Supplementary Information for

SPAGRM: Effectively controlling for sample relatedness in large-scale genome-wide association studies of longitudinal traits

Supplementary Note

1. Generalized estimation equations (GEE) for longitudinal traits

Compared to linear mixed effect models (LMM), generalized estimation equations (GEE) are another popular method for modeling longitudinal data[1]. A commonly used GEE model can be written as

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + g_i\beta_g + \varepsilon_{ij},$$

where y_{ij} is the j-th observation for i-th subject, g_i is the genotype of a single variant, $\boldsymbol{x}_{ij}$ $(p \times 1)$ covariates with fixed coefficient $\boldsymbol{\beta}$. Error term ε_{ij} follows a multivariate normal distribution with a mean vector of zeros and a variance-covariance matrix of a user-specified correlation matrix $\boldsymbol{V}_i$, including:

- Exchangeable correlation structure

$$\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}_{m_i \times m_i}$$

- Autoregressive correlation structure (order 1)

$$\begin{pmatrix} 1 & \rho & \cdots & \rho^{m_i-1} \\ \rho & 1 & \cdots & \rho^{m_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{m_i-1} & \rho^{m_i-2} & \cdots & 1 \end{pmatrix}_{m_i \times m_i}$$

where m_i is the number of observations for subject i, ρ is the parameter to be estimated for exchange and auto-regressive correlation structures[2-4]. We only compared these two most popular correlation structures in our analyses. Other correlation structures were not compared for specific reasons. For example, independent assumption may not hold in real data, which would perform poorly when the within-cluster correlation is large[2,4,5]; no available tools to fit the GEE with *m*-dependent correlation structure; it would take too much computation and memory usage to fit the GEE with unstructured correlation structure in real data analyses; and no prior knowledge to choose a fixed correlation structure.

Since SPA$_{GRM}$ method employs a retrospective framework, in which incorporating the random term to control for sample relatedness is optional, rather than required, when fitting null model. We fitted traditional GEE models without considering the random effect, and obtained model residuals under the null hypothesis $H_0: \beta_g = 0$. Like TrajGWAS analyses, we defined the score statistics for testing mean trajectories as

$$S_{\beta_g} = \sum_{i=1}^{n} \left[1_{m_i}^T (\boldsymbol{V}_i)^{-1} \hat{\boldsymbol{r}}_i\right] g_i = R_{\beta_g}^T G$$

Where $\hat{\boldsymbol{r}}_i = \boldsymbol{y}_i - \boldsymbol{X}_i \hat{\beta}$. We defined R_{β_g} as the model residuals and rescaled them to ensure sum of model residuals equal to zero before being input into the SPA$_{GRM}$ pipeline.

2. Details about the moment generating function $\boldsymbol{M_{S_i}(t)}$

Suppose that family i includes two related subjects and the score statistics $S_i = R_1 G_1 + R_2 G_2$. We let $G_1 = G_1^{(1)} + G_1^{(2)}$ and $G_2 = G_2^{(1)} + G_2^{(2)}$ where $G_j^{(k)} = 0,1, 1 \le j \le 2, 1 \le k \le 2$ is to encode one allele of the locus. The marginal distribution of $G_j^{(k)}$ is a Bernoulli distribution with a parameter of μ.

(0) Suppose that random variables $G_j^{(k)}$ are independent distributed to each other, then random variables G_1 and G_2 are independent and MGF is

$$M_{S_i}^{(0)}(t) = [1 - \mu + \mu e^{tR_1}]^2 \cdot [1 - \mu + \mu e^{tR_2}]^2$$

(1) Suppose that random variables $G_1^{(1)} = G_2^{(1)}$, and $G_1^{(1)}, G_1^{(2)}, G_2^{(2)}$ are identical and independent distributed, then they share one allele from the same ancestry at this locus and the MGF of S_i is

$$M_{S_i}^{(1)}(t) = \left[1 - \mu + \mu e^{t(R_1+R_2)}\right] \cdot [1 - \mu + \mu e^{tR_1}] \cdot [1 - \mu + \mu e^{tR_2}]$$

An intuitive example of this case is a pair of one parent and one offspring.

(2) Suppose that random variables $G_1^{(1)} = G_2^{(1)}$ and $G_1^{(2)} = G_2^{(2)}$, and $G_1^{(1)}, G_1^{(2)}$ are identical and independent distributed, then the two subjects share two alleles from the same ancestry at this locus and the MGF of S_i is

$$M_{S_i}^{(2)}(t) = \left[1 - \mu + \mu e^{t(R_1+R_2)}\right]^2$$

Under the assumption of no inbreeding, the estimated MGF of S_i is

$$M_{S_i}(t) = \delta^{(0)} \cdot M_{S_i}^{(0)}(t) + \delta^{(1)} \cdot M_{S_i}^{(1)}(t) + \delta^{(2)} \cdot M_{S_i}^{(2)}(t),$$

where non-negative values $\delta^{(0)}$, $\delta^{(1)}$, and $\delta^{(2)}$ are the IBD-sharing probabilities that two related individuals share $0,1,2$ alleles identity by descent at a locus, corresponding to the above three cases (0), (1), and (2), respectively.

Suppose that family i includes $n_i > 2$ individuals, Chow-Liu algorithm can be used to estimate MGF of $S_i = \sum_{j=1}^{n_i} R_{i_j} G_{i_j}$. Chow and Liu constructed approximation P' of a minimum difference in information[6]. The Kullback-Leibler divergence (also called relative entropy) between P and P' is derived as

$$D(P||P') = -\sum I(G_{i_{j_1}}, G_{i_{j_2}}) + \sum_{j=1}^{n_i} H(G_{i_j}) - H(\boldsymbol{G}_i)$$

where $I(G_{i_{j_1}}, G_{i_{j_2}}), 1 \le j_1 < j_2 \le n_i$ is the mutual information between $G_{i_{j_1}}$ and $G_{i_{j_2}}$, which is defined as

$$I(G_{i_{j_1}}, G_{i_{j_2}}) = \sum_{g_{i_{j_1}}, g_{i_{j_2}}} P(g_{i_{j_1}}, g_{i_{j_2}}) log\left(\frac{P(g_{i_{j_1}}, g_{i_{j_2}})}{P(g_{i_{j_1}})P(g_{i_{j_2}})}\right),$$

where $g_{i_{j_1}}, g_{i_{j_2}} = 0,1,2$. $I(G_{i_{j_1}}, G_{i_{j_2}})$ can be calculated if we know the IBD-sharing probabilities between subject i_{j_1} and i_{j_2}. And $H(G_{i_j})$ represents the information entropy of $G_{i_j}, 1 \le j \le n_i$, $H(\boldsymbol{G}_i)$ is the joint entropy of $\boldsymbol{G}_i = (G_1, G_2, \dots, G_{n_i})$. Since $\sum H(G_{i_j})$ and $H(\boldsymbol{G}_i)$ are independent of the $D(P||P')$, $\sum I(G_{i_{j_1}}, G_{i_{j_2}})$ determines the kullback-Leibler divergence between P and P'. So minimizing the kullback-Leibler divergence is equivalent to constructing a maximum-weight tree (MST) with the total branch weight $\sum I(G_{i_{j_1}}, G_{i_{j_2}})$. Simple algorithms like Prim's algorithm[7] can

easily construct a MST of $n_i - 1$ branches according to $n_i(n_i - 1)/2$ pairwise mutual information measures $I(G_{i_{j_1}}, G_{i_{j_2}}), 1 \le j_1 < j_2 \le n_i$. Once the MST is determined, all second-order terms $P(G_{i_{m_k}}|G_{i_{m_{h(k)}}})$ are also uniquely selected, so as

$$\hat{P}'(\boldsymbol{G}_i) \approx \prod\nolimits_{k=1}^{n_i} P(G_{i_{m_k}}|G_{i_{m_{h(k)}}}), \qquad 0 \le h(k) < k,$$

where $m_k, 1 \le k \le n_i$ are rearrangement of integers ranging from 1 to n_i, and the mapping $m_{h(k)}$ with $0 \le h(k) < k$ is called the dependence tree of the distribution. Define that $P(G_{i_{m_k}}|G_{i_0}) = P(G_{i_{m_k}})$.

After obtaining $\hat{P}'(\boldsymbol{G}_i)$, the MGF estimation of $S_i = \sum_{j=1}^{n_i} R_{i_j} G_{i_j}, n_i \ge 3$ is

$$M_{S_i}(t) \approx \sum\nolimits_{g_1, g_2, \ldots, g_{n_i}} \sum \widehat{\Pr}(G_{i_1} = g_1, \ldots, G_{i_{n_i}} = g_{n_i}) \cdot e^{t(R_{i_1} g_1 + \cdots + R_{i_{n_i}} g_{n_i})}$$

3. SPA$_{\text{GRM}}$ is valid to analyze study cohorts with a high proportion of relatedness

To mimic a real study cohort with a high proportion of relatedness, we excluded subjects unrelated to any other subjects and created a pseudo-cohort to conduct longitudinal GWAS for BMI. Among the 175,443 UKB white British participants that have longitudinal BMI measurements, 54,974 individuals were retained after excluding unrelated individuals to maximize the presence of cryptic relatedness (Supplementary Fig. 6c). We applied SPA$_{\text{GRM}}$, SPA$_{\text{GRM(INT)}}$, and SPA$_{\text{GRM(CCT)}}$ to approximately 23 million imputed variants with an imputation INFO score ≥ 0.6, MAF $\ge 2 \times 10^{-4}$ and Hardy-Weinberg equilibrium (HWE) p value $>1 \times 10^{-6}$. Quantile-quantile (QQ) plots for SPA$_{\text{GRM}}$-based methods didn't show evidence of inflation, and the genomic control inflation factors $\lambda = 0.92 - 0.96$ for these methods (Supplementary Fig. 9). Additionally, we applied SPA$_{\text{GRM}}$ with various maximum family size settings to this analysis, including 3, 5 (the default setting in SPA$_{\text{GRM}}$), 7, and 10. Supplementary Fig. 10 displayed that p values of SPA$_{\text{GRM}}$ using different maximum family size settings were almost the same on the -log10 scale, indicating that pedigree cuts had minimal impact on SPA$_{\text{GRM}}$'s performance, while rum time scaled linearly with the maximum family size.

