

# GWAS of longitudinal trajectories at biobank scale

## Authors

Seyoon Ko, Christopher A. German,  
Aubrey Jensen, ..., Janet S. Sinsheimer, Hua Zhou,  
Jin J. Zhou

## Correspondence

[jinjinzhou@ucla.edu](mailto:jinjinzhou@ucla.edu) (J.J.Z.),  
[huazhou@ucla.edu](mailto:huazhou@ucla.edu) (H.Z.)



Ko et al., 2022, The American Journal of Human Genetics 109, 433–445  
March 3, 2022 © 2022 American Society of Human Genetics.  
<https://doi.org/10.1016/j.ajhg.2022.01.018>



# GWAS of longitudinal trajectories at biobank scale

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

## 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.<sup>1–3</sup> 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.<sup>4–6</sup> For example, a recent LabWAS tool demonstrates the broad impact of using such "real world" measurements for genetic association studies.<sup>7</sup> 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.<sup>1,8–10</sup>

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<sup>11,12</sup> 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,<sup>13–15</sup> independent of the glycemic control of the mean; blood pressure variability has been associated with the increased risk of heart failure<sup>16</sup> and stroke.<sup>11</sup> Experimental research has revealed the biological basis of glycemic variability and diabetic kidney injury.<sup>17</sup> WS variability in reaction times has also been suggested as a leading endophenotype for neurocognitive disorders, such as attention deficit hyperactivity disorder and schizophrenia.<sup>18,19</sup> 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.<sup>20,21</sup>

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.<sup>22–25</sup> 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

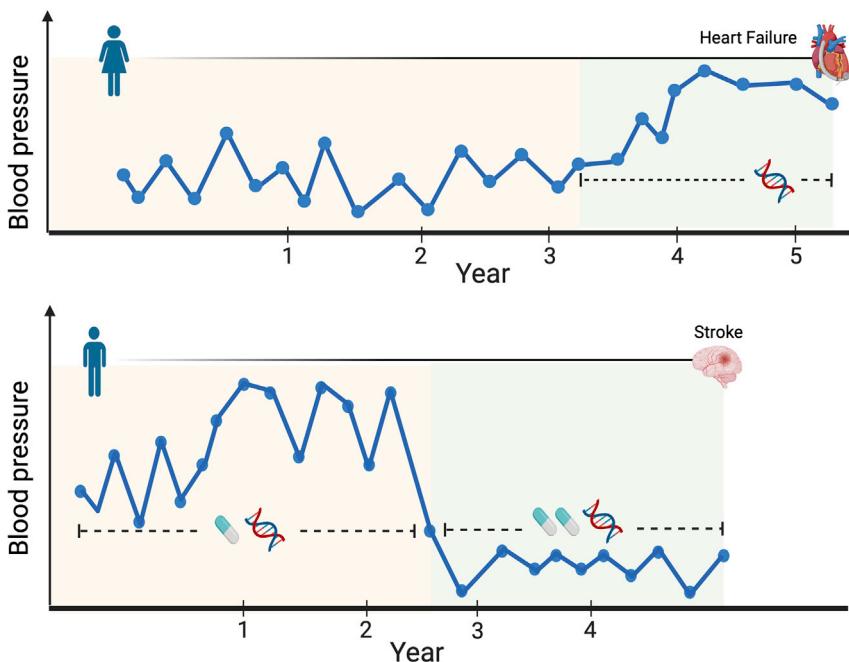
<sup>1</sup>Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>2</sup>Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>3</sup>Biostatistics and Research Decision Sciences, Merck & Co., Inc., Kenilworth, NJ 07033, USA; <sup>4</sup>Department of Epidemiology, Emory University, Atlanta, GA 30322, USA; <sup>5</sup>Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>6</sup>Department of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; <sup>7</sup>Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ 85721, USA

<sup>8</sup>These authors contributed equally.

\*Correspondence: [jinjinzhou@ucla.edu](mailto:jinjinzhou@ucla.edu) (J.J.Z.), [huazhou@ucla.edu](mailto: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)<sup>26</sup> ( $p$  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,<sup>27–29</sup> 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<sup>30</sup> 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.<sup>31</sup> 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<sup>32,33</sup> 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,<sup>34</sup> 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.<sup>1</sup> 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.<sup>12,35,36</sup> 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.<sup>31</sup> 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),<sup>37</sup> 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

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + g_i \boldsymbol{\beta}_g + \mathbf{z}_{ij}^T \boldsymbol{\gamma}_i + \epsilon_{ij}, \quad (\text{Equation 1})$$

where  $y_{ij}$  is individual  $i$ 's measurement at occasion  $j \in \{1, 2, \dots, n_i\}$ ,  $\mathbf{x}_{ij}$  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  $\boldsymbol{\beta}_g$ , and  $\mathbf{z}_{ij}$  is the  $q \times 1$  vector of covariates with corresponding random effects  $\boldsymbol{\gamma}_i$ . The WS variability is captured by the random terms  $\epsilon_{ij}$  with mean zero and inhomogeneous variance

$$\sigma_{\epsilon_{ij}}^2 = \exp(\mathbf{w}_{ij}^T \boldsymbol{\tau} + g_i \tau_g + \omega_i), \quad (\text{Equation 2})$$

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

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

$$\text{Var}\left(\begin{pmatrix} \gamma_i \\ \omega_i \end{pmatrix}\right) = \Sigma_{\gamma\omega} = \begin{pmatrix} \Sigma_\gamma & \sigma_{\gamma\omega} \\ \sigma_{\gamma\omega}^T & \sigma_\omega^2 \end{pmatrix}.$$

Covariates  $\mathbf{x}_{ij}$ ,  $\mathbf{z}_{ij}$ , and  $\mathbf{w}_{ij}$  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 ( $G \times E$ ) effects.

The models in [Equation 1](#) and [Equation 2](#) are similar to a multiple location scale model considered by Dzubur et al.,<sup>33</sup> who assume normality of the random effects  $(\gamma_i^T, \omega_i)^T$  and  $\epsilon_{ij}$  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,<sup>37</sup> 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<sup>32,33</sup> 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

$$\begin{aligned} \hat{\beta}, \hat{\beta}_g &= \arg \min_{\beta, \beta_g} \frac{1}{2} \sum_{i=1}^m (\mathbf{y}_i - \mathbf{X}_i \beta - g_i \beta_g)^T \left( \mathbf{V}_i^{(0)} \right)^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - g_i \beta_g) \\ \hat{\tau}, \hat{\tau}_g, \hat{\Sigma}_\gamma &= \arg \min_{\tau, \tau_g, \Sigma_\gamma} \frac{1}{2} \sum_{i=1}^m \text{tr} \left( \left( \mathbf{V}_i^{(0)} \right)^{-1} \mathbf{R}_i \left( \mathbf{V}_i^{(0)} \right)^{-1} \mathbf{R}_i \right), \end{aligned} \quad (\text{Equation 3})$$

where  $\mathbf{R}_i = (\mathbf{y}_i - \mathbf{X}_i \hat{\beta} - g_i \hat{\beta}_g)(\mathbf{y}_i - \mathbf{X}_i \hat{\beta} - g_i \hat{\beta}_g)^T - \mathbf{V}_i(\tau, \tau_g, \Sigma_\gamma)$ ,

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

and  $\mathbf{V}_i^{(0)} = \mathbf{V}_i(\tau^{(0)}, \tau_g^{(0)}, \Sigma_\gamma^{(0)})$  is an initial estimator of  $\text{Var}(\mathbf{Y}_i)$ . Model parameters are the mean fixed effects  $\beta$  and  $\beta_g$ , WS variance fixed effects  $\tau$  and  $\tau_g$ , and the random effects covariance  $\Sigma_\gamma$ . In the

special case  $\mathbf{V}_i^{(0)} = \mathbf{I}_{n_i}$ , 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  $\mathbf{Y}_i$ . Using an initial estimate  $\mathbf{V}_i^{(0)}$  improves the estimation efficiency of WiSER. In practice, we set the initial  $\mathbf{V}_i^{(0)}$  according to a least-squares estimator of  $\tau$  and  $\Sigma_\gamma$ . WiSER enjoys a double robustness property. It is robust to the misspecification of both the distribution of random effects  $(\gamma_i^T, \omega_i)^T$  and the distribution of  $\mathbf{Y}_i$  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.,<sup>37</sup> 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  $\theta \in \mathbb{R}$  be the genetic effect  $\beta_g$  or  $\tau_g$ . We are interested in testing the null hypothesis  $\theta_1 = 0$ . Let  $\theta_2 \in \mathbb{R}^{p+\ell+q(q+1)/2}$  collect all parameters in the null model. We first derive the score (gradient of the WiSER loss function)  $\psi_{H_1}$  under the full model and then evaluate it under the null model, i.e.,  $\psi_{H_1}(\bar{\theta})$ , where  $\bar{\theta} = (0, \hat{\theta}_2)$  and  $\hat{\theta}_2$  is the estimate under the null model. The generalized score test statistic<sup>38</sup> is

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

where  $\mathbf{V}_\psi$  is the variance of score  $\psi$ . The score test statistic  $T$  is asymptotically distributed as  $\chi_1^2$  under the null model. In [supplemental methods](#), section A, we show that the scores for testing  $\beta_g = 0$  and  $\tau_g = 0$  are

$$\mathbf{S}_{\beta_g} = \sum_{i=1}^m \left[ \mathbf{1}_{n_i}^T \left( \mathbf{V}_i^{(0)} \right)^{-1} \hat{\mathbf{R}}_i \right] g_i =: \mathbf{c}_{\beta_g}^T \mathbf{g}$$

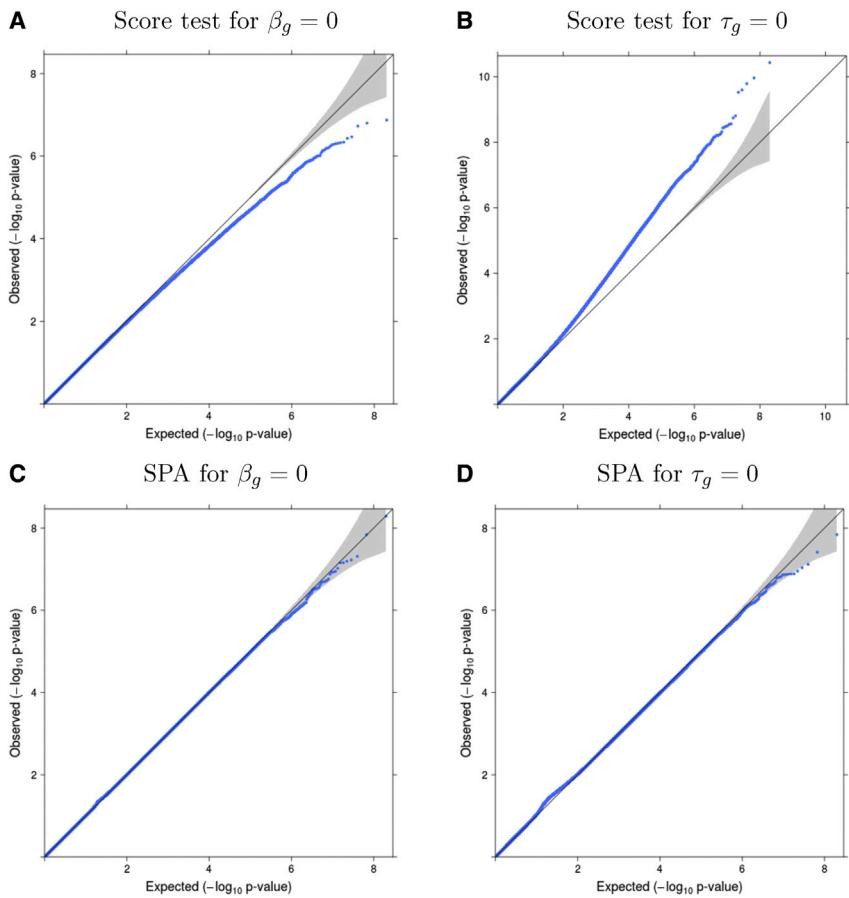
and

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

respectively. The quantities  $\hat{\mathbf{R}}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\beta}$ ,  $\hat{\mathbf{R}}_i$ ,  $\hat{\tau}$ , and  $\mathbf{V}_i^{(0)}$  are readily available from the fitted null model. Calculation of each score involves linear combination of the genotype dosages with the coefficient vector  $\mathbf{c}_{\beta_g}$  or  $\mathbf{c}_{\tau_g}$  pre-computed and cached. In [supplemental methods](#), section A, we show that the calculation of variance  $\hat{\mathbf{V}}_\psi$  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  $\text{MAF} < 0.01$ .<sup>39,40</sup> [Figures 2A](#) and [2B](#) show that, when testing a null variant with  $\text{MAF} = 0.01$ , the score test shows deflation in testing  $\beta_g$  and inflation in testing  $\tau_g$ . 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  $\beta_g$  (A) and for testing  $\tau_g$  (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  $\beta_g$  (C) and for testing  $\tau_g$  (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).<sup>39–43</sup> 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,<sup>44</sup> resulting in superior performance. For testing a single variant, we directly use the score,  $S_{\beta_g}$  and  $S_{\tau_g}$ , as the test statistic. Since the CGFs of  $S_{\beta_g}$  and  $S_{\tau_g}$  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.,<sup>39</sup> SPA is applied when  $|S_{\beta_g}| > r\sqrt{\text{Var}(\mathbf{c}_{\beta_g})\text{Var}(\mathbf{g})}$  and  $|S_{\tau_g}| > r\sqrt{\text{Var}(\mathbf{c}_{\tau_g})\text{Var}(\mathbf{g})}$  for testing  $\beta_g$  and  $\tau_g$ , 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  $S_{\tau_g}$  and  $S_{\beta_g}$  and then take their harmonic mean.<sup>45</sup>

## 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  $\gamma_i$  is intercept only so  $\mathbf{Z}_i$  is a single column of 1's.  $\mathbf{X}_i$  and  $\mathbf{W}_i$  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  $(\gamma_i, \omega_i)$  from the multivariate normal distribution with mean zero and covariance  $\Sigma_{\gamma\omega} = \begin{pmatrix} 2.0 & 0.0 \\ 0.0 & 0.1 \end{pmatrix}$

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.<sup>39</sup> We report type I error rates for testing the genetic contribution to both the mean,  $\beta_g$ , and the WS variance,  $\tau_g$ .

### Power

To evaluate the power for testing  $\tau_g$  and  $\beta_g$ , 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	$\beta_g$ score	$\beta_g$ SPA	$\tau_g$ score	$\tau_g$ 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  $5 \times 10^{-8}$  based on  $10^9$  simulation replicates. The score test shows inflated type I error at low minor allele frequencies (MAFs) for testing  $\tau_g$  where SPA (saddlepoint approximation) has proper type I error rates. Joint score and joint SPA are based on the harmonic means of the respective  $\beta_g$  and  $\tau_g$  p values.

a previously reported and validated semi-supervised approach<sup>46</sup> 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<sup>2</sup>, and age  $\times$  sex for both mean and WS variability; age and age<sup>2</sup> are treated as time-varying covariates. The selection of covariates is guided by previous GWAS analyses<sup>47,48</sup> 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”);<sup>49–51</sup> (2) adding a sensible constant (“shifting”) to the treated subjects;<sup>48,52–55</sup> and (3) censored normal regression.<sup>56</sup> Shifting and censored normal regression are often recommended for their superior performances over the indicator method.<sup>56</sup> In this paper, we use the shifting method if a sensible value for adjust-

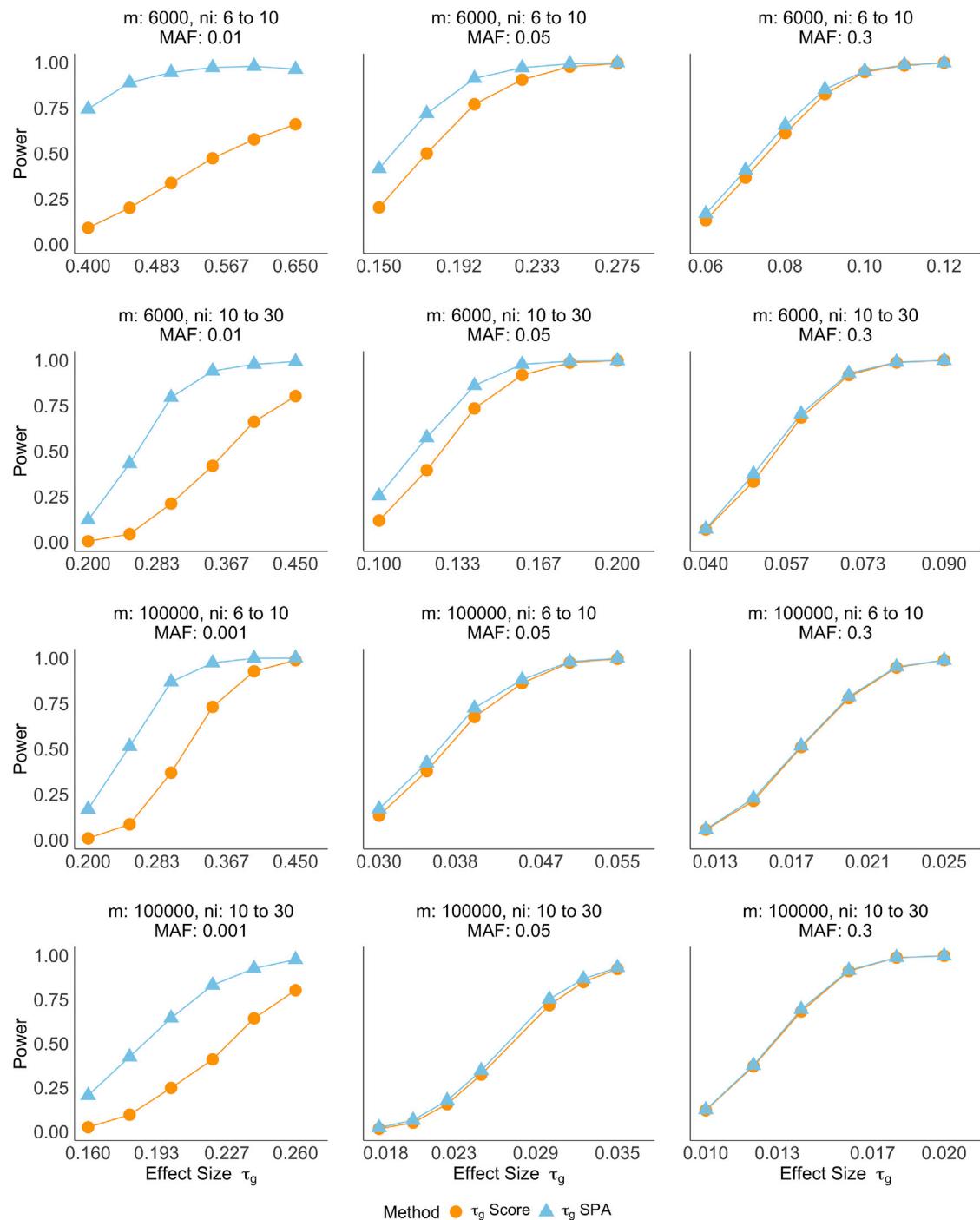
ment is available through previous studies and use the indicator method for others. We compare adjustment by shifting and adjustment by an additional covariate in [Figures S21–S24](#). For blood pressures, we add 15 mmHg for SBP and 10 mmHg for DBP<sup>55</sup> for subjects taking blood-pressure-lowering medication before standardization. For lipids, following previous GWAS analysis,<sup>47</sup> 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.,  $\beta_g$ , we create lists of previously reported genetic associations for each analyzed trait by using the GWAS Catalog<sup>57</sup> queried by the R package gwasrapidd<sup>58</sup> (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  $\alpha = 5 \times 10^{-8}$  threshold, based on  $10^9$  simulation replicates. At lower MAFs, the score test for  $\tau_g$  has substantially inflated type I errors, whereas SPA leads to well-calibrated type I error rates. Inflation in the score test for  $\tau_g$  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  $\beta_g$ , 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  $\beta_g$  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  $\tau_g$  or  $\beta_g$ . 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  $m = 6000, n_i = 6 \text{ to } 10, \text{MAF} = 0.01$  scenario. Additional QQ plots are presented in Figures S2–S5.

