Web-based Supplementary Materials for "Set-based Tests for Genetic Association in Longitudinal Studies" by

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

1. Detailed Proofs

1.1 Unbiasedness of the Estimating Equations

We show that the estimating function

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

is unbiased and the generalized score evaluated at $\gamma=0$, $U_\gamma(\boldsymbol{\beta},\eta,0)=(\boldsymbol{Y}-\boldsymbol{\mu})^T\boldsymbol{S}(\boldsymbol{I}-\eta\boldsymbol{T})(\boldsymbol{Y}-\boldsymbol{\mu})$ has mean 0 under H_0 and positive mean $\gamma E\{(\boldsymbol{Y}-\boldsymbol{\mu})^T\boldsymbol{S}^2(\boldsymbol{Y}-\boldsymbol{\mu})\}$ under $H_1:\gamma>0$. Below we denote $U_\gamma(\boldsymbol{\beta},\eta,\gamma)=\sum_{i,l}U_{\gamma,i,l}(\boldsymbol{\beta},\eta,\gamma)$ where

$$U_{\gamma,i,l}(\boldsymbol{\beta},\eta,\gamma)=\frac{\partial E(Y_{i,l}|\boldsymbol{Y}_{-(i,l)})}{\partial\gamma}\{Y_{i,l}-E(Y_{i,l}|\boldsymbol{Y}_{-(i,l)})\}.$$

Using an iterated expectation argument, we have

$$\begin{aligned}E\{U_{\gamma,i,l}(\boldsymbol{\beta},\eta,\gamma)\}&=E[E\{U_{\gamma,i,l}(\boldsymbol{\beta},\eta,\gamma)\}|\boldsymbol{Y}_{-(i,l)}]\\&=E[\frac{\partial E(Y_{i,l}|\boldsymbol{Y}_{-(i,l)})}{\partial\gamma}\{E(Y_{i,l}|\boldsymbol{Y}_{-(i,l)})-E(Y_{i,l}|\boldsymbol{Y}_{-(i,l)})\}]=0,\end{aligned}$$

where the second equality is because $\frac{\partial E(Y_{i,l}|\boldsymbol{Y}_{-(i,l)})}{\partial\gamma}$ is a function of $\boldsymbol{Y}_{-(i,l)}$. Therefore, under correct specification of the model, i.e., $E(\boldsymbol{Y}|\boldsymbol{Y}_-)=\boldsymbol{\mu}+(\eta\boldsymbol{T}+\gamma\boldsymbol{S})(\boldsymbol{Y}-\boldsymbol{\mu})$, the estimating

function $U_\gamma(\boldsymbol{\beta}, \eta, \gamma) = (\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}(\boldsymbol{I} - \eta\boldsymbol{T} - \gamma\boldsymbol{S})(\boldsymbol{Y} - \boldsymbol{\mu})$ is unbiased in the sense that it has expectation zero. Because of this unbiasedness, it follows that

$$E\{U_\gamma(\boldsymbol{\beta}, \eta, 0)\} = E\{U_\gamma(\boldsymbol{\beta}, \eta, \gamma)\} + \gamma E\{(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}^2(\boldsymbol{Y} - \boldsymbol{\mu})\} = \gamma E\{(\boldsymbol{Y} - \boldsymbol{\mu})^T \boldsymbol{S}^2(\boldsymbol{Y} - \boldsymbol{\mu})\}$$

which equals 0 under H_0 and is positive under $H_1 : \gamma > 0$. So a large value of $U_\gamma(\boldsymbol{\beta}, \eta, 0)$ supports the alternative hypothesis. Furthermore, one can show that

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

are both unbiased by similar argument.

1.2 Asymptotic Representation of Q_G

We note that, for both GR and IBS similarity, $\boldsymbol{S}$ can be written as $\boldsymbol{Z}\boldsymbol{Z}^T + \boldsymbol{C}$, where $\boldsymbol{C} = -diag(\boldsymbol{Z}\boldsymbol{Z}^T)$ and is needed because in the definition of $\boldsymbol{S}$, subjects are not compared to themselves in terms of genetic similarity. For example, for GR similarity, $\boldsymbol{Z}(n \times q)$, is the centered genotype matrix, i.e., each column of the genotype matrix $\boldsymbol{G}$, $\boldsymbol{G}_{,h}$, is now centered by the genotype population mean $2p_h$, and for IBS similarity, $\boldsymbol{Z}$ is an $n \times 3q$ matrix again with each element defined in terms of genotype, described in the next subsection. Here we prove the following result.

Result 1.

$$Q_G = \frac{(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})^T \boldsymbol{S}(\boldsymbol{I} - \widehat{\eta}\boldsymbol{T})(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})}{m} = \frac{(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})^T \boldsymbol{Z}\boldsymbol{Z}^T(\boldsymbol{I} - \eta_0\boldsymbol{T})(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})}{m} + c + o_p(1),$$

where $\widehat{\boldsymbol{\mu}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$; $\widehat{\eta}$ and $\widehat{\boldsymbol{\beta}}$ are the solution to estimating equations

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

Proof. We first note that

$$Q_G = \frac{(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})^T \boldsymbol{Z}\boldsymbol{Z}^T(\boldsymbol{I} - \widehat{\eta}\boldsymbol{T})(\boldsymbol{Y} - \widehat{\boldsymbol{\mu}})}{m} + \frac{1}{m}\sum_{i=1}^{m} c_i(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i)^T(\boldsymbol{I}_{n_i} - \widehat{\eta}\boldsymbol{T}_i)(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i), \quad (1)$$

where c_i is the (i,i)-th element of $\boldsymbol{C}$ which equals $-\sum_{h=1}^{q}(G_{i,h} - 2p_h)^2$ for GR similarity and $-2q$ for IBS similarity; $\boldsymbol{T}_i$ is the (i,i)-th block of $\boldsymbol{T}$; $\boldsymbol{I}_{n_i}$ is an $n_i \times n_i$ identity matrix.

The first term in equation (1) is an inner product of $\frac{1}{\sqrt{m}}\boldsymbol{Z}^T(\boldsymbol{I}-\widehat{\eta}\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})$ and $\frac{1}{\sqrt{m}}\boldsymbol{Z}^T(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})$. We show that

$$\begin{aligned}
& \frac{1}{\sqrt{m}}\boldsymbol{Z}^T(\boldsymbol{I}-\widehat{\eta}\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}}) \\
= & \frac{1}{\sqrt{m}}\boldsymbol{Z}^T(\boldsymbol{I}-\eta_0\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}}) - \frac{1}{\sqrt{m}}\boldsymbol{Z}^T\boldsymbol{T}(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})(\widehat{\eta}-\eta_0) \\
= & \frac{1}{\sqrt{m}}\boldsymbol{Z}^T(\boldsymbol{I}-\eta_0\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}}) - \sqrt{m}(\widehat{\eta}-\eta_0)\{\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{Z}_i^T\boldsymbol{T}_i(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)+\boldsymbol{o}_p(1)\} \\
= & \frac{1}{\sqrt{m}}\boldsymbol{Z}^T(\boldsymbol{I}-\eta_0\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}}) + \boldsymbol{o}_p(1), \text{ as } m\to\infty,
\end{aligned}$$

where the second equality follows by Taylor expansion at $\boldsymbol{\beta}$. Assuming that the number of repeated measurements of each subject is bounded, the estimator $\widehat{\eta}$ is $\sqrt{m}$-consistent for η under H_0 by the property of generalized estimating equations. Hence the last equality follows by the $\sqrt{m}$-consistency of $\widehat{\eta}$ and the weak law of large numbers. Therefore,

$$\frac{(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})^T\boldsymbol{Z}\boldsymbol{Z}^T(\boldsymbol{I}-\widehat{\eta}\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})}{m} = \frac{(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})^T\boldsymbol{Z}\boldsymbol{Z}^T(\boldsymbol{I}-\eta_0\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})}{m} + o_p(1).$$

Next we show that the second term in equation (1) asymptotically converges to a constant.

$$\begin{aligned}
& \frac{1}{m}\sum_{i=1}^{m} c_i(\boldsymbol{Y}_i-\widehat{\boldsymbol{\mu}}_i)^T(\boldsymbol{I}_{n_i}-\widehat{\eta}\boldsymbol{T}_i)(\boldsymbol{Y}_i-\widehat{\boldsymbol{\mu}}_i) \\
= & \frac{1}{m}\sum_{i=1}^{m} c_i(\boldsymbol{Y}_i-\widehat{\boldsymbol{\mu}}_i)^T(\boldsymbol{I}_{n_i}-\eta_0\boldsymbol{T}_i)(\boldsymbol{Y}_i-\widehat{\boldsymbol{\mu}}_i) - (\widehat{\eta}-\eta_0)\frac{1}{m}\sum_{i=1}^{m} c_i(\boldsymbol{Y}_i-\widehat{\boldsymbol{\mu}}_i)^T\boldsymbol{T}_i(\boldsymbol{Y}_i-\widehat{\boldsymbol{\mu}}_i) \\
= & \frac{1}{m}\sum_{i=1}^{m} c_i(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)^T(\boldsymbol{I}_{n_i}-\eta_0\boldsymbol{T}_i)(\boldsymbol{Y}_i-\boldsymbol{\mu}_i) - (\widehat{\eta}-\eta_0)\frac{1}{m}\sum_{i=1}^{m} c_i(\boldsymbol{Y}_i-\boldsymbol{\mu}_i)^T\boldsymbol{T}_i(\boldsymbol{Y}_i-\boldsymbol{\mu}_i) + o_p(1) \\
= & c + o_p(1),
\end{aligned}$$

where the second equality is again by Taylor expansion and the last equality by the weak law of large numbers. Summarizing results, we have finished proving the result:

$$Q_G = \frac{(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})^T\boldsymbol{S}(\boldsymbol{I}-\widehat{\eta}\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})}{m} = \frac{(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})^T\boldsymbol{Z}\boldsymbol{Z}^T(\boldsymbol{I}-\eta_0\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})}{m} + c + o_p(1).$$