4. Resampling method to demonstrate that SPA$_{GRM}$ is more powerful than TrajGWAS if the model residuals are highly unbalanced

In the simulation studies, we found that SPA$_{GRM}$ was more powerful than TrajGWAS when testing WS variability H_0: $\tau_g = 0$, even if all subjects were unrelated. This is not trivial as the sample sizes in SPA$_{GRM}$ and TrajGWAS analyses were the same. In this section, we used a resampling method to construct the empirical cumulative distribution function (CDF) of the score statistics, which was then used to calculate empirical p values corresponding to the observed score statistics. Subsequently, to demonstrate that TrajGWAS is conservative and SPA$_{GRM}$ is more accurate, we compare the p values from the resampling method, SPA$_{GRM}$, and TrajGWAS.

TrajGWAS and its empirical saddlepoint approximation use a perspective strategy to treat model residuals as random variables, whereas SPA$_{GRM}$ uses a retrospective strategy which treats genotypes as random variables to calibrate p values. Similarly, we used these two strategies to resample model residuals and genotypes. For the perspective resampling strategy, we resampled model residuals for 10^6 times and then calculated 10^6 score statistics. For the retrospective resampling strategy, we resampled genotypes for 10^6 times and then calculated 10^6 score statistics.

Supplementary Fig. 11a displays the distributions of model residuals when testing the mean profile H_0: $\beta_g = 0$ and WS variability H_0: $\tau_g = 0$. Supplementary Fig. 11b displays the resampling score statistics and their corresponding empirical p values. The empirical distributions of the resampling score statistics are the same for the retrospective strategy and perspective strategy, which is expected as these two sampling methods are actually equivalent. When testing the mean profile H_0: $\beta_g = 0$, the observed score statistics is 212.5. The corresponding p values from resampling method, TrajGWAS, and SPA$_{GRM}$ are very similar (from 4.69e-05 to 5.54e-05), which indicates that TrajGWAS and SPA$_{GRM}$ are both accurate if the model residuals are balanced. On the other hand, when testing for the WS variability H_0: $\tau_g = 0$, the observed score statistics is -19479.7 and the p values from resampling method is 2.52e-05. For SPA$_{GRM}$, the p value is 3.99e-05, which is close to that from resampling method, and for TrajGWAS, the p value is 3.30e-4, which is conservative.

TrajGWAS employs SPA to approximate the distribution of model residuals and then approximate the distribution of score statistics. Although SPA is more accurate than normal distribution approximation, it could be still invalid if the distribution of model residuals is extremely unbalanced, which further results in the power loss of TrajGWAS.

5. Genome-wide association studies of serum thyroid stimulating hormone

Thyroid stimulating hormone (TSH) plays an essential role in normal development and metabolism, and its secretion is linked to various diseases, including hyperthyroidism, thyroid nodules and tumors, and autoimmune thyroid diseases[8,9]. Zhou et al. conducted a GWAS meta-analysis using 119,715 individuals and identified 74 TSH-associated loci[10]. Recently, Williams et al. enlarged the sample size to 247,107 participants and increased the number of genetic associations to 260 loci[11]. The increased sample size is mainly due to the inclusion of TSH measurements from the UKB primary care data. However, Williams et al. transformed the longitudinal TSH records to a simpler quantitative trait by taking every individual's first non-missing TSH measurement, which would result in power loss.

In this study, we included 145,519 subjects in SPA_{GRM}, $SPA_{GRM(CCT)}$, and $Norm_{GRM}$ analysis, of which 121,916 (83.8%) unrelated subjects were analyzed using TrajGWAS. An average of 4.99 TSH records were measured for one subject. Supplementary Fig. 20 displays the results via Manhattan plots and QQ plots. Our longitudinal GWAS of TSH identified 211 (189+22) loci significantly associated with the mean level (loci identified by SPA_{GRM} + loci additionally identified by $SPA_{GRM(CCT)}$). Among these, 61 loci were exclusively detected by SPA_{GRM} and $SPA_{GRM(CCT)}$, whereas TrajGWAS only identified 3 additional loci.

We compared our results with the 148 loci that passed the p value threshold of 5e-8 using only the UKB primary care data in Williams et al.[11] Surprisingly, only 12 of 148 (8.1%) loci were located outside 500kb of the significant SNPs identified by SPA_{GRM} and $SPA_{GRM(CCT)}$. Meanwhile, based on the same criteria, 98 of 211 (46.4%) loci were exclusively identified by SPA_{GRM} and $SPA_{GRM(CCT)}$, demonstrating that the longitudinal

GWAS successfully identified more loci. In addition to 211 loci associated with the mean level of TSH, we also identified 86 loci linked to the WS variability.

Here, we highlighted several loci exclusively identified in our analysis to be associated with both the mean and WS variability of serum TSH. Most of the loci were located in genes previously reported to be associated with thyroid related diseases. For example, *BACH2* (rs597325, SPA_{GRM} p value = 1.02e-09 and 8.42e-10 for mean and WS variability, respectively) was identified to be associated with autoimmune thyroid disease (Graves' disease)[12]. It's worth noting that in a clinical case study of patients with thyroid nodules, specific fusion events involving *PPARG* (rs795002, SPA_{GRM} p value = 3.10e-28 & 2.20e-7 for mean and WS variability, respectively) were identified[13,14]. This is a significant characteristic for detecting the disease, and can also be indicative of thyroid cancer[14]. High expression of *GOT1* (rs10748781, SPA_{GRM} p value = 8.88e-30 & 1.06e-9 for mean and WS variability, respectively) in thyroid cancer has been associated with decreased patient survival, possibly due to *GOT1*'s ability to effectively promote cancer cell glycolysis and oxidation into the acidification pathway[15]. Knocking down the *PTPN11* (rs11066309, SPA_{GRM} p value = 9.76e-17 & 2.42e-24 for mean and WS variability, respectively) can prevent senescence in human thyroid cells over-expression the onco-genic protein BRAF V600E, suggesting that *PTPN11* may be a tumor suppressor gene targeted against thyroid cancer[16].

6. Simulation studies based on generalized estimation equations

We used three generative models to simulate longitudinal traits, including linear mixed model (LMM) as in model (3), and GEE with exchangeable and autoregressive working correlation structures. Based on model (5), we added an additional random effect b_i to characterize sample relatedness, and simulated error term ε_{ij} according to corresponding correlation matrices with the parameter $\rho = 0.8$. For type I error analysis, we simulated 100 sets of longitudinal traits without any causal genetic effects (i.e., $\beta_g = 0$) and conducted association tests on 100,000 common variants and 100,000 rare variants, respectively, resulting in 1×10^7 replications in each scenario (Supplementary Fig. 23). For empirical power analysis, we selected 5 common variants

and 5 rare variants as causal SNPs with $\beta_g = -\log10(\text{MAF}) \times 0.1$ for common variants and $\beta_g = -\log10(\text{MAF}) \times 0.02$ for rare variants, respectively. In each scenario, 200 sets of longitudinal traits were simulated, and a total of 1×10^3 replications were performed. The mean chi-square statistics are summarized in Supplementary Table 5. All other settings followed those described in the simulation studies in the main text.

7. Evaluation on accuracy of MGF estimation via the Chow-Liu algorithm

We used pedigree data consisting of 6250 4-member families as showed in Supplementary Fig. 1a. Genotypes with MAFs of 0.3, 0.01, and 0.001 were considered to represent common and rare variants. We considered residuals for testing $\beta_g = 0$ and $\tau_g = 0$, to mimic balanced and unbalanced phenotype distributions. The theoretical MGF estimation of score statistics can be expressed as $M_S(t) = \prod_{i=1}^{6250} M_{S_i}(t)$, where $S_i = \sum_{j=1}^{4} R_{i_j} G_{i_j}$ represents the score for i-th family. Since the family structure is known, $M_{S_i}(t)$ for 4-member families as showed in Supplementary Fig. 1a can be computed as

$$
\begin{aligned}
M_{S_i}(t) = \frac{1}{4}\left(1-\mu+\mu e^{t\left(R_{i_1}+R_{i_3}+R_{i_4}\right)}\right)\cdot\left(1-\mu+\mu e^{t\left(R_{i_2}+R_{i_3}+R_{i_4}\right)}\right) \\
\cdot\left(1-\mu+\mu e^{tR_{i_3}}\right)\cdot\left(1-\mu+\mu e^{tR_{i_4}}\right)+\frac{1}{4}\left(1-\mu+\mu e^{t\left(R_{i_1}+R_{i_3}+R_{i_4}\right)}\right) \\
\cdot\left(1-\mu+\mu e^{t\left(R_{i_2}+R_{i_3}\right)}\right)\cdot\left(1-\mu+\mu e^{t\left(R_{i_2}+R_{i_4}\right)}\right)\cdot\left(1-\mu+\mu e^{tR_{i_1}}\right) \\
+\frac{1}{4}\left(1-\mu+\mu e^{t\left(R_{i_2}+R_{i_3}+R_{i_4}\right)}\right)\cdot\left(1-\mu+\mu e^{t\left(R_{i_1}+R_{i_3}\right)}\right) \\
\cdot\left(1-\mu+\mu e^{t\left(R_{i_1}+R_{i_4}\right)}\right)\cdot\left(1-\mu+\mu e^{tR_{i_2}}\right) \\
+\frac{1}{4}\left(1-\mu+\mu e^{t\left(R_{i_1}+R_{i_3}\right)}\right)\cdot\left(1-\mu+\mu e^{t\left(R_{i_1}+R_{i_4}\right)}\right) \\
\cdot\left(1-\mu+\mu e^{t\left(R_{i_2}+R_{i_3}\right)}\right)\cdot\left(1-\mu+\mu e^{t\left(R_{i_2}+R_{i_4}\right)}\right),
\end{aligned}
$$

where μ represents MAF. Similarly, MGF estimation via the Chow-Liu algorithm can be expressed as $\widehat{M}_S(t) = \prod_{i=1}^{6250} \widehat{M}_{S_i}(t)$, where $\widehat{M}_{S_i}(t) = \sum \widehat{\Pr}(G_{i_1} = g_1, \dots, G_{i_4} = g_{n_i}) \cdot e^{t(R_{i_1} g_1 + \cdots + R_{i_4} g_4)}$. Normal distribution approximation only needs the mean and variance of score statistics

$$E(S) = 2 \cdot \mu \cdot R^T \cdot \mathbf{1}_n, \qquad Var(S) = 2\mu(1-\mu) \cdot R^T \cdot \Phi \cdot R,$$

where Φ is the GRM. Then MGF estimation via normal distribution approximation can be expressed as $\widehat{M}_S(t) = E(S) \cdot t + \frac{1}{2} \cdot Var(S) \cdot t^2$. MGF estimation of three methods across various genotype and phenotype distributions are displayed in Supplementary Fig. 29.