Power curves for testing  $\tau_g$  and  $\beta_g$  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**

Biomarker	Sample size	Median (IQR)	Mean (SD)	%	Mean (SD)	BMI
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  $\alpha = 5 \times 10^{-8}$  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

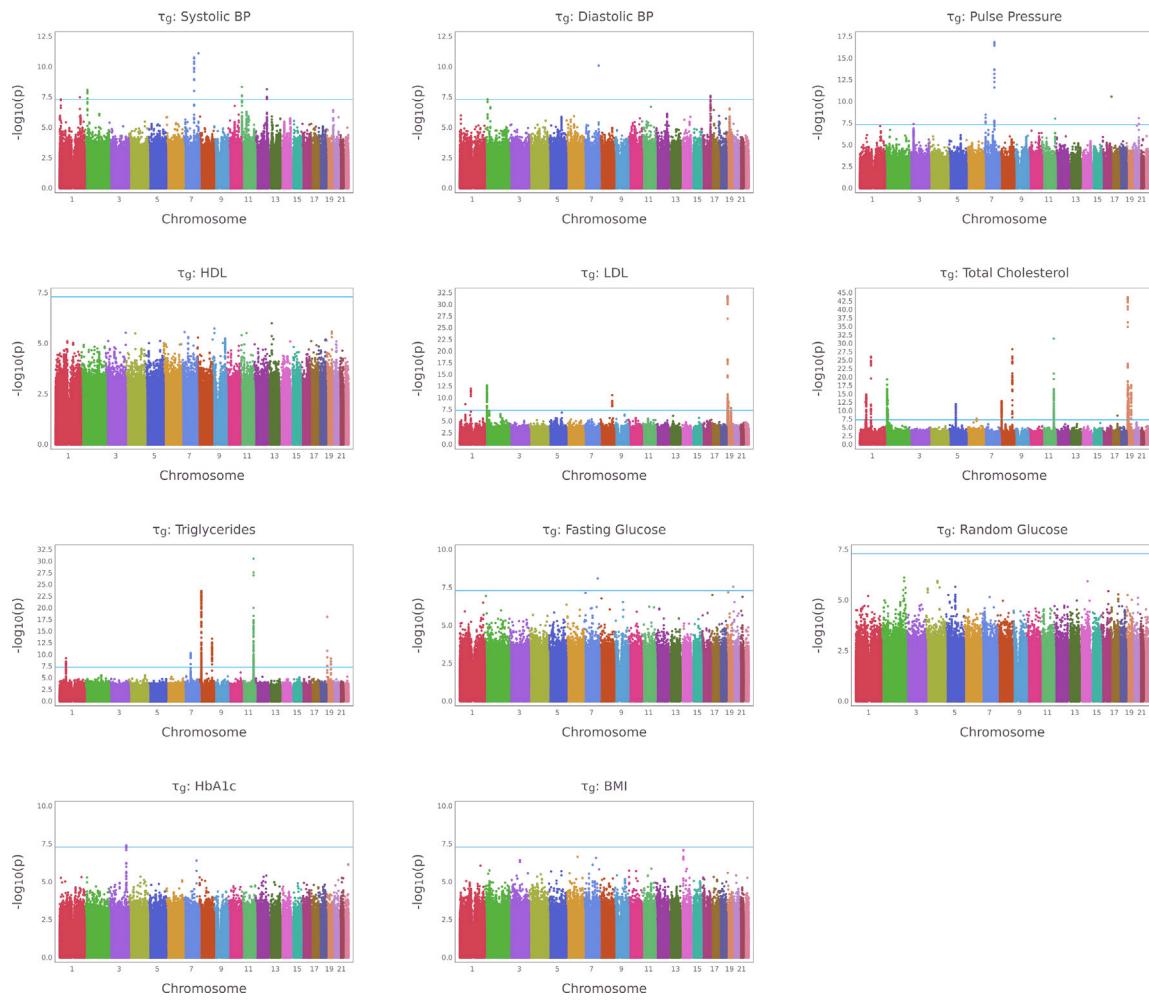
With careful implementation, each iteration of the optimization algorithm for fitting the WiSER null model scales linearly in the total number of observations  $n$ . For testing a single SNP, our score test with SPA scales linearly with the sample size  $m$ . Therefore, our TrajGWAS analysis based on 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,<sup>46,59</sup> 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  $\tau_g$  and Figure S17 for  $\beta_g$ ) 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 ( $\lambda_{GC}$ ) 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  $\tau_g$  for longitudinal markers in the UK Biobank study**

Manhattan plots for testing  $\tau_g$ , the effects of the WS variability, for 11 longitudinal biomarkers in the UK Biobank study. The blue line represents the genome-wide significance level,  $5 \times 10^{-8}$ .

mean, the median p values used for calculation of the genomic control factor may be miscalibrated.<sup>40</sup> Thus, even though many QQ plots appear normal for  $\tau_g$ , the reported  $\lambda_{GC}$  is inflated for some traits. To give a complete picture, we report  $\lambda_{GC}$  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 <sup>a</sup>	Num. of sig. loci for $\beta_g/\tau_g$ <sup>b,c</sup>	Num. of sig. SNPs for $\beta_g/\tau_g$ <sup>d</sup>	Num. of sig. SNPs in GWAS Catalog <sup>e</sup>	Replication rate $\beta_g$ <sup>f</sup>	Num. of sig. SNPs for $\beta_g/\tau_g > 1$ Mb from GWAS Catalog SNPs <sup>g</sup>
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

<sup>a</sup>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)."

<sup>b</sup>Significant SNPs for each biomarker are clumped via PLINK 1.9.<sup>61</sup> 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.

<sup>c</sup>The number of significant loci (after clumping).

<sup>d</sup>The number of significant SNPs ( $< 5 \times 10^{-8}$ ) on  $\beta_g$  and  $\tau_g$ .

<sup>e</sup>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).

<sup>f</sup>Percent of significant SNPs from GWAS Catalog that are nominally significant in  $\beta_g$  (p value  $< 0.05$ ) in the TrajGWAS analysis.

<sup>g</sup>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.<sup>60</sup>

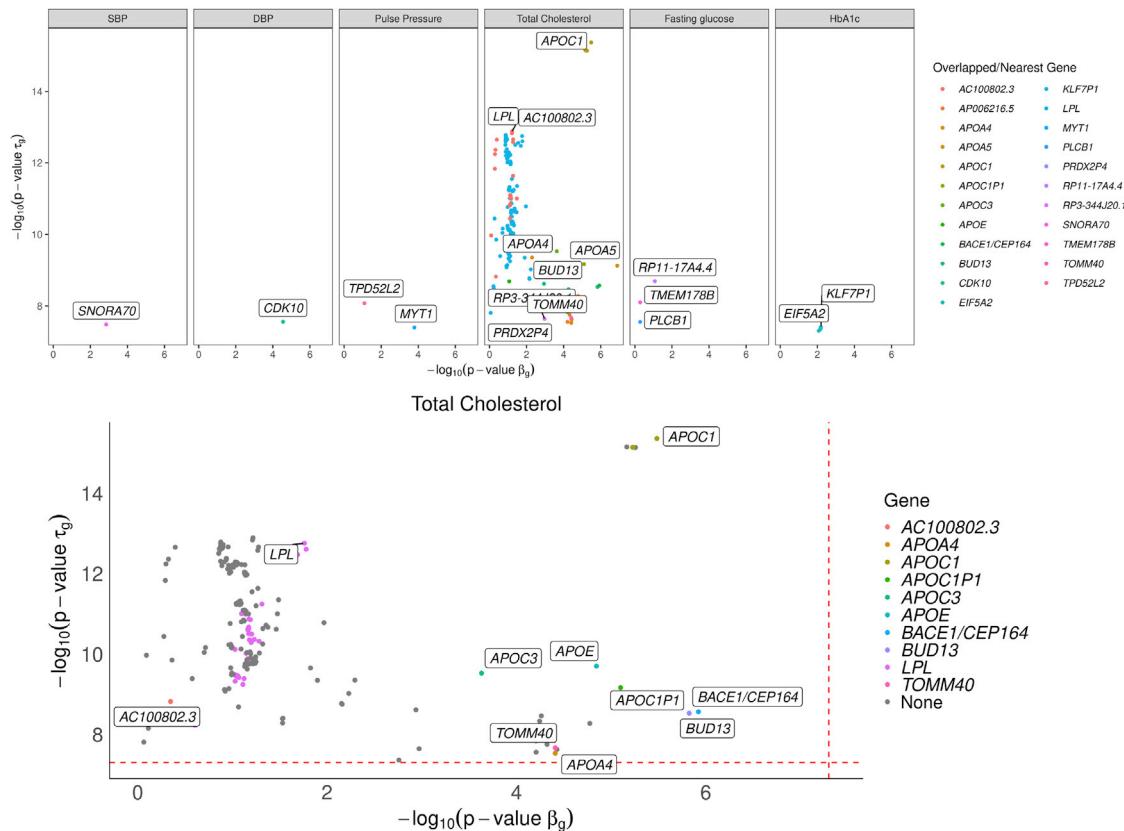
The majority of genes that affect WS variability of a trajectory also affect mean, but not always. Figure 5 highlights the 235 SNPs that are significantly associated with WS variability but not with mean levels with p values and gene annotations. Consistent with our simulation, with too few longitudinal measures, it is hard to detect  $\tau_g$  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  $\tau_g$ . 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.<sup>62</sup> Several GWASs have identified the association of *LPL* with different lipid-related phenotypes.<sup>63,64</sup> 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  $\tau_g$  on *LPL*. It shows there are big differences in the tail distri-

butions 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.<sup>65</sup> 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.<sup>66</sup>

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.<sup>25</sup> 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.<sup>25</sup> Our cohort yields the lowest p value of  $1.16 \times 10^{-102}$  for vQTL analysis. However,  $\tau_g$  for WS variability of TrajGWAS minimum p value in the same region is  $1.18 \times 10^{-4}$ , 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  $\tau_g$  but not for  $\beta_g$ .

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.<sup>30,32,33</sup> 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,<sup>45</sup> 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 variability. 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,<sup>14,16</sup> 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/ramiromagnog/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.I., 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). Lab-WAS: 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., Steinhorsdottir, 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). Tcell 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., Steinhorsdottir, 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 (pharmacogenetics). *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., Fritzsche, 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., Ordonez, 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. Bunielo, 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-Fontanau, 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.

**Supplemental information**

**GWAS of longitudinal trajectories at biobank scale**

**Seyoon Ko, Christopher A. German, Aubrey Jensen, Judong Shen, Anran Wang, Devan V. Mehrotra, Yan V. Sun, Janet S. Sinsheimer, Hua Zhou, and Jin J. Zhou**

# Supplemental Material for “GWAS of Longitudinal Trajectories at Biobank Scale”

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

---

<sup>1</sup>Department of Computational Medicine, University of California, Los Angeles, CA 90095, U.S.A.

<sup>2</sup>Department of Biostatistics, University of California, Los Angeles, CA 90095, U.S.A.

<sup>3</sup>Biostatistics and Research Decision Sciences, Merck & Co., Inc., Kenilworth, NJ 07033, U.S.A.

<sup>4</sup>Department of Epidemiology, Emory University, Atlanta, GA 30322, U.S.A.

<sup>5</sup>Department of Human Genetics, University of California, Los Angeles, CA 90095, U.S.A.

<sup>6</sup>Department of Medicine, University of California, Los Angeles, CA 90095, U.S.A.

<sup>7</sup>Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ 85721, U.S.A.

<sup>8</sup>These author contributed equally to this work.

\*Correspondence: jinjinzhou@ucla.edu

\*\*Correspondence: huazhou@ucla.edu

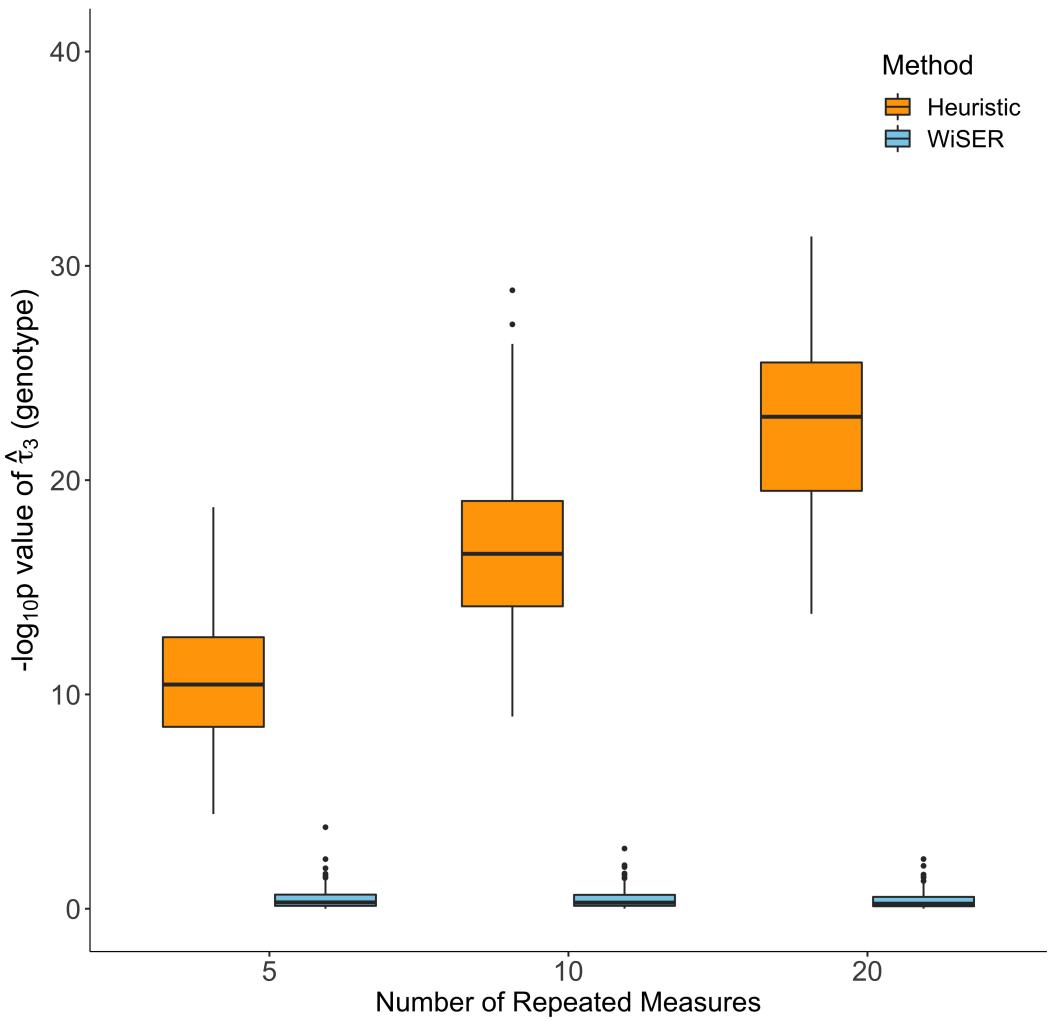
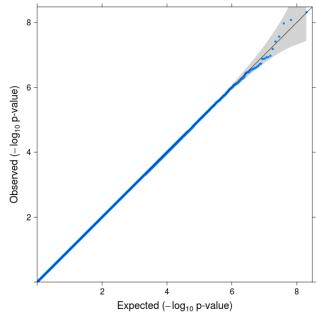


Figure S1: Boxplots of  $-\log_{10}(p$  values) for testing a null effect genotype by WiSER (adjusting for time-varying covariates) and the heuristic method of using standard deviation of the residuals as the outcome

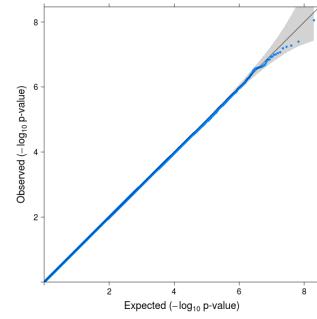
The sample size is  $m = 6,000$  and the number of replicates is 100. The heuristic method leads to incorrect inference – false positives due to systematically inflated p values.

Score test for  $\beta_g$  without SPA,  $m = 6,000$

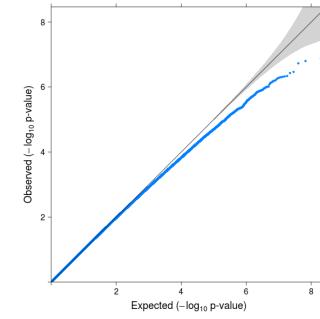
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0007$



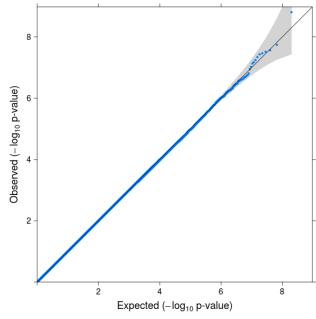
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0029$



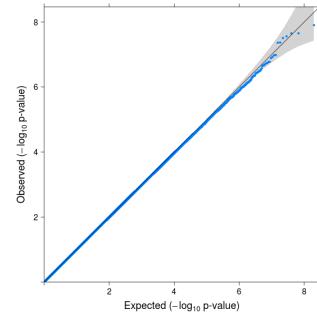
MAF=0.01,  $n_i = 6$  to 10,  $\lambda = 1.0116$



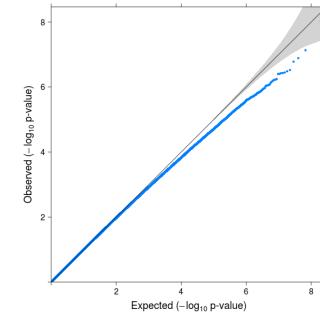
MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 1.0012$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0025$

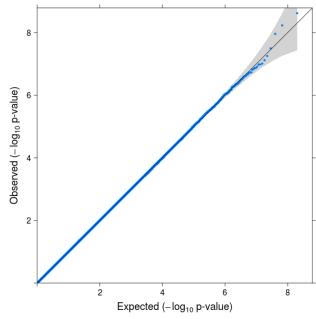


MAF=0.01,  $n_i = 10$  to 30,  $\lambda = 1.0113$

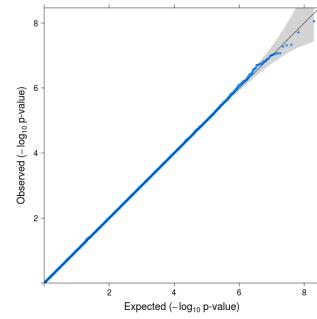


Score test for  $\beta_g$  with SPA,  $m = 6,000$

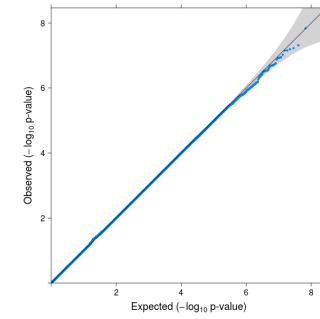
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0007$



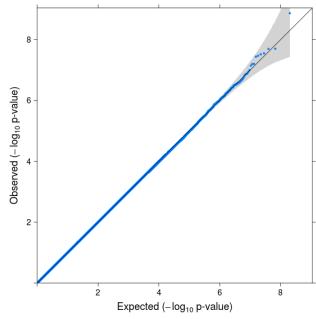
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0029$



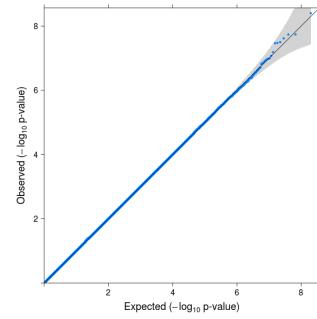
MAF=0.01,  $n_i = 6$  to 10,  $\lambda = 1.0116$



MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 1.0012$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0025$



MAF=0.01,  $n_i = 10$  to 30,  $\lambda = 1.0113$

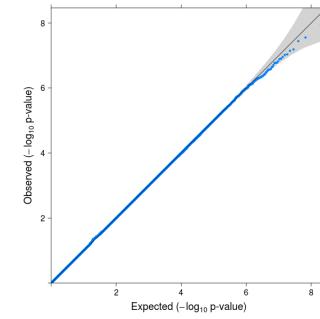
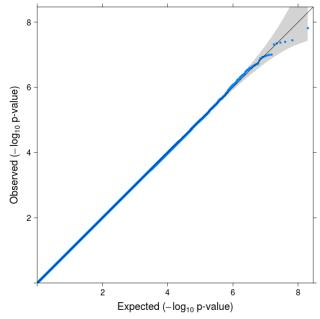


Figure S2: Quantile-Quantile (QQ) plots of p values for testing  $\beta_g$  from the score test and SPA of the simulation studies,  $m = 6,000$

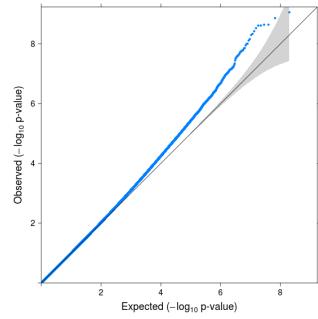
QQ plots from score test without SPA (row 1-2) and SPA (row 3-4) for testing  $\beta_g$ , where  $m = 6,000$ ,  $n_i = 6$  to 10 (row 1 and row 3) and  $n_i = 10$  to 30 (row 2 and row 4), MAF = 0.3 (column 1), 0.05 (column 2), and 0.01 (column 3), based on  $10^9$  replicates.

