GWAS of longitudinal trajectories at biobank scale

Seyoon Ko,[1,2,8] Christopher A. German,[2,8] Aubrey Jensen,[2] Judong Shen,[3] Anran Wang,[3] Devan V. Mehrotra,[3] Yan V. Sun,[4] Janet S. Sinsheimer,[1,2,5] Hua Zhou,[1,2,*] and Jin J. Zhou[2,6,7,*]

Summary

Biobanks linked to massive, longitudinal electronic health record (EHR) data make numerous new genetic research questions feasible. One among these is the study of biomarker trajectories. For example, high blood pressure measurements over visits strongly predict stroke onset, and consistently high fasting glucose and Hb1Ac levels define diabetes. Recent research reveals that not only the mean level of biomarker trajectories but also their fluctuations, or within-subject (WS) variability, are risk factors for many diseases. Glycemic variation, for instance, is recently considered an important clinical metric in diabetes management. It is crucial to identify the genetic factors that shift the mean or alter the WS variability of a biomarker trajectory. Compared to traditional cross-sectional studies, trajectory analysis utilizes more data points and captures a complete picture of the impact of time-varying factors, including medication history and lifestyle. Currently, there are no efficient tools for genome-wide association studies (GWASs) of biomarker trajectories at the biobank scale, even for just mean effects. We propose TrajGWAS, a linear mixed effect model-based method for testing genetic effects that shift the mean or alter the WS variability of a biomarker trajectory. It is scalable to biobank data with 100,000 to 1,000,000 individuals and many longitudinal measurements and robust to distributional assumptions. Simulation studies corroborate that TrajGWAS controls the type I error rate and is powerful. Analysis of eleven biomarkers measured longitudinally and extracted from UK Biobank primary care data for more than 150,000 participants with 1,800,000 observations reveals loci that significantly alter the mean or WS variability.

Introduction

Biomarker trajectories are important phenotypes that reflect the evolution of an individual's health or disease progression.[1–3] With the increasing use of electronic health records (EHRs) linked with biobanks, large scale and repeatedly measured EHR-based quantitative laboratory-derived phenotypes are becoming highly influential in genetic studies of human health.[4–6] For example, a recent LabWAS tool demonstrates the broad impact of using such "real world" measurements for genetic association studies.[7] LabWAS summarizes longitudinal measurements by taking the mean for analyses. Although proven to be robust, this approach may lose power by ignoring the many rich features in the whole trajectories. Identifying genetic and clinical factors associated with these longitudinal trajectories can quantify the susceptibility to the onset of disease and disease progression, which ultimately offers new opportunities for early clinical prevention.[1,8–10]

Besides mean level trajectory patterns, the biomarker fluctuations may also differ between individuals; some individuals show higher levels of variation around their mean than others (Figure 1). This intra-individual variability or within-subject (WS) variability[11,12] has been shown to be an important risk factor for disease. For example, among diabetes patients, visit-to-visit intra-individual fasting glucose variability is a risk factor for the development of vascular complications,[13–15] independent of the glycemic control of the mean; blood pressure variability has been associated with the increased risk of heart failure[16] and stroke.[11] Experimental research has revealed the biological basis of glycemic variability and diabetic kidney injury.[17] WS variability in reaction times has also been suggested as a leading endophenotype for neurocognitive disorders, such as attention deficit hyperactivity disorder and schizophrenia.[18,19] As the wearable devices gain more and more popularity, WS variability becomes a clinical metric of disease management, such as the glucose coefficient of variability output from the continuous glucose monitoring (CGM) device report.[20,21]

WS variability differs from the between-subject (BS) variability, which also has recently attracted much attention. Variance quantitative trait loci (vQTLs) analysis seeks to identify loci that show different trait variances among groups of individuals with different variant genotypes.[22–25] Such phenotypic variance heterogeneity can be caused by gene-by-environment interaction, selection, epistasis, or phantom vQTLs. vQTL analysis is typically performed on a cross-sectional cohort, while TrajGWAS requires longitudinal data. In contrast to vQTL, TrajGWAS investigates genetic contributions to the WS variability instead of BS variability. Thus, TrajGWAS and vQTL analyses can provide complementary insights into the etiology of a disease. As an interesting example, we

[1]Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [2]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90095, USA; [3]Biostatistics and Research Decision Sciences, Merck & Co., Inc., Kenilworth, NJ 07033, USA; [4]Department of Epidemiology, Emory University, Atlanta, GA 30322, USA; [5]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA; [6]Department of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [7]Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ 85721, USA
[8]These authors contributed equally
*Correspondence: jinjinzhou@ucla.edu (J.J.Z.), huazhou@ucla.edu (H.Z.)
https://doi.org/10.1016/j.ajhg.2022.01.018.
© 2022 American Society of Human Genetics.

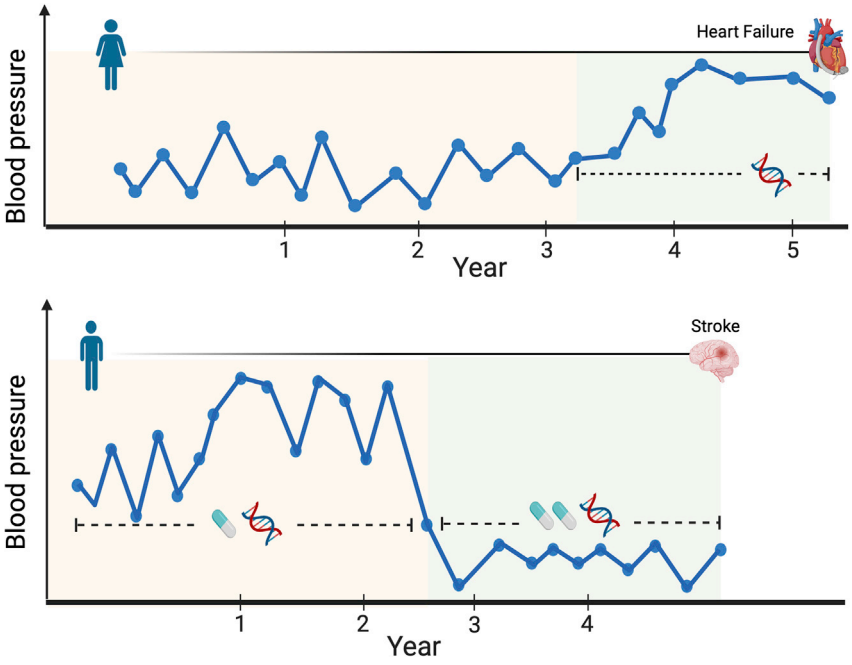

Figure 1. Illustration of TrajGWAS
TrajGWAS identifies the genetic factors that shift the mean or alter the within-subject (WS) variability of a biomarker trajectory (e.g., blood pressure measurements over visits), which changes with time-varying covariates (e.g., medication history) or time-invariant covariates (e.g., sex).

find that the well-known *FTO* (MIM: 610966) vQTL for body mass index (BMI)[26] (p value $= 1.16 \times 10^{-102}$) is not associated with the WS variability (p value $= 1.18 \times 10^{-4}$) at the genome-wide significance level.

Identifying genome-wide genetic contributions to longitudinal trajectories, including both mean and WS variability, is both methodologically and computationally challenging. Despite recent efforts,[27–29] no existing software is able to analyze massive longitudinal traits at the biobank scale. The linear mixed effect model (LMM) is a powerful and popular method for longitudinal data analysis. Generalizations such as the mixed-effects location scale model[30] allow for simultaneous modeling of the mean and variability of the longitudinal measurement, increase power, and reduce bias. It leverages information across individuals to produce more precise estimates.[31] However, the expensive numerical integration required in each iteration prohibits many modern data applications. For example, the run times of the full likelihood approach with MixWILD software[32,33] on two simulated datasets with 1,000 individuals and ten observations per individual ranged from 40 min to 10+ h depending on the different modeling assumptions being made. MLwiN,[34] a multi-level model (a type of mixed effect model), has been used to estimate the mean trajectories while accounting for the change in scale and variance of measures over time.[1] However, none of these tools were designed for modern genome-wide scans. The heuristic strategies being employed in practice involve a two-stage model: (1) summarize a subject-level measure of the variation of the longitudinal measurement such as standard deviation (SD), average real variability (ARV), or the coefficient of variation (CV); (2) model those as the responses with covariates.[12,35,36] This framework makes an implicit assumption that an individual's variability remains constant over time and cannot be affected by time-varying covariates (Figure 1). Yet intra-individual variability is affected by both time-varying (e.g., medication use or adherence to the treatment regime) and time-invariant features (e.g., sex and genes). Regressing the subject-level variability summaries on predictors leads to serious bias.[31] The simulation study in supplemental methods, section C, shows that, without properly adjusting for time-varying covariates, the heuristic method can seriously inflate the type I error.

Building upon our recent methods, within-subject variance estimator by robust regression (WiSER),[37] we derive an ultra-fast score test, which only requires fitting one null model across the whole genome-wide set. This testing strategy scales linearly in the number of individuals. We also develop and implement a saddlepoint approximation (SPA) for our score test to ensure well-controlled type I error rates for single rare variant testing with minor allele frequencies (MAFs) as low as 0.001.