8. Evaluation of quality control in longitudinal GWAS

Quality control (QC) for raw UKB primary care data is crucial for GWAS of longitudinal traits especially when testing the within-subject (WS) variability. Expanding upon the QC by Ko et al.[17], we also considered the presence of duplicated and outlier values in the raw UKB primary care data. (see the Methods section). We selected three longitudinal traits to evaluate the influence of QC: body mass index (BMI), fasting glucose, and high-density lipoprotein cholesterol (HDL). These three traits were also analyzed in Ko et al.[17] We applied TrajGWAS and SPA_{GRM} to these traits using the QC in Ko et al.[17] (named "previous QC") and our purposed QC pipeline (named "our QC"). In this section, we restricted our analysis to 9.4 million imputed variants with MAF > 0.01, as rare variants are likely not significant for WS variability.

Manhattan plots (Supplementary Fig. 32) replicate nearly the same results as in Ko et al.[17] when use the previous QC, and our purposed QC can well detect some signals associated with the WS variability that were overlooked by Ko et al.[17]. We highlight the *FTO* gene (SNP rs12149832, SPA_{GRM} p value = 2.12e-13) and *TCF7L2* gene (SNP rs36090025, SPA_{GRM} p value = 2.88e-22) are found to be associated with the WS variability of BMI and glucose levels, respectively. However, without a thorough QC pipeline, these two genes were not identified when testing for within-subject variability (p value > 1.2e-4)[17]. These two genes have been identified to be the variance quantitative trait loci (vQTLs) for corresponding traits[18,19]. Although vQTL analysis is typically performed on the cross-sectional data[20,21], our findings reveal some consistency between the vQTLs and loci associated with WS variability. Our purposed QC also detected several loci associated with the WS variability of HDL. Precise QC will not only ensure to identify WS variability related variants, but also enhance the power to test for the mean profile for some traits (e.g. fasting glucose).

9. SPA$_{GRM}$ gains statistical power through PGS adjustment

In this section, we demonstrated that SPA$_{GRM}$-PGS can further gain statistical power via incorporating polygenic scores (PGSs) as covariates. SPA$_{GRM}$-PGS is a two-stage-strategy that consists of calculating polygenic scores using the GWAS results of SPA$_{GRM}$ and then including the polygenic scores (PGSs) as fixed effects for a second round GWAS. Also, consistent with previous studies[22-24], leave-one-chromosome-out (LOCO) scheme was applied to construct LOCO-PGSs to avoid proximal contamination[25,26]. We focused the comparison on the mean effect, as only a few loci were associated with WS variability in real data analyses.

In stage 1, we constructed the LOCO-PGSs using a straightforward C+T method[27]. For longitudinal trait analyses, GWAS results of original SPA$_{GRM}$ were clumped using plink v1.9 (p1=p2=5e-8; r^2=0.01 and kb=2000). And later effect sizes of these index SNPs were estimated by WiSER[28] using the unrelated samples. LOCO-PGSs were constructed using index SNPs on the other 21 autosomal chromosomes and their corresponding effect sizes mentioned above. This resulted 22 sets of PGS values. In stage 2, the estimated LOCO-PGSs were then included into the null model as covariates. For longitudinal trait analyses, we incorporated calculated LOCO-PGSs into WiSER[28] to obtain model residuals. We obtained 22 sets of LOCO model residuals, each for each chromosome, and then passed them to step 2 for the association analyses via SPA$_{GRM}$.

Supplementary Fig. 33 displayed that PGS adjustment resulted in more significant p values at known peaks and an increased number of newly identified loci. For eGFR and TSH, SPA$_{GRM}$-PGS additionally identified 13 and 10 loci associated with the mean trajectory compared to SPA$_{GRM}$. For ferritin, only one new locus was exclusively discovered by SPA$_{GRM}$-PGS, probably due to the smaller sample size. Similar as mentioned above, we also used GEE models with exchangeable and autoregressive working correlation structures to fit the null model, and SPA$_{GRM}$-PRS still performed better (Supplementary Fig. 34).

10. SPA$_{GRM}$ applied to quantitative and binary traits

In the main text, we applied SPA$_{GRM}$ to longitudinal trait analyses and evaluated its performance via both simulation studies and UK Biobank real data analyses. To the best

of our knowledge, there is no available mixed model methods to incorporate GRM-related random effect into a multiple location scale model[29] to address family relatedness. Thus, we compared SPA$_{GRM}$ to the recently proposed TrajGWAS[17] which is limited to analyze unrelated subjects. The analysis results demonstrated that SPA$_{GRM}$ overperformed TrajGWAS, which suggests that SPA$_{GRM}$ can serve as a valuable solution to address family relatedness for the complex traits with complicated structure, especially if no relevant mixed model method is available.

As a universal analysis framework, SPA$_{GRM}$ is not limited to analyzing longitudinal traits. Here, we focused on the comparison between SPA$_{GRM}$ and many existing mixed model approaches[30-34] for both quantitative and binary traits. Our comparison has two main aims. We first compared SPA$_{GRM}$ with those sparse-GRM-based mixed model methods[32,34] to evaluate the influence of not incorporating the random effect. We then compared SPA$_{GRM}$-PGS with those dense-GRM-based mixed model methods[30,31,35] to evaluate whether PGS adjustment can approximate results from mixed model approaches.

The formal consistency of score statistics across various traits

For a diverse set of traits including binary and quantitative traits, the score statistics share a consistent format expressed as $S = R^T G$. Here, $G = (G_1, G_2, \dots, G_n)^T$ is the genotype vector of n samples and $R = (R_1, R_2, \dots, R_n)^T$ is the model residual vector whose definition varies depending on the type of traits. In this paper, in addition to longitudinal traits, we also evaluated SPA$_{GRM}$ using quantitative and binary traits. For linear and logistic regression, the score statistics share the general format $S = G^T(Y - \hat{\mu})$, where Y is phenotype vector and $\hat{\mu}$ is the predictor vector of Y under the null hypothesis[36]. The score statistics corresponding to mixed models can be found elsewhere[30,31].

In SPA$_{GRM}$, we consider the genotypes as random variables, while considering the residuals as fixed variables. Consequently, for different statistical models, SPA$_{GRM}$ can

employ a universal strategy to approximate the null distribution of score statistics and conduct the score testing. Thus, our discussion pertains to the consistent format of score statistics $S = R^T \cdot G$, without restricting it to a specific trait.

SPA$_{GRM}$ performs comparable to sparse-GRM-based mixed model methods

To comprehensively evaluate SPA$_{GRM}$, for both quantitative and binary traits, we fit the conventional models and the mixed models incorporating random effect in step 1 to calculate model residuals. For a quantitative trait, we fit a linear model (denoted as SPA$_{GRM}$-LM) and a linear mixed model (denoted as SPA$_{GRM}$-LMM). For a binary trait, we fit a generalized linear model with a logit link function (denoted as SPA$_{GRM}$-GLM) and a generalized linear mixed model (denoted as SPA$_{GRM}$-GLMM). Regardless of the traits and models, the model residuals were passed to step 2 for the association analyses in SPA$_{GRM}$.

The fastGWA-MLM and fastGWA-GLMM are popular mixed model methods that support sparse GRM. Sparse GRM was calculated using 227,437 LD-pruned high-quality genetic variants and elements below than 0.05 were set to zero. SPA$_{GRM}$-LMM and SPA$_{GRM}$-GLMM use linear mixed model and logistic mixed model to fit the null model and obtain the model residuals. In practice, we used the SAIGE[31] to fit the null mixed models and obtain corresponding model residuals. Consistent with fastGWA, we used sparse GRM to fit the null mixed models for SPA$_{GRM}$. SPA$_{GRM}$-LR and SPA$_{GRM}$-GLM use the linear regression and logistic regression to fit the null model and obtain the model residuals, respectively.

We applied SPA$_{GRM}$, fastGWA-MLM[34], and fastGWA-GLMM[32] to analyze four quantitative traits including height (UKB field ID: 50; $N = 407{,}443$), body mass index (BMI; UKB field ID: 21001; $N = 405{,}187$), alanine transaminase (ALT; UKB field ID: 30620; $N = 382{,}371$), and calcium (UKB field ID: 30680; $N = 356{,}043$), and four binary traits including coronary artery disease (PheCode 411; case-control ratio: 1:10); colorectal cancer (PheCode 153; case-control ratio: 1:64), glaucoma

(PheCode 365; case-control ratio: 1:67), and thyroid cancer (PheCode 193; case-control ratio: 1:710) with the same sample size $N = 408{,}510$. The analyses were restricted to 408,961 White British participants and 9.4 million imputed SNPs with MAF > 0.01. By default, we adjusted age, gender, and the first 10 principal components (PCs) in the null model.

Supplementary Figs. 35-37 display the analysis results via Manhattan plots and quantile-quantile (QQ) plots. Across all traits, the results show close consistency between fastGWA and SPA_{GRM}. SPA_{GRM} yielded comparable p values to fastGWA when incorporating the random effect in the null model (i.e., SPA_{GRM}-LMM and SPA_{GRM}-GLMM), though it exhibited slightly comprehensive p values for some variants. The results also revealed that including a random effect based on sparse GRM provided minimal to no power benefit for SPA_{GRM} in quantitative and binary traits, respectively.

The comparison is informative for the necessity to develop new mixed model-based GWAS methods. As a universal framework, SPA_{GRM} is not intended to replace existing mixed-model methods but is a preferred option to act as a supplementary solution, especially in situations mixed-model methods are not available, such as the longitudinal trait analysis. Using a sparse GRM in mixed models may offer little benefit for power improvement compared to SPA_{GRM}, which largely reduces the concerns that the universal framework may lose power.