Score test for  $\tau_g$  without SPA,  $m = 6,000$

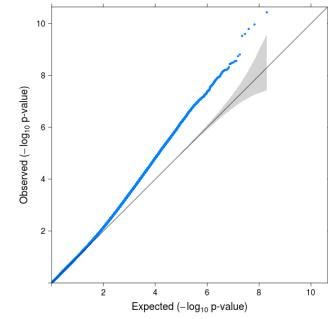
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0021$



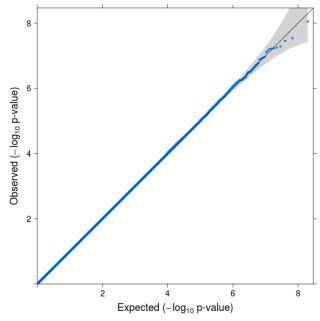
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0059$



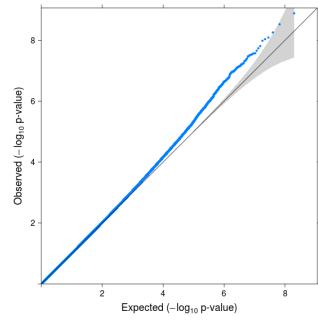
MAF=0.01,  $n_i = 6$  to 10,  $\lambda = 1.0241$



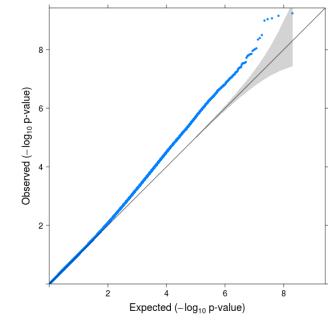
MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 1.0020$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0053$

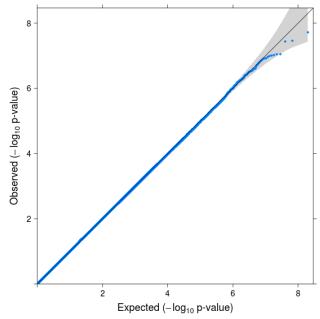


MAF=0.01,  $n_i = 10$  to 30,  $\lambda = 1.0225$

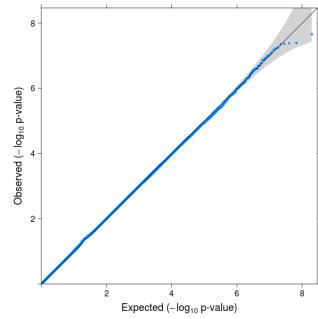


Score test for  $\tau_g$  with SPA,  $m = 6,000$

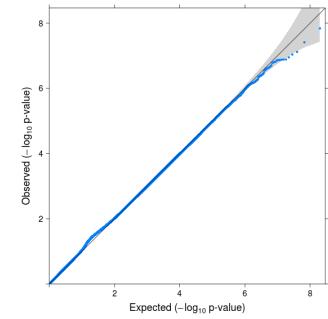
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0021$



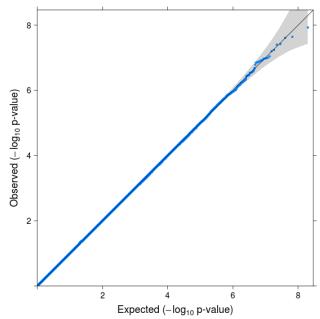
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0059$



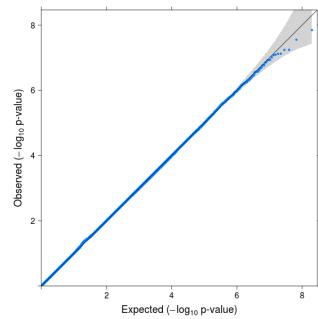
MAF=0.01,  $n_i = 6$  to 10,  $\lambda = 1.0241$



MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 1.0020$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0053$



MAF=0.01,  $n_i = 10$  to 30,  $\lambda = 1.0225$

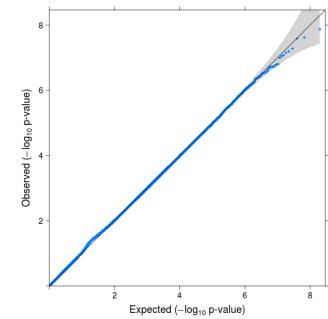
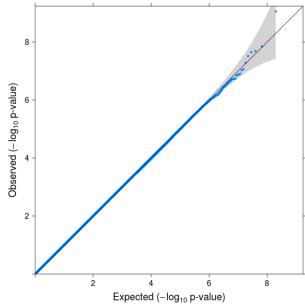


Figure S3: Quantile-Quantile (QQ) plots of p values for testing  $\tau_g$  from the score test and SPA of the simulation studies,  $m = 6,000$

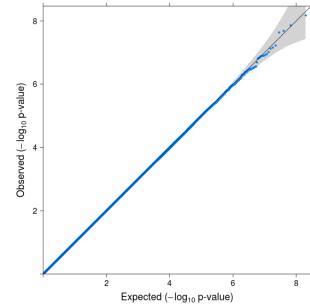
QQ plots of p values from score test (row 1-2) and SPA (row 3-4) for testing  $\tau_g$ , where  $m = 6,000$ ,  $n_i = 6$  to 10 (row 1 and row 3) and  $n_i = 10$  to 30 (row 2 and row 4), MAF = 0.3 (column 1), 0.05 (column 2), and 0.01 (column 3), based on  $10^9$  replicates.

Score test for  $\beta_g$  without SPA,  $m = 100,000$

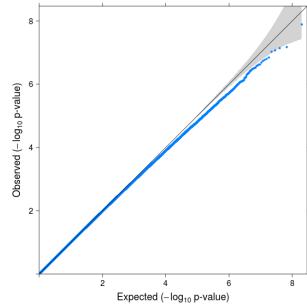
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0000$



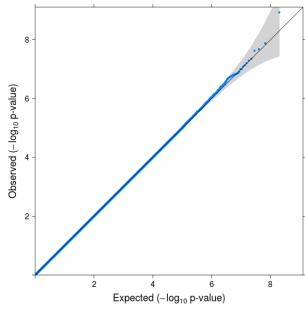
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0001$



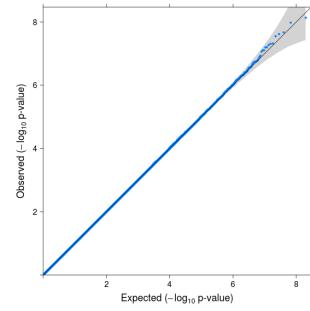
MAF=0.001,  $n_i = 6$  to 10,  $\lambda = 1.0063$



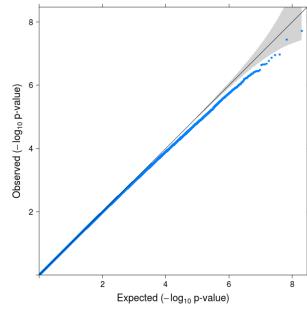
MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 0.9999$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0000$

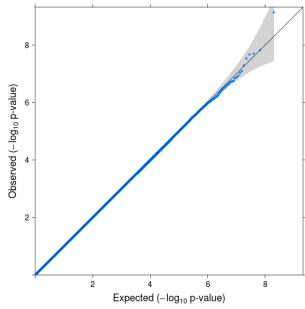


MAF=0.001,  $n_i = 10$  to 30,  $\lambda = 1.0064$

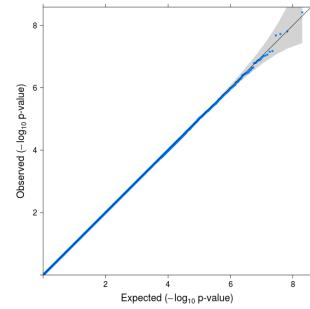


Score test for  $\beta_g$  with SPA,  $m = 100,000$

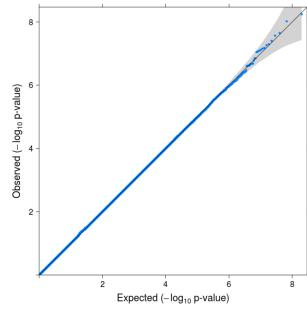
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0000$



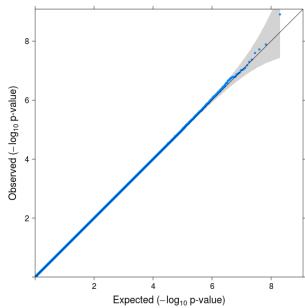
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0001$



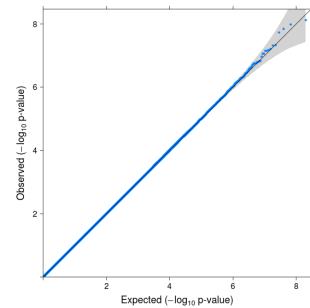
MAF=0.001,  $n_i = 6$  to 10,  $\lambda = 1.0063$



MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 0.9999$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0000$



MAF=0.001,  $n_i = 10$  to 30,  $\lambda = 1.0064$

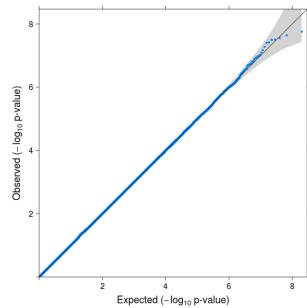
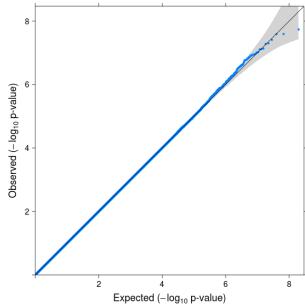


Figure S4: Quantile-Quantile (QQ) plots of p values for testing  $\beta_g$  from the score test and SPA of the simulation studies,  $m = 100,000$

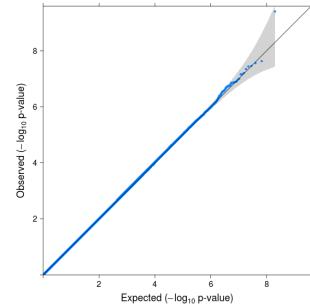
QQ plots of p values from score test (row 1-2) and SPA (row 3-4) for testing  $\beta_g$ , where  $m = 100,000$ ,  $n_i = 6$  to 10 (row 1 and row 3) and  $n_i = 10$  to 30 (row 2 and row 4), MAF = 0.3 (column 1), 0.05 (column 2), and 0.001 (column 3), based on  $10^9$  replicates.

Score test for  $\tau_g$  without SPA,  $m = 100,000$

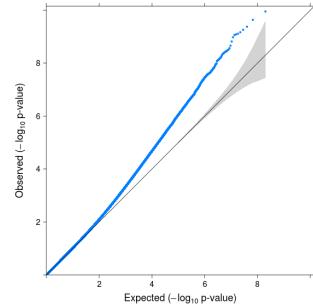
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0002$



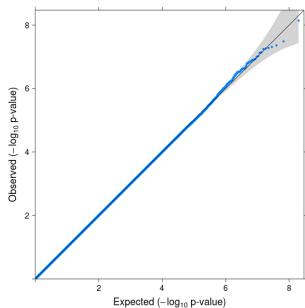
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0005$



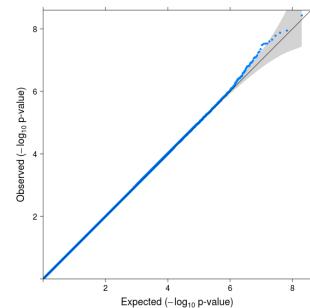
MAF=0.001,  $n_i = 6$  to 10,  $\lambda = 1.0148$



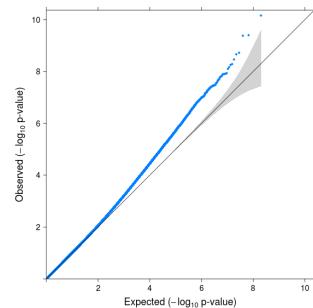
MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 0.9995$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0001$

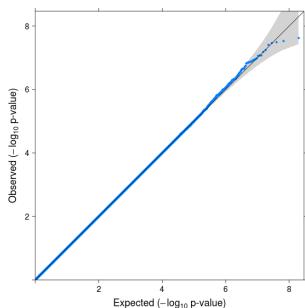


MAF=0.001,  $n_i = 10$  to 30,  $\lambda = 1.0131$

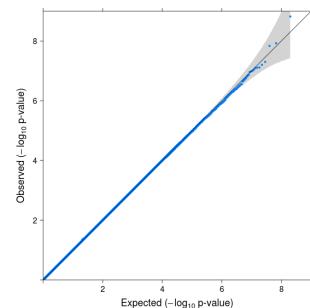


Score test for  $\tau_g$  with SPA,  $m = 100,000$

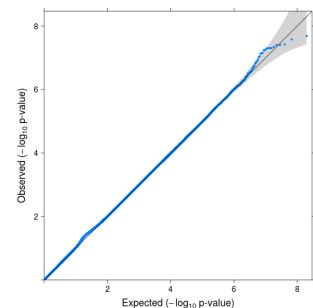
MAF=0.3,  $n_i = 6$  to 10,  $\lambda = 1.0002$



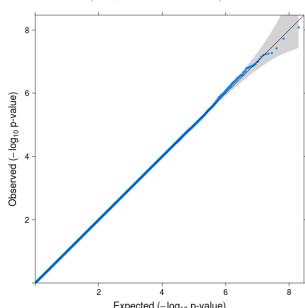
MAF=0.05,  $n_i = 6$  to 10,  $\lambda = 1.0005$



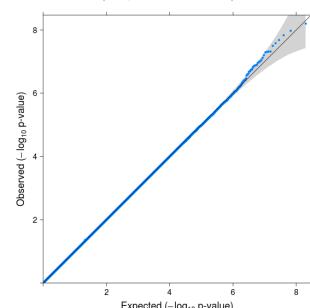
MAF=0.001,  $n_i = 6$  to 10,  $\lambda = 1.0148$



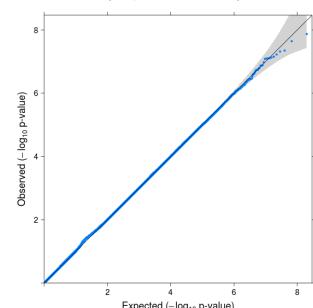
MAF=0.3,  $n_i = 10$  to 30,  $\lambda = 0.9995$



MAF=0.05,  $n_i = 10$  to 30,  $\lambda = 1.0001$



MAF=0.001,  $n_i = 10$  to 30,  $\lambda = 1.0121$



**Figure S5: Quantile-Quantile (QQ) plots of p values for testing  $\tau_g$  from the score test and SPA of the simulation studies,  $m = 10,000$**

QQ plots of p values from score test (row 1-2) and SPA (row 3-4) for testing  $\tau_g$ , where  $m = 10,000$ ,  $n_i = 6$  to 10 (row 1 and row 3) and  $n_i = 10$  to 30 (row 2 and row 4), MAF = 0.3 (column 1), 0.05 (column 2), and 0.001 (column 3), based on  $10^9$  replicates.

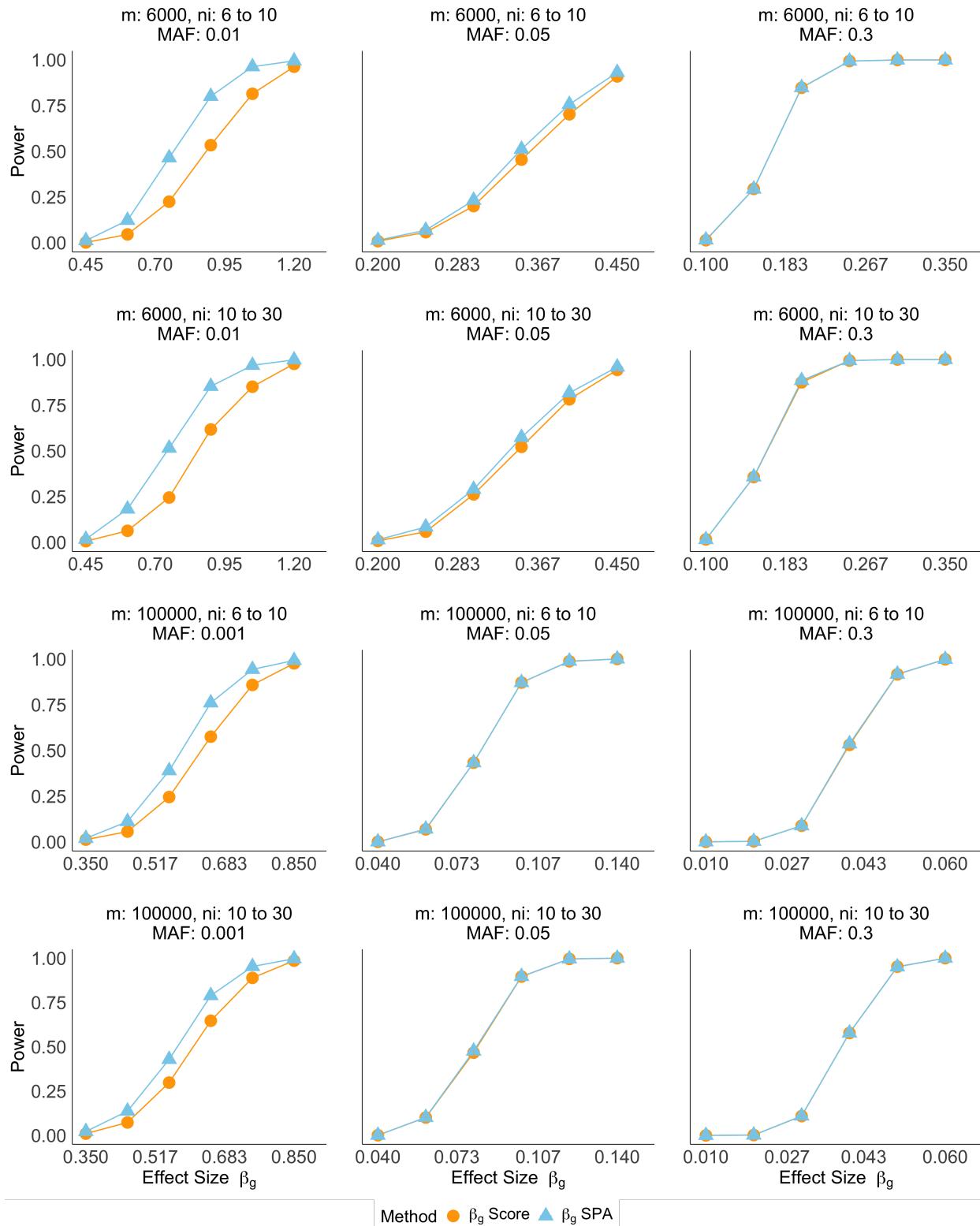


Figure S6: **Empirical powers of testing  $\beta_g$  with score test and SPA**

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.

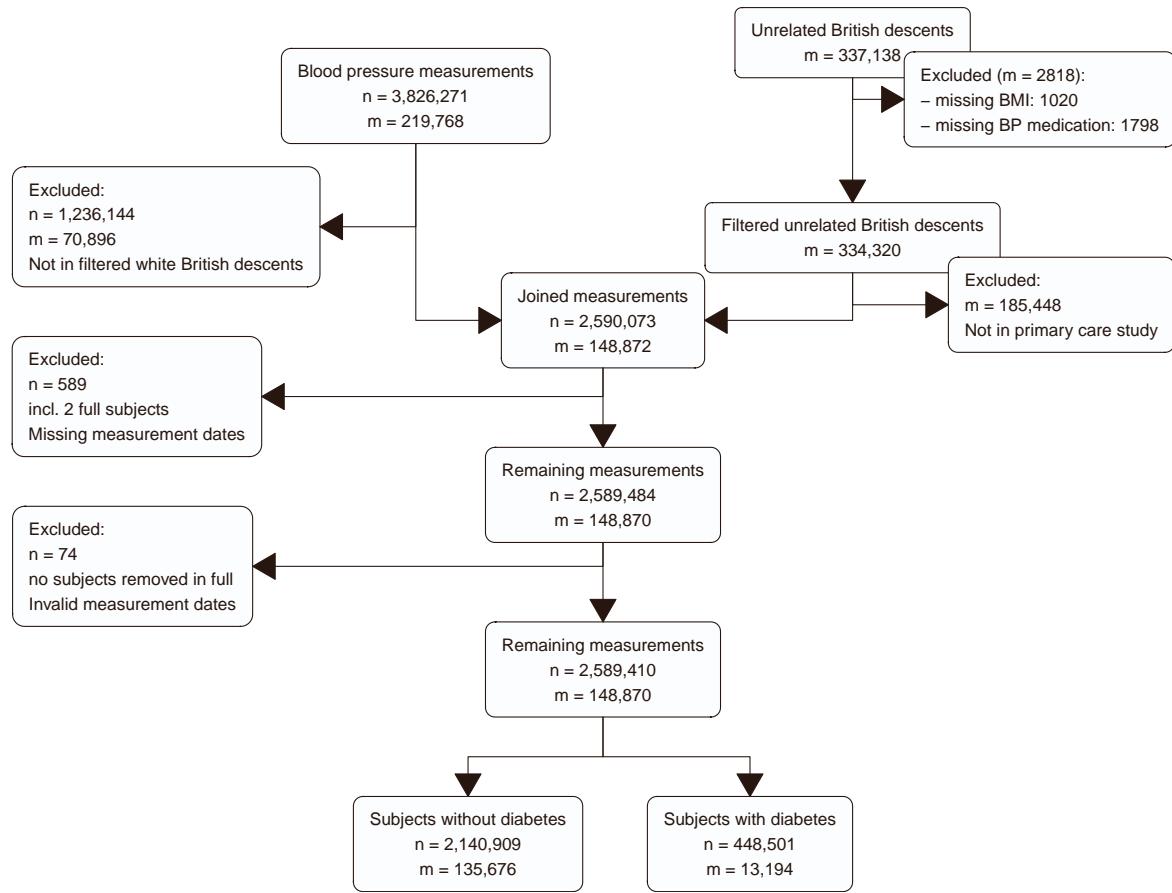


Figure S7: Cohort curation for blood pressure TrajGWAS analysis from UK Biobank primary care data.

