta_0, \widehat{\boldsymbol{\beta}}) = \sum_{i=1}^m \{\widetilde{\boldsymbol{Z}}_i(\eta_0)^T(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i)\}$, which is a summation of m terms each with expectation zero under the null regardless of the specified working correlation structure. Therefore, the summand is an unbiased estimating function for $\boldsymbol{\beta}$, and according to the theory of M-estimators (Stefanski and Boos, 2002), $\frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0, \widehat{\boldsymbol{\beta}})$ is asymptotically normal with a covariance matrix that can be robustly estimated by some sandwich covariance estimates, leading to robustness to misspecification of working correlation.

In Results 2 and 3 of Supplementary Materials, using the theory of M-estimation as well as distributions for quadratic forms, we show that Q_G has an asymptotic distribution

$$\frac{1}{2}\sum_{k=1}^{2dq} \lambda_k \chi_k^2 + c$$

under H_0, where c is a constant which does not affect the inference; χ_k^2's are i.i.d. Chi-square distributions with degree of freedom one; λ_k are eigenvalues of a $2dq \times 2dq$ matrix

$$\begin{pmatrix} \mathbf{0}_{dq} & \boldsymbol{I}_{dq} \\ \boldsymbol{I}_{dq} & \mathbf{0}_{dq} \end{pmatrix} \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ can be consistently estimated by a sandwich covariance estimate $\widehat{\boldsymbol{\Sigma}}$, defined in Result 2 of the Supplementary Materials. Moreover, the null distribution of Q_G only depends on the eigen-values of a $2dq \times 2dq$ matrix. As the number of variants in a target gene q is relatively small, it is computationally efficient and hence suitable for large scale GWAS. To obtain the p-value, Davies' method (1980) can be used as a computationally efficient way to analytically calculate the tail probability of a mixture of chi-squares by inverting the corresponding characteristic function.

2.4 *Testing for the Joint Effect of Gene and Gene-time Interaction*

As in a regression framework interaction effect is typically modeled using new variables defined as the product of two interacting factors, similarly, we can define interaction terms, $\boldsymbol{G}_i t_{i,l} = (G_{i,1}t_{i,l}, G_{i,2}t_{i,l}, \ldots, G_{i,q}t_{i,l})^T$, and treat them the same way as $\boldsymbol{G}_i$. Therefore the modified LGRF is given by:

$$\begin{aligned} Y_{i,l}|\boldsymbol{Y}_{-(i,l)} &= \boldsymbol{X}_{i,l}^T\boldsymbol{\beta} + \sum_{k\neq l} w(t_{i,k}, t_{i,l}; \boldsymbol{\eta})(Y_{i,k} - \boldsymbol{X}_{i,k}^T\boldsymbol{\beta}) + \gamma_1 \sum_{(j,k)\neq(i,l)} s_{i,j}(Y_{j,k} - \boldsymbol{X}_{j,k}^T\boldsymbol{\beta}) \\ &+ \gamma_2 \sum_{(j,k)\neq(i,l)} s_{il,jk}^{GT}(Y_{j,k} - \boldsymbol{X}_{j,k}^T\boldsymbol{\beta}) + \varepsilon_{i,l}, \end{aligned}$$

where $s_{il,jk}^{GT}$ is the similarity generated by gene-time interaction terms, similar to the genetic similarity; and γ_1 and γ_2 represent the genotype main effect and gene-time interaction effect, respectively. The IBS similarity is not suitable for the interaction terms because it is specifically designed for genetic variants/imputed dosage lying between 0 and 2. In the spirit of genetic relationship similarity, we define $s_{il,jk}^{GT} = \psi(\boldsymbol{G}_i t_{i,l}, \boldsymbol{G}_j t_{j,k}) = \sum_{h=1}^{q}(G_{i,h}t_{i,l} - \overline{G_h t})(G_{j,h}t_{j,k} - \overline{G_h t})$, where $\overline{G_h t} = \frac{1}{n}\sum_{(i,l)} G_{i,h}t_{i,l}$. Considering a working within-subject similarity matrix $\eta\boldsymbol{T}$ as before, in matrix form the model is written as

$$\boldsymbol{Y}|\boldsymbol{Y}_{-} = \boldsymbol{X}\boldsymbol{\beta} + (\eta\boldsymbol{T} + \gamma_1\boldsymbol{S} + \gamma_2\boldsymbol{S}_{GT})(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \tag{6}$$

where $\boldsymbol{S}_{GT}$ is the similarity matrix of the interaction terms with the (l,k)-th element of the (i,j)-th block $(n_i \times n_j)$ equal to $s^{GT}_{il,jk}$ except for the diagonal of $\boldsymbol{S}_{GT}$. Under this model, we can evaluate the joint effect of gene and gene-time interaction by testing $H_0^J : \gamma_1 = \gamma_2 = 0$.