SPA_{GRM} gains power by incorporating PGS into null model fitting

SPA_{GRM} can benefit from inclusion of related individuals via a retrospective approach and further gain statistical power through PGS adjustment. Here, we evaluated SPA_{GRM}-PGS in quantitative and binary trait analyses, and compared it with established mixed model methods[30,31,35]. We applied BOLT-LMM[30], SAIGE[31], REGENIE[35], SPA_{GRM}-PGS, and the original SPA_{GRM} to four quantitative and binary traits mentioned previously. By default, all these methods use a LOCO scheme (except for the original SPA_{GRM}). 227,437 high-quality genetic variants were used for BOLT-LMM and

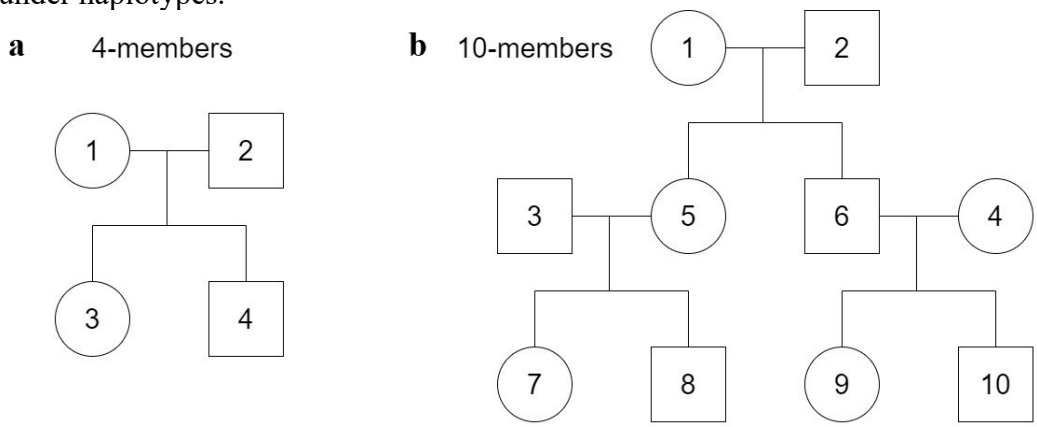
SAIGE to construct a full GRM, and for REGENIE to estimate polygenetic effects. We only reported the test statistics of BOLT-LMM-Inf (infinitesimal model), as we focused on the comparison under the infinitesimal assumption. Similar to longitudinal trait analyses, LOCO-PGS was constructed based on the GWAS results of SPA_{GRM}-LR and SPA_{GRM}-GLM for quantitative and binary traits, respectively. To evaluate the improvement of power, SPA_{GRM}-LR and SPA_{GRM}-GLM were also included in the comparation as the baseline.

Supplementary Figs. 38-40 display the results of quantitative and binary traits via Manhattan plots and quantile-quantile (QQ) plots. Results showed that SPA_{GRM}-PGS significantly gained powers for most of the quantitative traits and yielded limited power improvement for binary traits compared to SPA_{GRM}. For quantitative traits, SPA_{GRM}-PGS achieved very close p values compared with REGENIE, whereas BOLT-LMM was generally most powerful. All four methods showed good agreement for binary traits.

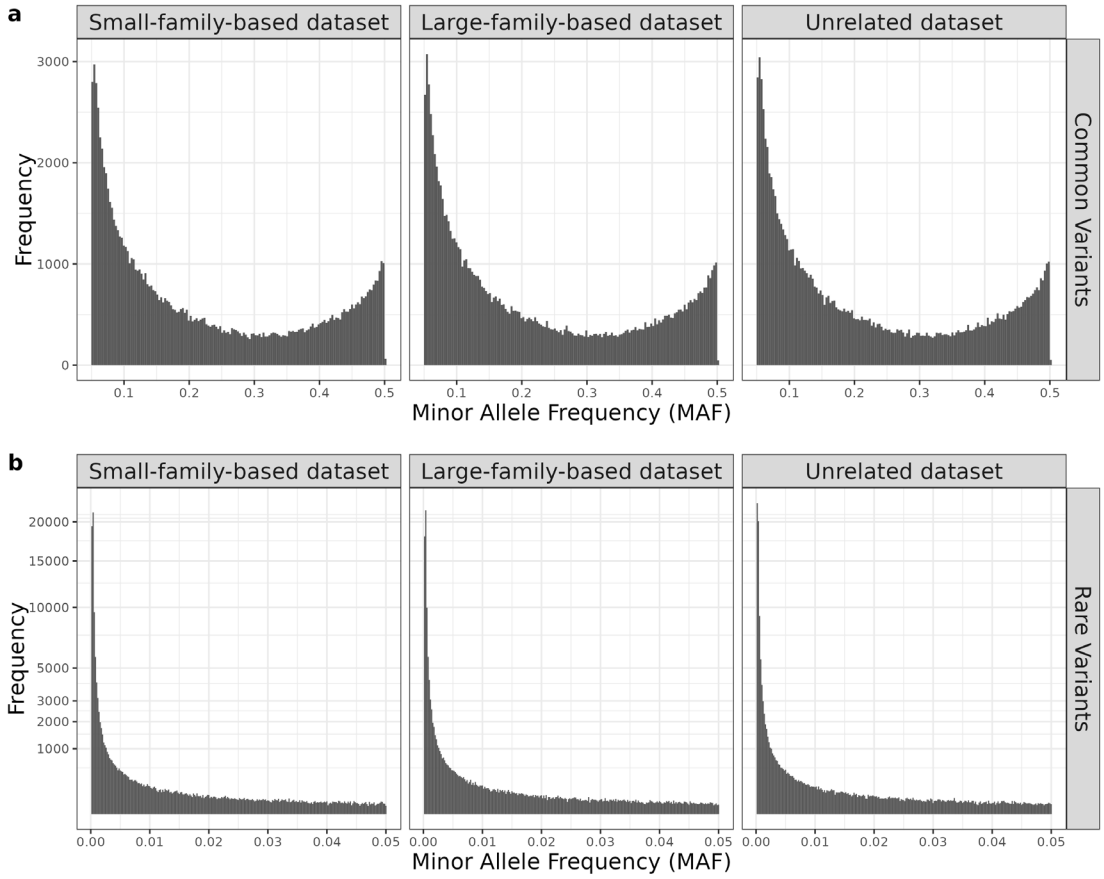
Computational efficiency of SPA_{GRM} analyzing quantitative and binary traits

In the main text, we evaluate the computational efficiency of SPA_{GRM} in analyzing longitudinal traits. Here we evaluate its computational efficiency in quantitative and binary trait analyses. The evaluations were performed on 9.4 million imputed SNPs using two real traits, including BMI ($N = 405{,}187$), and coronary artery disease ($N = 408{,}510$). All analyses were conduct on a CPU model of Intel(R) Xeon(R) Gold 6342 CPU @ 2.80GHz. The peak memory usage was less than 3.2 GB for these two traits. SPA_{GRM} performed slightly faster than REGENIE for a single trait analysis (231 hours vs 250 h for quantitative trait and 175 h vs 235 h for binary trait), and was also faster than SAIGE for binary trait (175 h vs 219 h). Notably, fastGWA demonstrated remarkable efficiency, being 62 (3.7 h) and 36 times (4.8 h) faster than SPA_{GRM} for quantitative and binary traits, respectively.

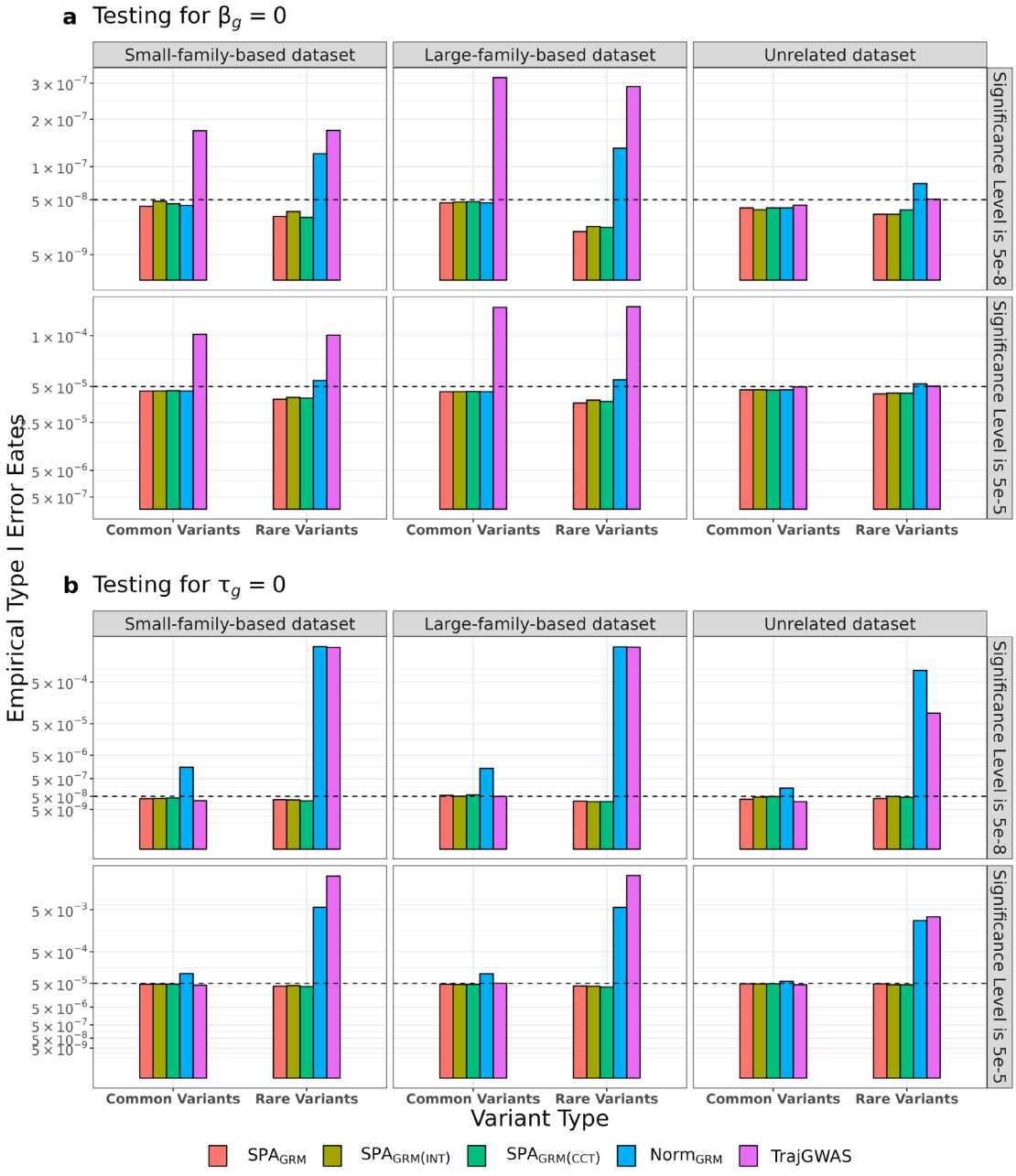
Supplementary Figure 1. Family structures in simulation data. A total of 25,000 related subjects are simulated using the below 2 scenarios. (**a**). 4-member families: 6,250 families, each of which includes all 4 members in the family. The members #1 and #2 are subjects with founder haplotypes. (**b**). 10-member families: 2,500 families, each of which includes 10 members in the family. The members #1-#4 are subjects with founder haplotypes.