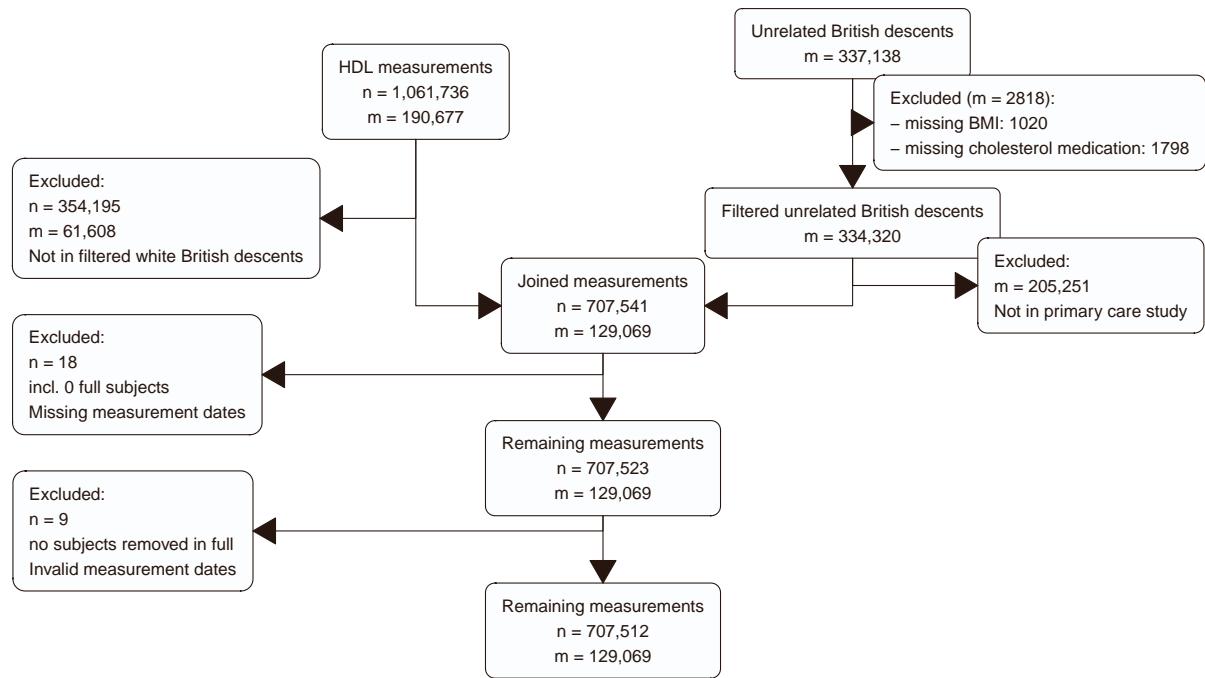


Figure S8: Cohort curation for HDL TrajGWAS analysis from UK Biobank primary care data.

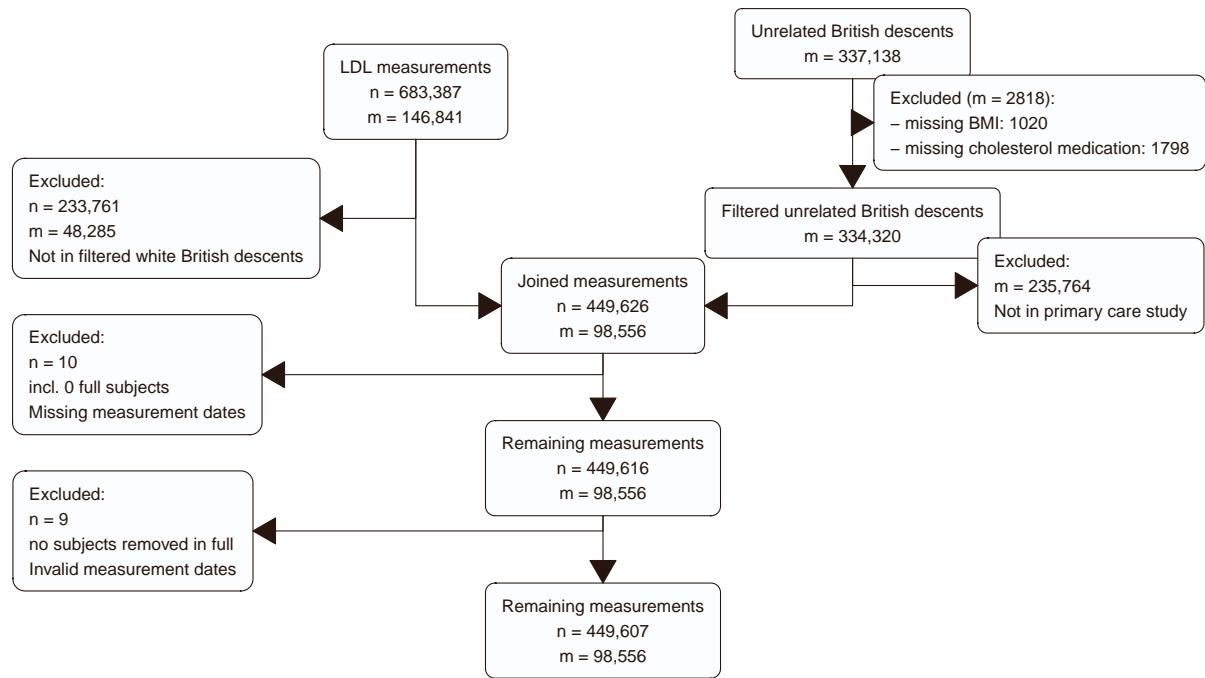
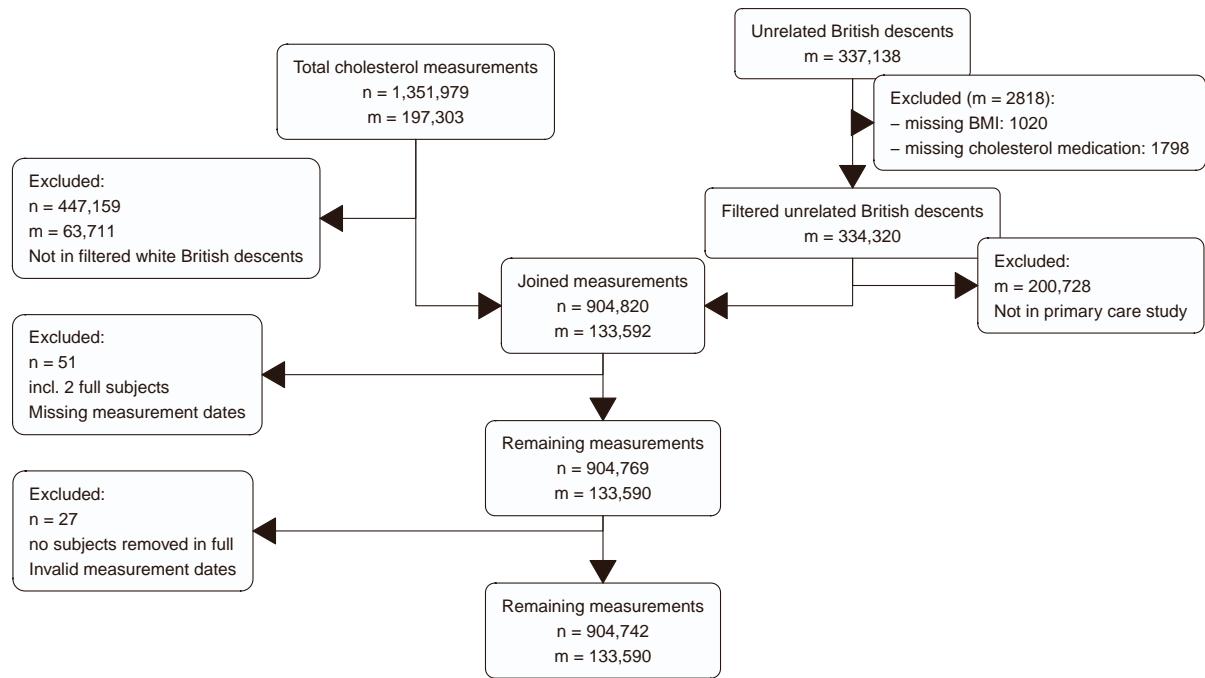


Figure S9: Cohort curation for LDL TrajGWAS analysis from UK Biobank primary care data.



**Figure S10: Cohort curation for total cholesterol TrajGWAS analysis from UK Biobank primary care data.**

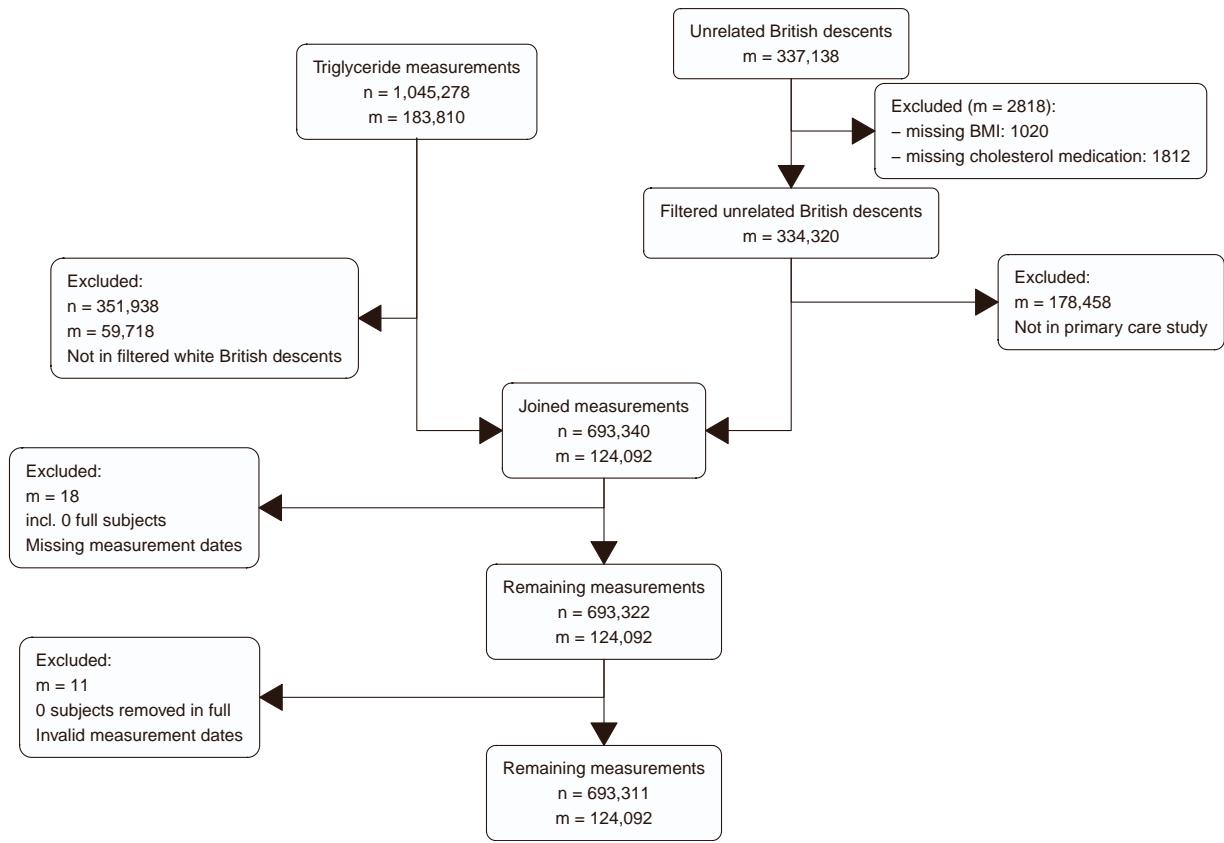


Figure S11: Cohort curation for triglycerides TrajGWAS analysis from UK Biobank primary care data.

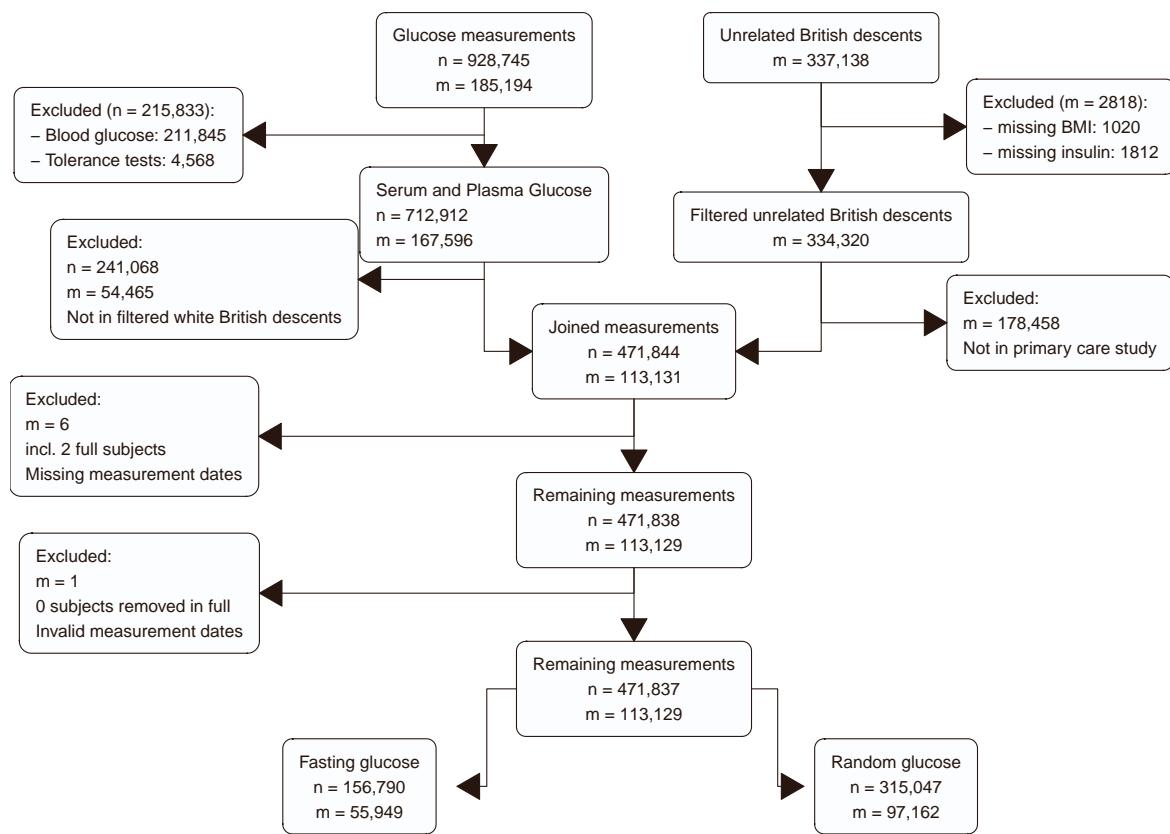


Figure S12: Cohort curation for glucose TrajGWAS analysis from UK Biobank primary care data.

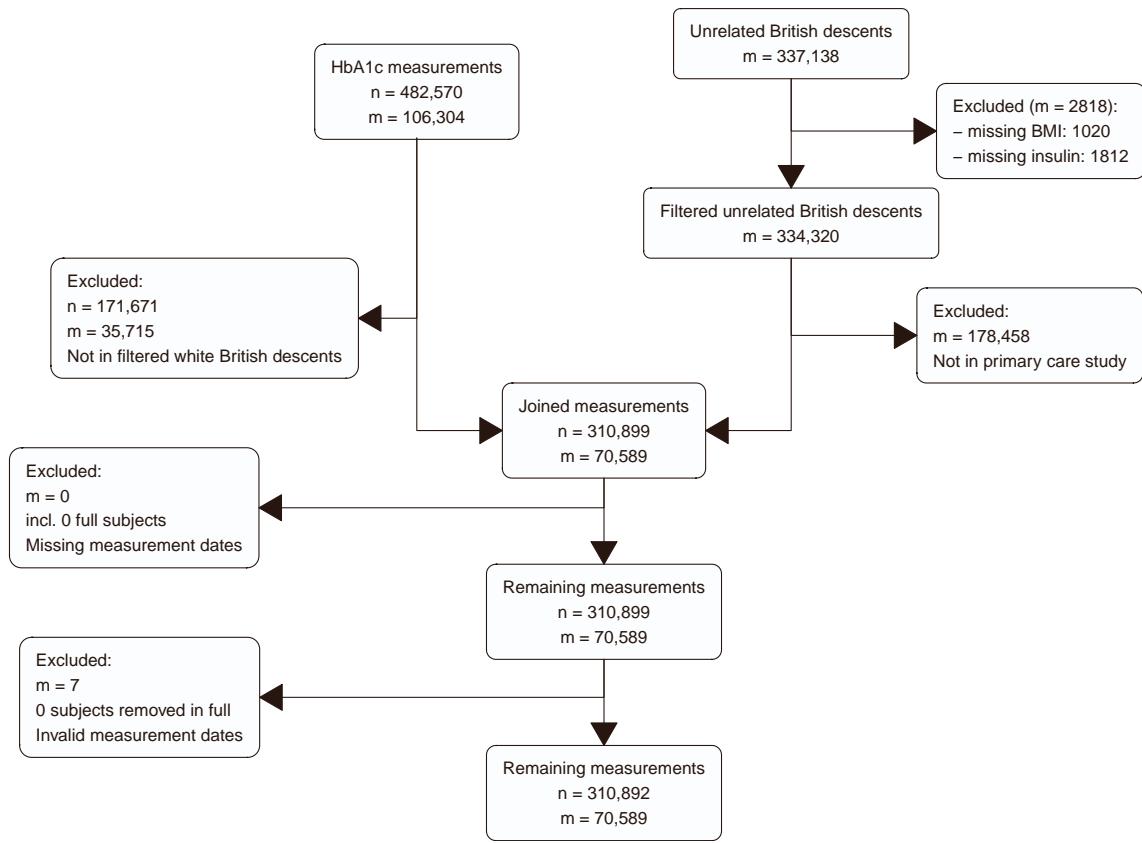


Figure S13: Cohort curation for HbA1c TrajGWAS analysis from UK Biobank primary care data.

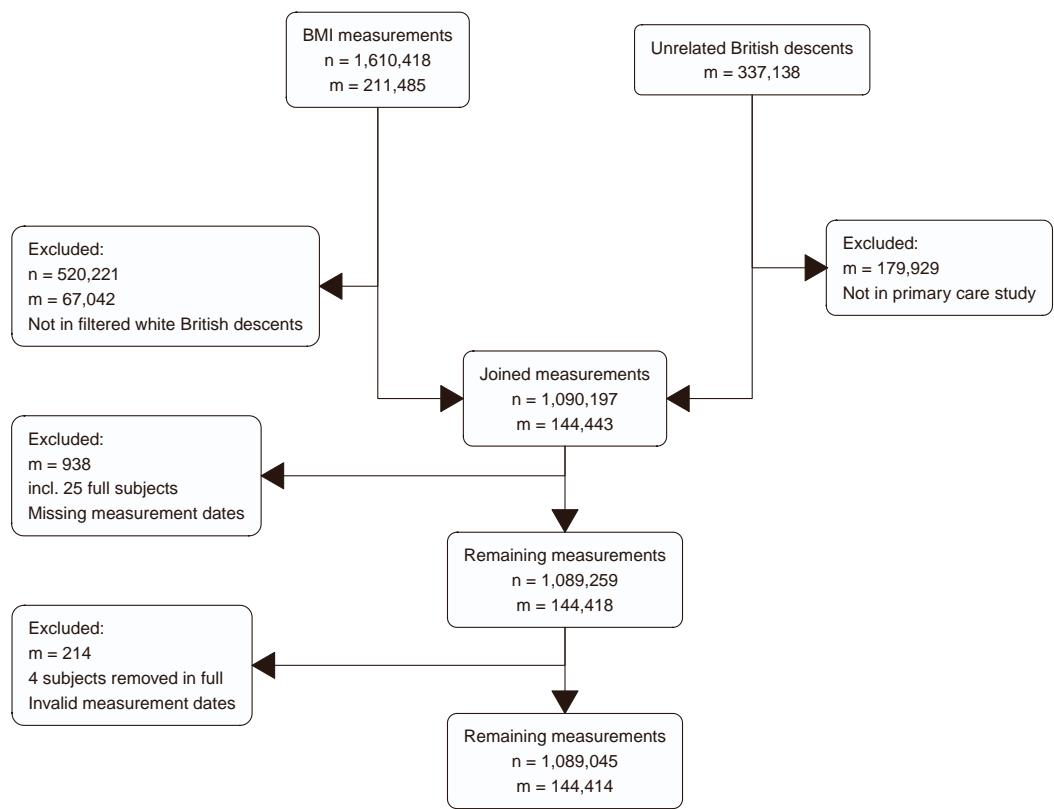


Figure S14: Cohort curation for BMI TrajGWAS analysis from UK Biobank primary care data.