Material and methods

An LMM framework for testing genetic effects on the trajectory mean and WS variability

Our modeling assumptions are as follows. Assume there are m independent individuals, individual i has n_i longitudinal measurements of a biomarker, and $n = \sum_{i=1}^{m} n_i$ is the total number of observations. Consider an LMM for modeling different sources of variation in a biomarker in the longitudinal setting

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + g_i \beta_g + \boldsymbol{z}_{ij}^T \boldsymbol{\gamma}_i + \varepsilon_{ij}, \qquad \text{(Equation 1)}$$

where y_{ij} is individual i's measurement at occasion $j \in \{1, 2, \ldots, n_i\}$, $\boldsymbol{x}_{ij}$ is the $p \times 1$ vector of regressors with corresponding regression coefficients $\boldsymbol{\beta}$, g_i is the genotype dosage of individual i with corresponding genetic mean effect β_g, and $\boldsymbol{z}_{ij}$ is the $q \times 1$ vector of covariates with corresponding random effects $\boldsymbol{\gamma}_i$. The WS variability is captured by the random terms ε_{ij} with mean zero and inhomogeneous variance

$$\sigma_{\varepsilon_{ij}}^2 = \exp\left(\boldsymbol{w}_{ij}^T \boldsymbol{\tau} + g_i \tau_g + \omega_i\right), \qquad \text{(Equation 2)}$$

where $\boldsymbol{w}_{ij}$ is the $\ell \times 1$ vector of covariates with corresponding fixed effects $\boldsymbol{\tau}$, τ_g is the genetic effect on the WS variability, and ω_i is a

random intercept. We assume that the random effects $(\boldsymbol{\gamma}_i^T, \omega_i)^T$ are independent of ε_{ij}, have mean zero, and have covariance

$$\mathrm{Var}\begin{pmatrix}\boldsymbol{\gamma}_i\\ \omega_i\end{pmatrix} = \boldsymbol{\Sigma}_{\gamma\omega} = \begin{pmatrix}\boldsymbol{\Sigma}_\gamma & \boldsymbol{\sigma}_{\gamma\omega}\\ \boldsymbol{\sigma}_{\gamma\omega}^T & \sigma_\omega^2\end{pmatrix}.$$

Covariates $\boldsymbol{x}_{ij}$, $\boldsymbol{z}_{ij}$, and $\boldsymbol{w}_{ij}$ typically contain an intercept and can include both time-invariant covariates, e.g., sex and baseline measurements, and time-varying covariates, e.g., age at measurement, medication history, and life-style indicators. Individuals can have varying numbers of observations, which do not need to be aligned.

Given a longitudinal biomarker of interest, our primary goal is to test (1) the mean effect of genotype, $H_0 : \beta_g = 0$, i.e., whether a genotype shifts the mean of the biomarker trajectory; (2) the WS variance effect of genotype, $H_0 : \tau_g = 0$, i.e., whether a genotype changes the WS variation of the biomarker trajectory around its mean; and (3) the joint effect, $H_0 : \beta_g = \tau_g = 0$, i.e., whether a genotype affects either the mean, or the WS variation, or both. Although for the models in Equation 1 and Equation 2 we use scalar g_i to represent a single genotype, our method and software can also test a group of genotypes or gene-by-environment (G × E) effects.

The models in Equation 1 and Equation 2 are similar to a multiple location scale model considered by Dzubur et al.,[33] who assume normality of the random effects $(\boldsymbol{\gamma}_i^T, \omega_i)^T$ and ε_{ij} and resort to the maximum likelihood estimation (MLE). Because each iteration of the MLE algorithm requires expensive numerical integration, it is not only distributionally restrictive but also computationally prohibitive. Both limitations prevent its application to genome-wide association studies (GWASs) of biobank data. Instead, we employ our recent estimation method, WiSER,[37] which is robust to the misspecification of the trait distribution (conditional on random effects) and the random effects distribution. The estimation algorithm is free of numerical integration and scales linearly in the total number of longitudinal measurements. For example, the run times of the full likelihood approach with the MixWILD software[32,33] on two simulated datasets with 1,000 individuals and ten observations per individual range from 40 min to 10+ h according to the different modeling assumptions being made, while WiSER takes less than half a second.

Briefly, the WiSER estimator is defined as

$$\widehat{\boldsymbol{\beta}}, \widehat{\beta}_g = \arg\min_{\boldsymbol{\beta},\beta_g} \frac{1}{2}\sum_{i=1}^m (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - g_i\beta_g)^T \left(\boldsymbol{V}_i^{(0)}\right)^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - g_i\beta_g)$$

$$\widehat{\boldsymbol{\tau}}, \widehat{\tau}_g, \widehat{\boldsymbol{\Sigma}}_\gamma = \arg\min_{\boldsymbol{\tau},\tau_g,\boldsymbol{\Sigma}_\gamma} \frac{1}{2}\sum_{i=1}^m \mathrm{tr}\left(\left(\boldsymbol{V}_i^{(0)}\right)^{-1}\boldsymbol{R}_i\left(\boldsymbol{V}_i^{(0)}\right)^{-1}\boldsymbol{R}_i\right), \quad \text{(Equation 3)}$$

where $\boldsymbol{R}_i = (\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - g_i\widehat{\beta}_g)(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - g_i\widehat{\beta}_g)^T - \boldsymbol{V}_i(\boldsymbol{\tau}, \tau_g, \boldsymbol{\Sigma}_\gamma)$,

$$\boldsymbol{V}_i(\boldsymbol{\tau}, \tau_g, \boldsymbol{\Sigma}_\gamma) = \begin{pmatrix}\exp(\boldsymbol{w}_{i1}^T\boldsymbol{\tau} + g_i\tau_g) & & \\ & \ddots & \\ & & \exp\left(\boldsymbol{w}_{in_i}^T\boldsymbol{\tau} + g_i\tau_g\right)\end{pmatrix} + \boldsymbol{Z}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{Z}_i^T, \quad \text{(Equation 4)}$$

and $\boldsymbol{V}_i^{(0)} = \boldsymbol{V}_i(\boldsymbol{\tau}^{(0)}, \tau_g^{(0)}, \boldsymbol{\Sigma}_\gamma^{(0)})$ is an initial estimator of $\mathrm{Var}(\boldsymbol{Y}_i)$. Model parameters are the mean fixed effects $\boldsymbol{\beta}$ and β_g, WS variance fixed effects $\boldsymbol{\tau}$ and τ_g, and the random effects covariance $\boldsymbol{\Sigma}_\gamma$. In the special case $\boldsymbol{V}_i^{(0)} = \boldsymbol{I}_{n_i}$, WiSER reduces to a method of moments (MoM) estimate because the objective functions in Equation 3 are simply the least-squares losses for the first two moments of $\boldsymbol{Y}_i$. Using an initial estimate $\boldsymbol{V}_i^{(0)}$ improves the estimation efficiency of WiSER. In practice, we set the initial $\boldsymbol{V}_i^{(0)}$ according to a least-squares estimator of $\boldsymbol{\tau}$ and $\boldsymbol{\Sigma}_\gamma$. WiSER enjoys a double robustness property. It is robust to the misspecification of both the distribution of random effects $(\boldsymbol{\gamma}_i^T, \omega_i)^T$ and the distribution of $\boldsymbol{Y}_i$ conditional on random effects. In TrajGWAS, we employ a score test that only requires fitting one null model, with $\beta_g = \tau_g = 0$, across the genome-wide tests. Compared with the Wald test proposed by German et al.,[37] which requires fitting WiSER for each genotype, it is much faster and enables fast longitudinal trajectory GWAS analysis at biobank scale.

Robust and scalable score testing

Let $\theta_1 \in \mathbb{R}$ be the genetic effect β_g or τ_g. We are interested in testing the null hypothesis $\theta_1 = 0$. Let $\boldsymbol{\theta}_2 \in \mathbb{R}^{p+q+q(q+1)/2}$ collect all parameters in the null model. We first derive the score (gradient of the WiSER loss function) $\boldsymbol{\psi}_{H_1}$ under the full model and then evaluate it under the null model, i.e., $\boldsymbol{\psi}_{H_1}(\tilde{\boldsymbol{\theta}})$, where $\tilde{\boldsymbol{\theta}} = (0, \widehat{\boldsymbol{\theta}}_2)$ and $\widehat{\boldsymbol{\theta}}_2$ is the estimate under the null model. The generalized score test statistic[38] is

$$T = \frac{1}{m}\left[\sum_{i=1}^m \boldsymbol{\psi}_{H_1,i}(\tilde{\boldsymbol{\theta}})\right]^T \boldsymbol{V}_{\boldsymbol{\psi}_{H_1}(\tilde{\boldsymbol{\theta}})}^{-1}\left[\sum_{i=1}^m \boldsymbol{\psi}_{H_1,i}(\tilde{\boldsymbol{\theta}})\right],$$

where $\boldsymbol{V}_{\boldsymbol{\psi}}$ is the variance of score $\boldsymbol{\psi}$. The score test statistic T is asymptotically distributed as χ_1^2 under the null model. In supplemental methods, section A, we show that the scores for testing $\beta_g = 0$ and $\tau_g = 0$ are

$$S_{\beta_g} = \sum_{i=1}^m \left[1_{n_i}^T\left(\boldsymbol{V}_i^{(0)}\right)^{-1}\widehat{\boldsymbol{r}}_i\right] g_i =: \boldsymbol{c}_{\beta_g}^T\boldsymbol{g}$$

and

$$S_{\tau_g} = -\sum_{i=1}^m \left\{1_{n_i}^T \mathrm{diag}\left[\begin{pmatrix}e^{\boldsymbol{w}_{i1}^T\widehat{\boldsymbol{\tau}}} & & \\ & \ddots & \\ & & e^{\boldsymbol{w}_{in_i}^T\widehat{\boldsymbol{\tau}}}\end{pmatrix}\left(\boldsymbol{V}_i^{(0)}\right)^{-1}\widehat{\boldsymbol{R}}_i\left(\boldsymbol{V}_i^{(0)}\right)^{-1}\right]\right\} g_i =: \boldsymbol{c}_{\tau_g}^T\boldsymbol{g}$$

respectively. The quantities $\widehat{\boldsymbol{r}}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{R}}_i$, $\widehat{\boldsymbol{\tau}}$, and $\boldsymbol{V}_i^{(0)}$ are readily available from the fitted null model. Calculation of each score involves linear combination of the genotype dosages with the coefficient vector $\boldsymbol{c}_{\beta_g}$ or $\boldsymbol{c}_{\tau_g}$ pre-computed and cached. In supplemental methods, section A, we show that the calculation of variance $\widehat{\boldsymbol{V}}_{\boldsymbol{\psi}}$ costs $O(1)$ flops. Therefore, forming each score test statistic costs $O(m)$ flops, where m is the number of individuals, usually much smaller than the total number of longitudinal measurements n. This extreme computational efficiency makes TrajGWAS easily scalable to biobank data with $10^5 \sim 10^6$ samples and millions of SNPs.