□

1.3 Robust Inference

We define $\widetilde{\boldsymbol{Z}}(\eta) = \{(\boldsymbol{I}-\eta\boldsymbol{T})\boldsymbol{Z}, \boldsymbol{Z}\}$, $\widetilde{\boldsymbol{X}}(\eta) = (\boldsymbol{I}-\eta\boldsymbol{T})\boldsymbol{X}$, and

$$\begin{aligned}\frac{1}{\sqrt{m}}\widetilde{\boldsymbol{R}}(\eta,\boldsymbol{\beta}) &= \frac{1}{\sqrt{m}}\left\{\begin{array}{c}\boldsymbol{R}_1(\eta,\boldsymbol{\beta})\\ \boldsymbol{R}_2(\eta,\boldsymbol{\beta})\end{array}\right\} = \frac{1}{\sqrt{m}}\left\{\begin{array}{c}\widetilde{\boldsymbol{Z}}(\eta)^T\\ \widetilde{\boldsymbol{X}}(\eta)^T\end{array}\right\}(\boldsymbol{Y}-\boldsymbol{\mu})\\ &= \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left\{\begin{array}{c}\widetilde{\boldsymbol{Z}}_i(\eta)^T\\ \widetilde{\boldsymbol{X}}_i(\eta)^T\end{array}\right\}(\boldsymbol{Y}_i-\boldsymbol{\mu}_i) = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{\boldsymbol{R}}_i(\eta,\boldsymbol{\beta}),\end{aligned}$$

where $\widetilde{\boldsymbol{Z}}_i(\eta)$, $\widetilde{\boldsymbol{X}}_i(\eta)$, $\boldsymbol{Y}_i$, $\boldsymbol{\mu}_i = \boldsymbol{X}_i^T\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{R}}_i(\eta,\boldsymbol{\beta})$ are the components corresponding to subject i respectively. We can rewrite $(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})^T\boldsymbol{Z}\boldsymbol{Z}^T(\boldsymbol{I}-\eta_0\boldsymbol{T})(\boldsymbol{Y}-\widehat{\boldsymbol{\mu}})/m$ as the quadratic form of $\frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0,\widehat{\boldsymbol{\beta}})$ by straightforward matrix algebra and have:

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

where $\boldsymbol{I}_{dq}$ is a $dq\times dq$ identity matrix and $\boldsymbol{0}_{dq}$ is a $dq\times dq$ matrix with all elements 0. In this subsection we prove the following and a directly followed results. We note that the proof does not rely on the normality assumption of outcomes $\boldsymbol{Y}$.

Result 2. *Under the $H_0:\gamma=0$,*

$$\frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0,\widehat{\boldsymbol{\beta}}) = \boldsymbol{A}\frac{1}{\sqrt{m}}\widetilde{\boldsymbol{R}}(\eta_0,\boldsymbol{\beta}_0) + \boldsymbol{o}_p(1) \Rightarrow N(0,\boldsymbol{\Sigma}),$$

where $\boldsymbol{A} = (\boldsymbol{I}_{2dq}, -E\{\frac{\partial\boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}^T}\}E^{-1}\{\frac{\partial\boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}^T}\})$, $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T$, and $\boldsymbol{D} = var\{\widetilde{\boldsymbol{R}}_i(\eta_0,\boldsymbol{\beta}_0)\}$. And $\boldsymbol{\Sigma}$ can be consistently estimated by the sandwich variance estimator $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{A}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{A}}^T$, where $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{D}}$ are the corresponding empirical counterpart defined below.

Proof. Note that $\widehat{\boldsymbol{\beta}}$ is the solution to $\boldsymbol{R}_2(\widehat{\eta},\boldsymbol{\beta}) = 0$, i.e., $\frac{1}{\sqrt{m}}\boldsymbol{R}_2(\widehat{\eta},\widehat{\boldsymbol{\beta}}) = 0$. We first show that $0 = \frac{1}{\sqrt{m}}\boldsymbol{R}_2(\widehat{\eta},\widehat{\boldsymbol{\beta}}) = \frac{1}{\sqrt{m}}\boldsymbol{R}_2(\eta_0,\widehat{\boldsymbol{\beta}}) + \boldsymbol{o}_p(1)$. It follows because, by Taylor expansion,

$$\frac{1}{\sqrt{m}}\boldsymbol{R}_2(\widehat{\eta},\widehat{\boldsymbol{\beta}}) = \frac{1}{\sqrt{m}}\boldsymbol{R}_2(\eta_0,\widehat{\boldsymbol{\beta}}) - \frac{1}{m}\sum_{i=1}^{m}\boldsymbol{X}_i^T\boldsymbol{T}_i(\boldsymbol{Y}_i-\boldsymbol{X}_i\widehat{\boldsymbol{\beta}})\sqrt{m}(\widehat{\eta}-\eta_0),$$

and note $\sqrt{m}(\widehat{\eta}-\eta_0)$ is bounded in probability and $\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{X}_i^T\boldsymbol{T}_i(\boldsymbol{Y}_i-\boldsymbol{X}_i\widehat{\boldsymbol{\beta}})$ converges in probability to $E\{\boldsymbol{X}_i^T\boldsymbol{T}_i(\boldsymbol{Y}_i-\boldsymbol{X}_i\boldsymbol{\beta}_0)\} = 0$, where $\boldsymbol{\beta}_0$ is the true parameter under H_0.

By Taylor expansion,

$$\begin{aligned}\frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0,\widehat{\boldsymbol{\beta}}) &= \frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0) + \frac{1}{\sqrt{m}}\frac{\partial\boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}^T}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \boldsymbol{o}_p(1),\\ 0 = \frac{1}{\sqrt{m}}\boldsymbol{R}_2(\eta_0,\widehat{\boldsymbol{\beta}}) + \boldsymbol{o}_p(1) &= \frac{1}{\sqrt{m}}\boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0) + \frac{1}{\sqrt{m}}\frac{\partial\boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}^T}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + \boldsymbol{o}_p(1),\end{aligned}$$

Plugging the second equation into the first,

$$\begin{aligned}\frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0,\widehat{\boldsymbol{\beta}}) &= \frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0) - \frac{\partial \boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\{\frac{\partial \boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\}^{-1}\frac{1}{\sqrt{m}}\boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0) + \boldsymbol{o}_p(1) \\ &= (\boldsymbol{I}_{2dq}, -\frac{\partial \boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\{\frac{\partial \boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\}^{-1})\frac{1}{\sqrt{m}}\widetilde{\boldsymbol{R}}(\eta_0,\boldsymbol{\beta}_0) + \boldsymbol{o}_p(1).\end{aligned}$$

It is easy to see that

$$\frac{1}{m}\frac{\partial \boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\{\frac{1}{m}\frac{\partial \boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\}^{-1} = E\{\frac{\partial \boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\}E^{-1}\{\frac{\partial \boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\} + \boldsymbol{o}_p(1). \qquad (2)$$

Thus

$$\begin{aligned}\frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0,\widehat{\boldsymbol{\beta}}) &= (\boldsymbol{I}_{2dq}, -E\{\frac{\partial \boldsymbol{R}_1(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\}E^{-1}\{\frac{\partial \boldsymbol{R}_2(\eta_0,\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T}\})\frac{1}{\sqrt{m}}\widetilde{\boldsymbol{R}}(\eta_0,\boldsymbol{\beta}_0) + \boldsymbol{o}_p(1) \\ &= \boldsymbol{A}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{\boldsymbol{R}}_i(\eta_0,\boldsymbol{\beta}_0) + \boldsymbol{o}_p(1),\end{aligned}$$

which by the central limit theory converges to a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T$. It is easy to check that, by the weak law of large numbers and the $\sqrt{m}$-consistency of $\widehat{\eta}$ and $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}$ can be consistently estimated by the sandwich variance estimator $\widehat{\boldsymbol{A}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{A}}^T$, where

$$\widehat{\boldsymbol{D}} = \frac{1}{m}\sum_{i=1}^{m}\widetilde{\boldsymbol{R}}_i(\widehat{\eta},\widehat{\boldsymbol{\beta}})\widetilde{\boldsymbol{R}}_i(\widehat{\eta},\widehat{\boldsymbol{\beta}})^T,$$

$$\widehat{\boldsymbol{A}} = (\boldsymbol{I}_{2dq}, -\frac{\partial \boldsymbol{R}_1(\widehat{\eta},\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T}\{\frac{\partial \boldsymbol{R}_2(\widehat{\eta},\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T}\}^{-1}).$$