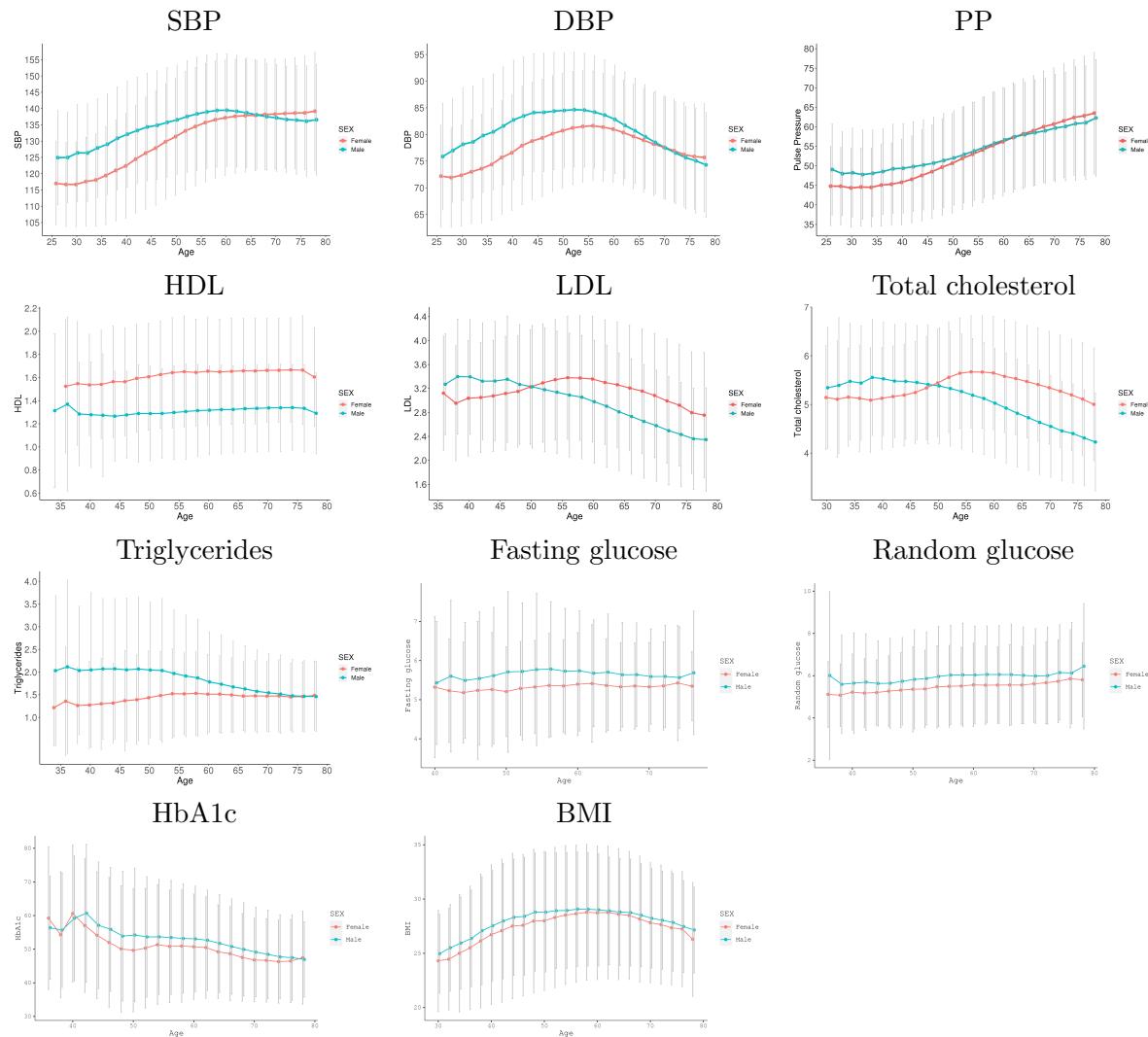
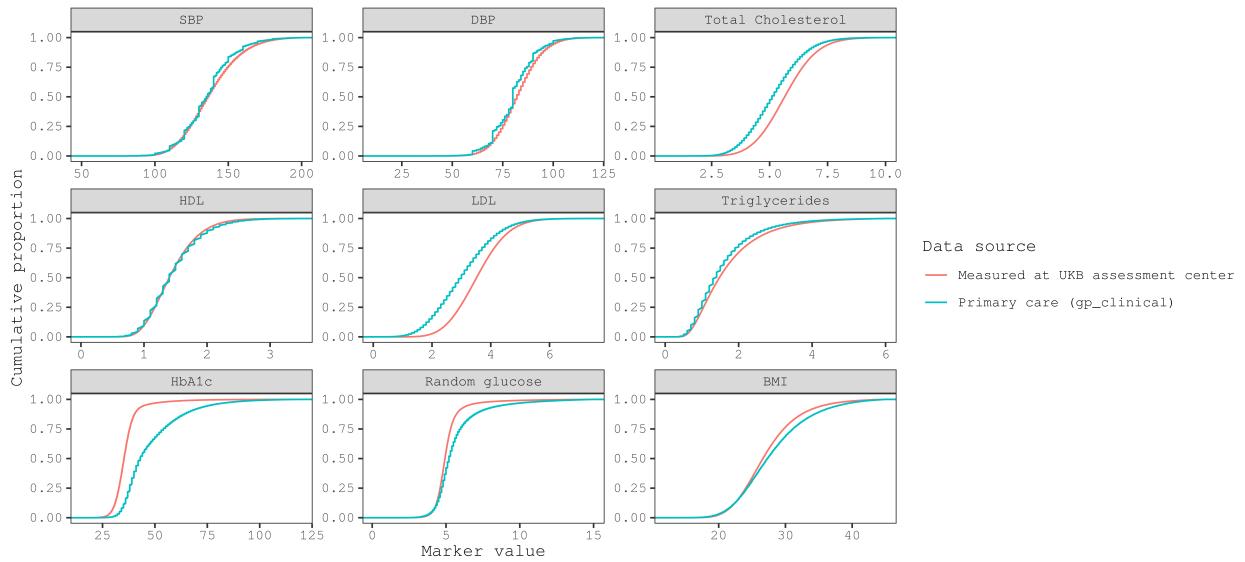
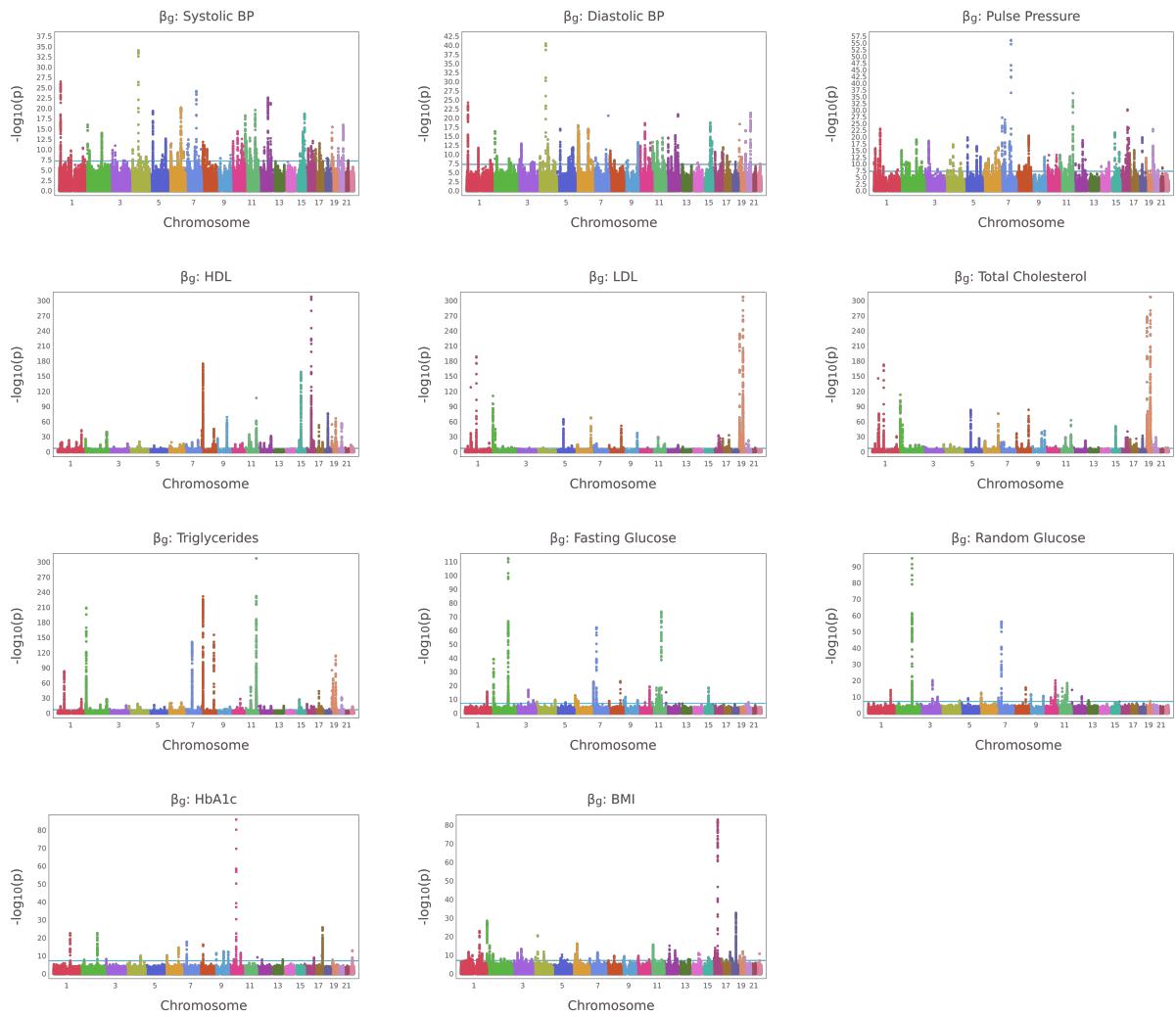


Figure S15: Mean profile plot over age groups for biomarker measures extracted from the UK Biobank primary care data.



**Figure S16: Empirical cumulative distribution functions (eCDFs) of biomarker measures from UK Biobank primary care data (blue curve) and assessment centers (red curve) respectively.**

Agreement between the eCDFs provides quality control for the extraction procedure from the primary care data.



**Figure S17: Manhattan plots for testing  $\beta_g$  for longitudinal markers in the UK Biobank study**

Manhattan plots for testing  $\beta_g$ , effects to the mean, for 11 longitudinal biomarkers in the UK Biobank study. The blue line represents the genome-wide significance level,  $5 \times 10^{-8}$ .

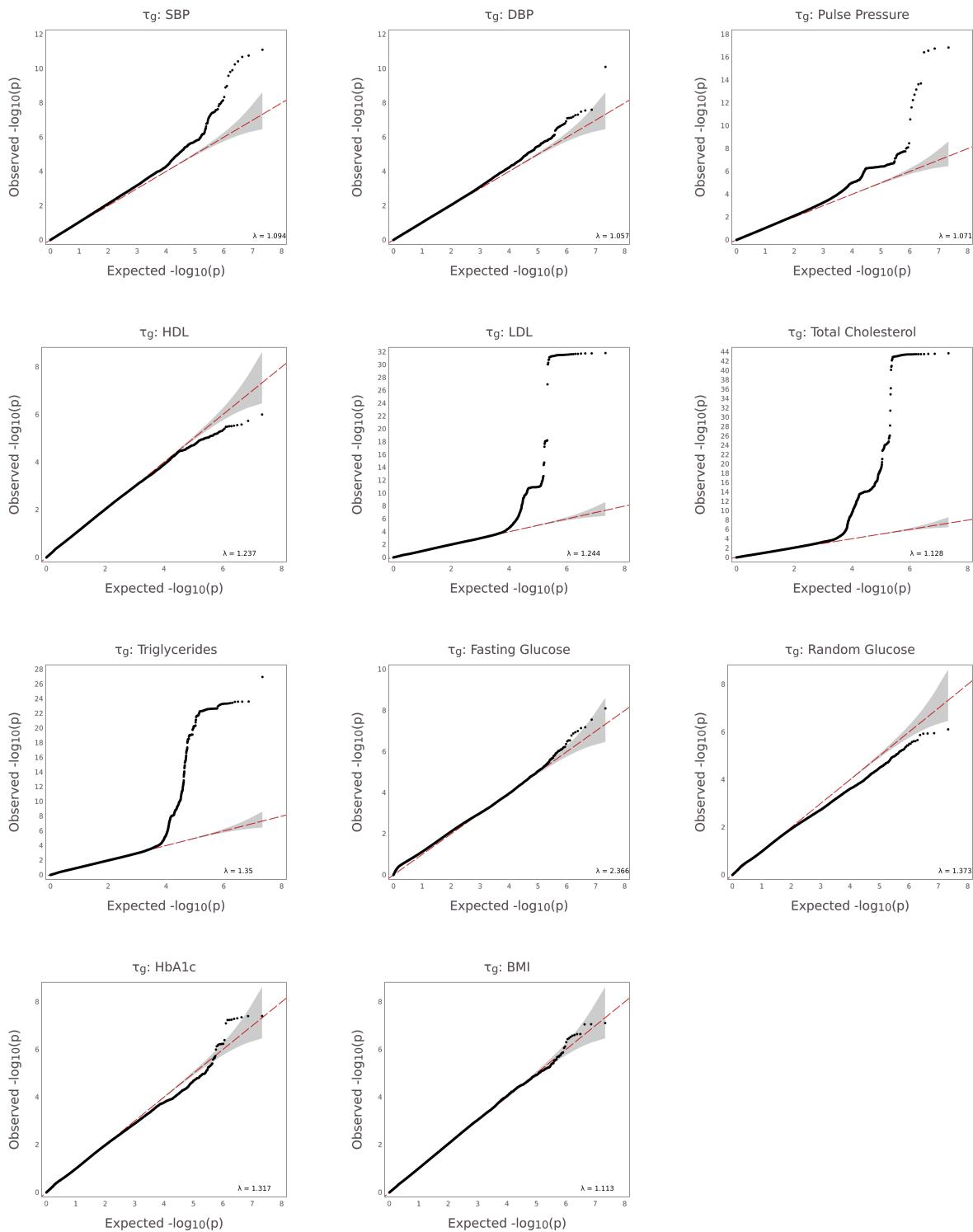


Figure S18: TrajGWAS QQ plots for testing  $\tau_g$ , effects on the WS variability, on 11 longitudinal biomarkers in the UK Biobank study

Genomic control factor,  $\lambda$ , is based on the median p value, where SPA is not applied. See Supplementary Table S3 for the  $\lambda$  values at different p value quantiles.

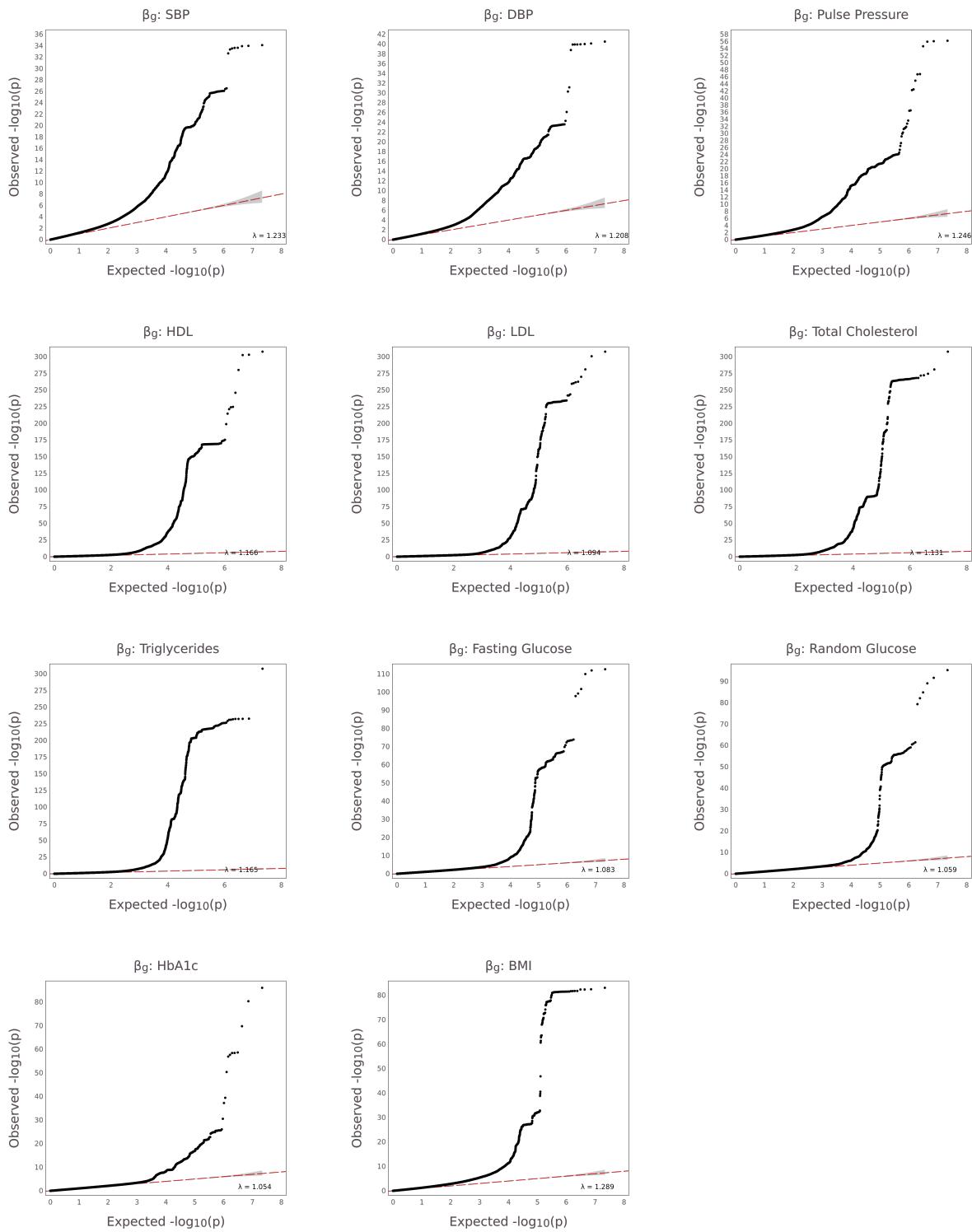


Figure S19: TrajGWAS QQ plots for  $\beta_g$ , effects to the mean, on 11 longitudinal biomarkers in the UK Biobank study

Genomic control factor,  $\lambda$ , is based off the median p value, where the score test is often applied. See Supplementary Table S3 for different p value quantile cutoffs.