Denoting $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^T$, following previous development, we construct two estimating function with respect to γ_1 and γ_2:

$$\begin{cases} U_{\gamma_1}(\boldsymbol{\beta}, \eta, \boldsymbol{\gamma}) = (\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}(\boldsymbol{I} - \eta\boldsymbol{T} - \gamma_1\boldsymbol{S} - \gamma_2\boldsymbol{S}_{GT})(\boldsymbol{Y} - \boldsymbol{\mu}) \\ U_{\gamma_2}(\boldsymbol{\beta}, \eta, \boldsymbol{\gamma}) = (\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}_{GT}(\boldsymbol{I} - \eta\boldsymbol{T} - \gamma_1\boldsymbol{S} - \gamma_2\boldsymbol{S}_{GT})(\boldsymbol{Y} - \boldsymbol{\mu}). \end{cases}$$

As before, evaluating the corresponding estimating functions at $H_0^J : \gamma_1 = \gamma_2 = 0$ leads to the following generalized score statistics

$$\begin{cases} Q_G = U_{\gamma_1}(\hat{\boldsymbol{\beta}}, \hat{\eta}, \boldsymbol{0})/m = (\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})^T \boldsymbol{S}(\boldsymbol{I} - \eta\boldsymbol{T})(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})/m \\ Q_{GT} = U_{\gamma_2}(\hat{\boldsymbol{\beta}}, \hat{\eta}, \boldsymbol{0})/m = (\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})^T \boldsymbol{S}_{GT}(\boldsymbol{I} - \eta\boldsymbol{T})(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})/m. \end{cases}$$

We propose to combine these two by:

$$Q_J = \alpha_G Q_G + \alpha_{GT} Q_{GT} = (\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})^T (\alpha_G \boldsymbol{S} + \alpha_{GT}\boldsymbol{S}_{GT})(\boldsymbol{I} - \eta\boldsymbol{T})(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})/m,$$

where $\alpha_G = \sqrt{\frac{v_{GT}^2}{v_{GT}^2 + v_G^2}}$ and $\alpha_{GT} = \sqrt{\frac{v_G^2}{v_{GT}^2 + v_G^2}}$; $v_G^2 = 2\mathrm{tr}(\boldsymbol{S}^2)$ and $v_G^2 = 2\mathrm{tr}(\boldsymbol{S}_{GT}^2)$ are proportional to the variance of U_{γ_1} and U_{γ_2} respectively. Though the choice of weights can be arbitrary depending on the need of assessing marginal or interaction effect, our weights are defined such that $\alpha_G Q_G$ and $\alpha_{GT} Q_{GT}$ have approximately equal variance. Defining $\boldsymbol{Z}_{GT}$ as the centered gene-interaction matrix, i.e., each gene-interaction term $G_{i,h}t_{i,l}$ is centered by the its mean $\overline{G_h t}$, $\widetilde{\boldsymbol{Z}}_J(\eta) = \{\alpha_G^{\frac{1}{2}}(\boldsymbol{I} - \eta\boldsymbol{T})\boldsymbol{Z}, \alpha_{GT}^{\frac{1}{2}}(\boldsymbol{I} - \eta\boldsymbol{T})\boldsymbol{Z}_{GT}, \alpha_G^{\frac{1}{2}}\boldsymbol{Z}, \alpha_{GT}^{\frac{1}{2}}\boldsymbol{Z}_{GT}\}$ and $\boldsymbol{R}_{J1}(\eta, \boldsymbol{\beta}) = \widetilde{\boldsymbol{Z}}_J(\eta)^T(\boldsymbol{Y} - \boldsymbol{\mu})$, we can rewrite the joint test statistic as a quadratic form:

$$Q_J = \frac{1}{2m}\boldsymbol{R}_{J1}(\eta_0, \widehat{\boldsymbol{\beta}})^T \begin{pmatrix} \boldsymbol{0}_{(d+1)q} & \boldsymbol{I}_{(d+1)q} \\ \boldsymbol{I}_{(d+1)q} & \boldsymbol{0}_{(d+1)q} \end{pmatrix} \boldsymbol{R}_{J1}(\eta_0, \widehat{\boldsymbol{\beta}}) + c_J + o_p(1),$$

where d is a constant depending on the chosen genetic similarity for the marginal genetic effect as in Section 2.4 and c_J is a constant similar to c. Although more complex, Q_J has an identical form as Q_G in Section 2.4. The inference follows directly from previous

development and therefore we omit the details. The proposed method does not test the gene-time interaction separately; instead, it improves the power of LGRF test by exploiting the potential interaction effect if exists.