Specifically, $\frac{\boldsymbol{R}_1(\widehat{\eta},\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T} = -\widetilde{\boldsymbol{Z}}(\widehat{\eta})^T\boldsymbol{X}$ and $\frac{\partial \boldsymbol{R}_2(\widehat{\eta},\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T} = -\widetilde{\boldsymbol{X}}(\widehat{\eta})^T\boldsymbol{X}$. □

Result 3. *Under regularity conditions, Q_G has an asymptotic distribution*

$$\frac{1}{2}\sum_{i=1}^{2dq}\lambda_i\chi_i^2 + c, \qquad (3)$$

where c is a constant which does not affect the inference; χ_i^2s are i.i.d. Chi-square distributions; λ_i are eigenvalues of a $2dq \times 2dq$ matrix

$$\begin{pmatrix} \boldsymbol{0}_{dq} & \boldsymbol{I}_{dq} \\ \boldsymbol{I}_{dq} & \boldsymbol{0}_{dq} \end{pmatrix}\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \\ \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \end{pmatrix};$$

$\boldsymbol{\Sigma}$ is defined above and can be consistently estimated by $\widehat{\boldsymbol{\Sigma}}$ as in Result 2; $\boldsymbol{\Sigma}_{11}$ is the first $dq \times dq$ block of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$, $\boldsymbol{\Sigma}_{22}$ can be defined similarly.

Proof. In the proof of Result 2, we have showed $\frac{1}{\sqrt{m}}\boldsymbol{R}_1(\eta_0, \widehat{\boldsymbol{\beta}}) \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma})$ under $H_0 : \gamma = 0$. Therefore,

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

is asymptotically distributed as

$$\frac{1}{2}\sum_{i=1}^{2dq} \lambda_i \chi_i^2 + c$$

by the property of quadratic form of normal random variables. In addition, $\boldsymbol{\Sigma}$ can be consistently estimated by $\widehat{\boldsymbol{\Sigma}}$ as we showed in Result 2.

□

1.4 Use of IBS metric

As discussed in the main paper, the Identity-by-state (IBS) similarity: $s_{i,j} = \rho(\boldsymbol{G}_i, \boldsymbol{G}_j) = \sum_{h=1}^{q}(2 - |G_{i,h} - G_{j,h}|)$, which has been commonly used, e.g., in SKAT, is an alternative choice to quantify genetic similarity in LGRF. However, the use of IBS kernel is limited by its computational inefficiency (Wu, et al., 2011), though they have recognized that IBS kernel usually has higher power than linear kernel in the presence of gene-gene interaction. We hereby propose a fast implementation of IBS metric in LGRF, and as a result both the robustness to misspecification of working correlation structure and computational efficiency can be achieved. Recall the genotype of a single genetic variant of subject i can be coded by {0, 1, 2}. We generate three pseudo-variables by

$$\begin{matrix} 0: \\ 1: \\ 2: \end{matrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 1 \\ 0 & \sqrt{2} & 0 \end{pmatrix} =: \boldsymbol{B}.$$

That is, the pseudo-variables are $(\sqrt{2}, 0, 0)$ if the genotype is 0; $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 1)$ if it is 1; $(0, \sqrt{2}, 0)$ if it is 2. The inner-product of two subjects' pseudo-variables exactly equal the IBS metric; that is,

$$\boldsymbol{B}\boldsymbol{B}^T = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

If we denote the pseudo-variables with respect to p genetic variants as $\boldsymbol{Z}_{IBS}$, an $n \times 3q$ matrix, the IBS metric between all pairs of subjects are $\boldsymbol{Z}_{IBS}\boldsymbol{Z}_{IBS}^T$. Therefore, genetic similarity in terms of the IBS metric can be represented as $\boldsymbol{S} = \boldsymbol{Z}_{IBS}\boldsymbol{Z}_{IBS}^T - \boldsymbol{C}$, where $\boldsymbol{C} = -diag(\boldsymbol{Z}_{IBS}\boldsymbol{Z}_{IBS}^T)$; again note the term $\boldsymbol{C}$ is due to that in the definition of $\boldsymbol{S}$, subjects are not compared to themselves in terms of genetic similarity. By using peudo-variables the computational efficiency increases dramatically, but is still slightly less compared with using the genetic relationship similarity because the number of variables increase from q to $3q$. This representation shows that the IBS metric corresponds to a linear model in SKAT with $3q$ pseudo variables, which actually does not model the interaction among genetic variants.

2. Additional Simulations

2.1 LGRF Run-time Simulation

To evaluate the computational performance of LGRF, we varied the number of total observations and recorded the running times for both fitting the null model and testing the target region C10orf107 (154 SNPs). The numbers of total observations were set to be 3000, 6000 and 10000, mimicking the total number of observations in CHN, HIS/AFA and CAU ethnic groups in MESA respectively. Supplementary Table 1 shows the running times for testing a region with 154 SNPs on a 2.67GHz Linux PC with an Intel Xeon X5650 processor. The numbers of total observations (n) are 3000, 6000 and 10000, akin to those observed in the CHN, HIS/AFA and CAU ethnic groups respectively in MESA. For a longitudinal study containing 3000 observations, like the CHN ethnic group in MESA, LGRF-G requires 2.0 seconds to fit the null model and 0.6 seconds to calculate the p-value for the entire target region. Since the null model only need to be fit once, the computational cost for testing K regions is approximately $0.6 \times K$ seconds. We expect that the computational cost will increase if number of SNPs in the region increases, but this number is usually bounded by the length of the region. For example, the largest candidate region in our analysis has 1026 SNPs. The LGRF-J test requires longer time for calculating p-value because additional interaction terms are explicitly included. On the other hand, the running time increases as the number of observations increases. If the number of total observations is increased to 10000, such as the CAU ethnic in MESA, LGRF-G requires 9.7 seconds to fit the null model and 1.9 seconds to compute the p-value.

Supplementary Table 1: Running times corresponding to different number of total observations. The running times for both fitting the null model and testing the target region C10orf107 (190 SNPs) on a 2.67GHz Linux PC with an Intel Xeon X5650 processor are showed in this table. The numbers of total observations (n) are 3000, 6000 and 10000, approximating the numbers corresponding to CHN, HIS/AFA and CAU ethnic groups respectively, as observed in MESA. CAU: Caucasians; AFA: African Americans; HIS: Hispanics; CHN: Asians of Chinese descent.

Number of Total	Fitting the	Calculating the P-value	
Observations (n)	Null Model	LGRF-G	LGRF-J
3000	2.0 seconds	0.6 seconds	2.0 seconds
6000	4.9 seconds	1.1 seconds	3.4 seconds
10000	9.7 seconds	1.9 seconds	5.0 seconds

2.2 Simulations Investigating Meta/Mega-analysis Strategies with a Multi-Ethnic Cohort

We additionally simulated scenarios where four ethnic groups shared the same set of causal variants versus different set of causal variants in the same target region to compare meta and mega analysis, and show the advantage of gene-level meta-analysis over single-SNP meta-analysis. The gene-level meta-analysis evaluated the region for each race ethnicity by LGRF-G and combine the p-values by fisher's method. The single-SNP meta-analysis approach used the popular meta-analysis tool METAL proposed by Willer et al. (2010). Each SNP was tested using GEE-G within each ethnicity and the four Z-scores converted from the four ethnic groups p-values were then combined to provide overall measures of significance. The minimum p-value was then adjusted for multiple testing by the Bonferroni correction. The mega-analysis was pooling the ethnic groups together and then applying LGRF-G and GEE-G using individual level data.

For each replicated dataset, four ethnic groups were randomly simulated from the CAU, AFA, HIS and CHN ethnic groups correspondingly, and gene region C10orf107 was chosen as the target region. The total number of subjects was 1000 and each subject had four repeated measurements. The sample sizes corresponding to the simulated ethnic groups were 400, 250, 220 and 130 respectively, proportional to those observed in MESA. A different SNP was randomly chosen to be causal within each ethnic group in the case of distinct effects, while four ethnic groups shared the same causal SNP in the common effect case. Specifically, the

true model is of form:

$$Y_{i,l} = \alpha_0 t_{i,l} + \alpha_1 G_{E,i} + \epsilon_{i,l}, t_{i,l} = 1, \ldots, r,$$

where $G_{E,i}$ is the genotype of subject i for the randomly selected causal SNP of the ethnicity E that subject i belongs to; $\alpha_0 = 12/r$, $\alpha_1 = 0.4$; r is the number of measurements per subject. The missingness indicators and other simulation parameters were almost the same as the power simulation scenario I where we considered a single causal SNP that had a marginal effect and the within-subject correlation structure was CS.