Saddlepoint approximation for rare variant testing

It is well-known that asymptotic score tests may yield deflated or inflated type I errors at stringent significance levels for rare variants with MAF < 0.01.[39,40] Figures 2A and 2B show that, when testing a null variant with MAF $= 0.01$, the score test shows deflation in testing β_g and inflation in testing τ_g. To calibrate the null

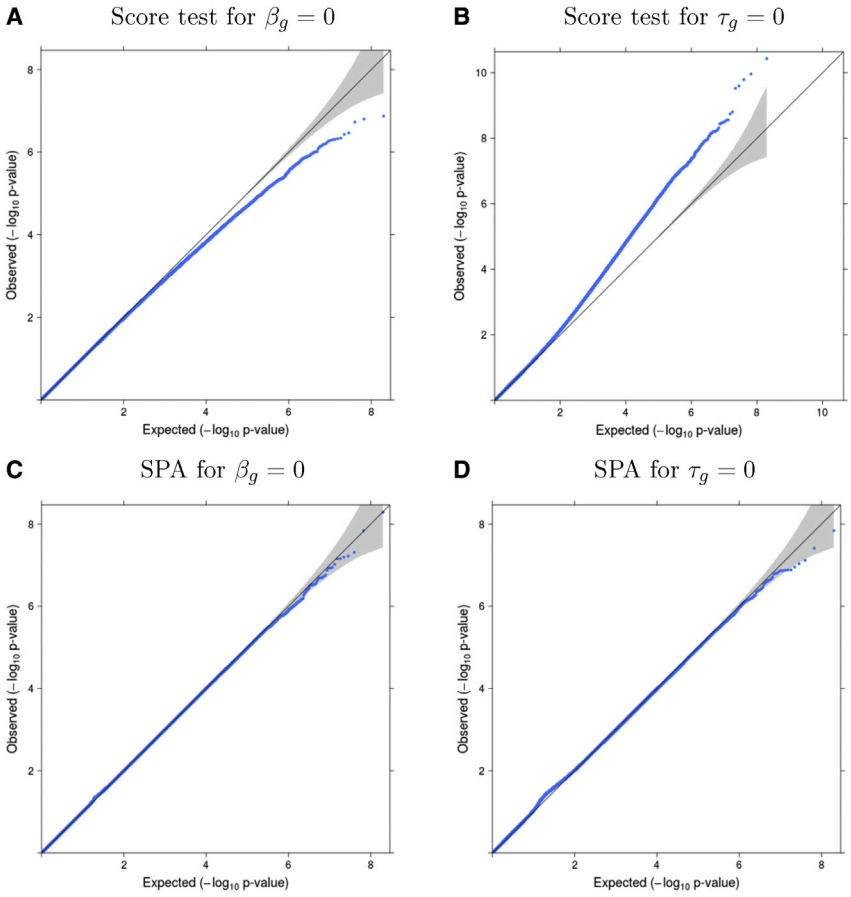

Figure 2. Quantile-quantile (QQ) plots of p values for simulation studies

(A and B) Quantile-quantile (QQ) plots of p values from the score test for testing β_g (A) and for testing τ_g (B), where $m = 6{,}000$, $n_i = 6$ to 10, and minor allele frequency (MAF) = 0.01, based on 10^9 replicates.

(C and D) QQ plots of SPA-corrected p values for testing β_g (C) and for testing τ_g (D); simulations are based on $m = 6{,}000$, $n_i = 6$ to 10, and minor allele frequency (MAF) = 0.01. SPA corrects the deflation or inflation that occurs in the score test at low MAFs. QQ plots for all simulation scenarios are shown in Figures S2–S5.

distribution for score statistics when testing rare variants, we apply a saddlepoint approximation (SPA).[39–43] This approach uses the entire cumulant generating function (CGF) to approximate the null distribution instead of the first two moments as in the normal approximation and Satterthwaite method,[44] resulting in superior performance. For testing a single variant, we directly use the score, S_{β_g} or S_{τ_g}, as the test statistic. Since the CGFs of S_{β_g} and S_{τ_g} do not have a simple closed-form expression, we use the empirical CGF based on the empirical moment generating function (MGF). Details are provided in the supplemental methods, section B. Because the normal approximation of the score test performs well near the mean of the distribution, to save on computation, we only apply SPA when the observed score statistic is large. Following Bi et al.,[39] SPA is applied when $\left|S_{\beta_g}\right| > r\sqrt{\mathrm{Var}(\boldsymbol{c}_{\beta_g})\mathrm{Var}(\boldsymbol{g})}$ and $\left|S_{\tau_g}\right| > r\sqrt{\mathrm{Var}(\boldsymbol{c}_{\tau_g})\mathrm{Var}(\boldsymbol{g})}$ for testing β_g and τ_g, respectively. In this paper, we use $r = 0.75$ for all analyses. A smaller value of r leads to more tests having SPA applied and increased computational time. For the joint test of null hypothesis $\tau_g = \beta_g = 0$, we compute p values for both S_{τ_g} and S_{β_g} and then take their harmonic mean.[45]

Simulations

We carry out simulations to evaluate type I error rates and power of TrajGWAS. For each subject, we generate the response according to the models in Equation 1 and Equation 2. In our simulations, the random mean effect $\boldsymbol{\gamma}_i$ is intercept only so $\boldsymbol{Z}_i$ is a single column of 1's. $\boldsymbol{X}_i$ and $\boldsymbol{W}_i$ contain a random time-invariant binary variable (0 or 1) in their second columns, a time-invariant standard normal variable in their third columns, and a time-varying standard normal variable in their fourth columns. The true regression coefficients are $\boldsymbol{\beta} = (10.0, 5.0, 0.5, -0.3)^T$ and $\boldsymbol{\tau} = (0.25, 0.3, -0.15, 0.1)^T$. We generate the random effects (γ_i, ω_i) from the multivariate normal distribution with mean zero and covariance $\boldsymbol{\Sigma}_{\gamma\omega} = \begin{pmatrix} 2.0 & 0.0 \\ 0.0 & 0.1 \end{pmatrix}$

For both type I error and power simulations, we consider 12 scenarios with different combinations of (1) sample sizes: $m = 6{,}000$ and $m = 100{,}000$, (2) number of repeated-measurements: $n_i = 6$ to 10 and $n_i = 10$ to 30, and (3) MAF: 0.01, 0.05 and 0.3 for $m = 6{,}000$ and 0.001, 0.05 and 0.3 for $m = 100{,}000$. Results of both the score test and SPA are reported.

Type I error

To evaluate type I error rates at genome-wide significance level $\alpha = 5 \times 10^{-8}$, for each scenario we generate 1,000 datasets each with 10^6 variants following Hardy-Weinberg equilibrium, yielding 10^9 total replicates.[39] We report type I error rates for testing the genetic contribution to both the mean, β_g, and the WS variance, τ_g.

Power

To evaluate the power for testing τ_g and β_g, we generate 100 datasets under the alternative model for each scenario. In each dataset, the alternative model uses the parameters in simulations and contains ten causal variants each with the same effect size, selected specific to each scenario in order to show the spread of power. We compare power of the score test and SPA at the significance level $\alpha = 5 \times 10^{-8}$.

Application to the UK Biobank study

We conduct TrajGWAS analysis by using longitudinal biomarker measures extracted from the UK Biobank primary care data, including systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse pressure (PP), high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, total cholesterol (TC), triglycerides, glucose (fasting and random), hemoglobin A1C (HbA1c), and body mass index (BMI). Record-level access to primary care data is obtained by requesting field 42040 ("GP clinical event records") from the UK Biobank showcase. We combine

Table 1. Empirical type I error rates for the simulation studies

Simulation conditions			Empirical type I error rate (standard error) $\times 10^{-8}$					
Sample size *m*	n_i	**MAF**	β_g **score**	β_g **SPA**	τ_g **score**	τ_g **SPA**	**Joint score**	**Joint SPA**
6,000	6 to 10	0.01	0.30 (0.17)	4.00 (0.63)	138.50 (3.72)	3.50 (0.59)	80.80 (2.84)	4.10 (0.64)
6,000	6 to 10	0.05	3.30 (0.57)	4.10 (0.64)	34.50 (1.86)	6.30 (0.79)	22.90 (1.51)	5.90 (0.77)
6,000	6 to 10	0.3	4.10 (0.64)	4.20 (0.65)	4.80 (0.69)	4.30 (0.66)	4.80 (0.69)	4.40 (0.66)
6,000	10 to 30	0.01	0.40 (0.20)	6.00 (0.77)	42.70 (2.07)	4.00 (0.63)	23.20 (1.52)	4.20 (0.65)
6,000	10 to 30	0.05	4.00 (0.63)	4.90 (0.70)	20.50 (1.43)	5.10 (0.71)	12.80 (1.13)	5.50 (0.74)
6,000	10 to 30	0.3	4.50 (0.67)	5.20 (0.72)	6.60 (0.81)	6.00 (0.77)	5.20 (0.72)	6.00 (0.77)
100,000	6 to 10	0.001	1.20 (0.35)	4.80 (0.69)	136.80 (3.70)	4.40 (0.66)	80.90 (2.84)	3.90 (0.62)
100,000	6 to 10	0.05	5.00 (0.71)	5.00 (0.71)	6.20 (0.79)	5.10 (0.71)	5.80 (0.76)	5.60 (0.75)
100,000	6 to 10	0.3	4.10 (0.64)	4.00 (0.63)	5.30 (0.73)	5.50 (0.74)	5.40 (0.73)	5.30 (0.73)
100,000	10 to 30	0.001	2.40 (0.49)	5.80 (0.76)	50.80 (2.25)	4.80 (0.69)	28.50 (1.69)	5.40 (0.73)
100,000	10 to 30	0.05	5.40 (0.73)	5.10 (0.71)	7.60 (0.87)	6.50 (0.81)	7.20 (0.85)	6.20 (0.79)
100,000	10 to 30	0.3	6.30 (0.79)	6.10 (0.78)	4.00 (0.63)	3.90 (0.62)	5.40 (0.73)	5.70 (0.75)

Empirical type I error rates (standard error) for the score test and SPA ($\times 10^{-8}$) at a significance level 5×10^{-8} based on 10^9 simulation replicates. The score test shows inflated type I error at low minor allele frequencies (MAFs) for testing τ_g where SPA (saddlepoint approximation) has proper type I error rates. Joint score and joint SPA are based on the harmonic means of the respective β_g and τ_g p values.

a previously reported and validated semi-supervised approach[46] and in-house extraction criteria to create clinical biomarker phenotypes. We matched and compared empirical cumulative distributions of extracted lab values from the primary care database and those provided through the UK Biobank assessment center to infer the measurement units and for further quality control (Figure S16). Detailed data extraction, unit conversion, and quality control procedures are documented in the supplemental methods, section D.