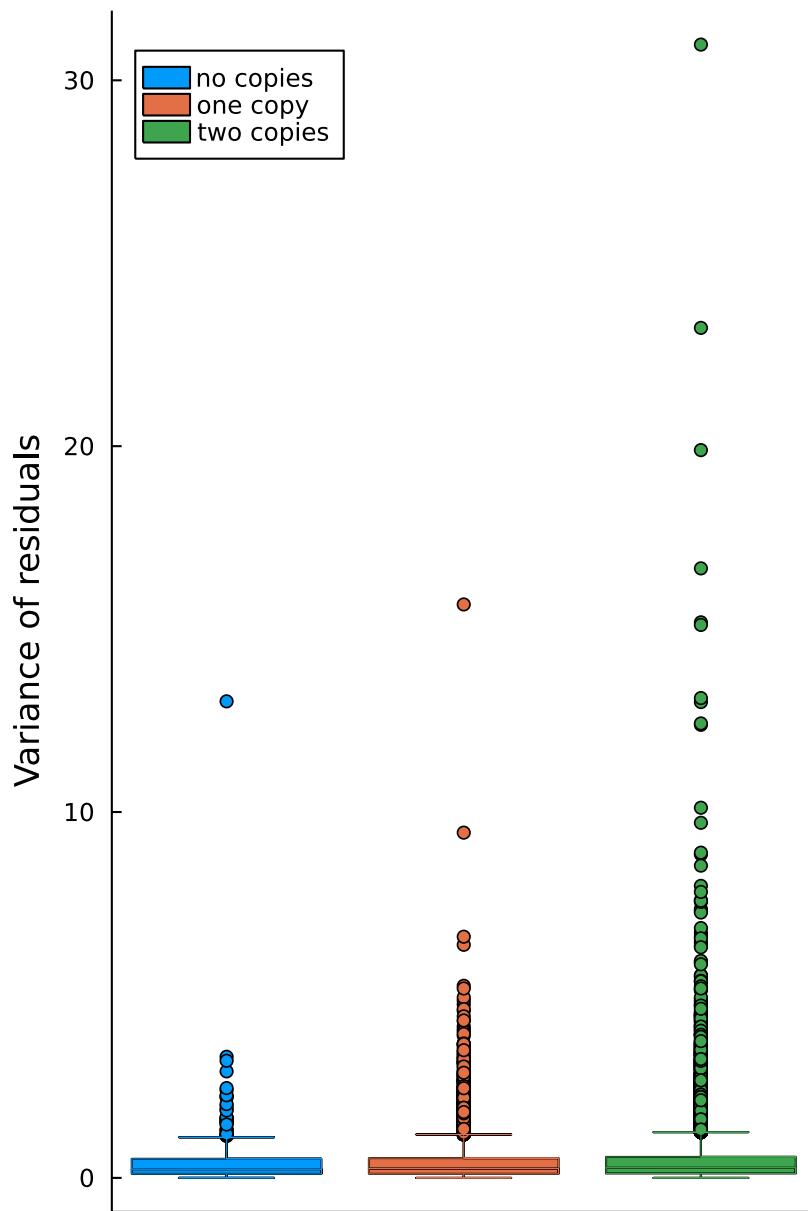
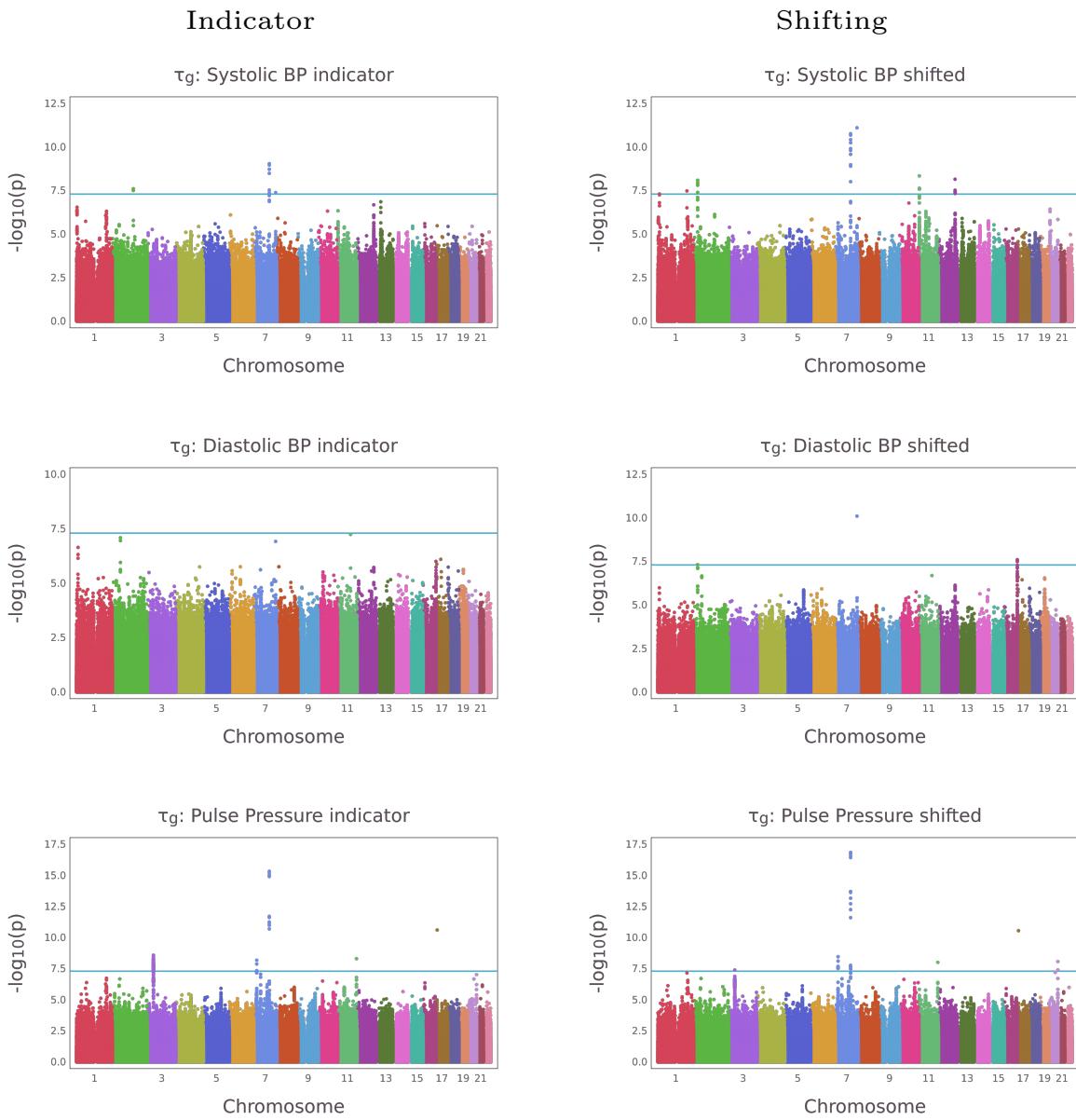


Figure S20: **Boxplots of within-sample variance of residuals of a SNP significant in terms of  $\tau_g$**

Boxplots of within-sample variance of residuals (regressed out mean part covariates other than the SNP count) of total cholesterol for different reference allele counts (0, 1, or 2) of rs6993414, the SNP with the lowest p value in terms of  $\tau_g$  on the *LPL* gene. Note that the phenotype is standardized before the analysis.



**Figure S21: Comparison of the medication adjustment methods in the blood pressure TrajGWAS analysis.**

We plot Manhattan plots for  $\tau_g$  in the blood pressure TrajGWAS analysis, effects to the WS variability for SBP, DBP, and PP in UK Biobank study. (Left) Medication adjusted by an additional indicator covariate reflecting on or off medication in both  $\beta_g$  and  $\tau_g$ . (Right) A sensible constant is added to the observed measures.

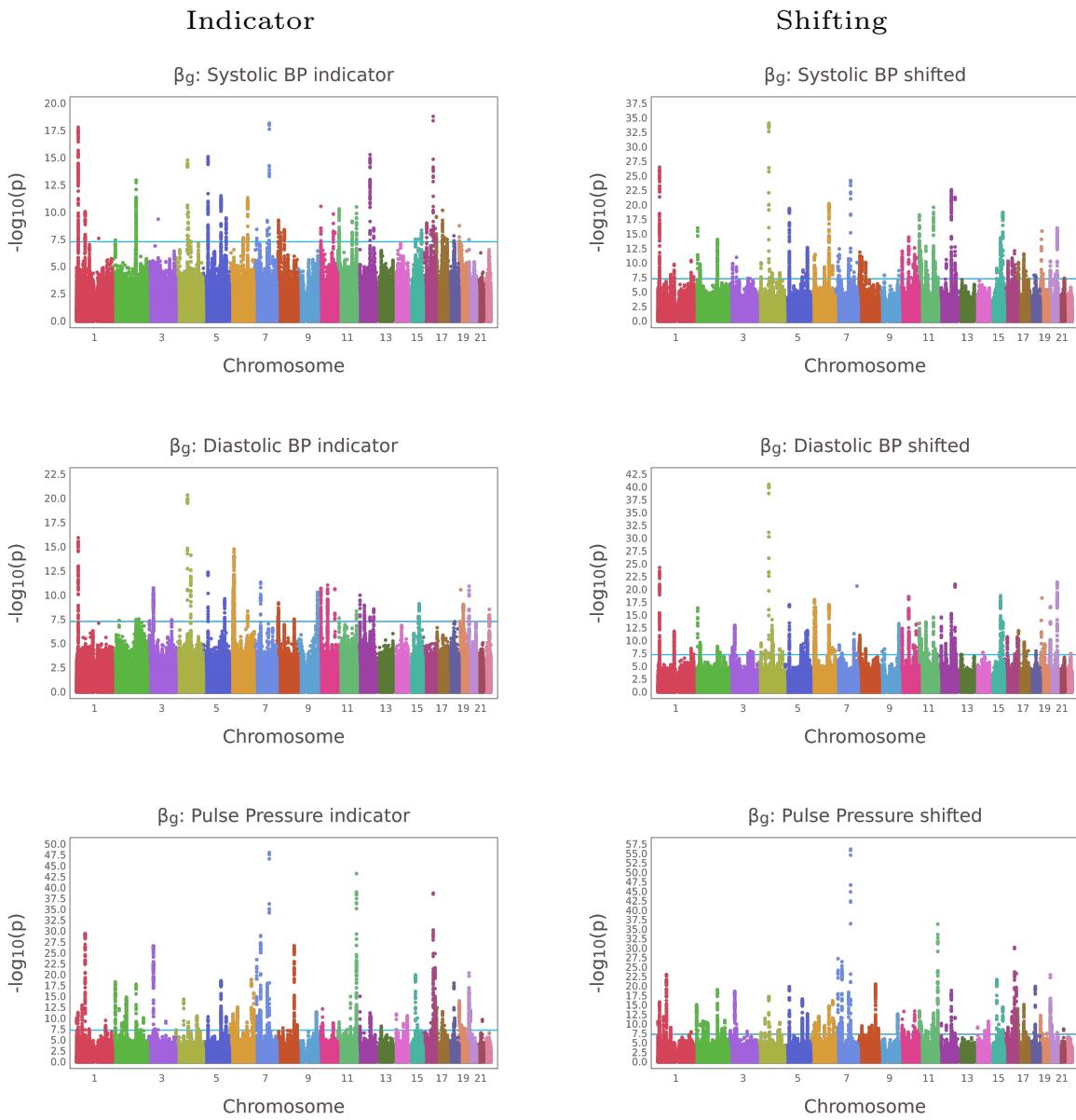
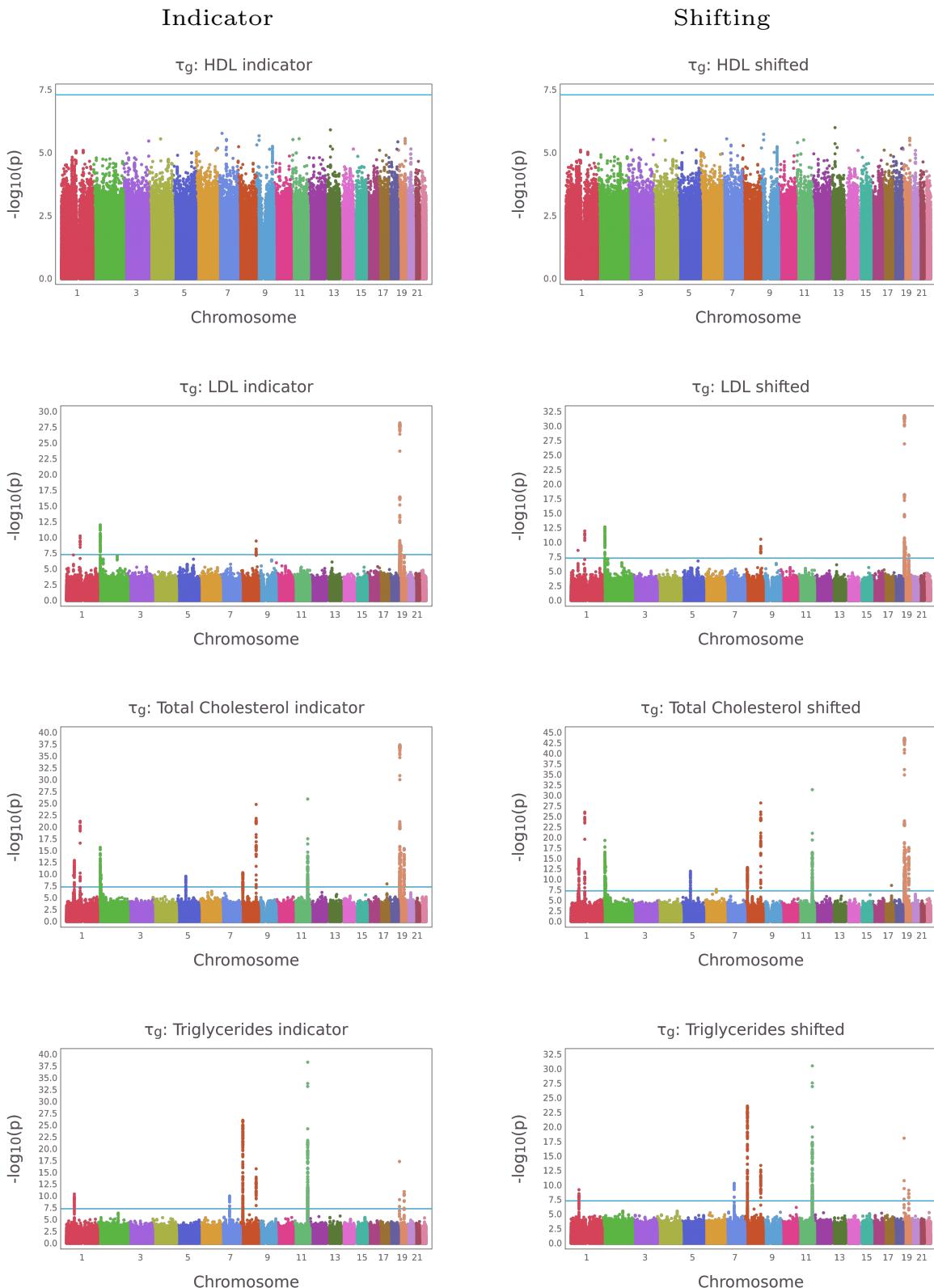


Figure S22: Comparisons of medication adjustment methods in the blood pressures TrajGWAS analysis.

We plot Manhattan plots for  $\beta_g$  in the TrajGWAS, effects to the mean for SBP, DBP, and PP, in UK Biobank study. (Left) Medication is adjusted by an additional indicator covariate to reflect on medication or not in both  $\beta_g$  and  $\tau_g$ . (Right) A sensible constant is added to the observed measures.



**Figure S23: Comparisons of the effects of medication adjustment in the lipid TrajGWAS analysis.**

We plot Manhattan plots for  $\beta_g$  in the TrajGWAS analysis, effects to the mean for HDL, LDL, total cholesterol, and triglyceride in UK Biobank study. (Left) Medication adjusted by an additional indicator covariate reflecting on and off medications in both  $\beta_g$  and  $\tau_g$ . (Right) A sensible constant is added to the observed measures.

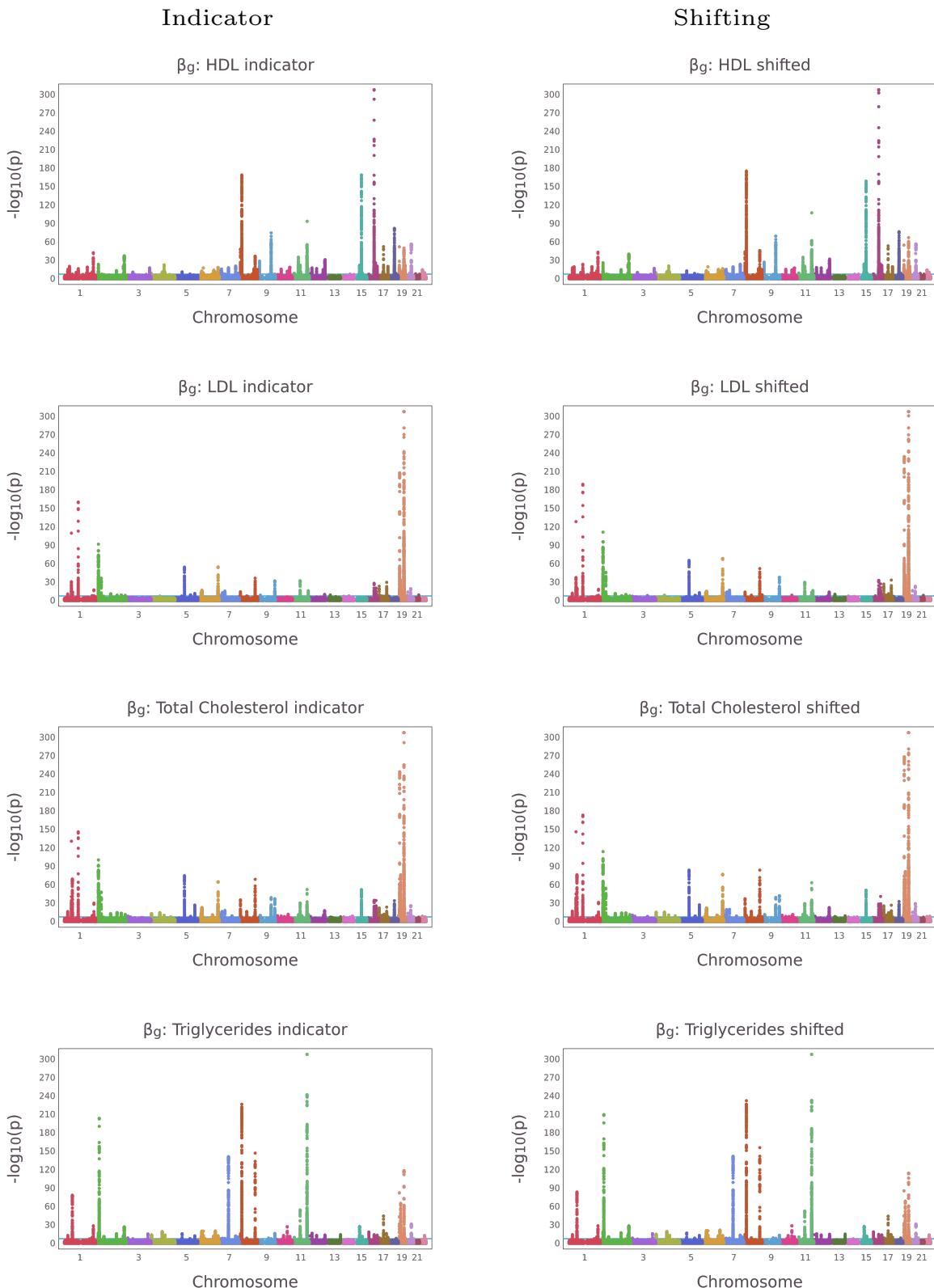


Figure S24: **Comparisons of the effects of medication adjustment in the lipid TrajGWAS analysis.**

We plot Manhattan plots for  $\beta_g$  in the TrajGWAS analysis, effects to the mean for HDL, LDL, total cholesterol, and triglyceride in UK Biobank study. (Left) Medication adjusted by an additional indicator covariate reflecting on or off medication in both  $\beta_g$  and  $\tau_g$ . (Right) A sensible constant is added to the observed measures.

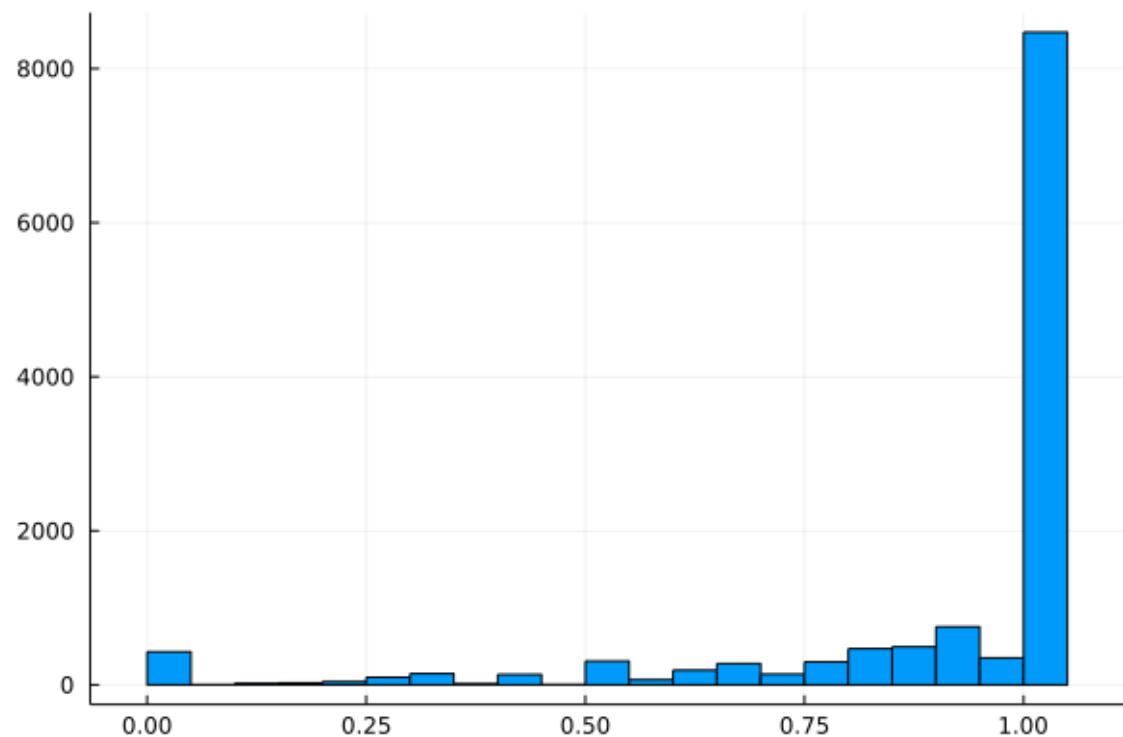
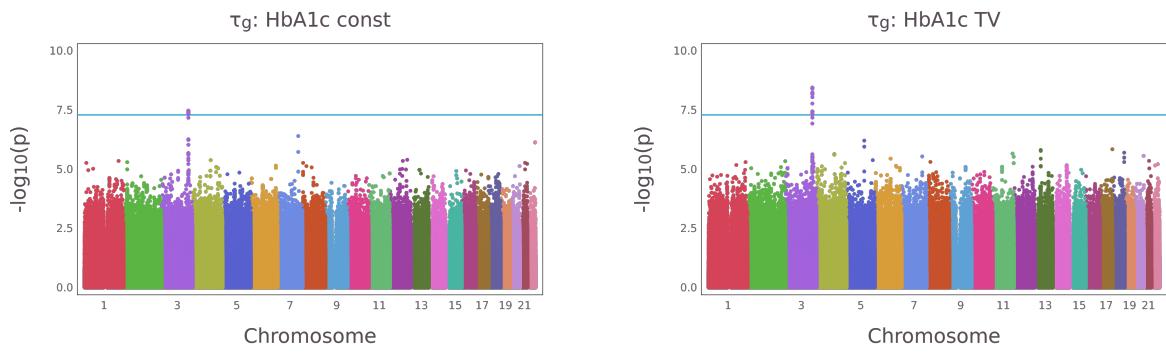


Figure S25: Histogram of the proportion of observation after the diagnosis of diabetes



**Figure S26: Manhattan plots for testing  $\tau_g$  for HbA1c with different adjustments for disease status**

Manhattan plots for testing  $\tau_g$ , the effects of the WS variability, for HbA1c. Adjustment for diabetes status is performed on both mean and WS variation component: (left) adjustment using a constant indicator, (right) adjustment using a time-varying indicator. The blue line represents the genome-wide significance level,  $5 \times 10^{-8}$ .