Supplementary Table 2 presents the comparisons. When the four ethnic groups have different causal variants, gene-level meta-analysis shows substantial higher power (0.832) than single-SNP meta-analysis (0.520). This is intuitive because single-SNP meta-analysis will dilute the signal of each causal variant as the strength is not uniform across each cohort at the SNP level. Moreover, a gene-level meta-analysis is preferred here than a mega-analysis using individual level data for the same reason that pooling the data together will dilute the signal. When the four ethnic groups have the same causal variant, gene-level meta-analysis achieves slightly lower power (0.724 vs. 0.782), because the signal was accumulated on the same variant while combining the four groups.

Supplementary Table 2: Power Studies for Meta/Mega-analysis when Causal Variants are Distinct/Common across Four Ethnic Groups. Each cell represents the empirical power from 500 replicates at level α=0.05. In each ethnic group, one randomly selected SNP is causal and has a marginal effect. LGRF-meta: use LGRF-G to test the region within ethnicity and combine the p-values by Fishers method. LGRF-mega: jointly tests the four ethnic groups by pooling the individual level data. GEE-meta: use GEE-G to test each SNP within ethnicity and combine the p-values by METAL proposed by Willer, et al. (2010). GEE-mega: test each SNP using the individual level data of four ethnic groups jointly.

Causal Variants	Meta-analysis		Mega-analysis	
	LGRF-meta	GEE-meta	LGRF-mega	GEE-mega
Distinct across ethnic groups	0.832	0.520	0.614	0.558
Common across ethnic groups	0.724	0.782	0.754	0.820

2.3 Impact of Genetic Similarity Metrics

We evaluated the impact of the genetic similarity metrics in three main scenarios considered in table 3-5 of the main text and summarize the result in Supplementary Table 3. The number of repeated measurements per subject is six, and there are 400 subjects in each replicate. The correlation structure among repeated measurements is compound symmetric. Detailed parameters are same as the three power simulation settings in the main text respectively. The IBS similarity has analogous performance as genetic relationship in the simulation studies considered in the paper.

Supplementary Table 3: Power Studies for evaluating the impact of the genetic similarity metric: Each cell represents the empirical power from 500 replicates at level α=0.05. The number of repeated measurements per subject is six, and there are 400 subjects in each replicate. The correlation structure among repeated measurements is compound symmetric. The parameter configurations are same as the three simulation settings (Tables 3-5) in the main text. 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. Both LGRF and LGRF-J use the genetic similarity metric. LGRF-G-IBS: the LGRF test for the marginal effect of a gene using the identity-by state (IBS) similarity. LGRF-J-IBS: the LGRF test for the joint effect of gene and gene-time interaction. The genetic main effect is modeled using IBS similarity.

Simulation Scenario	Marginal Association Test		Joint Association Test	
	LGRF-G	LGRF-G-IBS	LGRF-J	LGRF-J-IBS
Single SNP Marginal Effect	0.426	0.438	0.417	0.364
Single SNP×Time effect	0.326	0.318	0.430	0.486
Multiple SNPs Combined Effect	0.330	0.324	0.354	0.344

2.4 Further Evaluation of the Power Difference

Given the analogous quadratic form of the LGRF score test and SKAT test, we expect they will have similar power if we have a longitudinal version of SKAT even if they are developed from two different perspective. To confirm this, we applied the LGRF test to the average of repeated measurements to three main scenarios considered in table 3-5 of the main text and present the result in Supplementary Table 4. We also included the GenRF test for comparison. The number of repeated measurements per subject is six, and there are 400 subjects in each replicate. The correlation structure among repeated measurements is compound symmetric. Detailed parameters are same as the three simulation settings in the main text respectively. We observed that GenRF, SKAT and LGRF have comparable power in the scenarios considered here when they are all applied to the average of repeated measurements. This shows that the longitudinal design is the main reason of the power difference.

Supplementary Table 4: Power Studies including GenRF: Each cell represents the empirical power from 500 replicates at level α=0.05. The number of repeated measurements per subject is six, and there are 400 subjects in each replicate. The correlation structure among repeated measurements is compound symmetric. Detailed parameters are same as the three simulation settings (Table 3-5) in the main text respectively. LGRF-G: the LGRF test for the marginal effect of a gene using longitudinal dats. LGRF-Avg.: the LGRF test applied to the average of repeated measurements. GenRF-Avg.: the GenRF test applied to the average of repeated measurements. SKAT-Avg.: the SKAT test applied to the average of repeated measurements.

Causal Effect		Based on Average		
	LGRF-G	LGRF-Avg.	GenRF-Avg.	SKAT-Avg.
Single SNP Marginal Effect	0.426	0.290	0.287	0.290
Single SNP×Time effect	0.326	0.196	0.174	0.186
Multiple SNPs Combined Effect	0.330	0.214	0.196	0.208

2.5 Type - I Error Rate Evaluation at Lower Significance levels

We further evaluated LGRF-G (the LGRF test for the marginal effect of a gene) at a lower α level using 2.5×10^7 replicates. The smaller α level (2.5×10^{-6}) considered here reflects the scenario of a genome-wide gene-level analysis where we have approximately 20,000 genes in total. Other parameters are held same as the type I error simulations in the main text (with an α level of 0.001). We present the results in the Supplementary Table 5. As we expected, LGRF-G tends to be conservative in these scenarios due to the use of sandwich estimator as in regular GEE, which has been known to be slightly conservative.

Supplementary Table 5: Type-I error rate evaluation at small α level. Each cell represents the empirical type-I error rate of LGRF-G (the LGRF test for the marginal effect of a gene) based on 2.5×10^7 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. The working correlation assumed in LGRF-G is CS.

Type-I Error Rate						
	Four Repeated Measurements (600 Subjects)					
$\alpha =$	0.05	10^{-3}	10^{-4}	10^{-5}	2.5×10^{-6}	10^{-6}
Ind.	0.0494	9.03×10^{-4}	7.79×10^{-5}	5.20×10^{-6}	9.20×10^{-7}	5.20×10^{-7}
CS	0.0493	8.98×10^{-4}	7.95×10^{-5}	5.20×10^{-6}	6.39×10^{-7}	4.33×10^{-7}
AR1	0.0494	8.98×10^{-4}	7.76×10^{-5}	5.12×10^{-6}	6.26×10^{-7}	4.63×10^{-7}

2.6 Power Evaluation at Lower Significance levels

We evaluated how power changes at $\alpha = 0.001$ and $\alpha = 2.5 \times 10^{-6}$ when the number of subjects increases from 1200 to 6000, and each subject has four repeated measurements. The α level considered here either approximates the scenario in our data analysis, in which we consider a replication study with 29 regions, or reflect the scenario of a genome-wide gene-level analysis where we have approximately 20,000 genes in total. The simulation scenario is similar to Table 5 in the main text, where 10 out of the 154 SNPs in the region 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 parameters are held same as Table 5 in the main text except the sample size, such that the total variation in the outcome explained by the SNPs (including gene-time interaction) is approximately 1.5% - 2.0%. We present the results in Supplementary Table 6 and 7. We observed that the relative power difference is similar to what we showed in Table 5 of the main text.

Supplementary Table 6: Power comparisons when the number of subjects ranges from 1,200 to 6,000 and four repeated measurements were recorded. 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=0.001. 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					
			Ind.		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
1200	0.18	0.17	0.10	0.08	0.06
1800	0.32	0.30	0.21	0.15	0.11
2400	0.40	0.42	0.26	0.25	0.20
3600	0.63	0.65	0.49	0.48	0.41
6000	0.82	0.82	0.71	0.71	0.65
			CS		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
1200	0.24	0.24	0.17	0.11	0.08
1800	0.37	0.37	0.23	0.19	0.15
2400	0.50	0.53	0.38	0.32	0.29
3600	0.70	0.70	0.58	0.54	0.50
6000	0.83	0.84	0.77	0.76	0.73
			AR1		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
1200	0.32	0.36	0.19	0.17	0.13
1800	0.52	0.54	0.35	0.32	0.28
2400	0.60	0.62	0.48	0.44	0.39
3600	0.80	0.84	0.65	0.67	0.65
6000	0.90	0.91	0.84	0.86	0.85
			RR		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
1200	0.49	0.47	0.28	0.31	0.23
1800	0.62	0.62	0.44	0.47	0.40
2400	0.75	0.74	0.55	0.59	0.53
3600	0.86	0.85	0.75	0.81	0.77
6000	0.94	0.93	0.88	0.91	0.91