3. Illustration in MESA

We refer to the LGRF test for the marginal effect of a gene as LGRF-G and the joint test as LGRF-J. We illustrate the proposed methods using data from the Multi-Ethnic Study of Atherosclerosis (MESA). MESA is a collaborative longitudinal study initiated in July 2000 to investigate the prevalence, correlates, and progression of subclinical cardiovascular disease (CVD) (Bild, et al., 2002). From 2000 to 2007, four examinations of blood pressure (BP) were conducted over 18- to 24-month periods. We aimed to replicate the findings (29 significant SNPs associated with blood pressure) of the International Consortium for Blood Pressure (ICBP) (International Consortium for Blood Pressure Genome-Wide Association Studies, 2011) by a region based analysis. A total of 6361 subjects consisting of 2526 Caucasians (CAU), 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. The longitudinal summaries and characteristics of the study population, descriptive statistics are provided in Supplementary Tables 8 - 11. For this analysis, we used SNPs that have been directly genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 or imputed as per MESA protocol. Imputation was performed using the IMPUTE 2.1.0 program (Marchini, et al., 2007) in conjunction with HapMap Phase I and II reference panels (CEU+YRI+CHB+JPT, release 22 - NCBI Build 36 for African-, Chinese- and Hispanic-American participants; CEU, release 24 - NCBI Build 36 for European Americans). We selected genomic regions around the 29 index SNPs that have demonstrated significant (p-value $< 10^{-9}$) by the ICBP. Each genomic region was defined according to the following criteria: when the index SNP fell

within a gene, we selected all SNPs within the gene +/- 5kb and adopted the gene's name. When the index SNP fell outside of a gene, we selected the index SNP plus all SNPs +/- 50kb and name the region after the index SNP. All SNPs are included in the analysis without any minor allele frequency filters. We applied LGRF-G and LGRF-J using longitudinal outcomes and SKAT using the average value of repeated measures to test the association between each candidate region and BPs (sBP and dBP) for the four ethnic groups separately, adjusting for age, gender, BMI and top two principal components (PCs) to correct for potential within-ethnicity stratification. Since only the first two principal components were associated with either systolic or diastolic blood pressure in any ethnicity at $p < 0.01$ (Supplementary Table 7), we only included these two principal components as adjustment variables. We adjusted the measured blood pressures for participants taking anti-hypertension medications using the standard procedure of adding 10 mmHg to systolic blood pressure and 5 mmHg to diastolic blood pressure (Cui, et al., 2003). The SKAT was implemented with a linear kernel and equal weights on the SNPs. Based on the p-values of the stratified analysis, a meta-analysis was done by Fisher's method.

We analyzed 29 regions with details summarized in the Supplementary Tables 12 - 21. The LGRF-G test results in comparable or smaller p-values than SKAT using average outcomes in most cases. We expect LGRF-J to have higher power than LGRF-G when there exists gene-time interaction, but lower power when there is no such interaction. In the MESA example, the LGRF-J test has smaller p-values than LGRF-G in relatively few instances (for example association of C10orf107 with diastolic blood pressure in Table 1), but larger p-values than LGRF-G in general. This may indicate that gene-time interaction does not have sufficient contribution to the marginal gene-level association in most cases. Table 1 shows the results of the top two associations between sBP/dBP and candidate regions. The top two regions were selected according to the p-values of LGRF-G in meta-analysis using Fisher's combined

probability test. The region indexed by $rs13082711$ emerged as the most strongly associated region. The meta-analysis p-values of LGRF-G are 8.69×10^{-4} for sBP and 6.25×10^{-4} for dBP. Another suggestive association identified by LGRF-G that is consistent with the ICBP analysis is between dBP and $C10orf107$ (p-value= 9.71×10^{-4}), and LGRF-J exhibited a smaller p-value for this association (p-value= 8.64×10^{-4}).

[Table 1 about here.]