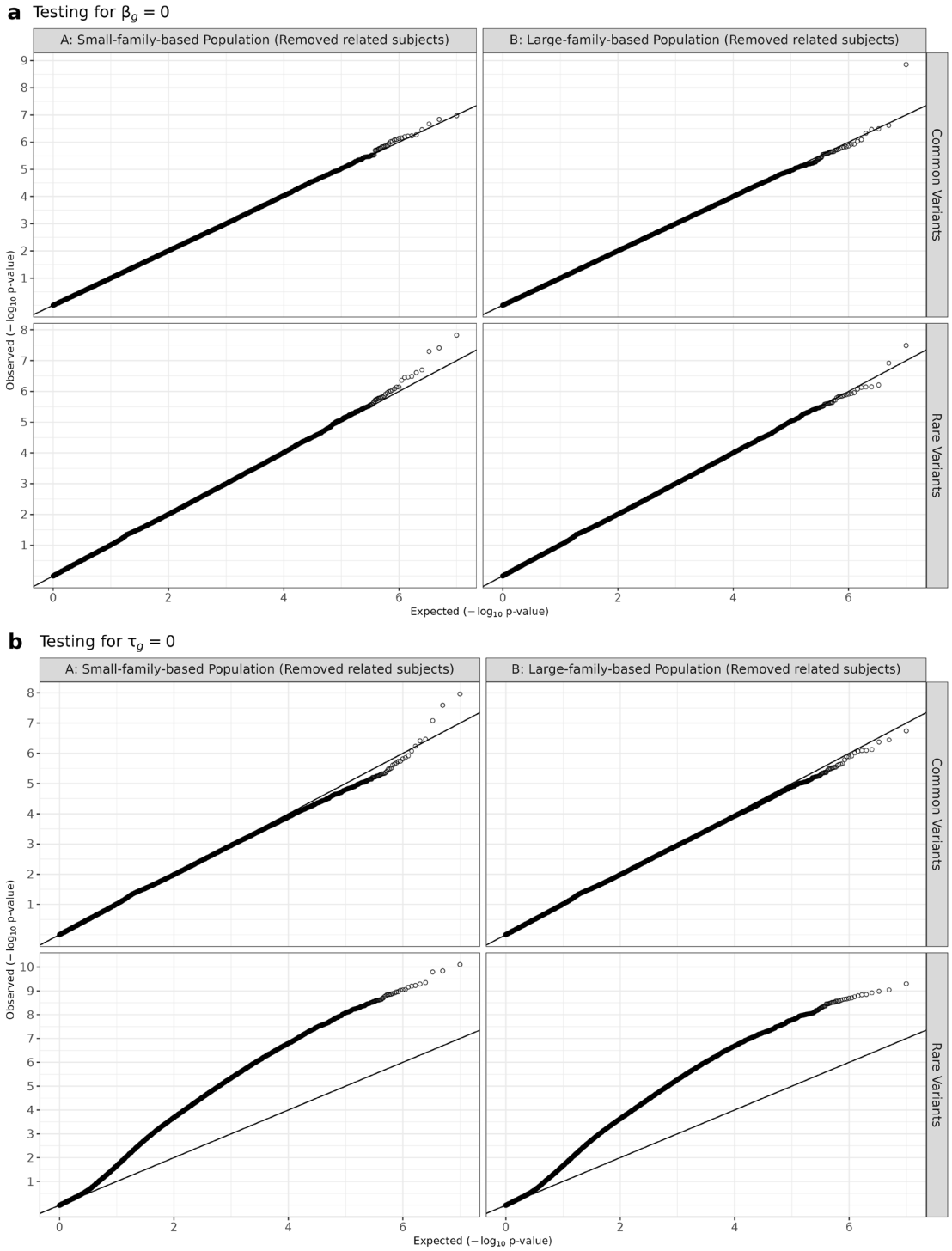
Supplementary Figure 2. Distributions of MAFs for simulated common and rare variants. (**a**). MAFs of common variants range from 0.05 to 0.5; and (**b**). MAFs of rare variants range from 0.0002 to 0.05.



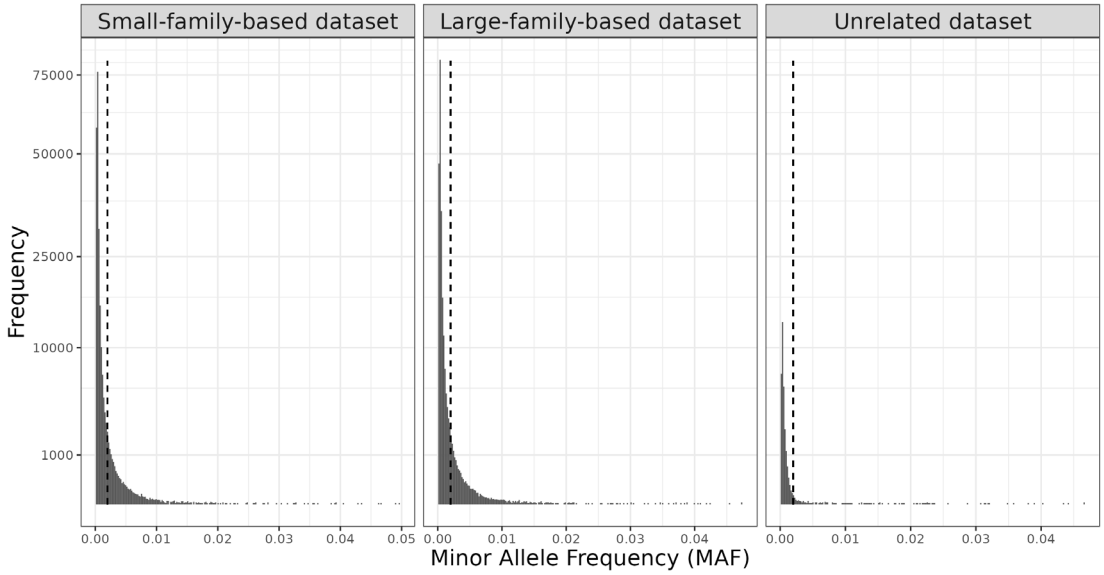
Supplementary Figure 3. Empirical type I error rates of SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, Norm$_{GRM}$, and TrajGWAS methods based on 10^9 replications. Top and bottom plots represent empirical type I error rates for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. From left to right, the plots consider three family relatedness of small-family-based dataset; large-family-based dataset; and unrelated dataset. We compared the empirical type I error rates at significance level of 5×10^{-8} and 5×10^{-5}.



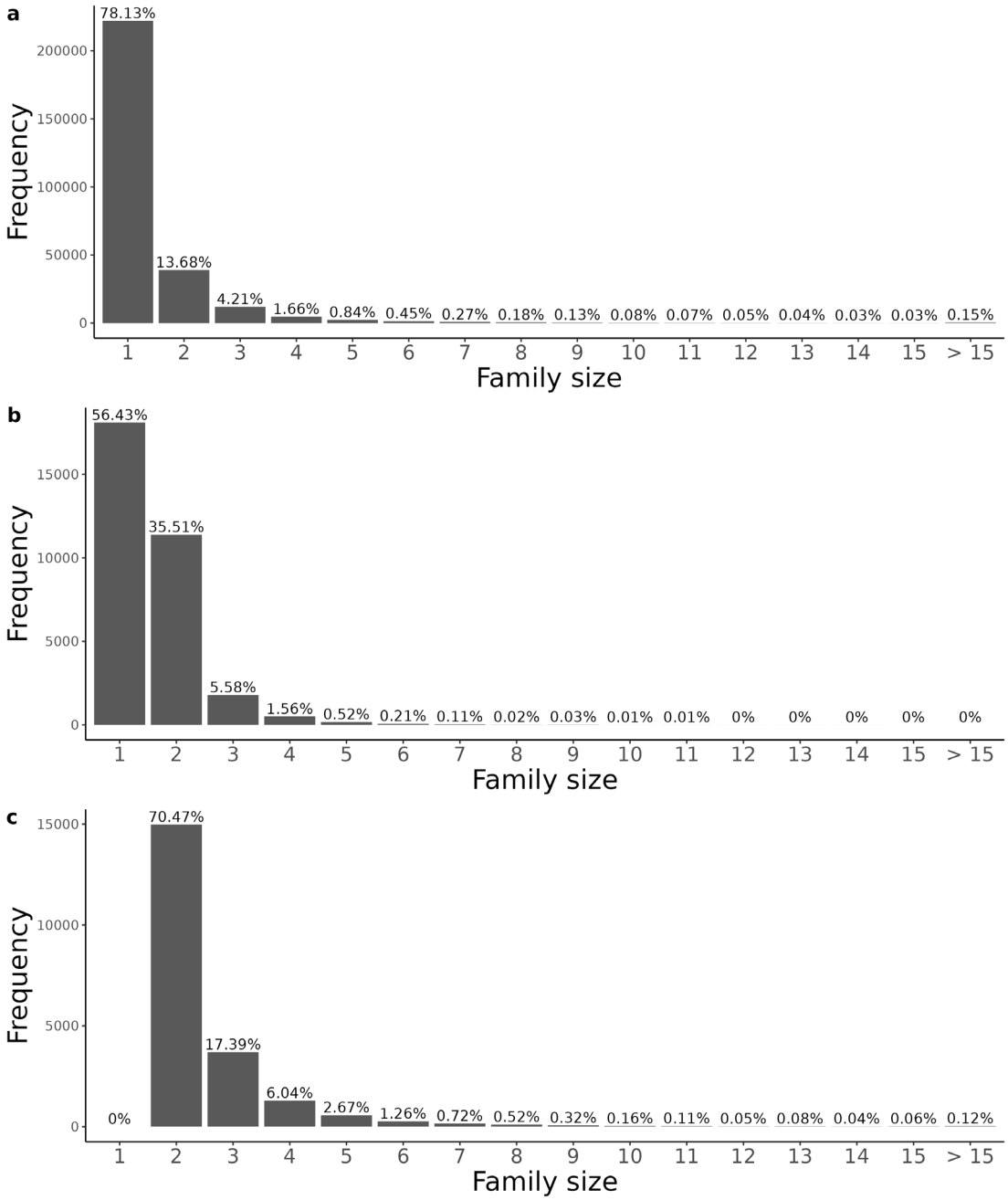
Supplementary Figure 4. Quantile-quantile (QQ) plots of p values of TrajGWAS method in small-family based dataset and large-family based dataset when related subjects were removed. Top and bottom plots represent QQ plots of p values of TrajGWAS for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. From left to right, the plots consider two datasets of small-family-based dataset (removed related subjects) and large-family-based dataset (removed related subjects). From top to bottom in each subplot, plots consider 10^5 common variants and rare variants. In each scenario, $10^7 = 100 \times 10^5$ tests were conducted. The inflation stemmed from the ultra-rare variants with MAF less than 0.002 (see **Supplementary Fig. 5**).