Supplementary Table 7: Power comparisons when the number of subjects ranges from 3,600 to 9,600 and four repeated measurements were recorded. Randomly selected multiple SNPs are causal and have both marginal and interaction effects. Each cell represents the empirical power from 500 replicates at level $\alpha = 2.5 \times 10^{-6}$. 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					
			Ind.		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
3600	0.25	0.23	0.17	0.19	0.15
4800	0.44	0.45	0.31	0.37	0.31
6000	0.51	0.49	0.35	0.41	0.37
7200	0.62	0.62	0.48	0.57	0.53
9600	0.78	0.78	0.65	0.74	0.70
			CS		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
3600	0.31	0.32	0.21	0.26	0.22
4800	0.46	0.47	0.36	0.42	0.40
6000	0.58	0.57	0.47	0.53	0.51
7200	0.66	0.67	0.57	0.63	0.60
9600	0.80	0.79	0.74	0.78	0.76
			AR1		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
3600	0.42	0.46	0.31	0.38	0.35
4800	0.60	0.64	0.44	0.55	0.52
6000	0.72	0.74	0.60	0.67	0.66
7200	0.77	0.78	0.68	0.76	0.76
9600	0.84	0.86	0.77	0.83	0.82
			RR		
	LGRF-G	LGRF-J	SKAT-Avg.	GEE-G	GEE-J
3600	0.58	0.57	0.42	0.54	0.50
4800	0.71	0.71	0.54	0.70	0.65
6000	0.81	0.80	0.69	0.80	0.78
7200	0.89	0.89	0.79	0.88	0.86
9600	0.89	0.90	0.84	0.91	0.89

3. Descriptive Statistics of MESA

Supplementary Table 8: Gender distribution of MESA subjects across site and race. Each cell represents the number of subject in the corresponding category. WFU: Wake Forest University, Winston Salem, NC; COL: Columbia University, New York, NY; JHU: Johns Hopkins University, Baltimore, MD; UMN: University of Minnesota, Twin Cities, MN; NWU: Northwestern University, Chicago, IL; UCLA: University of California Lost Angeles, Los Angeles, CA.

	Gender		
Site	Female	Male	All
WFU	528	464	992
COL	536	434	970
JHU	556	488	1044
UMN	532	518	1050
NWU	551	508	1059
UCLA	666	648	1314
All	3369	3060	6429

	Gender		
Race	Female	Male	All
White/Caucasian	1321	1206	2527
Chinese American	394	381	775
Black/African-American	906	771	1677
Hispanic	748	702	1450
All	3369	3060	6429

Supplementary Table 9: Longitudinal summary of blood pressure phenotypes and covariates we adjusted for in MESA across four exams. Sd: standard deviation. N: number of subject. sBP: systolic blood pressure. dBP: diastolic blood pressure. BMI: body mass index.

	Exam 1 (24 months)			Exam 2 (18 months)			Exam 3 (18 months)			Exam 4 (24 months)		
	Mean	Sd	N	Mean	Sd	N	Mean	Sd	N	Mean	Sd	N
sBP (mm Hg)	126.51	21.55	6427	124.33	20.79	5898	123.16	20.58	5619	123.59	20.56	5399
dBP (mm Hg)	71.82	10.27	6427	70.37	10.09	5898	69.69	9.94	5619	69.61	10.05	5399
BMI (kg/m^2)	28.3	5.47	6429	28.33	5.48	5889	28.28	5.51	5621	28.38	5.58	5402
Age (years)	62.22	10.24	6429	63.69	10.1	5900	64.99	9.99	5628	66.51	9.94	5505

Supplementary Table 10: Sensitivity analysis of the top 5 principal components (PCs) in MESA. Each cell represents the p-value. The analysis was done by fitting a multivariate linear regression using exam 1 data in MESA.

Race	PC1	PC2	PC3	PC4	PC5
White/Caucasian	0.0002	0.3502	0.2895	0.8484	0.0315
Chinese American	< 0.0001	0.0439	0.7882	0.9913	0.1982
Black/African-American	0.0029	0.3976	0.1406	0.3066	0.3808
Hispanic	0.9428	0.0306	0.4939	0.1760	0.6313

4. Detailed Data Analysis of MESA

This section reports the full data analysis results of analyzing 29 candidate regions using the data from MESA. In addition to LGRF tests and SKAT, we also carried out an individual SNP based analysis (MinP) by testing every SNP in the region using GEE and adjusting the minimum p-value for multiple testing correction by multiplying with the effective number of independent tests explaining 99.95% variation (Gao, Starmer, and Martin, 2008). This proportion was determined by simulation such that the type-I error rate is neither inflated nor conservative (data not shown). We note that the preservation of nominal type-I error levels cannot be ensured because the real data scenarios can be very different from the simulated data due to the various effect sizes, proportions of causal variants, LD structures and so on. Supplementary Table 12 - 19 show the results of multi-ethnic groups analysis and table 20 - 21 show the results of meta-analysis.

References

[1] Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26(17),** 2190–2191.

Supplementary Table 11: Chromosomal Region Information for the 29 regions considered in the MESA analysis.

Region Name	Chromosome	Start	End	Index SNP	Nearest Gene	Coded Allele Frequency
MOV10	1	113012286	113049891	rs2932538	MOV10	0.75
rs13082711	3	27462913	27562913	rs13082711	SLC4A7	0.78
MECOM	3	170278981	170869100	rs419076	MECOM	0.47
SLC39A8	4	103386221	103576438	rs13107325	SLC39A8	0.05
GUCY1A3	4	156802313	156877951	rs13139571	GUCY1A3,GUCY1B3	0.76
rs1173771	5	32800785	32900785	rs1173771	NPR3,C5orf23	0.6
rs11953630	5	157727980	157827980	rs11953630	EBF1	0.37
HFE	6	26190488	26211550	rs1799945	HFE	0.14
rs805303	6	31674345	31774345	rs805303	BAT2,BAT5	0.61
rs4373814	10	18409978	18509978	rs4373814	CACNB2	0.55
PLCE1	10	95738736	96083139	rs932764	PLCE1	0.44
rs7129220	11	10257114	10357114	rs7129220	ADM	0.89
ARHGAP42	11	100058594	100371866	rs633185	FLJ32810,TMEM133	0.28
FES	15	89222929	89245010	rs2521501	FURIN,FES	0.31
GOSR2	17	42350482	42465002	rs17608766	GOSR2	0.86
rs1327235	20	10867030	10967030	rs1327235	JAG1	0.46
rs6015450	20	57134512	57234512	rs6015450	GNAS,EDN3	0.12
MTHFR	1	11763367	11794564	rs17367504	MTHFR,NPPB	0.15
ULK4	3	41258094	41983926	rs3774372	ULK4	0.83
rs1458038	4	81333747	81433747	rs1458038	FGF5	0.29
CACNB2	10	18464612	18875804	rs1813353	CACNB2	0.68
C10orf107	10	63087725	63201530	rs4590817	C10orf107	0.84
NT5C2	10	104830930	104948046	rs11191548	CYP17A1,NT5C2	0.91
PLEKHA7	11	16751418	16997566	rs381815	PLEKHA7	0.26
ATP2B1	12	88500959	88632208	rs17249754	ATP2B1	0.84
SH2B3	12	110323135	110378810	rs3184504	SH2B3	0.47
rs10850411	12	113822179	113922179	rs10850411	TBX5,TBX3	0.7
rs1378942	15	72814420	72914420	rs1378942	CYP1A1,ULK3	0.35
ZNF652	17	44716567	44799834	rs12940887	ZNF652	0.38

Supplementary Table 12: CAU ethnic group sBP (2526 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	42	0.13535	0.07635	0.09202	0.23265
2	MTHFR	1	11763367	11794564	72	0.01945	0.05276	0.05533	0.12314
3	rs13082711	3	27462913	27562913	111	0.01381	0.00466	0.00520	0.00778
4	MECOM	3	170278981	170869100	1026	1.00000	0.06022	0.10837	0.17882
5	ULK4	3	41258094	41983926	1000	0.22846	0.71848	0.70611	0.71814
6	SLC39A8	4	103386221	103576438	249	1.00000	0.86085	0.80668	0.81795
7	GUCY1A3	4	156802313	156877951	134	1.00000	0.53377	0.44202	0.33637
8	rs1458038	4	81333747	81433747	145	0.96055	0.22415	0.10866	0.05626
9	rs1173771	5	32800785	32900785	212	1.00000	0.33519	0.43248	0.54914
10	rs11953630	5	157727980	157827980	181	1.00000	0.89904	0.71726	0.86028
11	HFE	6	26190488	26211550	36	0.54141	0.72512	0.71851	0.68847
12	rs805303	6	31674345	31774345	146	0.00190	0.13020	0.08986	0.07329
13	rs4373814	10	18409978	18509978	284	0.59315	0.17728	0.20288	0.59591
14	PLCE1	10	95738736	96083139	401	1.00000	0.58603	0.61264	0.82326
15	CACNB2	10	18464612	18875804	902	0.38080	0.07818	0.10274	0.18348
16	C10orf107	10	63087725	63201530	190	1.00000	0.88840	0.92763	0.97798
17	NT5C2	10	104830930	104948046	113	0.82815	0.07243	0.08205	0.12620
18	rs7129220	11	10257114	10357114	178	1.00000	0.36771	0.45173	0.66651
19	ARHGAP42	11	100058594	100371866	715	1.00000	0.78860	0.71982	0.82962
20	PLEKHA7	11	16751418	16997566	464	0.16019	0.16659	0.17554	0.19694
21	ATP2B1	12	88500959	88632208	169	1.00000	0.22803	0.29551	0.56748
22	SH2B3	12	110323135	110378810	45	0.48835	0.48136	0.34815	0.53985
23	rs10850411	12	113822179	113922179	260	1.00000	0.43682	0.44055	0.51080
24	FES	15	89222929	89245010	18	1.00000	0.18458	0.28370	0.54740
25	rs1378942	15	72814420	72914420	84	0.00797	0.00186	0.00185	0.00233
26	GOSR2	17	42350482	42465002	138	0.37358	0.34244	0.37069	0.35589
27	ZNF652	17	44716567	44799834	79	1.00000	0.49696	0.57930	0.61687
28	rs1327235	20	10867030	10967030	313	0.21181	0.30059	0.28128	0.33724
29	rs6015450	20	57134512	57234512	180	0.27332	0.61897	0.57733	0.66472