4. Simulation Studies

We evaluated three classes of methods: (a) the proposed multi-marker tests for longitudinal data: LGRF-G, LGRF-J; (b) a multi-marker test in cross-sectional studies using the average of the repeated measures as a single outcome: SKAT; and (c) single-marker tests for longitudinal outcomes: namely, GEE, adjusted by the Bonferroni correction. Specifically, in LGRF-G, LGRF-J and GEE, a working compound symmetric correlation structure was used, and SKAT was implemented with equal weights on the SNPs. Classes (b) and (c) represent two commonly used strategies in practice as currently no multi-marker tests are available for longitudinal data and the specific method (SKAT and GEE) is chosen to be the representative in each class, recognizing that multiple other alternatives in each class exist. Additional simulation studies with respect to the impact of different genetic similarity measures, further evaluation of the power gain using a longitudinal design, use of LGRF in a meta-analysis, and evaluation of type-I error rate and power at lower significance levels are showed in the Supplementary Tables 2 - 7.

For each replicated dataset, subjects were randomly selected from the Caucasian (CAU) ethnic group in MESA, and the variants commonly existing in all four ethnicities (154 SNPs) in genotype region $C10orf107$ are included as the target region. We varied the number of repeated measurements to be 4, 6 and 8, and number of subjects 600, 400 and 300

respectively, keeping total number of observations as 2400. Assuming missing completely at random, we first simulated the complete data, and then a missingness indicator with fixed drop-out rate of 4% at each exam was applied approximating what is observed in the MESA study.

4.1 *Type-I Error Simulations*

We evaluated the type-I error rate at level $\alpha = 0.05$, 0.01, and 0.001 using 100000 replicates. Data are generated from the model:

$$Y_{i,l} = \alpha_0 t_{i,l} + \epsilon_{i,l}, t_{i,l} = 1, \ldots, r, \tag{7}$$

where $\alpha_0 = \frac{12}{r}$; r is the number of measurements per subject; $\boldsymbol{\epsilon}_i = (\epsilon_{i,1}, \ldots, \varepsilon_{i,r})^T$ independently follows multivariate normal distribution with four types of covariance matrices:

- Independent (Ind.): $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2_{ind}\boldsymbol{I}_r)$.
- Auto-regressive of order 1 (AR1): $\boldsymbol{\epsilon}_i \sim N(0, \boldsymbol{\Sigma}_{AR})$, where $\boldsymbol{\Sigma}_{AR}$ is an $r \times r$ matrix and its (l, k) element is $\rho^{|l-k|}\sigma^2_{AR}$.
- Compound symmetry (CS): $\epsilon_{i,l} = b_i + \epsilon^*_{i,l}$, $\epsilon^*_{i,l} \sim N(0, \sigma^2_{error})$, $b_i \sim N(0, \sigma^2_{CS})$, where $\epsilon^*_{i,l}$ and b_i are independent.
- Mixed model with a random intercept and a random slope (RR): $\epsilon_{i,l} = b_{1i} + b_{2i}t_{i,l}/r + \epsilon^*_{i,l}$, $\epsilon^*_{i,l} \sim N(0, \sigma^2_{error})$, $b_{1i}, b_{2i} \sim N(0, \sigma^2_{RR})$, where $\epsilon^*_{i,l}$, b_{1i} and b_{2i} are independent.

Where σ^2_{ind}=16; $\sigma^2_{AR} = 6$, $\rho = 0.6$; $\sigma^2_{error} = 2.25$; $\sigma^2_{CS} = 2.25$; $\sigma^2_{RR} = 1$. The missingness indicator was then applied to the simulated data with 4% drop-out rate. The empirical type-I error rates are presented in Table 2. LGRF-G and LGRF-J both have well controlled type-I error rates under all scenarios, even if the true correlation is not the assumed working correlation "CS". The tests also have valid type-I error rates at low α-levels (0.01 and 0.001). The simulation results demonstrate that, consistent with the asymptotic result, the proposed methods are robust to misspecification of within-subject correlation in finite samples. We

note that the proposed methods tend to be slightly conservative at lower significance levels (Supplementary Table 5) due to the use of sandwich estimator as in regular GEE.

[Table 2 about here.]