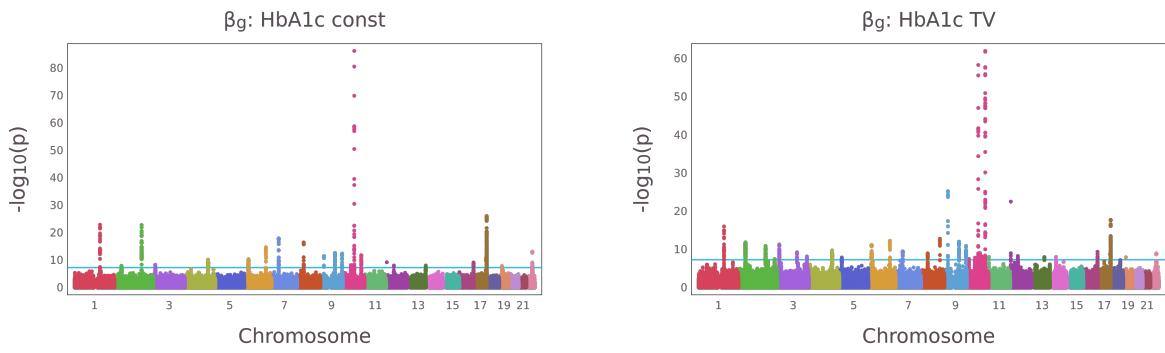


Figure S27: **Manhattan plots for testing  $\beta_g$  for HbA1c with different adjustments for disease status**

Manhattan plots for testing  $\beta_g$ , the effects of the WS variability, for HbA1c. Adjustment for diabetes status is performed on both mean and WS variation component: (left) adjustment using a constant indicator, (right) adjustment using a time-varying indicator. The blue line represents the genome-wide significance level,  $5 \times 10^{-8}$ .

**Table S1: Variables in the UK Biobank primary care clinical event records table (`clinical_table`) by source and the unified UK Biobank variable names.**

Column name	Description
<code>eid</code>	Participant identifier
<code>data_provider</code>	1 = England (Vision), 2 = Scotland, 3 = England (TPP), 4 = Wales
<code>event_dt</code>	Date clinical code was entered
<code>read_2</code>	Read v2
<code>read_3</code>	CTV3 (Read v3)
<code>value1</code>	Value recorded 1
<code>value2</code>	Value recorded 2
<code>value3</code>	Value recorded 3

**Table S2: Clinical terms used to extract biomarkers from UK Biobank primary care data and the UK Biobank field numbers being compared to in Figure S16.**

Biomarker	Units	Terminology	Terms	UK Biobank field
Systolic Blood Pressure	mmHg	Read v2/CTV3	246.., 246[012345679BCDEFGJNQSWYdg], G20.., XaF4[FLO], XaJ2[EG], XaKFx	93
Diastolic Blood Pressure	mmHg	Read v2/CTV3	246.., 246[01234567ABCDEFGPRTVXcg], G20.., XaF4[Sab], XaJ2[FH], XaKFw	94
Total Cholesterol	mmol/L	Read v2/CTV3	44P.., 44P[12349HJKZ], XE2eD, XaJe9, XSK14, XaLux, XaFs9, XaIRd	30690
HDL	mmol/L	Read v2/CTV3	XaEVr, X772M, 44d[23], 44P[5BC]	30760
LDL	mmol/L	Read v2/CTV3	XaEVs, 44d[45], 44P[6DE]	30780
Triglycerides	mmol/L	Read v2/CTV3	44e, X772O, 44Q[.12345Z], XE2q9	30870
Random Glucose	mmol/L	Read v2/CTV3	44f.., 44f0.., 44g.., 44g0.., 44TA.., XM0ly	30740
Fasting Glucose	mmol/L	Read v2/CTV3	44f1, 44g1	
HbA1c	mmol/mol	Read v2/CTV3	XaPbt, XaERp, X772q, 42W.., 42W[12345Z], 44TB.	30750
Height	m	Read v2/CTV3	229.., 229[1234Z]	50
Weight	kg	Read v2/CTV3	1622.., 22A.., 22A[1234567AZ], X76CG, XE1h4, XM01G, Xa7wI	21002
BMI	kg/m <sup>2</sup>	Read v2/CTV3	22K.., 22K[12345678], XaCDR, XaJJH, XaJqk, XaZcl	21001

“[.]” represents any one letter or number among the content provided in “[.]” can be used.

Table S3: **Genomic control factor  $\lambda$  at various p value quantiles for the UK Biobank analyses.**

Because SPA is only applied to score test statistics in the right tail, the genomic control factor calculated based on the median p values, which is not SPA adjusted, may appear inflated.

Phenotype	Parameter	Genomic control factor at $q^{th}$ p value quantile		
		$q = 0.5$ (median)	$q = 0.05$	$q = 0.01$
Systolic blood pressure	$\beta_g$	1.23	1.32	1.48
	$\tau_g$	1.09	1.07	1.07
Diastolic blood pressure	$\beta_g$	1.21	1.29	1.48
	$\tau_g$	1.06	1.04	1.03
Pulse pressure	$\beta_g$	1.25	1.34	1.50
	$\tau_g$	1.07	1.06	1.07
HDL	$\beta_g$	1.166	1.253	1.451
	$\tau_g$	1.237	1.023	1.013
LDL	$\beta_g$	1.094	1.12	1.193
	$\tau_g$	1.244	1.011	1.007
Total Cholesterol	$\beta_g$	1.131	1.203	1.386
	$\tau_g$	1.128	1.04	1.055
Triglycerides	$\beta_g$	1.165	1.229	1.419
	$\tau_g$	1.35	1.01	0.985
Random Glucose	$\beta_g$	1.08	1.08	1.10
	$\tau_g$	1.19	1.01	1.01
Fasting Glucose	$\beta_g$	1.08	1.09	1.12
	$\tau_g$	2.37	1.12	1.07
HbA1c	$\beta_g$	1.05	1.05	1.06
	$\tau_g$	1.32	0.99	0.99
BMI	$\beta_g$	1.29	1.39	1.53
	$\tau_g$	1.11	1.02	1.02

Table S4: Previously reported traits associated with novel SNPs of systolic blood pressure from NHGRI GWAS Catalog.

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
BCL2	18	1 / 0	Body mass index High light scatter reticulocyte count Modified Stumvoll Insulin Sensitivity Index (BMI interaction) Modified Stumvoll Insulin Sensitivity Index (model adjusted for BMI) Reticulocyte count Systolic blood pressure Triglyceride levels Triglycerides Type 2 diabetes Waist-hip ratio Waist-to-hip ratio adjusted for BMI Waist-to-hip ratio adjusted for BMI (adjusted for smoking behaviour) Waist-to-hip ratio adjusted for BMI x sex x age interaction (4df test) Waist-to-hip ratio adjusted for body mass index

Table S5: Previously reported traits associated with novel SNPs of diastolic blood pressure from NHGRI GWAS Catalog.

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
CLOCK	4	3 / 0	Height Red blood cell count Waist-to-hip ratio adjusted for BMI x sex interaction Waist-to-hip ratio adjusted for BMI x sex x age interaction (4df test)
PDCL2	4	1 / 0	Height
HAND2	4	1 / 0	Atrial fibrillation
HCG15	6	1 / 0	Depression (broad)
SAR1P1	6	1 / 0	Blood protein levels
XXbac-BPG308J9.3	6	4 / 0	Breast cancer Heel bone mineral density Intelligence (MTAG) Schizophrenia
OR11A1	6	1 / 0	Estimated glomerular filtration rate in diabetes
HLA-F	6	1 / 0	Psoriatic arthritis
HLA-F-AS1	6	2 / 0	Mixed cellularity Hodgkin lymphoma Psoriatic arthritis
HLA-V	6	1 / 0	Sarcoidosis (Lofgren's syndrome vs non-Lofgren's syndrome)
MCCD1P1	6	1 / 0	Non-small cell lung cancer
HCG4P5	6	1 / 0	Hemoglobin levels
HLA-A	6	2 / 0	Blood protein levels Hemoglobin levels
ZNRD1	6	1 / 0	AIDS progression Autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia (combined)
VDR	12	1 / 0	Cardiovascular disease Medication use (diuretics)
SEPT9	17	1 / 0	Cardiovascular disease Heel bone mineral density Medication use (diuretics) Systolic blood pressure

**Table S6: Previously reported traits associated with novel SNPs of pulse pressure from NHGRI GWAS Catalog.**

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
EBF2	8	2 / 0	Cardiovascular disease Medication use (agents acting on the renin-angiotensin system) Systolic blood pressure
SORCS3	10	1 / 0	Systolic blood pressure

**Table S7: Previously reported traits associated with novel SNPs of HDL from NHGRI GWAS Catalog.**

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
BCL2	18	1 / 0	Body mass index High light scatter reticulocyte count Modified Stumvoll Insulin Sensitivity Index (BMI interaction) Modified Stumvoll Insulin Sensitivity Index (model adjusted for BMI) Reticulocyte count Systolic blood pressure Triglyceride levels Triglycerides Type 2 diabetes Waist-hip ratio Waist-to-hip ratio adjusted for BMI Waist-to-hip ratio adjusted for BMI (adjusted for smoking behaviour) Waist-to-hip ratio adjusted for BMI x sex x age interaction (4df test) Waist-to-hip ratio adjusted for body mass index

Table S8: Previously reported traits associated with novel SNPs of total cholesterol from NHGRI GWAS Catalog.

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
ZNRD1	6	1 / 0	AIDS progression Autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia (combined)
TRIM31	6	1 / 0	Beta-2 microglobulin plasma levels
TRIM31-AS1	6	2 / 0	Beta-2 microglobulin plasma levels Cold sores
TRIM15	6	1 / 0	Blood protein levels
PAIP1P1	6	1 / 0	Help-seeking from a GP (without major depressive disorder symptoms) Help-seeking from a psychiatrist
TRIM26	6	5 / 0	Asthma (adult onset) Asthma (childhood onset) Autism spectrum disorder, attention deficit-hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia (combined) Help-seeking from a GP (without major depressive disorder symptoms) Help-seeking from a psychiatrist Medication use (anti-inflammatory and antirheumatic products, non-steroids) Schizophrenia
HCG17	6	3 / 0	General cognitive ability LDL cholesterol levels
LPL	8	0 / 17	Apolipoprotein A1 levels Apolipoprotein B levels Cholesterol efflux capacity Coronary artery disease Eosinophil counts HDL Cholesterol - Triglycerides (HDLC-TG) HDL cholesterol HDL cholesterol levels HDL cholesterol levels in current drinkers HDL cholesterol levels x alcohol consumption (drinkers vs non-drinkers) interaction (2df) HDL cholesterol levels x alcohol consumption (regular vs non-regular drinkers) interaction (2df) HDL cholesterol levels x long total sleep time interaction (2df test) HDL cholesterol levels x short total sleep time interaction (2df test) Height High density lipoprotein cholesterol levels Lipid metabolism phenotypes Lipid traits Lipoprotein phospholipase A2 activity in cardiovascular disease Low density lipoprotein cholesterol levels Medication use (HMG CoA reductase inhibitors) Metabolic syndrome Metabolic syndrome (bivariate traits) Metabolite levels Parental longevity (father's age at death or father's attained age) Peripheral artery disease Red cell distribution width Serum metabolite levels Serum metabolite levels (CMS) Triacylglyceride levels Triglyceride levels Triglyceride levels in current drinkers Triglyceride levels x alcohol consumption (drinkers vs non-drinkers) interaction (2df) Triglyceride levels x alcohol consumption (regular vs non-regular drinkers) interaction (2df) Triglycerides

AC100802.3

8

0 / 1

Triglycerides x physical activity interaction (2df test)  
Triglycerides-Blood Pressure (TG-BP)  
Type 2 diabetes  
Type 2 diabetes (adjusted for BMI)  
Waist-hip ratio  
Waist-to-hip ratio adjusted for BMI  
High density lipoprotein cholesterol levels

---

Table S9: Previously reported traits associated with novel SNPs of triglycerides from NHGRI GWAS Catalog.

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
NR5A2	1	1 / 0	Waist-hip ratio
CCND2	12	1 / 0	Appendicular lean mass Birth weight Body mass index Cardiovascular disease Heel bone mineral density Height Low density lipoprotein cholesterol levels Medication use (HMG CoA reductase inhibitors) Medication use (drugs used in diabetes) Offspring birth weight Pulse pressure Systolic blood pressure Triglyceride levels Type 2 diabetes Waist circumference adjusted for body mass index

**Table S10: Previously reported traits associated with novel SNPs of random glucose from NHGRI GWAS Catalog.**

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
HK1	10	6 / 0	Apolipoprotein B levels Glycated hemoglobin levels HDL cholesterol levels Hematocrit Hemoglobin Hemoglobin A1c levels Hemoglobin concentration Hemoglobin levels High light scatter reticulocyte count High light scatter reticulocyte percentage of red cells Immature fraction of reticulocytes LDL cholesterol levels Low density lipoprotein cholesterol levels Mean corpuscular hemoglobin Mean corpuscular volume Red blood cell count Red blood cell traits Red cell distribution width Reticulocyte count Reticulocyte fraction of red cells Appendicular lean mass Birth weight Body mass index Cardiovascular disease Heel bone mineral density Height Low density lipoprotein cholesterol levels Medication use (HMG CoA reductase inhibitors) Medication use (drugs used in diabetes) Offspring birth weight Pulse pressure Systolic blood pressure Triglyceride levels Type 2 diabetes Waist circumference adjusted for body mass index
CCND2	12	1 / 0	Appendicular lean mass Birth weight Body mass index Cardiovascular disease Heel bone mineral density Height Low density lipoprotein cholesterol levels Medication use (HMG CoA reductase inhibitors) Medication use (drugs used in diabetes) Offspring birth weight Pulse pressure Systolic blood pressure Triglyceride levels Type 2 diabetes Waist circumference adjusted for body mass index
CCND2-AS1	12	1 / 0	Appendicular lean mass Birth weight Body mass index Cardiovascular disease Heel bone mineral density Height Low density lipoprotein cholesterol levels Medication use (HMG CoA reductase inhibitors) Medication use (drugs used in diabetes) Offspring birth weight Pulse pressure Systolic blood pressure Triglyceride levels Type 2 diabetes Waist circumference adjusted for body mass index

Table S11: Previously reported traits associated with novel SNPs of fasting glucose from NHGRI GWAS Catalog.

Gene	Chromosome	Number of SNPs ( $\beta_g / \tau_g$ )	Trait(s)
WFS1	4	1 / 0	Type 2 diabetes
ACSL1	4	4 / 0	Type 2 diabetes
GPSM1	9	9 / 0	Birth weight Insulinogenic index Offspring birth weight Type 2 diabetes Type 2 diabetes (adjusted for BMI) Waist-to-hip ratio adjusted for BMI White blood cell count Apolipoprotein B levels Glycated hemoglobin levels HDL cholesterol levels Hematocrit Hemoglobin Hemoglobin A1c levels Hemoglobin concentration Hemoglobin levels High light scatter reticulocyte count High light scatter reticulocyte percentage of red cells Immature fraction of reticulocytes LDL cholesterol levels Low density lipoprotein cholesterol levels Mean corpuscular hemoglobin Mean corpuscular volume Red blood cell count Red blood cell traits Red cell distribution width Reticulocyte count Reticulocyte fraction of red cells Birth weight Type 2 diabetes
HK1	10	2 / 0	Appendicular lean mass Birth weight Body mass index Cardiovascular disease Heel bone mineral density Height Low density lipoprotein cholesterol levels Medication use (HMG CoA reductase inhibitors) Medication use (drugs used in diabetes) Offspring birth weight Pulse pressure Systolic blood pressure Triglyceride levels Type 2 diabetes Waist circumference adjusted for body mass index Appendicular lean mass Birth weight Body mass index Cardiovascular disease Heel bone mineral density Height Low density lipoprotein cholesterol levels Medication use (HMG CoA reductase inhibitors) Medication use (drugs used in diabetes) Offspring birth weight Pulse pressure Systolic blood pressure Triglyceride levels Type 2 diabetes Waist circumference adjusted for body mass index Balding type 1 Depressive symptoms (MTAG) Fasting blood glucose adjusted for BMI Heel bone mineral density Height Male-pattern baldness Resting heart rate Type 2 diabetes Type 2 diabetes (adjusted for BMI)
KCNQ1	11	2 / 0	
CCND2	12	1 / 0	
CCND2-AS1	12	1 / 0	
ZHX3	20	2 / 0	

## Supplementary Methods

### A WiSER Score Test

In this section we derive the score test for the WiSER model. Let  $\mathbf{X}_{i,1} \in \mathbb{R}^{n_i \times r_1}$  be the covariates to be tested for the mean component of subject  $i$ , and  $\mathbf{W}_{i,1} \in \mathbb{R}^{n_i \times r_2}$  be the covariates to be tested for the WS variability of subject  $i$ . Let  $r = r_1 + r_2$  and  $\tilde{\boldsymbol{\theta}} = (\mathbf{0}_r, \hat{\boldsymbol{\theta}}_2)^T \in \mathbb{R}^{r+p+\ell+q(q+1)/2}$ , where  $\hat{\boldsymbol{\theta}}_2 = (\boldsymbol{\beta}^T, \boldsymbol{\tau}^T, \text{vech}(\mathbf{L}_\gamma^T))^T$  is the WiSER estimator under the null model, where  $\text{vech}$  is the half-vectorization function and  $\mathbf{L}_\gamma$  is the Cholesky factor of  $\boldsymbol{\Sigma}_\gamma$ . We use  $\psi_{H_1} \in \mathbb{R}^r$  and  $\psi_{H_0} \in \mathbb{R}^{p+\ell+q(q+1)/2}$  to represent the gradient vectors derived under the full model and the null model respectively. Then the chi-square score test statistic for testing  $\boldsymbol{\theta}_1 = \mathbf{0}_r$  takes the form

$$S = \frac{1}{m} \left[ \psi_{H_1}(\tilde{\boldsymbol{\theta}}) \right]^T \mathbf{V}_{\psi_{H_1}(\tilde{\boldsymbol{\theta}})}^{-1} \left[ \sum_i \psi_{H_1}(\tilde{\boldsymbol{\theta}}) \right],$$

where  $\mathbf{V}_{\psi_{H_1}(\tilde{\boldsymbol{\theta}})}$  is the covariance matrix of score  $\psi_{H_1}(\tilde{\boldsymbol{\theta}})$ . The score test statistic  $S$  is asymptotically distributed as  $\chi_r^2$  under  $H_0$ . The score is given by  $\psi_{H_1}(\tilde{\boldsymbol{\theta}}) = \sum_{i=1}^n \psi_{iH_1}(\tilde{\boldsymbol{\theta}})$ , where  $\psi_{iH_1}(\tilde{\boldsymbol{\theta}})$  is estimated by

$$\hat{\psi}_{iH_1}(\tilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{X}_{i,1}^T [\mathbf{V}_i^{(0)}]^{-1} \hat{\mathbf{r}}_i \\ -\mathbf{W}_{i,1}^T \text{diag} \left[ \begin{pmatrix} e^{\mathbf{w}_{i1}^T \hat{\boldsymbol{\tau}}} & & \mathbf{X}_{i,1}^T [\mathbf{V}_i^{(0)}]^{-1} \hat{\mathbf{r}}_i \\ & \ddots & \\ & & e^{\mathbf{w}_{in_i}^T \hat{\boldsymbol{\tau}}} \end{pmatrix} \left( \mathbf{V}_i^{(0)} \right)^{-1} \hat{\mathbf{R}}_i \left( \mathbf{V}_i^{(0)} \right)^{-1} \right] \end{pmatrix} \in \mathbb{R}^r,$$

where  $\hat{\mathbf{r}}_i$ ,  $\hat{\mathbf{R}}_i$ , and  $\hat{\boldsymbol{\tau}}$  are from the fitted null model. For genotypes, which is assumed to be constant through time, it reduces to a scalar  $S_{\beta_g}$  for testing  $\beta_g = 0$  ( $r_1 = 1$  and  $r_2 = 0$ ), and  $S_{\tau_g}$  for testing  $\tau_g = 0$  ( $r_1 = 0$  and  $r_2 = 1$ ).