For each GWAS, we use the standardized biomarker phenotypes for TrajGWAS analysis by subtracting the overall mean from each measurement and dividing by the standard deviation and we adjust for ten principal components (PCs) on the mean component. Using the PCs to adjust both the mean and WS variance makes no differences for the final results. Each biomarker uses a different covariate adjustment scheme, which is detailed in supplemental methods, section E. In general, we adjust for sex, age, age^2, and age × sex for both mean and WS variability; age and age^2 are treated as time-varying covariates. The selection of covariates is guided by previous GWAS analyses[4,47,48] and the mean profile plots are shown in the Figure S15. Non-significant covariates in the null model are then removed from the GWAS analysis. In addition, we include self-reported diabetes status as a time-fixed covariate for glycemic measures (HbA1c and random and fasting glucose). Diabetes status included as a time-varying indicator is also explored (supplemental methods, section F). Summary of the covariates included and adjustments made for medication is summarized in supplemental methods, section E.

Controlling the effect of medication on the biomarkers is important in the analysis. Most widely used methods for such adjustments are (1) treatment modeled as an additional covariate ("indicator");[49–51] (2) adding a sensible constant ("shifting") to the treated subjects;[48,52–55] and (3) censored normal regression.[56] Shifting and censored normal regression are often recommended for their superior performances over the indicator method.[56] In this paper, we use the shifting method if a sensible value for adjustment is available through previous studies and use the indicator method for others. We compare adjustment by shifting and adjustment by an additional covariate in Figures S21–S24. For blood pressures, we add 15 mmHg for SBP and 10 mmHg for DBP[55] for subjects taking blood-pressure-lowering medication before standardization. For lipids, following previous GWAS analysis,[47] we add 0.208 mmol/L for triglycerides, 1.347 mmol/L for total cholesterol, 1.290 mmol/L for LDL cholesterol, and subtract 0.060 mmol/L for HDL cholesterol for participants on lipid-controlling treatments. For glycemic measures (HbA1c and random and fasting glucose), a sensible value for adjustment was not available, so they are adjusted with the indicator method.

To evaluate and compare the genetic association of trajectory means, i.e., β_g, we create lists of previously reported genetic associations for each analyzed trait by using the GWAS Catalog[57] queried by the R package gwasrapidd[58] (curated on 7/8/2021). We search the catalog for phenotypes matching our analyzed biomarkers by using syntax, "efo_trait=," and keep SNPs with p value less than genome-wide significance level $< 5 \times 10^{-8}$.

Results

Simulation

Table 1 reports the empirical type I error rates of the score test and SPA at an $\alpha = 5\times10^{-8}$ threshold, based on 10^9 simulation replicates. At lower MAFs, the score test for τ_g has substantially inflated type I errors, whereas SPA leads to well-calibrated type I error rates. Inflation in the score test for τ_g at less common alleles (MAF = 0.05) is large for smaller sample sizes and fewer repeated measures. The amount of type I error inflation decreases as the MAF and the number of repeated measures increase. For β_g, the score test is conservative at lower MAFs and SPA corrects the type I error in the right direction. For common

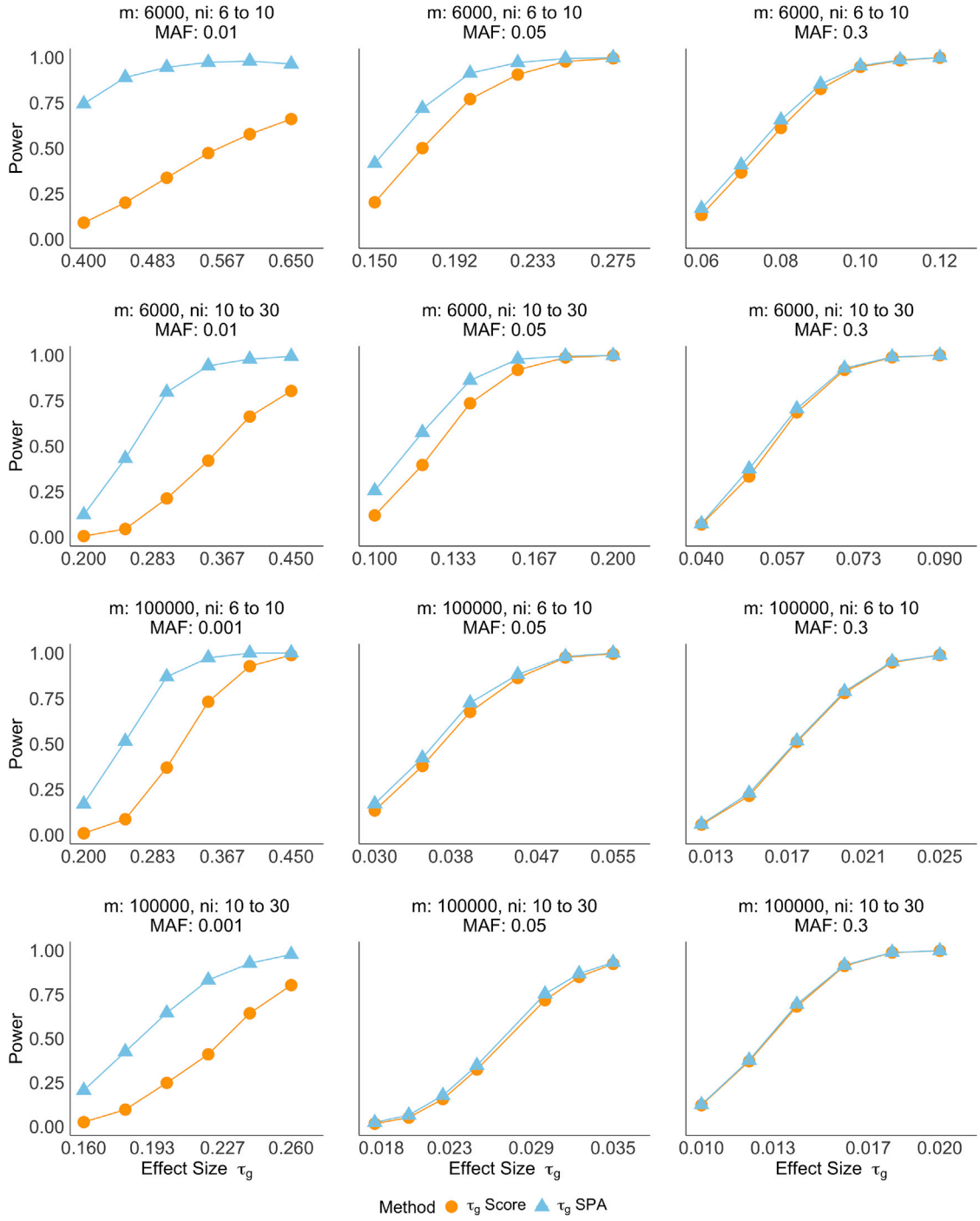

Figure 3. Empirical powers of testing β_g with score test and SPA
Each row contains the same sample size m and number of observations per individual n_i. Power is evaluated at the significance level $\alpha = 5 \times 10^{-8}$. Each scenario is based on 1,000 replicates.

alleles such as MAF = 0.3, score test and SPA do not differ much in the type I error rates for either τ_g or β_g. Overall SPA has appropriate type I error at the $\alpha = 5 \times 10^{-8}$ significance level across all scenarios. Figure 2 illustrates how SPA corrects type I error in both directions by displaying QQ plots from a random sample of 100 million replicates of the $m = 6000, n_i = 6$ to 10, MAF = 0.01 scenario. Additional QQ plots are presented in Figures S2–S5.

Power curves for testing τ_g and β_g across 12 scenarios are displayed in Figure 3 and Figure S6, respectively. Although the score test is unable to adequately control type I error for rare variants, we still report power based on the nominal power at the $\alpha = 5 \times 10^{-8}$ significance level. Using the empirical significance levels estimated from the type I error simulations would result in even lower power for the score test than what is shown in the figure. SPA achieves higher

Table 2. Sample size m, the number of repeated-measurements n_i, and summary statistics for eleven biomarkers extracted from the UK Biobank primary care data

	m	n_i	y_{ij}	Female	Age	BMI
Biomarker	**Sample size**	**Median (IQR)**	**Mean (SD)**	**%**	**Mean (SD)**	**Mean (SD)**
SBP (mmHg)	148,870	12 (6, 24)	135.0 (15.3)	54.1	56.0 (8.7)	27.5 (4.8)
DBP (mmHg)	148,870	12 (6, 24)	81.0 (8.7)	54.1	56.0 (8.7)	27.5 (4.8)
PP (mmHg)	148,870	12 (6, 24)	53.9 (9.6)	54.1	56.0 (8.7)	27.5 (4.8)
HDL (mmol/L)	129,069	4 (2, 8)	1.5 (0.4)	53.1	59.5 (7.8)	27.7 (4.8)
LDL (mmol/L)	98,556	3 (1, 6)	3.2 (0.9)	52.3	59.3 (7.8)	27.8 (4.8)
Total cholesterol (mmol/L)	133,590	5 (2, 10)	5.4 (0.9)	53.3	58.7 (7.9)	27.6 (4.8)
Triglycerides (mmol/L)	124,092	4 (2, 8)	1.6 (1.0)	48.1	60.6 (7.8)	28.6 (5.0)
Fasting glucose (mmol/L)	55,949	2 (1, 3)	5.5 (1.4)	47.7	60.4 (7.6)	28.7 (5.1)
Random glucose (mmol/L)	97,162	2 (1, 4)	5.7 (2.1)	51.9	59.8 (8.2)	28.5 (5.1)
HbA1c (%)	70,589	2 (1, 4)	6.7 (1.4)	43.4	62.4 (7.8)	30.4 (5.7)
BMI	144,414	5 (3, 9)	28.3 (5.7)	54.9	57.3 (9.9)	–

The eleven biomarkers are as follows: SBP, systolic blood pressure; DBP, diastolic blood pressure; PP, pulse pressure = SBP – DBP; HDL, high-density lipoprotein; LDL, low-density lipoprotein; total cholesterol; triglycerides; random glucose; fasting glucose; HbA1c, hemoglobin A1C; and BMI, body mass index.

power at the $\alpha = 5\times10^{-8}$ significance level than the score test when the MAF is low, but the power of the two methods are nearly identical for common variants and large sample sizes. In conjunction with the type I error results, this indicates that SPA is able to better model the tail of the test statistic distributions for rare variants. When the variants are common, both approaches converge to the same results.