4.2 *Power Simulations*

In the first set of power simulations, one out of 154 SNPs was randomly set to be causal. We evaluated two distinct scenarios where the effect of the single causal SNP is manifested through: 1. its marginal association with outcome, without any gene-time interaction; 2. its interaction with time (SNP $\times$ Time interaction). The data was generated respectively:

$$1.\ \text{Gene marginal effect} : Y_{i,l} = \alpha_0 t_{i,l} + \alpha_1 G_i + \epsilon_{i,l}, t_{i,l} = 1, \ldots, r, \tag{8}$$

$$2.\ \text{Gene-time interaction} : Y_{i,l} = \alpha_0 t_{i,l} + \alpha_2 G_i t_{i,l} + \epsilon_{i,l}, t_{i,l} = 1, \ldots, r, \tag{9}$$

where G_i is the genotype of subject i for the randomly selected causal SNP; $\alpha_0 = 12/r$, $\alpha_1 = 0.4$ and $\alpha_2 = 0.6/r$; r is the number of measurements per subject. To mimic the real data scenario, α_1 and α_2 were elicited based on fitting single SNP models with and without gene-time interaction to MESA data. We chose a large α_0 in our simulation studies to illustrate the power gain that can be expected from a longitudinal design with strong time trend in the mean outcome levels compared to using the average of repeated measures. We recognize that smaller values of α_0 will lead to smaller power differences.

In the second set of simulations, ten out of 154 were randomly set to be causal each time. Among them, six SNPs have only marginal effects, three have both marginal and interaction effects and the remaining one has only an interaction effect. The true model is of the form:

$$Y_{i,l} = \alpha_0 t_{i,l} + \alpha_1^* \sum_{1 \leqslant k \leqslant 9} G_{i,k} + \alpha_2^* \sum_{7 \leqslant k \leqslant 10} G_{i,k} t_{i,l} + \epsilon_{i,l}, t_{i,l} = 1, \ldots, r.$$

Where $G_{i,k}$ is the genotype of subject i on the k-th randomly selected causal SNP. The coefficients are proportional to α_1 and α_2: $\alpha_1^* = \alpha_1/10 = 0.04$ and $\alpha_2^* = \alpha_2/10 = 0.06/r$, such that the empirical powers are differentiable.

Two important points are illustrated by this simulation: 1. the advantage of incorporating longitudinal information over using only the average outcome; 2. The use of multi-marker tests over single-marker tests. The proposed multi-marker tests using the longitudinal outcome have larger power than SKAT using the average of outcomes, as the proposed tests use the whole trajectory of longitudinal outcomes as opposed to only information contained in the average. When the number of repeated measurements increases, the power becomes more distinct. Not surprisingly, LGRF-J test has slightly lower power than LGRF-G because gene-time interaction does not exist in these scenarios.

[Table 3 about here.]

When the causal SNP has only an interaction effect (Table 4), the relative performance of the methods using repeated measures compared with the one using average outcome is more distinct. In addition, the joint test LGRF-J is able to further enhance power in these scenarios because it incorporates the gene-time interaction explicitly. We note that the power difference between LGRF and SKAT using average outcome is mainly attributed to the longitudinal design rather than the difference between genetic random field model and SKAT (Supplementary Table 4).

[Table 4 about here.]

We also note that the proposed multi-marker tests have larger power than single-marker tests using GEE with Bonferroni correction (Tables 3-5), consistent with results found in cross-sectional studies where advantages of multi-marker tests over single-marker tests have been demonstrated repeatedly. The advantage in power is more substantial when there are multiple causal SNPs (Table 5) than when there is only one causal SNP (Tables 3-4).

[Table 5 about here.]

5. Discussion

We extended the genetic random field model to the longitudinal setting and developed generalized score type tests to test the joint association between a set of genetic variants and a repeatedly measured phenotype. Besides the advantages of region-based tests over single-marker tests in cross-sectional studies, the LGRF model is able to utilize all the repeated measurements, incorporate gene-time interaction explicitly and result in higher power. As in GenRF, LGRF models the joint association using a single parameter by considering the similarity in phenotype induced by genetic similarity. A main challenge in modeling longitudinal data is to account for within-subject correlation and correlation is conceptually viewed and modeled in a unified way as the joint genetic association in LGRF. Furthermore, the specified correlation structure is treated as a working assumption in inference and the resulting LGRF tests are robust to misspecification.

LGRF tests are generalized score tests that only need to fit the model under the null hypothesis, which is irrelevant to the target region. Users can fit the null model once and test all regions without repeatedly fitting the model. In addition, the computational cost of LGRF mainly depends on the fixed number of variants in the region but not the sample size. This property improves the computational efficiency dramatically (see Supplementary Table 1) especially when the target region is small, for example if investigators are only interest in the exon.