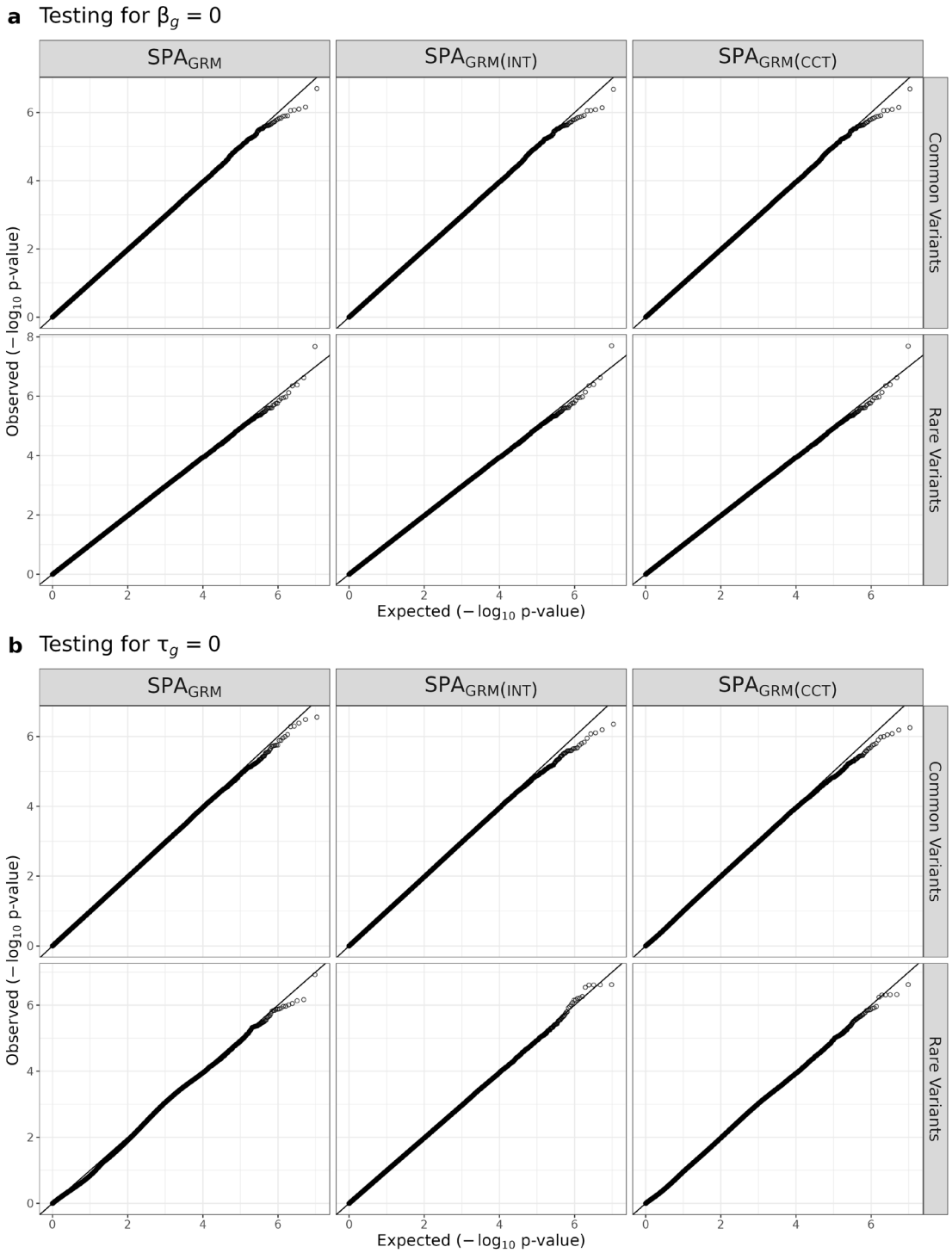
Supplementary Figure 5. Distributions of MAFs of rare variants with TrajGWAS p values $< 5 \times 10^{-5}$ for testing $\tau_g = 0$ in scenario 1 (i.e., the true $\beta_g = \tau_g = 0$). MAFs of rare variants range from 0.0002 to 0.05. From left to right, the plots consider three family relatedness of small-family-based dataset; large-family-based dataset; and unrelated dataset. $100 \times 10^5 = 1 \times 10^7$ tests were conducted in each scenario. The dashed line represents MAF = 0.002. TrajGWAS produced inflation when testing rare variants in terms of τ_g.



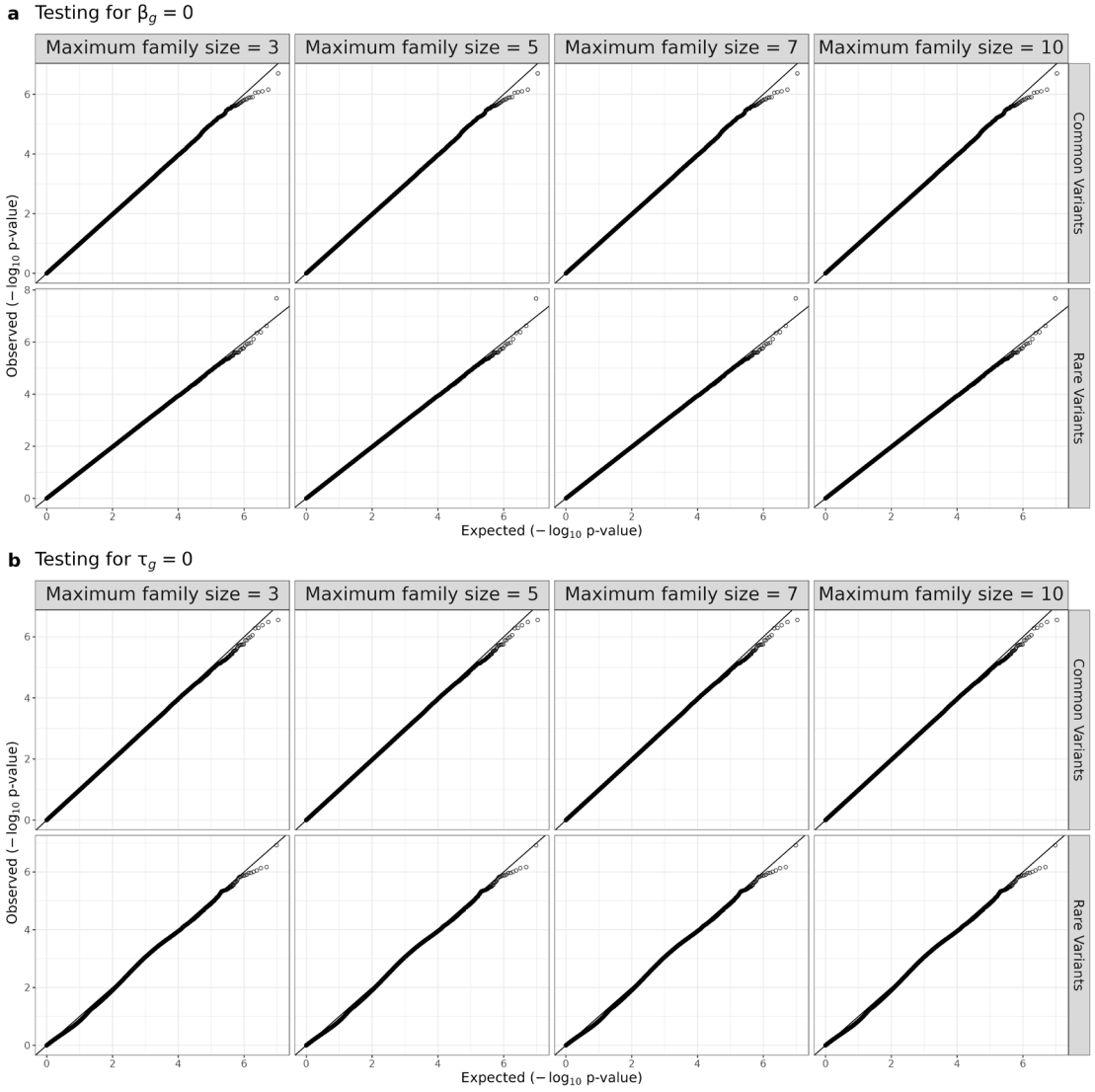
Supplementary Figure 6. Histograms of family sizes for UKB white British participants. (**a**) the full set of 408,961 UKB white British participants. (**b**) 50,000 selectively sampled UKB white British participants with a higher proportion of relatives. We first excluded unrelated individuals of 408,961 white British participants, resulting in 186,760 related samples; then we randomly sampled these 50,000 individuals from above related samples. (**c**) 54,974 UKB white British participants with longitudinal BMI measurements, excluding unrelated individuals. Pairs of subjects with genetic relatedness > 0.05 are grouped in the same family. Family size is defined as the number of members in a family. Each bar represents the proportion of this type of families to all families.