Supplementary Table 13: CAU ethnic group dBP (2526 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	42	0.46110	0.30341	0.27777	0.46583
2	MTHFR	1	11763367	11794564	72	0.00661	0.00130	0.00103	0.00355
3	rs13082711	3	27462913	27562913	111	0.31048	0.17037	0.17740	0.11849
4	MECOM	3	170278981	170869100	1026	0.04833	0.00513	0.02790	0.02207
5	ULK4	3	41258094	41983926	1000	0.04744	0.96234	0.91018	0.88880
6	SLC39A8	4	103386221	103576438	249	0.34061	0.55479	0.53973	0.28137
7	GUCY1A3	4	156802313	156877951	134	1.00000	0.47233	0.56610	0.87517
8	rs1458038	4	81333747	81433747	145	1.00000	0.29768	0.18177	0.25125
9	rs1173771	5	32800785	32900785	212	1.00000	0.66696	0.85676	0.89440
10	rs11953630	5	157727980	157827980	181	0.63602	0.23363	0.26583	0.11110
11	HFE	6	26190488	26211550	36	0.54645	0.34051	0.44308	0.45808
12	rs805303	6	31674345	31774345	146	0.30204	0.06300	0.04414	0.05236
13	rs4373814	10	18409978	18509978	284	0.16493	0.29760	0.32476	0.62033
14	PLCE1	10	95738736	96083139	401	1.00000	0.80312	0.82058	0.79770
15	CACNB2	10	18464612	18875804	902	0.08906	0.16648	0.14358	0.15338
16	C10orf107	10	63087725	63201530	190	0.02939	0.02024	0.02833	0.04118
17	NT5C2	10	104830930	104948046	113	1.00000	0.33286	0.25671	0.28467
18	rs7129220	11	10257114	10357114	178	0.66707	0.07883	0.08443	0.30574
19	ARHGAP42	11	100058594	100371866	715	1.00000	0.71191	0.54805	0.68127
20	PLEKHA7	11	16751418	16997566	464	0.61074	0.07355	0.08251	0.13227
21	ATP2B1	12	88500959	88632208	169	0.43572	0.75429	0.75579	0.61656
22	SH2B3	12	110323135	110378810	45	0.04899	0.16641	0.11498	0.21386
23	rs10850411	12	113822179	113922179	260	1.00000	0.35572	0.32875	0.36983
24	FES	15	89222929	89245010	18	1.00000	0.70865	0.74509	0.82268
25	rs1378942	15	72814420	72914420	84	0.25829	0.04784	0.02814	0.03862
26	GOSR2	17	42350482	42465002	138	1.00000	0.82360	0.99022	0.99637
27	ZNF652	17	44716567	44799834	79	0.96038	0.65212	0.85261	0.84157
28	rs1327235	20	10867030	10967030	313	1.00000	0.40429	0.40808	0.57369
29	rs6015450	20	57134512	57234512	180	0.39288	0.10504	0.09364	0.16543

Supplementary Table 14: AFA ethnic group sBP (1611 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	25	0.41614	0.29182	0.24842	0.31362
2	MTHFR	1	11763367	11794564	60	1.00000	0.74027	0.76483	0.76851
3	rs13082711	3	27462913	27562913	82	1.00000	0.68060	0.67502	0.63146
4	MECOM	3	170278981	170869100	841	1.00000	0.92017	0.91854	0.95864
5	ULK4	3	41258094	41983926	740	1.00000	0.93804	0.92316	0.82468
6	SLC39A8	4	103386221	103576438	218	1.00000	0.56093	0.39974	0.67028
7	GUCY1A3	4	156802313	156877951	111	0.61661	0.05508	0.05925	0.07713
8	rs1458038	4	81333747	81433747	122	1.00000	0.42951	0.50710	0.42752
9	rs1173771	5	32800785	32900785	167	1.00000	0.97818	0.95217	0.96964
10	rs11953630	5	157727980	157827980	153	0.81745	0.25415	0.28025	0.41883
11	HFE	6	26190488	26211550	32	0.12438	0.03934	0.04456	0.08545
12	rs805303	6	31674345	31774345	132	1.00000	0.20036	0.19171	0.15740
13	rs4373814	10	18409978	18509978	235	1.00000	0.90417	0.88062	0.80690
14	PLCE1	10	95738736	96083139	313	0.00837	0.16311	0.31370	0.13983
15	CACNB2	10	18464612	18875804	741	1.00000	0.92649	0.85592	0.92046
16	C10orf107	10	63087725	63201530	157	1.00000	0.54669	0.55742	0.46793
17	NT5C2	10	104830930	104948046	89	0.12760	0.00497	0.00747	0.01635
18	rs7129220	11	10257114	10357114	147	1.00000	0.92541	0.90917	0.89739
19	ARHGAP42	11	100058594	100371866	580	0.86629	0.55717	0.40753	0.57759
20	PLEKHA7	11	16751418	16997566	386	0.20144	0.12597	0.12380	0.06895
21	ATP2B1	12	88500959	88632208	127	0.99723	0.33890	0.32429	0.34161
22	SH2B3	12	110323135	110378810	39	0.57348	0.12702	0.09567	0.19792
23	rs10850411	12	113822179	113922179	214	1.00000	0.94877	0.89189	0.84952
24	FES	15	89222929	89245010	14	0.64001	0.60704	0.59015	0.51314
25	rs1378942	15	72814420	72914420	70	1.00000	0.19290	0.18940	0.20472
26	GOSR2	17	42350482	42465002	121	1.00000	0.82218	0.80143	0.89642
27	ZNF652	17	44716567	44799834	60	1.00000	0.97062	0.96289	0.90895
28	rs1327235	20	10867030	10967030	187	1.00000	0.70146	0.70000	0.64716
29	rs6015450	20	57134512	57234512	101	0.31179	0.20503	0.23915	0.29266

Supplementary Table 15: AFA ethnic group dBP (1611 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	25	0.64720	0.61927	0.53021	0.64604
2	MTHFR	1	11763367	11794564	60	0.88212	0.87692	0.96197	0.97343
3	rs13082711	3	27462913	27562913	82	0.33043	0.02325	0.02634	0.02223
4	MECOM	3	170278981	170869100	841	0.97042	0.59415	0.47938	0.43362
5	ULK4	3	41258094	41983926	740	0.65391	0.15723	0.21763	0.50748
6	SLC39A8	4	103386221	103576438	218	1.00000	0.62477	0.52559	0.77894
7	GUCY1A3	4	156802313	156877951	111	1.00000	0.08871	0.08911	0.16266
8	rs1458038	4	81333747	81433747	122	1.00000	0.31844	0.24816	0.17665
9	rs1173771	5	32800785	32900785	167	1.00000	0.76603	0.76637	0.81074
10	rs11953630	5	157727980	157827980	153	0.28363	0.28390	0.42828	0.49530
11	HFE	6	26190488	26211550	32	0.39060	0.08901	0.12339	0.10605
12	rs805303	6	31674345	31774345	132	0.05749	0.06843	0.08120	0.03396
13	rs4373814	10	18409978	18509978	235	1.00000	0.97314	0.80425	0.61203
14	PLCE1	10	95738736	96083139	313	1.00000	0.43693	0.44077	0.38918
15	CACNB2	10	18464612	18875804	741	0.54101	0.88589	0.82264	0.91039
16	C10orf107	10	63087725	63201530	157	0.09040	0.01524	0.01296	0.01055
17	NT5C2	10	104830930	104948046	89	1.00000	0.39532	0.41655	0.65896
18	rs7129220	11	10257114	10357114	147	0.26467	0.43629	0.36792	0.29269
19	ARHGAP42	11	100058594	100371866	580	0.58105	0.14785	0.26547	0.42156
20	PLEKHA7	11	16751418	16997566	386	1.00000	0.51775	0.59301	0.29514
21	ATP2B1	12	88500959	88632208	127	1.00000	0.09377	0.11849	0.14564
22	SH2B3	12	110323135	110378810	39	0.14582	0.58411	0.65584	0.25757
23	rs10850411	12	113822179	113922179	214	0.26957	0.86843	0.87413	0.90939
24	FES	15	89222929	89245010	14	1.00000	0.66729	0.64717	0.43090
25	rs1378942	15	72814420	72914420	70	1.00000	0.47473	0.49868	0.56505
26	GOSR2	17	42350482	42465002	121	0.39961	0.12266	0.17179	0.38108
27	ZNF652	17	44716567	44799834	60	1.00000	0.20299	0.21907	0.16900
28	rs1327235	20	10867030	10967030	187	1.00000	0.51377	0.72879	0.71563
29	rs6015450	20	57134512	57234512	101	0.32563	0.16282	0.17482	0.17651