To calculate  $\hat{\mathbf{V}}_{\psi_{H_1}}(\tilde{\boldsymbol{\theta}})$ , we note that from Boos and Stefanski Chapter 7,<sup>1</sup>  $\mathbf{V}_{\psi_{H_1}}(\tilde{\boldsymbol{\theta}})$  can be represented by

$$\mathbf{V}_{\psi_{H_1}}(\tilde{\boldsymbol{\theta}}) = \mathbf{B}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{B}_{21} - \mathbf{B}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{B}_{22}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \in \mathbb{R}^{r \times r},$$

where

$$\mathbf{A}(\tilde{\boldsymbol{\theta}}) = \mathbb{E} \left[ -\frac{\partial \psi}{\partial \boldsymbol{\theta}} \right] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

and

$$\mathbf{B}(\tilde{\boldsymbol{\theta}}) = \mathbb{E} \left[ \psi \psi^T \right] = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}.$$

Here  $\psi = \begin{bmatrix} \psi_{H_1} \\ \psi_{H_0} \end{bmatrix}$ ,  $\mathbf{A}_{11}$ ,  $\mathbf{B}_{11} \in \mathbb{R}^{r \times r}$ ,  $\mathbf{A}_{12}$ ,  $\mathbf{B}_{12} \in \mathbb{R}^{r \times [p+\ell+q(q+1)/2]}$ , and  $\mathbf{A}_{22}$ ,  $\mathbf{B}_{22} \in \mathbb{R}^{[p+\ell+q(q+1)/2] \times [p+\ell+q(q+1)/2]}$ .  $\hat{\mathbf{A}}_{22}^{-1}$  and  $\hat{\mathbf{A}}_{22}^{-1}\hat{\mathbf{B}}_{22}\hat{\mathbf{A}}_{22}^{-1}$  are readily available from the estimation and inference of the null model.

$\mathbf{A}_{22}$  is estimated from the Fisher Information Matrix of the null model and  $\mathbf{B}_{22}$  is estimated from the null model as

$$\hat{\mathbf{B}}_{22} = \frac{1}{m} \sum_{i=1}^m \hat{\psi}_{iH_0}(\hat{\boldsymbol{\theta}}_2) \hat{\psi}_{iH_0}^T(\hat{\boldsymbol{\theta}}_2) \in \mathbb{R}^{[p+\ell+q(q+1)/2] \times [p+\ell+q(q+1)/2]},$$

where  $\psi_{iH_0}(\hat{\boldsymbol{\theta}}_2)$  is the score vector under the null model for sample  $i$ . We estimate  $\mathbf{B}_{11}$  by

$$\hat{\mathbf{B}}_{11} = \frac{1}{m} \sum_{i=1}^m \hat{\psi}_{iH_1}(\tilde{\boldsymbol{\theta}}) \hat{\psi}_{iH_1}^T(\tilde{\boldsymbol{\theta}}) \in \mathbb{R}^{r \times r}$$

and  $\mathbf{B}_{12}$  by

$$\hat{\mathbf{B}}_{12} = \frac{1}{m} \sum_{i=1}^m \hat{\psi}_{iH_1}(\tilde{\boldsymbol{\theta}}) \hat{\psi}_{iH_0}^T(\tilde{\boldsymbol{\theta}}) \in \mathbb{R}^{r \times [p+\ell+q(q+1)/2]}.$$

Let

$$\widehat{\mathbf{A}}_{i,21} = \begin{pmatrix} \mathbf{X}_i^T [\mathbf{V}_i^{(0)}]^{-1} \mathbf{X}_{i,1} & & \\ & \mathbf{W}_i^T \begin{pmatrix} e^{w_{i1}^T \hat{\boldsymbol{\tau}}} & & \\ & \ddots & \\ & & e^{w_{in_i}^T \hat{\boldsymbol{\tau}}} \end{pmatrix} (\mathbf{V}_i^{(0)})^{-1} \widehat{\mathbf{R}}_i (\mathbf{V}_i^{(0)})^{-1} \begin{pmatrix} e^{w_{i1}^T \hat{\boldsymbol{\tau}}} & & \\ & \ddots & \\ & & e^{w_{in_i}^T \hat{\boldsymbol{\tau}}} \end{pmatrix} \mathbf{W}_{i,1} \\ \mathbf{O}_{\ell \times r_1} & 2\mathbf{C}_q^T \cdot \left( \boldsymbol{\ell}_\gamma^T \mathbf{Z}_i^T (\mathbf{V}_i^{(0)})^{-1} \odot \mathbf{Z}_i^T (\mathbf{V}_i^{(0)})^{-1} \right) \cdot \begin{pmatrix} e^{w_{i1}^T \hat{\boldsymbol{\tau}}} & & \\ & \ddots & \\ & & e^{w_{in_i}^T \hat{\boldsymbol{\tau}}} \end{pmatrix} \cdot \mathbf{W}_{i,1} \\ \mathbf{O}_{[q(q+1)/2] \times r_1} & \end{pmatrix} \in \mathbb{R}^{[p+\ell+q(q+1)/2] \times r},$$

where  $\odot$  represents the Khatri-Rao (column-wise Kronecker) product, and  $\mathbf{C}_q \in \mathbb{R}^{q^2 \times q(q+1)/2}$  is the copy matrix such that  $\mathbf{C}_q \text{vec}(\mathbf{M}) = \text{vec}(\mathbf{M})$ , the vectorization of  $\mathbf{M}$ . The estimates  $\hat{\boldsymbol{\tau}}$  and  $\widehat{\mathbf{R}}_i$  are from the fitted null model. We estimate  $\mathbf{A}_{21} = \mathbf{A}_{12}^T$  by

$$\widehat{\mathbf{A}}_{21} = \frac{1}{m} \sum_{i=1}^m \widehat{\mathbf{A}}_{i,21}.$$

## B Derivation of saddlepoint approximation (SPA)

The score test statistic for testing  $\beta_g$  is

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

where  $\mathbf{g}$  is the normalized genotype vector. The score test statistic for testing  $\tau_g$  is

$$S_{\tau_g} = - \sum_{i=1}^m g_i \left[ \mathbf{1}_{n_i}^T \text{diag} \left[ \begin{pmatrix} e^{w_{i1}^T \hat{\boldsymbol{\tau}}} & & \\ & \ddots & \\ & & e^{w_{in_i}^T \hat{\boldsymbol{\tau}}} \end{pmatrix} (\mathbf{V}_i^{(0)})^{-1} \widehat{\mathbf{R}}_i (\mathbf{V}_i^{(0)})^{-1} \right] \right] =: \mathbf{g}^T \mathbf{c}_{\tau_g}. \quad (2)$$

First, we construct the empirical cumulant generating function (CGF) of  $\mathbf{c} \in \{\mathbf{c}_{\tau_g}, \mathbf{c}_{\beta_g}\}$ . The empirical moment generating function (MGF) of  $\mathbf{c}$  is

$$\widehat{M}_0(z) := \frac{1}{m} \sum_{i=1}^m \exp(c_i z)$$

with the first two derivatives

$$\widehat{M}'_0(z) = \frac{1}{m} \sum_{i=1}^m c_i \exp(c_i z), \quad \widehat{M}''_0(z) = \frac{1}{m} \sum_{i=1}^m c_i^2 \exp(c_i z).$$

The empirical CGF of  $\mathbf{c}$  is defined as the logarithm of empirical MGF,  $\widehat{K}_0(z) = \log \widehat{M}_0(z)$ . Its first two derivatives are

$$\widehat{K}'_0(z) = \frac{\widehat{M}'_0(z)}{\widehat{M}_0(z)}, \quad \widehat{K}''_0(z) = \frac{\widehat{M}''_0(z) \widehat{M}_0(z) - [\widehat{M}'_0(z)]^2}{[\widehat{M}_0(z)]^2}.$$

Then, the empirical CGF and the first two derivatives of the observed score  $S = \sum_{i=1}^m g_i c_i$  are

$$\widehat{K}(z) = \sum_{i=1}^m \widehat{K}_0(g_i z), \quad \widehat{K}'(z) = \sum_{i=1}^m g_i \widehat{K}'_0(g_i z), \quad \text{and } \widehat{K}''(z) = \sum_{i=1}^m g_i^2 \widehat{K}''_0(g_i z).$$

To apply the saddlepoint approximation for an observed score  $S = s$ , we first find a “saddlepoint”  $\zeta$  such that  $\hat{K}'(\zeta) = s$  and retrieve

$$\begin{aligned}\omega &= \text{sign}(\zeta)\sqrt{2(\zeta s - \hat{K}(\zeta))}, \\ \nu &= \zeta\sqrt{\hat{K}''(\zeta)}.\end{aligned}$$

The cumulative distribution function of  $S$  at  $s$  is approximated by

$$P(S < s) \approx \Phi\left(\omega + \frac{1}{\omega} \log\left(\frac{\nu}{\omega}\right)\right),$$

where  $\Phi(\cdot)$  denotes the cumulative distribution function of the standard normal distribution<sup>2</sup>. The p value for the score test with the statistic  $S = s$  is given by  $P(S < -|s|) + (1 - P(S < |s|))$ .

## C Heuristic method inflates the type I error

Failure to properly control for time-varying covariates that are correlated with genotypes can lead to biased results. We conduct a small scale simulation in order to demonstrate this. The heuristic approach involves regressing per-subject residual standard deviations on a set of covariates. We simulate data from Model (1) and Model (2). We set  $\beta_{\text{true}} = (0.1, 6.5, 0.0, 1.0, 5.0)^T$ ,  $\tau_{\text{true}} = (0.0, 0.3, 0.0, 0.5, 0.25)^T$ , and the covariance of  $(\gamma_i, \omega_i)$  to be

$$\Sigma_{\gamma\omega} = \begin{pmatrix} 2.0 & 0.0 & 0.2 \\ 0.0 & 1.2 & 0.1 \\ 0.2 & 0.1 & 1.0 \end{pmatrix}.$$

For each subject, the mean level covariates  $\mathbf{X}$  are the same as the WS variability covariates  $\mathbf{W}$ .  $W_1$  is the intercept.  $W_2$  and  $W_3$  are time-invariant covariates with effect sizes  $\tau_2 = 0.3$  and  $\tau_3 = 0.0$  respectively.  $W_2$  acts as sex and is drawn from Bernoulli(0.5) per subject.  $W_3$  is the simulated genotype with MAF = 0.3 following Hardy-Weinberg equilibrium.  $W_4$  is a time-varying covariate with effect size  $\tau_4 = 0.5$ ; it is correlated with  $W_3$  (the genotype) with its entries generated from  $N(0, 0.5) \cdot I\{W_3 < 0\} + N(0, 2) \cdot I\{W_3 > 0\}$ .  $W_5$  is time-varying with entries generated from independent standard normals. Since the heuristic method involves one outcome per subject, time-varying covariates cannot be incorporated as-is, so we use their per-subject mean to control for them.

Figure S1 displays the  $-\log_{10}(p \text{ value})$  of  $\tau_3$  (the genotype with no effect on WS variability) based on 100 replicates per scenario with a sample size of 6,000. The number of repeated measurements per subject range from 5 to 20. The heuristic method leads to significantly inflated type I error. Inadequately controlling for time-varying covariates results in false positives when using standard deviation of the residuals as the outcome.

## D Clinical measurement extraction from the UK Biobank primary data

We extract HbA1c data using the HbA1c code list by Denaxas et al.<sup>3</sup>, omitting the terms: 42c and 44TC (tests of HbA1), XE24t (no occurrences), and X80U3 (no values accompanying term). We convert between DCCT (Diabetes Control and Complications Trial) align and IFCC (International Federation of Clinical Chemistry) standardized measurements, and keep values between 4-18% (20-173 mmol/mol). The distribution of resulting values is then compared against that of HbA1c available from UK Biobank field 30750.

To extract blood pressure records, we require values to be specified as systolic or diastolic in pairs. We look for codes relating to blood pressure or primary hypertension in “read v2” and “CTV3 (read v3)” dictionaries, and exclude codes related to ambulatory care or hypertension secondary to a transient cause. The terms included are 246.., 2461 - 2467, 2469, 246[A-G], 246J, 246N, 246[P-S], 246[V-Y], 246c, 246d, 246g, XaF4F, XaF4L, XaF4O, XaF4S,

XaF4a, XaF4b, XaJ2[E-H], XaKFw, XaKFX, and G20.. All units are assumed to be mmHg. Blood pressure readings are assigned as diastolic or systolic according to the attached term, where possible. Otherwise, where two unique values are given for an individual and date, the higher is assumed to be the systolic blood pressure, while the lower is assumed to be the diastolic blood pressure. For all pairs of blood pressure readings, systolic blood pressure is required to lie between 45 and 300, diastolic blood pressure is required to be greater than 30 mmHg but less than the associated systolic blood pressure. The distribution of resulting values is compared against that of systolic and diastolic blood pressure from UK Biobank field numbers 93 and 94, respectively.

HDL cholesterol is extracted from primary care using the code list prepared by Denaxas et al.<sup>3</sup>, with the addition of terms XaEVr (Plasma HDL cholesterol level), 44d2 (Plasma random HDL cholesterol level), 44d3 (Plasma fasting HDL cholesterol level), and X772M (High density lipoprotein cholesterol level). Units are assumed to be mmol/L, and the records (n=819) that explicitly indicate a different unit are excluded. Values are required to be less than 10. The distribution of resulting values is compared against that of HDL cholesterol values from UK Biobank field number 30760.

LDL cholesterol is extracted from primary care using terms 44P6, 44PD, 44PE, 44d4, 44d5, and XaEVs. Units are assumed to be mmol/L, and the records (n=21) that explicitly indicate a different unit are excluded. Values are required to be less than 30. The distribution of resulting values is compared against that of LDL cholesterol values from UK Biobank field number 30780.

Total cholesterol is extracted from primary care using the code list prepared by Denaxas et al.<sup>3</sup>, with the addition of terms XaFs9 (Fasting cholesterol level) and XaIRd (Plasma total cholesterol level). Units are assumed to be mmol/L. Records (n=45) that explicitly indicate a different unit are excluded. Values are required to be less than 30. The distribution of resulting values is compared against that of total cholesterol values from UK Biobank field number 30690. Triglyceride records are extracted from primary care using the code list prepared by Denaxas et al.<sup>3</sup>, with the addition of the code X772O (Triglyceride level) and the code prefix 44e (Plasma triglyceride level). Units are assumed to be mmol/L, and the records (n=249) that explicitly indicate a different unit are excluded. Values are required to be less than 30. The distribution of resulting values is compared against that of triglyceride levels from UK Biobank field number 30870.

Serum and plasma random glucose levels are extracted from primary care data using codes 44TA, 44f., 44f0, 44g., 44g0, and XM0ly. Serum and plasma fasting glucose levels are extracted using codes 44f1 and 44g1. Units are assumed to be mmol/L, and records (n=39) that explicitly indicate a different unit are excluded. Values are required to be less than 60. The distribution of random glucose is compared against that of glucose from UK Biobank field number 30740.

BMI is extracted from primary care using codes XaCDR, XaJJH, XaJqk, XaZcl, and prefix 22K. Values are required to lie between 12 and 75 kg/m<sup>2</sup>. Height records are extracted using code prefix 229, and values are required to lie between 125 and 210 cm. Weight records are extracted using codes X76CG, XM01G, XE1h4, Xa7wI, 1622, and prefix 22A. Values are required to lie between 30 and 200 kg. Height, weight, and BMI measures are then matched by individual and date. Missing values for height are filled in from previous or subsequent measurements, where possible. BMI is calculated from height and weight at each date. If a value for BMI is already recorded for that individual and date, and the reported BMI differs from the calculated BMI by more than 1.5 (0.27% of records), both measures are excluded, otherwise, the calculated BMI is retained. If a BMI record is reported for a date where height or weight is not available, the reported BMI is retained.

## E Covariate adjustment for $\beta_g$ and $\tau_g$ in UK Biobank TrajGWAS analyses

- Blood pressures, i.e., SBP, DBP, and PP (if on medication, add 15 mmHg for SBP, 10 mmHg for DBP, and 5 mmHg for PP<sup>4</sup>)
  - $\beta_g$ : age, age<sup>2</sup>, sex, age×sex, 10 PCs;

- $\tau_g$ : age, age<sup>2</sup>, sex, age×sex
- HDL (if on medication, -0.060 mmol/L<sup>5</sup>)
  - $\beta_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, 10 PCs
  - $\tau_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex
- LDL (if on medication, add 1.290 mmol/L<sup>5</sup>)
  - $\beta_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, 10 PCs
  - $\tau_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex
- Total cholesterol (if on medication, add 1.347 mmol/L<sup>5</sup>)
  - $\beta_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, 10 PCs
  - $\tau_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex
- Triglycerides (if on medication, add 0.208 mmol/L<sup>5</sup>)
  - $\beta_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, cholesrol\_drug, 10 PCs
  - $\tau_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, cholesrol\_drug
- Glucose, i.e., fasting glucose and random glucose
  - $\beta_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, self\_insulin, diabetes\_status, 10 PCs
  - $\tau_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, diabetes\_status
- HbA1c
  - $\beta_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, diabetes\_status, self\_insulin, 10 PCs
  - $\tau_g$ : age, age<sup>2</sup>, BMI, sex, age×BMI, age×sex, diabetes\_status
- BMI
  - $\beta_g$ : age, age<sup>2</sup>, sex, age×sex, 10 PCs
  - $\tau_g$ : age, age<sup>2</sup>, sex, age×sex

## F Considerations of Diabetes Diagnosis in TrajGWAS Analysis

One of the advantages of **TrajGWAS** is that it can incorporate time-varying covariates, e.g., diseases developed after the first observation. To fully explore how the occurrence of the related disease influences the genetic contribution to WS variability, we now include disease status as a time-varying covariates. We extracted first occurrence of diabetes from their primary care, hospital records, and registry data and include it as a time-varying covariate when analyzing HbA1c, given that HbA1c has abundant sample size as well as the larger number of repeated measurements than glucose measures. We compare the results with the those when including disease status as a time-fixed covariate.

For time-varying indicator, we use 0 for measurements before first diagnosis of diabetes, and 1 for measurements after the diagnosis. Percentage of observations after the diagnosis averaged over subjects is 95.8%. 66.6% of the subjects had all the observations after the diagnosis, and 85.1% had at least 75% of the observations after the diagnosis. Histogram of proportion of observation after diagnosis is shown in Figure S25. The results with time-varying indicator have magnified some of the signals compared to constant indicator, as shown in Figures S26-S27. For  $\beta_g$ , the SNPs magnified in Chromosome 2 maps to *THADA* (OMIM: 611800)<sup>6,7</sup>, the one on Chromosome 8 to *SLC30A8* (OMIM: 611145)<sup>8,9</sup>, and the last peak of Chromosome 10 to *TCF7L2* (OMIM: 602228), all of which are reported to be highly correlated to type 2 diabetes, glycemic changes, and beta-cell function.

## Supplemental References

1. Boos, D. D. and Stefanski, L. A. (2013). *Essential Statistical Inference*. Springer Texts in Statistics, Springer, New York. Theory and methods.
2. Barndorff-Nielsen, O. E. (1990). Approximate interval probabilities. *Journal of the Royal Statistical Society: Series B (Methodological)* *52*, 485–496.
3. 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.
4. 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.
5. Wu, J., Province, M. A., Coon, H., Hunt, S. C., Eckfeldt, J. H., Arnett, D. K., Heiss, G., Lewis, C. E., Ellison, R. C., Rao, D. C. *et al.* (2007). An investigation of the effects of lipid-lowering medications: genome-wide linkage analysis of lipids in the hypergen study. *BMC Genetics* *8*, 1–9.
6. Simonis-Bik, A. M., Nijpels, G., van Haeften, T. W., Houwing-Duistermaat, J. J., Boomsma, D. I., Reiling, E., van Hove, E. C., Diamant, M., Kramer, M. H., Heine, R. J. *et al.* (2010). Gene variants in the novel type 2 diabetes loci cdc123/camk1d, thada, adamts9, bcl11a, and mtnr1b affect different aspects of pancreatic  $\beta$ -cell function. *Diabetes* *59*, 293–301.
7. Thomsen, S. K., Ceroni, A., van de Bunt, M., Burrows, C., Barrett, A., Scharfmann, R., Ebner, D., McCarthy, M. I., and Gloyn, A. L. (2016). Systematic functional characterization of candidate causal genes for type 2 diabetes risk variants. *Diabetes* *65*, 3805–3811.
8. Flannick, J., Thorleifsson, G., Beer, N. L., Jacobs, S. B., Grarup, N., Burtt, N. P., Mahajan, A., Fuchsberger, C., Atzmon, G., Benediktsson, R. *et al.* (2014). Loss-of-function mutations in slc30a8 protect against type 2 diabetes. *Nature Genetics* *46*, 357–363.
9. Wang, T., Liu, H., Wang, L., Huang, T., Li, W., Zheng, Y., Heianza, Y., Sun, D., Leng, J., Zhang, S. *et al.* (2016). Zinc-associated variant in slc30a8 gene interacts with gestational weight gain on postpartum glycemic changes: a longitudinal study in women with prior gestational diabetes mellitus. *Diabetes* *65*, 3786–3793.