Computational efficiency

With careful implementation, each iteration of the optimization algorithm for fitting the WiSER null model scales linearly in the total number of observations n. For testing a single SNP, our score test with SPA scales linearly with the sample size m. Therefore, our TrajGWAS analysis based on WiSER can be applied to longitudinal genetic association analysis at biobank scale. For example, analyzing SBP for 10,805,717 imputed variants on all autosomal chromosomes takes about 150 central processing unit (CPU) h with SPA and 139 CPU h without SPA. The computation is split into 16 chunks per chromosome, resulting in 352 separate computational jobs that can run simultaneously on computing clusters. Under these conditions, each job runs within an hour with and without SPA.

Real data analysis

About 44% of the 500,000 UK Biobank participants are linked to their primary care EHR data. These EHR data are recorded with four controlled clinical terminologies: (1) Read version 2 (Read v2); (2) Clinical Terms Version 3 (CTV3); (3) British National Formulary (BNF); and (4) the Dictionary of Medicines and Devices (DM+D). Only Read v2 and CTV3 are relevant for biomarker extraction. Using previously validated algorithms,[46,59] we generate unified lists of Read v2 and CTV3 clinical terms, and extract measurements for all biomarkers from the clinical event records (gp_clinical table). Terms used for extraction are shown in Table S2. Ten longitudinal clinical measurements are extracted: blood pressures (SBP and DBP), HDL, LDL, total cholesterol, triglycerides, blood glucose (fasting and random), HbA1c, and BMI (supplemental methods, section D). Extracted records cover 55,000 to 150,000 participants. The flowcharts for creating the cohort for each biomarker are displayed in Figures S7–S14. There are more repeated-measures of SBP and DBP (median (IQR) = 12 (6, 24)) than of the lipid values (e.g., median (IQR) = 4 (2, 8) for HDL). See Table 2 for details. Taking blood pressure as an example, we exclude observations with no date or invalid date information, or missing BMI measures at recruitment, resulting in 2,598,484 observations. The sample size for GWAS analysis ranges from 55,949 (fasting glucose) to 148,870 (blood pressure). Patterns of the mean profile over age groups vary across different biomarker groups (Figure S15). DBP, LDL, and total cholesterol show strong non-linear, age-dependent trends.

We then apply TrajGWAS to UK Biobank imputed genetic data among European ancestry for these ten longitudinal clinical measures and one derived phenotype pulse pressure (PP = SBP − DBP). SNPs with MAF greater than 0.002 and imputation quality score (infoscore or r^2) greater than 0.3 are included in the analyses. The Manhattan plots (Figure 4 for τ_g and Figure S17 for β_g) and quantile-quantile (QQ) plots (Figures S18 and S19) show that TrajGWAS successfully identifies a large number of loci. Concordant with the simulation study, the QQ plots suggest that SPA controls type I error rates well. Highly polygenic traits with a larger number of associated variants have, on average, larger genomic control factor (λ_{GC}) values (Figures S18 and S19). Additionally, since SPA is not applied when the score statistics are close to the

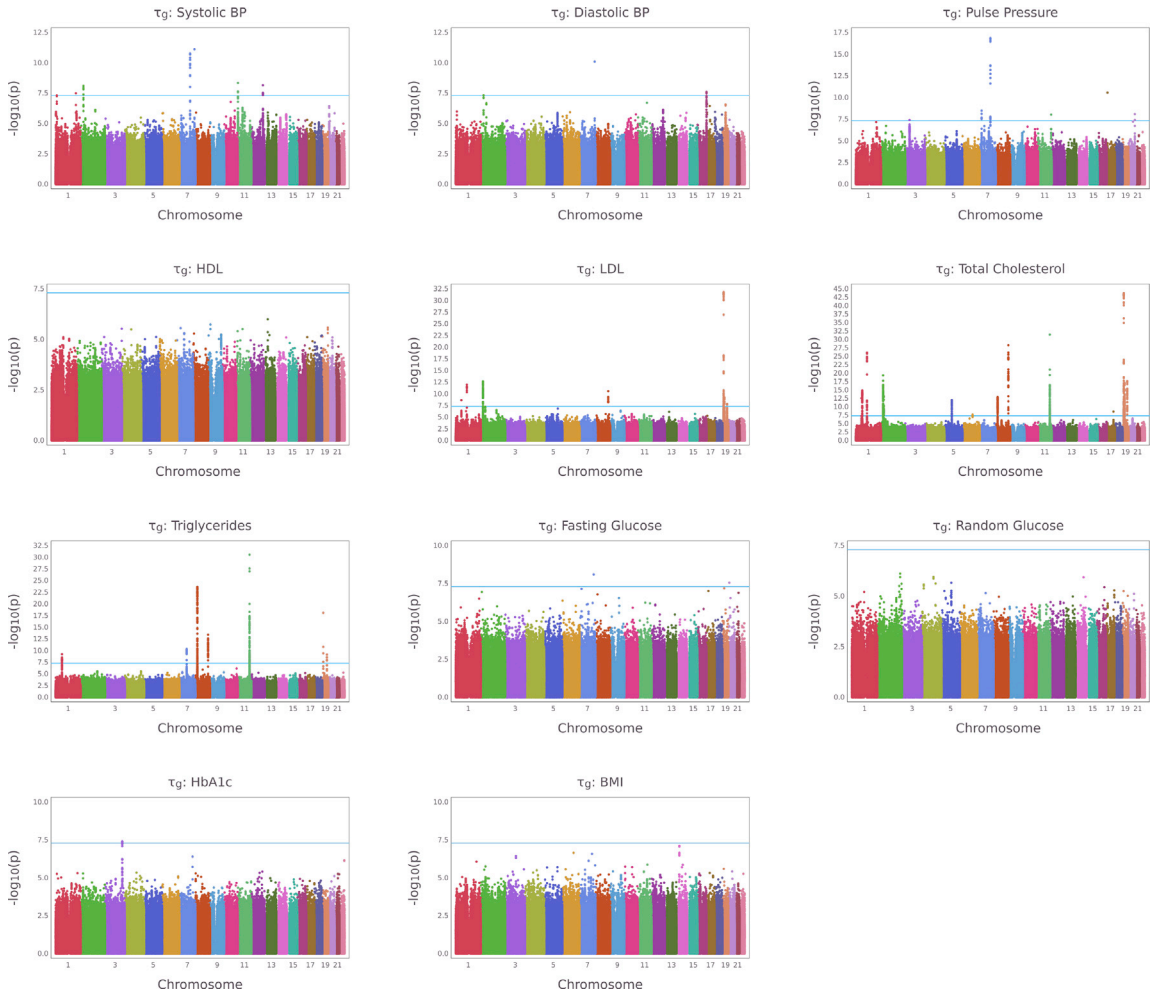

Figure 4. Manhattan plots for testing τ_g for longitudinal markers in the UK Biobank study
Manhattan plots for testing τ_g, the effects of the WS variability, for 11 longitudinal biomarkers in the UK Biobank study. The blue line represents the genome-wide significance level, 5×10^{-8}.

mean, the median p values used for calculation of the genomic control factor may be miscalibrated.[40] Thus, even though many QQ plots appear normal for τ_g, the reported λ_{GC} is inflated for some traits. To give a complete picture, we report λ_{GC} calculated at different p value quantiles for each trait in Table S3.