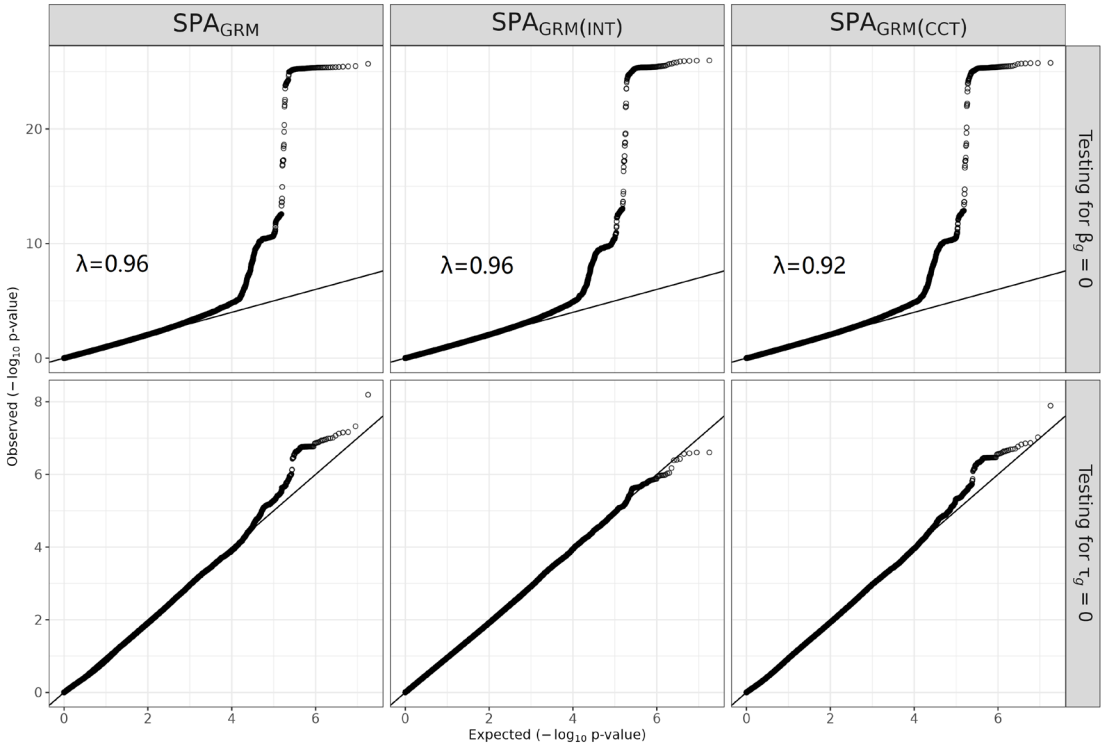
Supplementary Figure 7. Quantile-quantile (QQ) plots of p values of SPA$_{GRM}$-based methods SPA$_{GRM}$, SPA$_{GRM(INT)}$, and SPA$_{GRM(CCT)}$ in presence of cryptic sample relatedness. Top and bottom plots represent QQ plots of p values of SPA$_{GRM}$, SPA$_{GRM(INT)}$, and SPA$_{GRM(CCT)}$ for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. From top to bottom in each subplot, plots consider 100,000 common variants and 100,000 rare variants. In each scenario, $1 \times 10^7 = 100 \times 10^5$ tests were conducted.



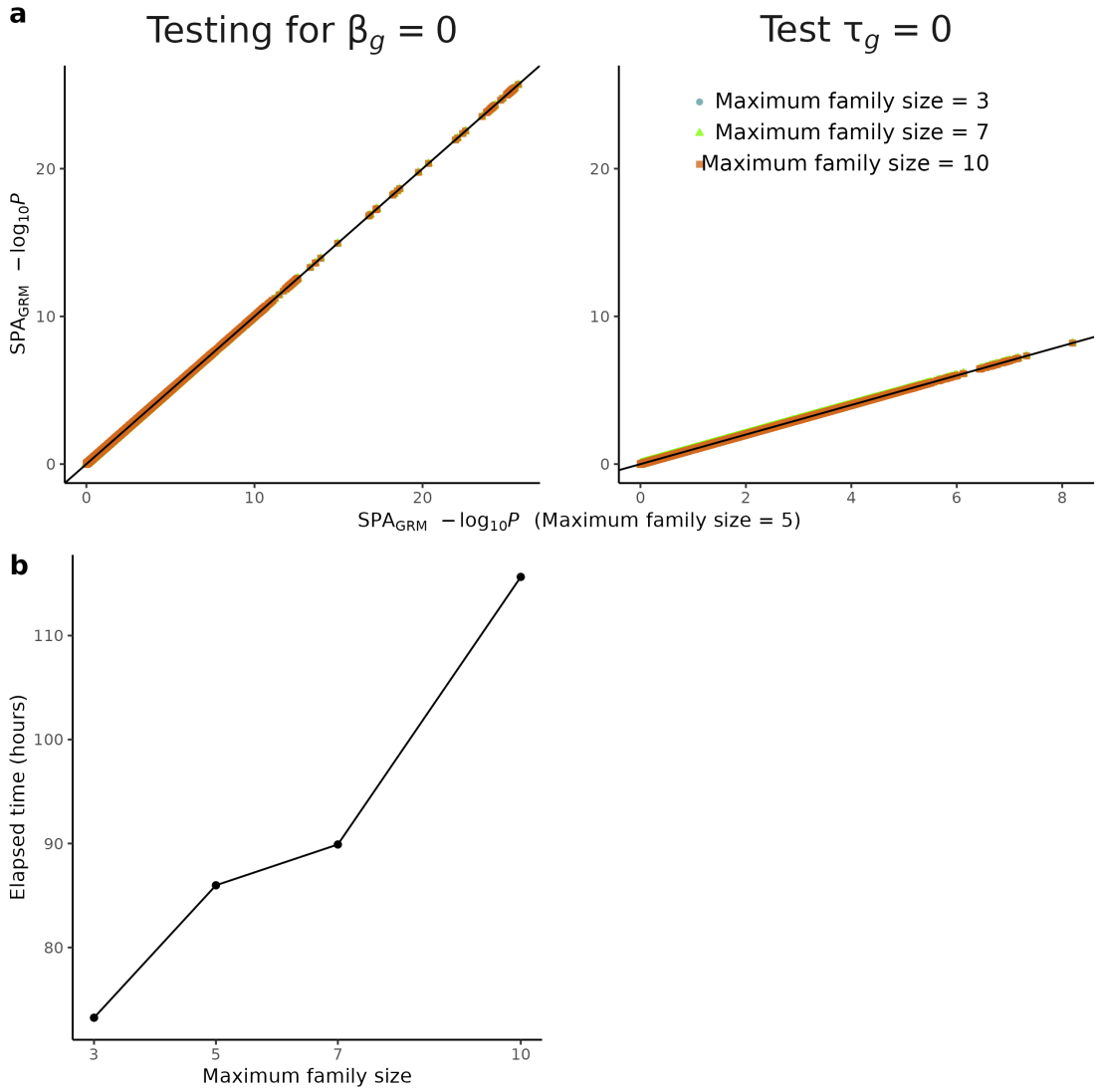
Supplementary Figure 8. Quantile-quantile (QQ) plots of p values of SPA$_{GRM}$ under various maximum family size settings in presence of cryptic sample relatedness. Top and bottom plots display QQ plots of p values of SPA$_{GRM}$ method for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. Within each subplot, the maximum family size considered, from left to right, are 3, 5 (the default setting in SPA$_{GRM}$), 7, and 10. From top to bottom in each subplot, plots consider 100,000 common variants and 100,000 rare variants. In each scenario, $1 \times 10^7 = 100 \times 10^5$ tests were conducted.



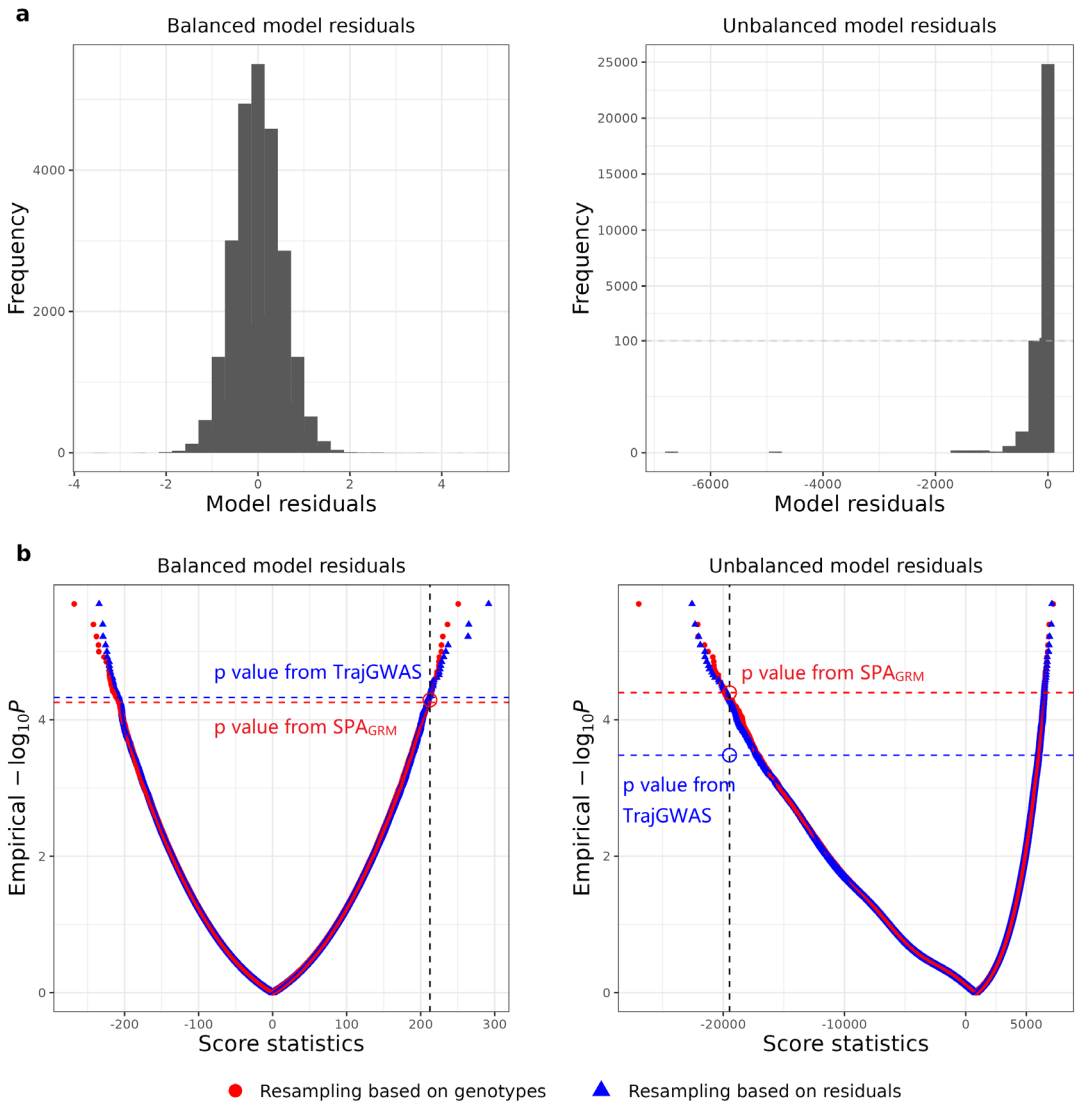
Supplementary Figure 9. Quantile-quantile (QQ) plots of GWAS results of SPA$_{GRM}$, SPA$_{GRM(INT)}$, and SPA$_{GRM(CCT)}$ using a real phenotype BMI. Totally there are 175,443 UKB white British participants with longitudinal BMI measurements. After excluding unrelated individuals, 54,974 individuals remain for this analysis to maximize the presence of cryptic relatedness (see **Supplementary Fig. 6**). From left to right, from top to bottom, plots represent QQ plots of p values of SPA$_{GRM}$-based methods for testing $\beta_g = 0$ and $\tau_g = 0$. The inflation in testing for $\beta_g = 0$ primarily due to pleiotropy, as indicated by genomic control inflation factors $\lambda = 0.92 - 0.96$ for these methods.