We note that not only the longitudinal outcomes precisely describe the phenotype progression, considering time varying exposure and its interaction with genotype may also improve the discovery process. However, an analysis using the average outcome and a single measure of exposure will lose the longitudinal features of the time varying exposure variables and their correlations, reducing the rich exposure and outcome data to an aggregate summary measure. In the spirit of multi-marker based tests for gene-environment interaction, such as

GESAT (Lin, et al., 2013), we expect that a potential future extension of LGRF towards separately testing gene-time or gene-environment interaction in longitudinal studies with time dependent covariates may enhance the discovery process. Finally, the proposed test is only valid when the data is missing completely at random as in GEE (Zeger and Liang, 1986). Future work extending the method to cases other than this will be of interest.

6. Supplementary Materials

Web Appendices referenced in Sections 2, 3, 4 and 5, and the R code implementing the method are available with this paper at the Biometrics website on Wiley Online Library. The code and an illustrative example are also available at: http://sitemaker.umich.edu/statzihuai/longitudinal_genetic_random_field__lgrf_.

Acknowledgements

This work was supported by NSF DMS 1406712 and NIH/NIEHS grant ES020811 (B.M.), NIH/NHLBI grant R00HL113164 (S.L.), NIH/NHLBI HL101161 (A.D., S.K., B.M., J.S.), and NIMHHD Grant 2P60MD002249 Center for Integrative Approaches to health Disparities (CIAHD) (A.D., S.K., J.S.). 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 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 and CTSA UL1-RR-024156. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0.

References

Adler, R.J. and Taylor, J.E. (2007). Random Fields and Geometry. Springer, New York.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B* **36(2),** 192–236.

Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Roux, A.V.D., Folsom, A.R., et al. (2002). Multi-ethnic study of atherosclerosis: objectives and design. *American Journal of Epidemiology* **156(9),** 871–881.

Boos, D.D. (1992). On generalized score tests. *The American Statistician* **46(4),** 327–333.

Cui, J. S., Hopper, J. L., and Harrap, S. B. (2003). Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension* **41(2),** 207–210.

Cressie, N. (1993). Statistics for spatial data. Wiley, New York.

Davies, R. (1980). The distribution of a linear combination of chi-square random variables. *Applied Statistics* **29,** 323–333.

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.

Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. (2011). Applied longitudinal analysis, 2nd Edition. John Wiley & Sons.

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

He, Z., Zhang, M., Zhan, X., and Lu, Q. (2014). Modeling and testing for joint association using a genetic random field model. *Biometrics*, doi: 10.1111/biom.12160.

International Consortium for Blood Pressure Genome-Wide Association Studies (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478(7367),** 103–109.

Li, M., He, Z., Zhang, M., Zhan, X., Wei, C., Elston, R.C., and Lu, Q. (2014). A generalized

genetic random field method for the genetic association analysis of sequencing data. *Genetic Epidemiology* **38,** 242–253.

Lin, X., Lee S., Christiani, D.C., and Lin, X. (2013). Test for the Interaction between a Genetic Marker Set and Environment in Generalized Linear Models. *Biostatistics* **14(4),** 667–681.

Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics* **39(7),** 906–913.

Morris, A.P., and Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genetic Epidemiology* **34(2),** 188–193.

Qu, A., Lindsay, B.G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87(4),** 823–836.

Stefanski, L.A. and Boos, D.D. (2002). The calculus of M-estimation. *The American Statistician* **56,** 29–38.

Tzeng, J.Y., Zhang, D., Chang, S.M., Thomas, D.C., and Davidian, M. (2009). Gene-Trait Similarity Regression for Multimarker-based Association Analysis. *Biometrics* **65(3),** 822–832.

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(1),** 82–93.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88(1),** 76–82.

Zeger, S.L., and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 121–130.

Received – 2014. *Revised* – –. *Accepted* – –.

Table 1

Analysis of Multi-Ethnic Study of Atherosclerosis (MESA) data: top two regions associated with systolic blood pressure/diastolic blood pressure. Each cell shows the p-value. CAU: Caucasians; AFA: African Americans; HIS: Hispanics; CHN: Asians of Chinese descent. Meta: Meta-analysis combining the results of four ethnic groups using Fisher's combined probability test. LGRF-G: the LGRF test for the marginal effect of a gene. LGRF-J: the LGRF test for the joint effect of gene and gene-time interaction. The working correlation assumed in LGRF is compound symmetric. SKAT-Avg.: cross-sectional SKAT using the average value of repeated measurements as the outcome. The column "SNPs" shows the total number of typed and imputed SNPs in each ethnic group.