Supplementary Table 16: HIS ethnic group sBP (1449 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	25	0.21720	0.19764	0.14669	0.22319
2	MTHFR	1	11763367	11794564	60	1.00000	0.30060	0.26727	0.05594
3	rs13082711	3	27462913	27562913	82	0.30355	0.03073	0.02657	0.04527
4	MECOM	3	170278981	170869100	841	0.19305	0.13142	0.12776	0.04830
5	ULK4	3	41258094	41983926	740	1.00000	0.82240	0.85875	0.63704
6	SLC39A8	4	103386221	103576438	218	1.00000	0.96378	0.97076	0.94457
7	GUCY1A3	4	156802313	156877951	111	0.05460	0.16260	0.19888	0.27068
8	rs1458038	4	81333747	81433747	122	0.18943	0.37949	0.31693	0.41744
9	rs1173771	5	32800785	32900785	167	0.06474	0.11461	0.06462	0.21419
10	rs11953630	5	157727980	157827980	153	1.00000	0.27397	0.31941	0.38667
11	HFE	6	26190488	26211550	32	0.94388	0.42017	0.28997	0.26090
12	rs805303	6	31674345	31774345	132	0.30084	0.32730	0.28558	0.45846
13	rs4373814	10	18409978	18509978	235	0.07423	0.05311	0.03130	0.03711
14	PLCE1	10	95738736	96083139	313	0.50239	0.13823	0.11782	0.06340
15	CACNB2	10	18464612	18875804	742	0.27413	0.12414	0.07441	0.09030
16	C10orf107	10	63087725	63201530	157	0.00208	0.00795	0.00330	0.00233
17	NT5C2	10	104830930	104948046	89	1.00000	0.49682	0.39424	0.59208
18	rs7129220	11	10257114	10357114	147	0.02707	0.21695	0.21966	0.25168
19	ARHGAP42	11	100058594	100371866	580	0.69250	0.45838	0.30851	0.13242
20	PLEKHA7	11	16751418	16997566	387	0.64844	0.25484	0.22700	0.08798
21	ATP2B1	12	88500959	88632208	127	1.00000	0.47628	0.68267	0.65226
22	SH2B3	12	110323135	110378810	39	1.00000	0.95814	0.95237	0.98013
23	rs10850411	12	113822179	113922179	214	1.00000	0.70291	0.56957	0.51005
24	FES	15	89222929	89245010	14	0.47022	0.28032	0.32165	0.34811
25	rs1378942	15	72814420	72914420	70	0.95797	0.40943	0.52693	0.34465
26	GOSR2	17	42350482	42465002	121	0.25193	0.50482	0.45614	0.48149
27	ZNF652	17	44716567	44799834	60	1.00000	0.39857	0.32511	0.48670
28	rs1327235	20	10867030	10967030	187	1.00000	0.55602	0.65414	0.61712
29	rs6015450	20	57134512	57234512	101	1.00000	0.62157	0.55568	0.18535

Supplementary Table 17: HIS ethnic group dBP (1449 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	25	1.00000	0.74680	0.87073	0.63924
2	MTHFR	1	11763367	11794564	60	0.34353	0.08581	0.05394	0.02890
3	rs13082711	3	27462913	27562913	82	0.00459	0.00577	0.00861	0.03495
4	MECOM	3	170278981	170869100	841	0.33805	0.22671	0.25818	0.20534
5	ULK4	3	41258094	41983926	740	1.00000	0.90854	0.95999	0.90103
6	SLC39A8	4	103386221	103576438	218	1.00000	0.53937	0.36558	0.07364
7	GUCY1A3	4	156802313	156877951	111	0.11861	0.30924	0.36171	0.40988
8	rs1458038	4	81333747	81433747	122	0.47361	0.25443	0.24657	0.25349
9	rs1173771	5	32800785	32900785	167	0.02616	0.54847	0.38889	0.71940
10	rs11953630	5	157727980	157827980	153	1.00000	0.36324	0.37470	0.46645
11	HFE	6	26190488	26211550	32	1.00000	0.86865	0.92151	0.93876
12	rs805303	6	31674345	31774345	132	0.16574	0.37206	0.39796	0.52797
13	rs4373814	10	18409978	18509978	235	1.00000	0.47485	0.32155	0.39683
14	PLCE1	10	95738736	96083139	313	1.00000	0.54681	0.53568	0.41218
15	CACNB2	10	18464612	18875804	742	0.88106	0.32142	0.24189	0.51498
16	C10orf107	10	63087725	63201530	157	0.05212	0.02343	0.01039	0.00814
17	NT5C2	10	104830930	104948046	89	1.00000	0.89105	0.76728	0.60572
18	rs7129220	11	10257114	10357114	147	0.12894	0.04984	0.06132	0.07062
19	ARHGAP42	11	100058594	100371866	580	0.74441	0.13852	0.09680	0.15075
20	PLEKHA7	11	16751418	16997566	387	0.10437	0.60733	0.49726	0.23367
21	ATP2B1	12	88500959	88632208	127	0.24094	0.40082	0.45522	0.42751
22	SH2B3	12	110323135	110378810	39	1.00000	0.96279	0.96176	0.98044
23	rs10850411	12	113822179	113922179	214	1.00000	0.73359	0.55539	0.52110
24	FES	15	89222929	89245010	14	1.00000	0.80443	0.75815	0.84079
25	rs1378942	15	72814420	72914420	70	0.17513	0.40532	0.29054	0.17715
26	GOSR2	17	42350482	42465002	121	0.11577	0.52576	0.53297	0.42869
27	ZNF652	17	44716567	44799834	60	1.00000	0.45140	0.37263	0.64254
28	rs1327235	20	10867030	10967030	187	1.00000	0.77211	0.87579	0.94684
29	rs6015450	20	57134512	57234512	101	1.00000	0.86702	0.91764	0.80125

Supplementary Table 18: CHN ethnic group sBP (775 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	26	0.15059	0.36309	0.28488	0.24989
2	MTHFR	1	11763367	11794564	59	1.00000	0.60830	0.55853	0.71237
3	rs13082711	3	27462913	27562913	79	0.10178	0.03019	0.01908	0.04958
4	MECOM	3	170278981	170869100	822	0.64821	0.39291	0.54738	0.75053
5	ULK4	3	41258094	41983926	722	1.00000	0.44015	0.36834	0.49278
6	SLC39A8	4	103386221	103576438	213	1.00000	0.56075	0.50357	0.71217
7	GUCY1A3	4	156802313	156877951	107	0.22165	0.48432	0.46386	0.40801
8	rs1458038	4	81333747	81433747	120	0.36911	0.18223	0.24495	0.27591
9	rs1173771	5	32800785	32900785	159	1.00000	0.36743	0.18806	0.18593
10	rs11953630	5	157727980	157827980	150	1.00000	0.97636	0.94505	0.85109
11	HFE	6	26190488	26211550	31	0.59685	0.20983	0.26259	0.33061
12	rs805303	6	31674345	31774345	130	1.00000	0.88966	0.82054	0.73643
13	rs4373814	10	18409978	18509978	229	1.00000	0.31388	0.33109	0.39644
14	PLCE1	10	95738736	96083139	305	1.00000	0.72016	0.63652	0.82586
15	CACNB2	10	18464612	18875804	714	0.26030	0.02177	0.02603	0.00944
16	C10orf107	10	63087725	63201530	154	1.00000	0.46945	0.46612	0.38714
17	NT5C2	10	104830930	104948046	88	0.01400	0.81579	0.85689	0.93057
18	rs7129220	11	10257114	10357114	142	1.00000	0.50385	0.49822	0.32532
19	ARHGAP42	11	100058594	100371866	554	0.02062	0.16038	0.13582	0.17357
20	PLEKHA7	11	16751418	16997566	378	0.41303	0.24477	0.14763	0.27488
21	ATP2B1	12	88500959	88632208	122	0.63404	0.39300	0.36684	0.18625
22	SH2B3	12	110323135	110378810	36	0.21587	0.37482	0.41033	0.42779
23	rs10850411	12	113822179	113922179	211	0.89265	0.97007	0.92754	0.90111
24	FES	15	89222929	89245010	14	1.00000	0.22070	0.22339	0.32847
25	rs1378942	15	72814420	72914420	70	1.00000	0.89694	0.87975	0.93642
26	GOSR2	17	42350482	42465002	116	0.99234	0.17330	0.15325	0.23567
27	ZNF652	17	44716567	44799834	59	0.39036	0.24913	0.47506	0.58936
28	rs1327235	20	10867030	10967030	184	1.00000	0.20434	0.18397	0.24773
29	rs6015450	20	57134512	57234512	96	0.18157	0.07655	0.08302	0.08470