Next, we compare associations identified by TrajGWAS with those reported in the GWAS Catalog. We extract association results from the GWAS Catalog by using the Experimental Factor Ontology (EFO) trait labels and keep the unique associations, i.e., SNPs, with p value $< 5 \times 10^{-8}$. The number of associations from TrajGWAS analysis is shown in the second and third columns of Table 3. Data in the GWAS Catalog are mapped to genome assembly GRCh38, while UK Biobank SNPs are mapped to GRCh37. We remove the queried SNPs with no genomic coordinates and convert GWAS Catalog associations to genome assembly GRCh37. The numbers of associated SNPs are shown in the fourth column of Table 3. Using the associations reported in the GWAS Catalog as positive controls, we evaluate whether SNPs associated with the mean from our TrajGWAS analysis can replicate previous findings (fifth column of Table 3). For eight out of eleven markers, we have replication rates higher than 80%, validating high quality of EHR-based biomarker phenotyping and TrajGWAS analysis. The analysis of HbA1c has the lowest replication rate 59.65%. This may be due to the relatively small sample size among all biomarkers and the differences in distribution of HbA1c measures from EHR (see Figure S16). The last column of Table 3 lists the numbers of SNPs TrajGWAS identifies as "novel," i.e., not in linkage disequilibrium (LD) with the existing SNPs in the GWAS Catalog (defined as being greater than one megabase from any SNP in the GWAS Catalog). Tables S4–S11 provide additional annotations for these "novel" SNPs. As an example, for total cholesterol, there are 177 and 209 SNPs associated with mean and WS variability that are at least 1 Mb away from the existing reported GWAS Catalog SNPs, respectively. Additional annotations shown in Table S8 demonstrate that the majority of SNPs reported to be "novel" for total cholesterol are relevant to lipids traits as well as psychiatric disorders. These

Table 3. Summary of TrajGWAS results

Biomarker[a]	Num. of sig. loci for β_g/τ_g[b,c]	Num. of sig. SNPs for β_g/τ_g[d]	Num. of sig. SNPs in GWAS Catalog[e]	Replication rate β_g[f]	Num. of sig. SNPs for β_g/τ_g > 1 Mb from GWAS Catalog SNPs[g]
SBP	269/8	4,720/32	1,738	82.48%	0/1
DBP	368/3	7,374/5	917	79.65%	615/0
PP	371/8	6,895/32	876	89.34%	93/0
HDL	1,443/0	14,068/0	1,953	88.62%	24/0
LDL	826/23	8,160/434	1,654	83.57%	0/0
Total cholesterol	1,356/92	16,002/1,525	1,270	95.06%	177/209
Triglycerides	1,172/55	11,796/820	1,693	87.57%	2/0
Fasting glucose	166/2	1,540/2	110	86.67%	144/2
Random glucose	81/0	824/0	256	73.63%	10/0
HbA1c	73/1	1,820/4	651	59.65%	11/0
BMI	220/0	3,651/0	3,507	86.95%	1/0

[a]Experimental Factor Ontology (EFO) trait labels (see web resources) used for query are as follows: SBP, "systolic blood pressure (EFO_0006335)"; DBP, "diastolic blood pressure (EFO_0006336)"; PP, "pulse pressure measurement (EFO_0005763)"; HDL, "high-density lipoprotein cholesterol measurement (EFO_0004612)"; LDL, "low-density lipoprotein cholesterol measurement (EFO_0004611)"; total cholesterol, "total cholesterol measurement (EFO_0004574)"; triglycerides, "triglyceride measurement (EFO_0004530)"; HbA1c, "HbA1c measurement (EFO_0004541)"; fasting glucose, "fasting blood glucose measurement (EFO_0004465)"; random glucose, "glucose measurement (EFO_0004468)"; BMI, "body mass index (EFO_0004340)" and "longitudinal BMI measurement (EFO_0005937)."
[b]Significant SNPs for each biomarker are clumped via PLINK 1.9.[61] Index variants are chosen greedily starting with the SNPs with lowest p value among those SNPs having p value $\leq 5 \times 10^{-8}$. Sites that are < 250 kb away from an index variant and $r^2 > 0.5$ with the index variant are assigned to that index variant's clump.
[c]The number of significant loci (after clumping).
[d]The number of significant SNPs ($< 5 \times 10^{-8}$) on β_g and τ_g.
[e]Number of GWAS Catalog SNPs with p value $< 5 \times 10^{-8}$ (all SNPs are converted to genome build 37; the SNPs with no genomic coordinates are removed; and GWAS Catalog stores the most significant SNP from each independent locus).
[f]Percent of significant SNPs from GWAS Catalog that are nominally significant in β_g (p value < 0.05) in the TrajGWAS analysis.
[g]The number of SNPs at least 1 megabase (Mb) away from any previously reported SNP for the given biomarker in the GWAS Catalog.

findings are consistent with the possibility of a disease-specific lipid pathway underlying the pathophysiology of psychiatric disorders.[60]

The majority of genes that affect WS variability of a trajectory also affect mean, but not always. Figure 5 highlights the 235 SNPs that are significantly associated with WS variability but not with mean levels with p values and gene annotations. Consistent with our simulation, with too few longitudinal measures, it is hard to detect τ_g at genome-wide significance level, e.g., random glucose (median $n_i = 2$), fasting glucose (median $n_i = 2$), and HbA1c (median $n_i = 2$). For traits with median $n_i > 4$, there are signals in τ_g. In particular, TrajGWAS identifies a genome-wide significant association between WS variability of total cholesterol and variants in the *LPL* gene (MIM: 609708), whereas they are not associated with the mean values. *LPL* is a protein-coding gene for lipoprotein lipase, which is expressed in heart, muscle, and adipose tissue. Severe mutations that cause *LPL* deficiency result in type I hyperlipoproteinemia, while less extreme mutations in *LPL* are linked to many disorders of lipoprotein metabolism.[62] Several GWASs have identified the association of *LPL* with different lipid-related phenotypes.[63,64] Figure S20 displays a boxplot of within-sample variance of residuals for subjects with 0, 1, and 2 copies of reference allele of rs6993414, the most significant SNP in terms of τ_g on *LPL*. It shows there are big differences in the tail distributions between them. Other examples include the association between HbA1c WS variability and the *EIF5A2* gene (MIM: 605782). *EIF5A2* is a protein-coding gene associated with type 2 diabetes and cancer.[65] Interestingly, a variant, rs8192675, and its proxies show the strongest association with HbA1c response to metformin; its LD block covered three genes and *EIF5A2* is one of them.[66]

TrajGWAS differs from the vQTL, which is predominantly used among cross-sectional studies and for G × E interaction screening. For a BMI analysis adjusted for age, sex, and ten PCs with the OSCA software (see web resources), 13 of the 22 vQTLs previously reported in Wang et al.[25] have a significant vQTL on the same gene (p value $< 2 \times 10^{-9}$) in our cohort. One well-known vQTL for BMI is the *FTO* gene, and variants in this gene are previously found to be associated with BS variance of BMI with very low p values.[25] Our cohort yields the lowest p value of 1.16×10^{-102} for vQTL analysis. However, τ_g for WS variability of TrajGWAS minimum p value in the same region is 1.18×10^{-4}, showing no significant SNP association with WS variability.

Discussion

We provide a genome-wide trajectory analysis tool, TrajGWAS, for simultaneous testing of genetic effects on the

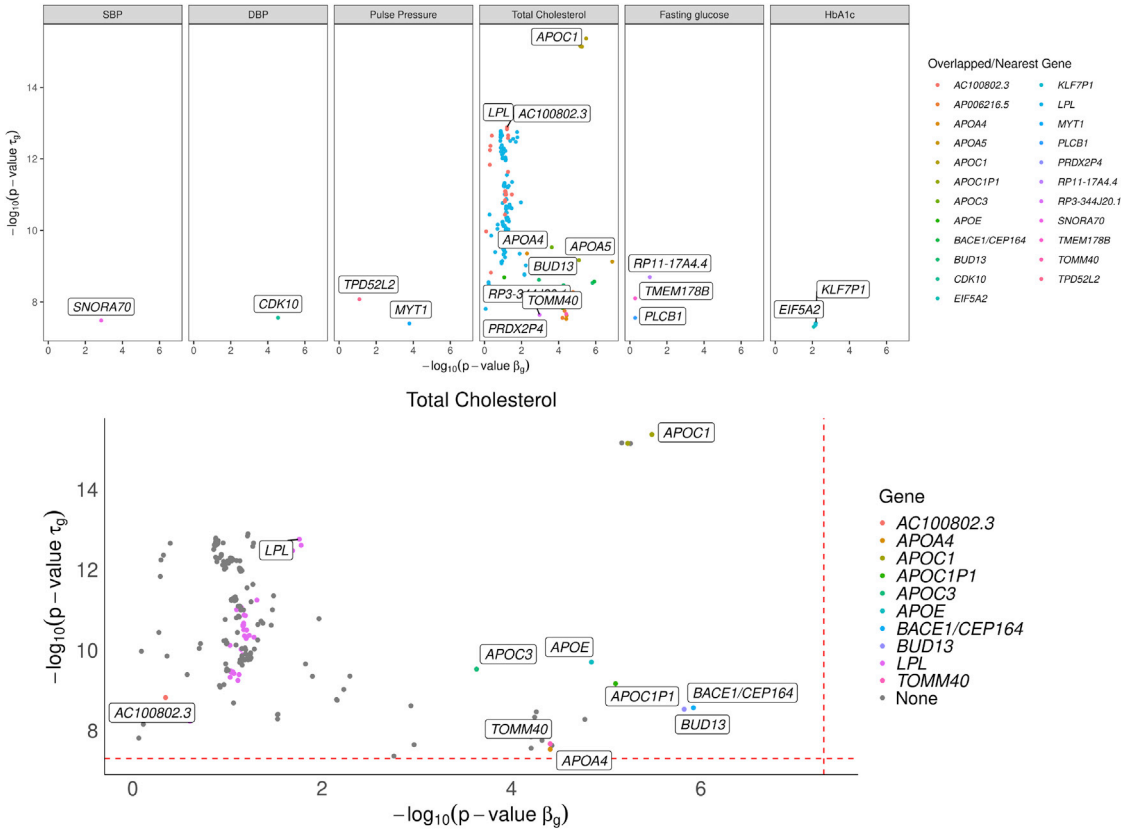

Figure 5. SNPs that significantly change the WS variability while not significantly shifting the mean
SNPs that significantly change the WS variability of longitudinal biomarkers (top) and total cholesterol (TC) trajectory (bottom) while not significantly shifting the mean of their trajectories. Each dot is a SNP that passes the genome-wide significance level (dashed line) for τ_g but not for β_g.

mean and WS variability of a longitudinal biomarker for biobank-scale studies. The method relies on a mixed-effects location scale model but has several advantages over existing methods. For example, the likelihood-based approach for fitting the mixed effect location scale model requires computationally intensive numerical integration, making it infeasible to implement for genome-wide scans of biobank data.[30,32,33] TrajGWAS relies on M-estimation asymptotics and is both computationally efficient and robust to distributional assumptions. It also does not assume the WS variability is constant and can capture and control for the effects of time-varying covariates such as medication usage and age. We use empirical SPA to calibrate p values so that type I error rates can be well controlled for rarer variants and when the number of repeated measures is small. Through extensive simulation studies and application to UK Biobank data, we demonstrate that TrajGWAS scales well for millions of markers, hundreds of thousands of individuals, and multiple random effects while retaining well-controlled type I error rates and power. One limitation of the SPA approach is that its construction only works for a single univariate hypothesis. Thus, for the joint test $\beta_g = \tau_g = 0$, we resort to the less satisfactory harmonic mean approach,[45] which might compromise power.