Systolic Blood Pressure								
	Region Indexed by $rs13082711$				Region Indexed by $rs1378942$			
	SNPs	LGRF-G	LGRF-J	SKAT-Avg.	SNPs	LGRF-G	LGRF-J	SKAT-Avg.
CAU	111	0.0052	0.0078	0.0047	84	0.0019	0.0023	0.0019
AFA	82	0.6750	0.6315	0.6806	70	0.1894	0.2047	0.1929
HIS	82	0.0267	0.0453	0.0307	70	0.5269	0.3446	0.4094
CHN	79	0.0191	0.0496	0.0302	70	0.8798	0.9364	0.8969
Meta	-	0.0009	0.0036	0.0013	-	0.0258	0.0248	0.0222
Diastolic Blood Pressure								
	Region Indexed by $rs13082711$				$C10orf107$			
	SNPs	LGRF-G	LGRF-J	SKAT-Avg.	SNPs	LGRF-G	LGRF-J	SKAT-Avg.
CAU	111	0.1774	0.1185	0.1704	190	0.0283	0.0412	0.0202
AFA	82	0.0263	0.0222	0.0233	157	0.0129	0.0106	0.0152
HIS	82	0.0086	0.0349	0.0058	157	0.0104	0.0081	0.0234
CHN	79	0.0292	0.0713	0.0308	154	0.5361	0.4998	0.4757
Meta	-	0.0006	0.0024	0.0004	-	0.0010	0.0009	0.0015

Table 2

Type-I Error Rate Corresponding to Different Within-Subject Correlation Structures. Each cell represents the empirical type-I error rate evaluated at α=0.05, 0.01 and 0.001 based on 100000 replicates. The total number of observations is 2,400 and repeated measurements per subject were generated in the same follow-up period according to different correlation structures. Ind.: the repeated measurements are independent. CS: the correlation is compound symmetric. AR1: the repeated measurements follow a first-order auto-regressive model. RR: observations follow a mixed model with a random intercept and a random slope. LGRF-G: the LGRF test for the marginal effect of a gene. LGRF-J: the LGRF test for the joint effect of gene and gene-time interaction. The working correlation assumed in LGRF is CS.

Type-I Error Rate						
Four Repeated Measurements (600 Subjects)						
	LGRF-G			LGRF-J		
α =	0.05	0.01	0.001	0.05	0.01	0.001
Ind.	0.0495	0.0096	0.0008	0.0493	0.0097	0.0008
CS	0.0493	0.0099	0.0009	0.0491	0.0096	0.0009
AR1	0.0499	0.0097	0.0009	0.0507	0.0097	0.0009
RR	0.0497	0.0094	0.0009	0.0498	0.0096	0.0008
Six Repeated Measurements (400 Subjects)						
	LGRF-G			LGRF-J		
α =	0.05	0.01	0.001	0.05	0.01	0.001
Ind.	0.0501	0.0096	0.0009	0.0501	0.0093	0.0010
CS	0.0501	0.0097	0.0009	0.0488	0.0089	0.0008
AR1	0.0485	0.0093	0.0009	0.0494	0.0097	0.0008
RR	0.0497	0.0096	0.0010	0.0500	0.0095	0.0009
Eight Repeated Measurements (300 Subjects)						
	LGRF-G			LGRF-J		
α =	0.05	0.01	0.001	0.05	0.01	0.001
Ind.	0.0488	0.0091	0.0008	0.0483	0.0091	0.0007
CS	0.0484	0.0092	0.0010	0.0488	0.0090	0.0007
AR1	0.0474	0.0090	0.0008	0.0471	0.0089	0.0009
RR	0.0492	0.0095	0.0008	0.0485	0.0091	0.0008

Table 3

Power comparisons when **one** *randomly selected SNP is causal and has* **a marginal effect**. *Each cell represents the empirical power from 500 replicates at level α=0.05. The total number of observations is 2,400 and repeated measurements were recorded in the same follow-up period. Ind.: the repeated measurements are independent. CS: the correlation is compound symmetric. AR1: the repeated measurements follow a first-order auto-regressive model. RR: observations follow a mixed model with a random intercept and a random slope. LGRF-G: the LGRF test for the marginal effect of a gene. LGRF-J: the LGRF test for the joint effect of gene and gene-time interaction. The working correlation assumed in LGRF is CS. SKAT-Avg.: cross-sectional SKAT using the average value of repeated measurements as the outcome. GEE-G: test the marginal association by GEE. GEE-J: jointly test the marginal association and gene-time interaction by GEE. These single-marker tests were implemented by testing every SNP in the region and adjusting the minimum p-value by the Bonferroni correction.*