Supplementary Table 19: CHN ethnic group dBP (775 subjects) Analysis. Four ethnic groups were analyzed separately using LGRF, SKAT and MinP. The analysis was done under the adjustment of age, gender, BMI and top two principle components to correct for potential within-ethnicity stratification. MinP: testing every SNP with GEE in the region and adjusting the minimum p-value by the effective number of variants explaining 99.95% of the genotype variation in the region. SKAT was applied to the average value of repeated measurements.

	name	chr	start	end	# of SNPs	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	26	0.17209	0.31239	0.26490	0.37278
2	MTHFR	1	11763367	11794564	59	1.00000	0.40381	0.36832	0.35052
3	rs13082711	3	27462913	27562913	79	0.03003	0.03078	0.02917	0.07132
4	MECOM	3	170278981	170869100	822	1.00000	0.48813	0.40892	0.44618
5	ULK4	3	41258094	41983926	722	0.06193	0.24029	0.18058	0.37429
6	SLC39A8	4	103386221	103576438	213	1.00000	0.78774	0.60471	0.68886
7	GUCY1A3	4	156802313	156877951	107	0.97396	0.29995	0.31670	0.38890
8	rs1458038	4	81333747	81433747	120	1.00000	0.41763	0.58483	0.58023
9	rs1173771	5	32800785	32900785	159	1.00000	0.40452	0.33038	0.29395
10	rs11953630	5	157727980	157827980	150	1.00000	0.99585	0.88124	0.85264
11	HFE	6	26190488	26211550	31	0.44131	0.13491	0.18534	0.32330
12	rs805303	6	31674345	31774345	130	0.53269	0.22116	0.21903	0.49052
13	rs4373814	10	18409978	18509978	229	0.83391	0.33075	0.30770	0.37043
14	PLCE1	10	95738736	96083139	305	0.90713	0.97003	0.93120	0.89614
15	CACNB2	10	18464612	18875804	714	0.12518	0.01679	0.01644	0.00552
16	C10orf107	10	63087725	63201530	154	1.00000	0.47566	0.53613	0.49981
17	NT5C2	10	104830930	104948046	88	1.00000	0.76354	0.70835	0.74797
18	rs7129220	11	10257114	10357114	142	1.00000	0.60813	0.64742	0.53630
19	ARHGAP42	11	100058594	100371866	554	0.10876	0.58452	0.58966	0.77740
20	PLEKHA7	11	16751418	16997566	378	1.00000	0.14965	0.08344	0.17324
21	ATP2B1	12	88500959	88632208	122	0.24554	0.21366	0.20688	0.06752
22	SH2B3	12	110323135	110378810	36	0.92189	0.73576	0.65778	0.61232
23	rs10850411	12	113822179	113922179	211	1.00000	0.80558	0.98446	0.98229
24	FES	15	89222929	89245010	14	1.00000	0.09102	0.09950	0.16090
25	rs1378942	15	72814420	72914420	70	1.00000	0.59380	0.70537	0.73863
26	GOSR2	17	42350482	42465002	116	0.82658	0.18860	0.18656	0.18634
27	ZNF652	17	44716567	44799834	59	0.23437	0.84935	0.94165	0.94194
28	rs1327235	20	10867030	10967030	184	1.00000	0.85964	0.78668	0.82811
29	rs6015450	20	57134512	57234512	96	0.38382	0.29070	0.35086	0.32896

Supplementary Table 20: Meta-Analysis of sBP (6361 subjects) in MESA. Four ethnic groups were combined using Fisher's method.

	name	chr	start	end	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	0.12662	0.11615	0.08422	0.20122
2	MTHFR	1	11763367	11794564	0.44532	0.27329	0.25608	0.19276
3	rs13082711	3	27462913	27562913	0.04979	0.00129	0.00087	0.00359
4	MECOM	3	170278981	170869100	0.84271	0.16447	0.26964	0.25385
5	ULK4	3	41258094	41983926	0.93729	0.94506	0.96538	0.90941
6	SLC39A8	4	103386221	103576438	1.00000	0.95245	0.88354	0.98117
7	GUCY1A3	4	156802313	156877951	0.27964	0.14522	0.14894	0.16461
8	rs1458038	4	81333747	81433747	0.71395	0.26335	0.20692	0.16141
9	rs1173771	5	32800785	32900785	0.70582	0.38030	0.22572	0.46261
10	rs11953630	5	157727980	157827980	0.99994	0.69307	0.69145	0.83252
11	HFE	6	26190488	26211550	0.58656	0.15250	0.14972	0.22746
12	rs805303	6	31674345	31774345	0.06037	0.28225	0.20031	0.19631
13	rs4373814	10	18409978	18509978	0.61972	0.15804	0.12702	0.27193
14	PLCE1	10	95738736	96083139	0.20492	0.31685	0.38811	0.24973
15	CACNB2	10	18464612	18875804	0.51404	0.02932	0.02661	0.02366
16	C10orf107	10	63087725	63201530	0.13607	0.12543	0.07493	0.04878
17	NT5C2	10	104830930	104948046	0.11075	0.02387	0.03046	0.09402
18	rs7129220	11	10257114	10357114	0.51322	0.58221	0.62433	0.64354
19	ARHGAP42	11	100058594	100371866	0.36079	0.55124	0.35965	0.34083
20	PLEKHA7	11	16751418	16997566	0.30172	0.10267	0.07077	0.04177
21	ATP2B1	12	88500959	88632208	0.99872	0.38877	0.48796	0.48406
22	SH2B3	12	110323135	110378810	0.69064	0.46967	0.36973	0.62359
23	rs10850411	12	113822179	113922179	0.99999	0.96044	0.92497	0.91949
24	FES	15	89222929	89245010	0.96616	0.26903	0.35591	0.55000
25	rs1378942	15	72814420	72914420	0.28300	0.02224	0.02575	0.02478
26	GOSR2	17	42350482	42465002	0.78478	0.49338	0.45838	0.57622
27	ZNF652	17	44716567	44799834	0.98444	0.63856	0.76786	0.88685
28	rs1327235	20	10867030	10967030	0.92765	0.48760	0.48532	0.55831
29	rs6015450	20	57134512	57234512	0.40123	0.24997	0.25722	0.17082

Supplementary Table 21: Meta-Analysis of dBP (6361 subjects) in MESA. Four ethnic groups were combined using Fisher's method.

	name	chr	start	end	MinP	SKAT	LGRF-G	LGRF-J
1	MOV10	1	113012286	113049891	0.65418	0.61873	0.56224	0.72836
2	MTHFR	1	11763367	11794564	0.13322	0.00934	0.00557	0.00854
3	rs13082711	3	27462913	27562913	0.00434	0.00041	0.00063	0.00241
4	MECOM	3	170278981	170869100	0.40581	0.04256	0.10761	0.07974
5	ULK4	3	41258094	41983926	0.12987	0.55612	0.56460	0.87757
6	SLC39A8	4	103386221	103576438	0.97592	0.87205	0.69878	0.34243
7	GUCY1A3	4	156802313	156877951	0.82750	0.19607	0.24412	0.47640
8	rs1458038	4	81333747	81433747	0.99279	0.32604	0.26011	0.26061
9	rs1173771	5	32800785	32900785	0.50600	0.82381	0.76341	0.87893
10	rs11953630	5	157727980	157827980	0.90491	0.48791	0.58455	0.46902
11	HFE	6	26190488	26211550	0.78654	0.18630	0.31383	0.39228
12	rs805303	6	31674345	31774345	0.11321	0.04401	0.04039	0.05240
13	rs4373814	10	18409978	18509978	0.86003	0.62699	0.50342	0.67279
14	PLCE1	10	95738736	96083139	1.00000	0.90959	0.90494	0.82606
15	CACNB2	10	18464612	18875804	0.23328	0.07493	0.05311	0.04745
16	C10orf107	10	63087725	63201530	0.02302	0.00146	0.00097	0.00086
17	NT5C2	10	104830930	104948046	1.00000	0.77595	0.68184	0.76498
18	rs7129220	11	10257114	10357114	0.47707	0.08902	0.09892	0.18138
19	ARHGAP42	11	100058594	100371866	0.63454	0.44668	0.29560	0.56021
20	PLEKHA7	11	16751418	16997566	0.70240	0.18356	0.13426	0.11532
21	ATP2B1	12	88500959	88632208	0.50290	0.25040	0.29796	0.15524
22	SH2B3	12	110323135	110378810	0.26183	0.71943	0.63767	0.55636
23	rs10850411	12	113822179	113922179	0.95580	0.90671	0.88300	0.89773
24	FES	15	89222929	89245010	1.00000	0.56643	0.57727	0.63884
25	rs1378942	15	72814420	72914420	0.62576	0.23689	0.16497	0.16428
26	GOSR2	17	42350482	42465002	0.58833	0.32515	0.41807	0.53758
27	ZNF652	17	44716567	44799834	0.93545	0.65153	0.70855	0.76768
28	rs1327235	20	10867030	10967030	1.00000	0.86045	0.92321	0.97171
29	rs6015450	20	57134512	57234512	0.64413	0.20782	0.23223	0.28420