Although originally motivated by the study of longitudinal biomarkers, TrajGWAS is also applicable to genome-wide scans of multiple, correlated phenotypes. The flexible LMM framework is apt to capture the correlations between traits and yields correct and powerful inference. TrajGWAS can also be used as a scanning tool by only testing SNPs that pass a threshold with the much slower but more powerful likelihood-based approaches. Although this paper focuses on genetic effects for the mean and WS variability, many studies are also interested in BS variance. It is possible to adapt this framework for modeling BS variability, but it comes at the cost of excluding random slopes in the model that are important in many situations.

Our findings raise a potential red flag for some existing Mendelian randomization (MR) analyses. A core assumption in MR is that the genetic determinant used as an instrument, G, only affects the outcome, Y, through the exposure, X (no horizontal pleiotropy). Many studies use mean levels of measurements as the exposure (e.g., blood pressure and cholesterol levels). This assumption may be violated in cases where (1) the outcome is associated with WS variability of the exposure independent of mean levels, such as blood pressure and glucose variability,[14,16] and (2) variants that affect both mean and WS variability are used as instruments. In our TrajGWAS analysis, we

find many SNPs that affect the mean also affect the WS variability. This suggests that the causal effects of the exposures on the outcomes estimated through these MRs may be biased because of a failure to account for the effect of the genetic determinant on the outcome acting through a second exposure (WS variability). This application gap may also provide an opportunity for new MR method development by considering both exposures.

Our method can incorporate time-varying covariates adjustment for both mean and WS variability. It makes controlling for disease status and medication usage over time possible, which sometimes increases the power (supplemental methods, section F). However, caution must be taken when considering disease and medication covariate adjustment. As medications types or disease status may be reversely correlated with biomarkers, the true genetic susceptibility can be obscured. How to best account for these effects remains an important question in future EHR-based longitudinal biomarker studies. One possible direction is a joint model that can model the biomarker trajectory, while simultaneously learning the association between disease trajectory (e.g., comorbidity events).

In conclusion, we present an ultra-efficient biobank-scale trajectories analysis tool that makes EHR-derived longitudinal traits analysis possible at very large scales. By modeling both mean effects and within subject variability, our method can provide insights that are not evident when the effects of genetic variants are only considered for the mean.

Data and code availability

The data that support the findings of this study are available from UK Biobank repositories. The UK Biobank data are retrieved under Project ID: 48152. Data are available at https://www.ukbiobank.ac.uk with the permission of the UK Biobank. The code generated during this study are available at https://github.com/OpenMendel/TrajGWAS.jl. GWAS summary statistics are available at https://kose-y.github.io/TrajGWAS-resources/.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.01.018.

Acknowledgments

This research was partially funded by grants from the National Research Foundation of Korea (NRF) (Basic Science Research Program, 2020R1A6A3A03037675, S.K.), the National Institute of General Medical Sciences (R35GM141798, J.S.S. and H.Z.), the National Human Genome Research Institute (R01HG009120, J.S.S.; R01HG006139, H.Z. and J.J.Z.), the National Science Foundation (DMS-1264153, J.S.S.; DMS-2054253, H.Z. and J.J.Z.), the National Institute of Diabetes and Digestive and Kidney Disease (K01DK106116, J.J.Z.; R01DK125187, Y.V.S.), and the National Heart, Lung, and Blood Institute (R21HL150374, J.J.Z.).

Declaration of interests

The authors declare no competing interests.

Received: August 10, 2021
Accepted: January 25, 2022
Published: February 22, 2022

Web resources

Experimental Factor Ontology, https://www.ebi.ac.uk/ols/ontologies/efo
GWAS Catalog, https://www.ebi.ac.uk/gwas/home
gwasrapidd R package, https://github.com/ramiromagno/gwasrapidd
OSCA software, https://cnsgenomics.com/software/osca
UK Biobank, https://www.ukbiobank.ac.uk/

References

1. Khera, A.V., Chaffin, M., Wade, K.H., Zahid, S., Brancale, J., Xia, R., Distefano, M., Senol-Cosar, O., Haas, M.E., Bick, A., et al. (2019). Polygenic prediction of weight and obesity trajectories from birth to adulthood. Cell *177*, 587–596.e9.
2. Tanaka, T., Basisty, N., Fantoni, G., Candia, J., Moore, A.Z., Biancotto, A., Schilling, B., Bandinelli, S., and Ferrucci, L. (2020). Plasma proteomic biomarker signature of age predicts health and life span. eLife *9*, e61073.
3. Kerschbaum, J., Rudnicki, M., Dzien, A., Dzien-Bischinger, C., Winner, H., Heerspink, H.L., Rosivall, L., Wiecek, A., Mark, P.B., Eder, S., et al. (2020). Intra-individual variability of eGFR trajectories in early diabetic kidney disease and lack of performance of prognostic biomarkers. Sci. Rep. *10*, 19743.
4. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., et al. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. *50*, 1514–1523, ~.
5. Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat. Genet. *50*, 390–400.
6. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. *20*, 467–484.
7. Goldstein, J.A., Weinstock, J.S., Bastarache, L.A., Larach, D.B., Fritsche, L.G., Schmidt, E.M., Brummett, C.M., Kheterpal, S., Abecasis, G.R., Denny, J.C., and Zawistowski, M. (2020). LabWAS: Novel findings and study design recommendations from a meta-analysis of clinical labs in two independent biobanks. PLoS Genet. *16*, e1009077.
8. Alves, A.C., De Silva, N.M.G., Karhunen, V., Sovio, U., Das, S., Taal, H.R., Warrington, N.M., Lewin, A.M., Kaakinen, M., Cousminer, D.L., et al. (2019). GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. Sci. Adv. *5*, eaaw3095.
9. Xu, K., Li, B., McGinnis, K.A., Vickers-Smith, R., Dao, C., Sun, N., Kember, R.L., Zhou, H., Becker, W.C., Gelernter, J., et al. (2020). Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. Nat. Commun. *11*, 5302.