Power: Single SNP Marginal Effect					
Four Repeated Measurements (600 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	**0.42**	0.39	0.34	0.26	0.19
CS	**0.53**	0.49	0.43	0.41	0.33
AR1	**0.46**	0.45	0.38	0.32	0.28
RR	**0.58**	0.55	0.46	0.50	0.43
Six Repeated Measurements (400 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	**0.48**	0.47	0.31	0.29	0.26
CS	0.40	**0.41**	0.28	0.28	0.23
AR1	**0.41**	0.38	0.29	0.26	0.21
RR	**0.51**	0.48	0.35	0.42	0.37
Eight Repeated Measurements (300 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	**0.40**	0.39	0.25	0.29	0.23
CS	**0.36**	0.35	0.25	0.22	0.18
AR1	**0.36**	0.36	0.22	0.23	0.21
RR	**0.49**	0.45	0.24	0.34	0.30

Table 4

Power comparisons when **one** *randomly selected SNP is causal and has* **only a gene-time interaction effect**. *Each cell represents the empirical power from 500 replicates at level α=0.05. The total number of observations is 2,400 and repeated measurements were recorded in the same follow-up period. Ind.: the repeated measurements are independent. CS: the correlation is compound symmetric. AR1: the repeated measurements follow a first-order auto-regressive model. RR: observations follow a mixed model with a random intercept and a random slope. LGRF-G: the LGRF test for the marginal effect of a gene. LGRF-J: the LGRF test for the joint effect of gene and gene-time interaction. The working correlation assumed in LGRF is CS. SKAT-Avg.: cross-sectional SKAT using the average value of repeated measurements as the outcome. GEE-G: test the marginal association by GEE. GEE-J: jointly test the marginal association and gene-time interaction by GEE. These single-marker tests were implemented by testing every SNP in the region and adjusting the minimum p-value by the Bonferroni correction.*

Power: Single SNP×Time Effect					
Four Repeated Measurements (600 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	0.38	**0.39**	0.29	0.21	0.20
CS	0.48	**0.54**	0.36	0.33	0.46
AR1	0.41	**0.49**	0.34	0.27	0.34
RR	0.53	**0.57**	0.39	0.42	0.50
Six Repeated Measurements (400 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	0.38	**0.43**	0.20	0.21	0.23
CS	0.33	**0.44**	0.19	0.17	0.37
AR1	0.31	**0.39**	0.21	0.16	0.21
RR	0.42	**0.50**	0.25	0.27	0.38
Eight Repeated Measurements (300 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	0.32	**0.36**	0.16	0.16	0.19
CS	0.25	**0.36**	0.16	0.12	0.30
AR1	0.25	**0.35**	0.14	0.13	0.16
RR	0.35	**0.44**	0.16	0.16	0.31

Table 5

Power comparisons when randomly selected **multiple** *SNPs are causal and have* **both marginal and interaction effects**. *Each cell represents the empirical power from 500 replicates at level α=0.05. The total number of observations is 2,400 and repeated measurements were recorded in the same follow-up period. Ind.: the repeated measurements are independent. CS: the correlation is compound symmetric. AR1: the repeated measurements follow a first-order auto-regressive model. RR: observations follow a mixed model with a random intercept and a random slope. LGRF-G: the LGRF test for the marginal effect of a gene. LGRF-J: the LGRF test for the joint effect of gene and gene-time interaction. The working correlation assumed in LGRF is CS. SKAT-Avg.: cross-sectional SKAT using the average value of repeated measurements as the outcome. GEE-G: test the marginal association by GEE. GEE-J: jointly test the marginal association and gene-time interaction by GEE. These single-marker tests were implemented by testing every SNP in the region and adjusting the minimum p-value by the Bonferroni correction.*

Power: Multiple SNPs Combined Effect					
Four Repeated Measurements (600 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	**0.36**	0.36	0.25	0.13	0.09
CS	**0.50**	0.49	0.37	0.19	0.18
AR1	**0.43**	0.42	0.35	0.19	0.17
RR	**0.60**	0.60	0.46	0.36	0.29
Six Repeated Measurements (400 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	**0.37**	0.36	0.21	0.15	0.11
CS	0.33	**0.35**	0.21	0.12	0.10
AR1	**0.32**	0.32	0.22	0.13	0.10
RR	**0.46**	0.43	0.24	0.22	0.15
Eight Repeated Measurements (300 Subjects)					
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
Ind.	0.30	**0.30**	0.17	0.11	0.11
CS	0.27	**0.29**	0.18	0.09	0.11
AR1	0.26	**0.28**	0.14	0.08	0.08
RR	0.40	**0.41**	0.20	0.19	0.15