10. Gabryszewski, S.J., Chang, X., Dudley, J.W., Mentch, F., March, M., Holmes, J.H., Moore, J., Grundmeier, R.W., Hakonarson, H., and Hill, D.A. (2021). Unsupervised modeling and genome-wide association identify novel features of allergic march trajectories. J. Allergy Clin. Immunol. *147*, 677–685.e10.
11. Rothwell, P.M., Howard, S.C., Dolan, E., O'Brien, E., Dobson, J.E., Dahlöf, B., Sever, P.S., and Poulter, N.R. (2010). Prognostic significance of visit-to-visit variability, maximum systolic blood pressure, and episodic hypertension. Lancet *375*, 895–905.
12. Ivarsdottir, E.V., Steinthorsdottir, V., Daneshpour, M.S., Thorleifsson, G., Sulem, P., Holm, H., Sigurdsson, S., Hreidarsson, A.B., Sigurdsson, G., Bjarnason, R., et al. (2017). Effect of sequence variants on variance in glucose levels predicts type 2 diabetes risk and accounts for heritability. Nat. Genet. *49*, 1398–1402.
13. Zhou, J.J., Schwenke, D.C., Bahn, G., Reaven, P.; and VADT Investigators (2018a). Glycemic variation and cardiovascular risk in the veterans affairs diabetes trial. Diabetes Care *41*, 2187–2194.
14. Zhou, J.J., Coleman, R., Holman, R.R., and Reaven, P. (2020). Long-term glucose variability and risk of nephropathy complication in UKPDS, ACCORD and VADT trials. Diabetologia *63*, 2482–2485.
15. Zhou, J.J., Koska, J., Bahn, G., and Reaven, P. (2021). Fasting glucose variation predicts microvascular risk in accord and vadt. J. Clin. Endocrinol. Metab. *106*, 1150–1162.
16. Nuyujukian, D.S., Koska, J., Bahn, G., Reaven, P.D., Zhou, J.J.; and VADT Investigators (2020). Blood pressure variability and risk of heart failure in ACCORD and the VADT. Diabetes Care *43*, 1471–1478.
17. Forbes, J.M., McCarthy, D.A., Kassianos, A.J., Baskerville, T., Fotheringham, A.K., Giuliani, K.T.K., Grivei, A., Murphy, A.J., Flynn, M.C., Sullivan, M.A., et al. (2021). T cell expression and release of kidney injury molecule-1 in response to glucose variations initiates kidney injury in early diabetes. Diabetes *70*, 1754–1766.
18. Castellanos, F.X., and Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. Nat. Rev. Neurosci. *3*, 617–628.
19. Pinar, A., Hawi, Z., Cummins, T., Johnson, B., Pauper, M., Tong, J., Tiego, J., Finlay, A., Klein, M., Franke, B., et al. (2018). Genome-wide association study reveals novel genetic locus associated with intra-individual variability in response time. Transl. Psychiatry *8*, 207.
20. Battelino, T., Danne, T., Bergenstal, R.M., Amiel, S.A., Beck, R., Biester, T., Bosi, E., Buckingham, B.A., Cefalu, W.T., Close, K.L., et al. (2019). Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care *42*, 1593–1603.
21. Ceriello, A. (2020). Glucose variability and diabetic complications: is it time to treat? Diabetes Care *43*, 1169–1171.
22. Hulse, A.M., and Cai, J.J. (2013). Genetic variants contribute to gene expression variability in humans. Genetics *193*, 95–108.
23. Ayroles, J.F., Buchanan, S.M., O'Leary, C., Skutt-Kakaria, K., Grenier, J.K., Clark, A.G., Hartl, D.L., and de Bivort, B.L. (2015). Behavioral idiosyncrasy reveals genetic control of phenotypic variability. Proc. Natl. Acad. Sci. USA *112*, 6706–6711.
24. Forsberg, S.K., Andreatta, M.E., Huang, X.-Y., Danku, J., Salt, D.E., and Carlborg, Ö. (2015). The multi-allelic genetic architecture of a variance-heterogeneity locus for molybdenum concentration in leaves acts as a source of unexplained additive genetic variance. PLoS Genet. *11*, e1005648.
25. Wang, H., Zhang, F., Zeng, J., Wu, Y., Kemper, K.E., Xue, A., Zhang, M., Powell, J.E., Goddard, M.E., Wray, N.R., et al. (2019). Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. Sci. Adv. *5*, eaaw3538.
26. Yang, J., Loos, R.J., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Mägi, R., et al. (2012). *FTO* genotype is associated with phenotypic variability of body mass index. Nature *490*, 267–272.
27. Sikorska, K., Montazeri, N.M., Uitterlinden, A., Rivadeneira, F., Eilers, P.H., and Lesaffre, E. (2015). GWAS with longitudinal phenotypes: performance of approximate procedures. Eur. J. Hum. Genet. *23*, 1384–1391.
28. Sikorska, K., Lesaffre, E., Groenen, P.J.F., Rivadeneira, F., and Eilers, P.H.C. (2018). Genome-wide analysis of large-scale longitudinal outcomes using penalization—GALLOP algorithm. Sci. Rep. *8*, 6815.
29. Wang, Z., Wang, N., Wang, Z., Jiang, L., Wang, Y., Li, J., and Wu, R. (2020). HiGwas: how to compute longitudinal GWAS data in population designs. Bioinformatics *36*, 4222–4224.
30. Hedeker, D., Mermelstein, R.J., and Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of Ecological Momentary Assessment (EMA) data. Biometrics *64*, 627–634.
31. Barrett, J.K., Huille, R., Parker, R., Yano, Y., and Griswold, M. (2019). Estimating the association between blood pressure variability and cardiovascular disease: An application using the ARIC Study. Stat. Med. *38*, 1855–1868.
32. Hedeker, D., and Nordgren, R. (2013). MIXREGLS: A program for mixed-effects location scale analysis. J. Stat. Softw. *52*, 1–38.
33. Dzubur, E., Ponnada, A., Nordgren, R., Yang, C.-H., Intille, S., Dunton, G., and Hedeker, D. (2020). MixWILD: A program for examining the effects of variance and slope of time-varying variables in intensive longitudinal data. Behav. Res. Methods *52*, 1403–1427.
34. Charlton, C., Rasbash, J., Browne, W., Healy, M., and Cameron, B. (2019). MLwiN version 3.03 (Centre for Multilevel Modelling, University of Bristol).
35. Smit, R.A.J., Jukema, J.W., Postmus, I., Ford, I., Slagboom, P.E., Heijmans, B.T., Le Cessie, S., and Trompet, S. (2018). Visit-to-visit lipid variability: Clinical significance, effects of lipid-lowering treatment, and (pharmaco) genetics. J. Clin. Lipidol. *12*, 266–276.e3.
36. Yadav, S., Cotlarciuc, I., Munroe, P.B., Khan, M.S., Nalls, M.A., Bevan, S., Cheng, Y.-C., Chen, W.-M., Malik, R., McCarthy, N.S., et al. (2013). Genome-wide analysis of blood pressure variability and ischemic stroke. Stroke *44*, 2703–2709.
37. German, C.A., Sinsheimer, J.S., Zhou, J., and Zhou, H. (2021). WiSER: Robust and scalable estimation and inference of within-subject variances from intensive longitudinal data. Biometrics. Published online June 18, 2021. https://doi.org/10.1111/biom.13506.
38. Boos, D.D. (1992). On generalized score tests. Am. Stat. *46*, 327–333.
39. Bi, W., Fritsche, L.G., Mukherjee, B., Kim, S., and Lee, S. (2020). A fast and accurate method for genome-wide time-to-event data analysis and its application to UK biobank. Am. J. Hum. Genet. *107*, 222–233.

40. Dey, R., Schmidt, E.M., Abecasis, G.R., and Lee, S. (2017). A fast and accurate algorithm to test for binary phenotypes and its application to PheWAS. Am. J. Hum. Genet. *101*, 37–49.
41. Daniels, H.E. (1980). Exact saddlepoint approximations. Biometrika *67*, 59–63.
42. Lugannani, R., and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. Adv. Appl. Probab. *12*, 475–490.
43. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018b). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. *50*, 1335–1341.
44. Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. Biometrics *2*, 110–114.
45. Wilson, D.J. (2019). The harmonic mean *p*-value for combining dependent tests. Proc. Natl. Acad. Sci. USA *116*, 1195–1200.
46. Denaxas, S., Shah, A.D., Mateen, B.A., Kuan, V., Quint, J.K., Fitzpatrick, N., Torralbo, A., Fatemifar, G., and Hemingway, H. (2020). A semi-supervised approach for rapidly creating clinical biomarker phenotypes in the UK Biobank using different primary care EHR and clinical terminology systems. JAMIA Open *3*, 545–556.
47. Yusuf, S., Bosch, J., Dagenais, G., Zhu, J., Xavier, D., Liu, L., Pais, P., López-Jaramillo, P., Leiter, L.A., Dans, A., et al. (2016). Cholesterol lowering in intermediate-risk persons without cardiovascular disease. N. Engl. J. Med. *374*, 2021–2031.
48. Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I., et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat. Genet. *50*, 1412–1425.
49. Brand, E., Wang, J.-G., Herrmann, S.-M., and Staessen, J.A. (2003). An epidemiological study of blood pressure and metabolic phenotypes in relation to the Gbeta3 C825T polymorphism. J. Hypertens. *21*, 729–737.
50. Matsubara, M., Kikuya, M., Ohkubo, T., Metoki, H., Omori, F., Fujiwara, T., Suzuki, M., Michimata, M., Hozawa, A., Katsuya, T., et al. (2001). Aldosterone synthase gene (CYP11B2) C-334T polymorphism, ambulatory blood pressure and nocturnal decline in blood pressure in the general Japanese population: the Ohasama Study. J. Hypertens. *19*, 2179–2184.
51. O'Donnell, C.J., Lindpaintner, K., Larson, M.G., Rao, V.S., Ordovas, J.M., Schaefer, E.J., Myers, R.H., and Levy, D. (1998). Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not women in the Framingham Heart Study. Circulation *97*, 1766–1772.
52. Cui, J., Hopper, J.L., and Harrap, S.B. (2002). Genes and family environment explain correlations between blood pressure and body mass index. Hypertension *40*, 7–12.
53. Cui, J.S., Hopper, J.L., and Harrap, S.B. (2003). Antihypertensive treatments obscure familial contributions to blood pressure variation. Hypertension *41*, 207–210.
54. Warren, H.R., Evangelou, E., Cabrera, C.P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J.P., et al. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat. Genet. *49*, 403–415.
55. Nierenberg, J.L., Anderson, A.H., He, J., Parsa, A., Srivastava, A., Cohen, J.B., Saraf, S.L., Rahman, M., Rosas, S.E., Kelly, T.N., et al. (2021). Association of blood pressure genetic risk score with cardiovascular disease and CKD progression: Findings from the CRIC study. Kidney360 *2*, 1251–1260.
56. Tobin, M.D., Sheehan, N.A., Scurrah, K.J., and Burton, P.R. (2005). Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. Stat. Med. *24*, 2911–2935.
57. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47* (D1), D1005–D1012.
58. Magno, R., and Maia, A.-T. (2020). gwasrapidd: an R package to query, download and wrangle GWAS catalog data. Bioinformatics *36*, 649–650.
59. Denaxas, S., Gonzalez-Izquierdo, A., Direk, K., Fitzpatrick, N.K., Fatemifar, G., Banerjee, A., Dobson, R.J.B., Howe, L.J., Kuan, V., Lumbers, R.T., et al. (2019). UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. J. Am. Med. Inform. Assoc. *26*, 1545–1559.
60. Andreassen, O.A., Djurovic, S., Thompson, W.K., Schork, A.J., Kendler, K.S., O'Donovan, M.C., Rujescu, D., Werge, T., van de Bunt, M., Morris, A.P., et al. (2013). Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am. J. Hum. Genet. *92*, 197–209.
61. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.
62. Pingitore, P., Lepore, S.M., Pirazzi, C., Mancina, R.M., Motta, B.M., Valenti, L., Berge, K.E., Retterstøl, K., Leren, T.P., Wiklund, O., and Romeo, S. (2016). Identification and characterization of two novel mutations in the *LPL* gene causing type I hyperlipoproteinemia. J. Clin. Lipidol. *10*, 816–823.
63. Davis, J.P., Huyghe, J.R., Locke, A.E., Jackson, A.U., Sim, X., Stringham, H.M., Teslovich, T.M., Welch, R.P., Fuchsberger, C., Narisu, N., et al. (2017). Common, low-frequency, and rare genetic variants associated with lipoprotein subclasses and triglyceride measures in Finnish men from the METSIM study. PLoS Genet. *13*, e1007079.
64. Tabassum, R., Rämö, J.T., Ripatti, P., Koskela, J.T., Kurki, M., Karjalainen, J., Palta, P., Hassan, S., Nunez-Fontarnau, J., Kiiskinen, T.T.J., et al. (2019). Genetic architecture of human plasma lipidome and its link to cardiovascular disease. Nat. Commun. *10*, 4329.
65. Wu, G.-Q., Xu, Y.-M., and Lau, A.T.Y. (2020). Recent insights into eukaryotic translation initiation factors 5A1 and 5A2 and their roles in human health and disease. Cancer Cell Int. *20*, 142.
66. Zhou, K., Yee, S.W., Seiser, E.L., van Leeuwen, N., Tavendale, R., Bennett, A.J., Groves, C.J., Coleman, R.L., van der Heijden, A.A., Beulens, J.W., et al. (2016). Variation in the glucose transporter gene *SLC2A2* is associated with glycemic response to metformin. Nat. Genet. *48*, 1055–1059.