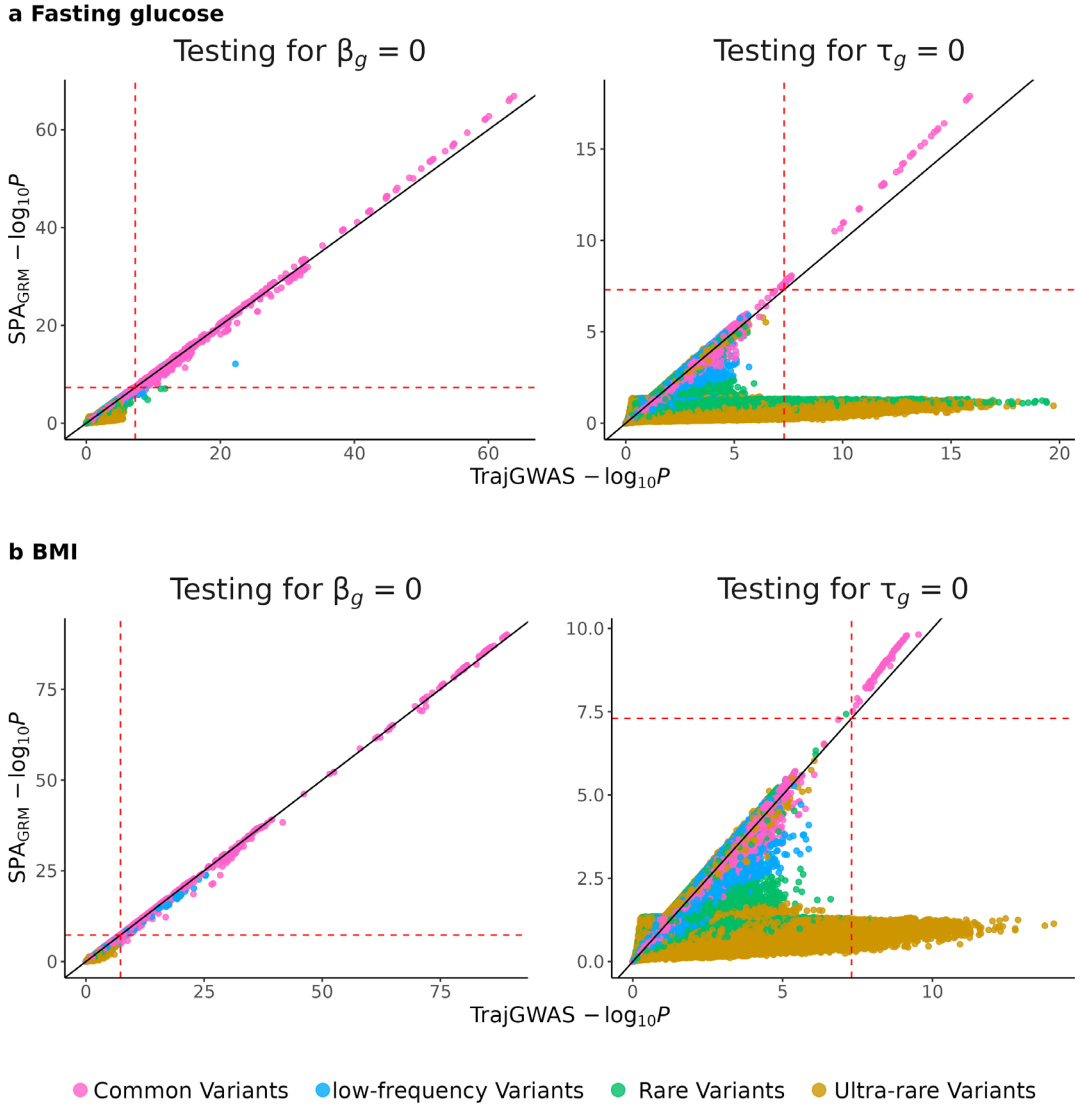
Supplementary Figure 10. Comparison of SPA_{GRM} using various maximum family size settings using a real phenotype BMI. (**a**) Scatter plots comparing p values of SPA_{GRM} using various maximum family size settings. (**b**) Run time of SPA_{GRM} using various maximum family size settings. Totally there are 175,443 UKB white British participants that have longitudinal BMI measurements. After excluding unrelated individuals, 54,974 individuals remain for the SPA_{GRM} analysis to maximize the presence of cryptic relatedness (see **Supplementary Fig. 6**). Left and right plots represent scatter plots of p values of SPA_{GRM} method for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. The x-axis represents the -log10 p values of SPA_{GRM} using the default setting (maximum family size equal to 5), while the y-axis represents the -log10 p values of SPA_{GRM} with maximum family sizes of 3, 7, and 10. Dots can be overlapped because using different maximum family size settings has minimal impact on SPA_{GRM}' performance.



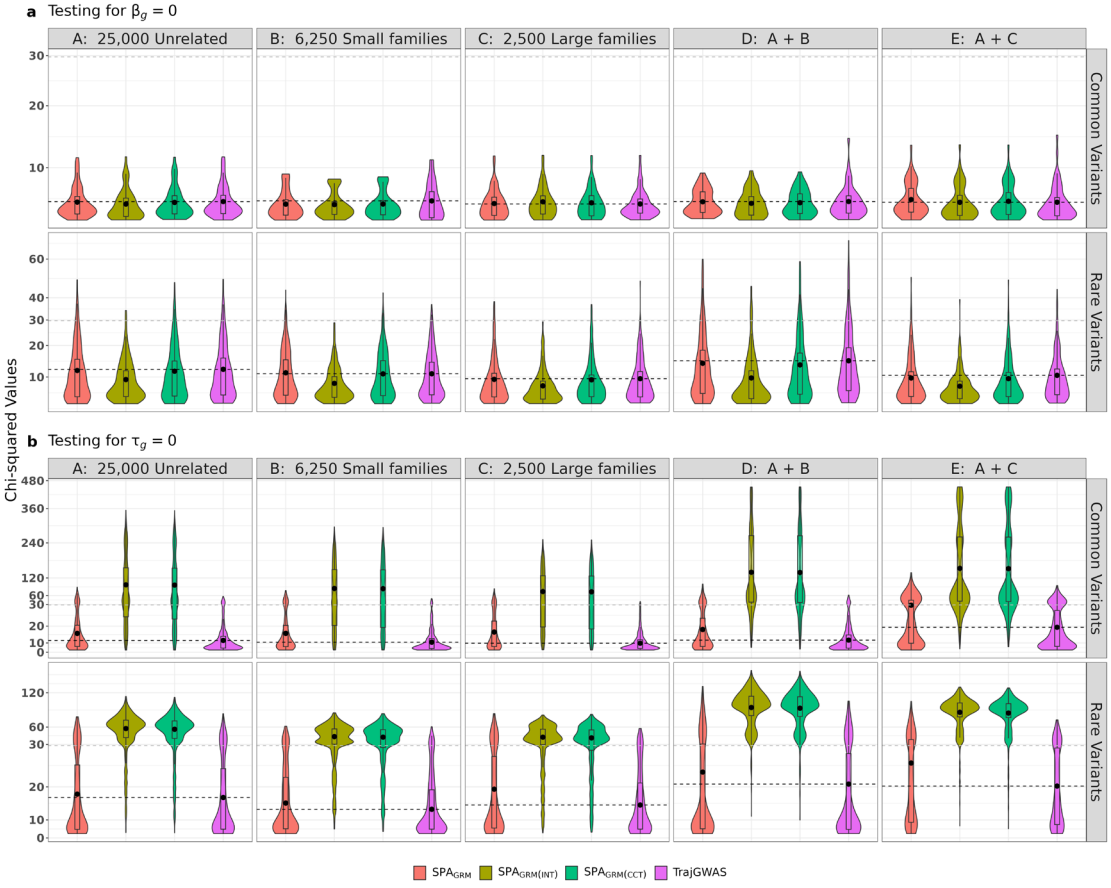
Supplementary Figure 11. Simulation results to demonstrate TrajGWAS produces conservative p values when model residuals are high unbalanced. (**a**) represents the histograms of balanced and unbalanced residuals representing model residuals obtained from mean and WS variability formulas, respectively. The grey dashed horizontal line represents the breakpoint for the different scaling of the y axis. (**b**) presents the empirical $-log_{10}P$ values estimated from resampling methods based on genotypes and model residuals, respectively. The dashed vertical line represents the realized score statistic. The red dashed horizontal line represents SPA_{GRM} produced $-log_{10}P$, and the blue dashed horizontal line represents TrajGWAS produced $-log_{10}P$.



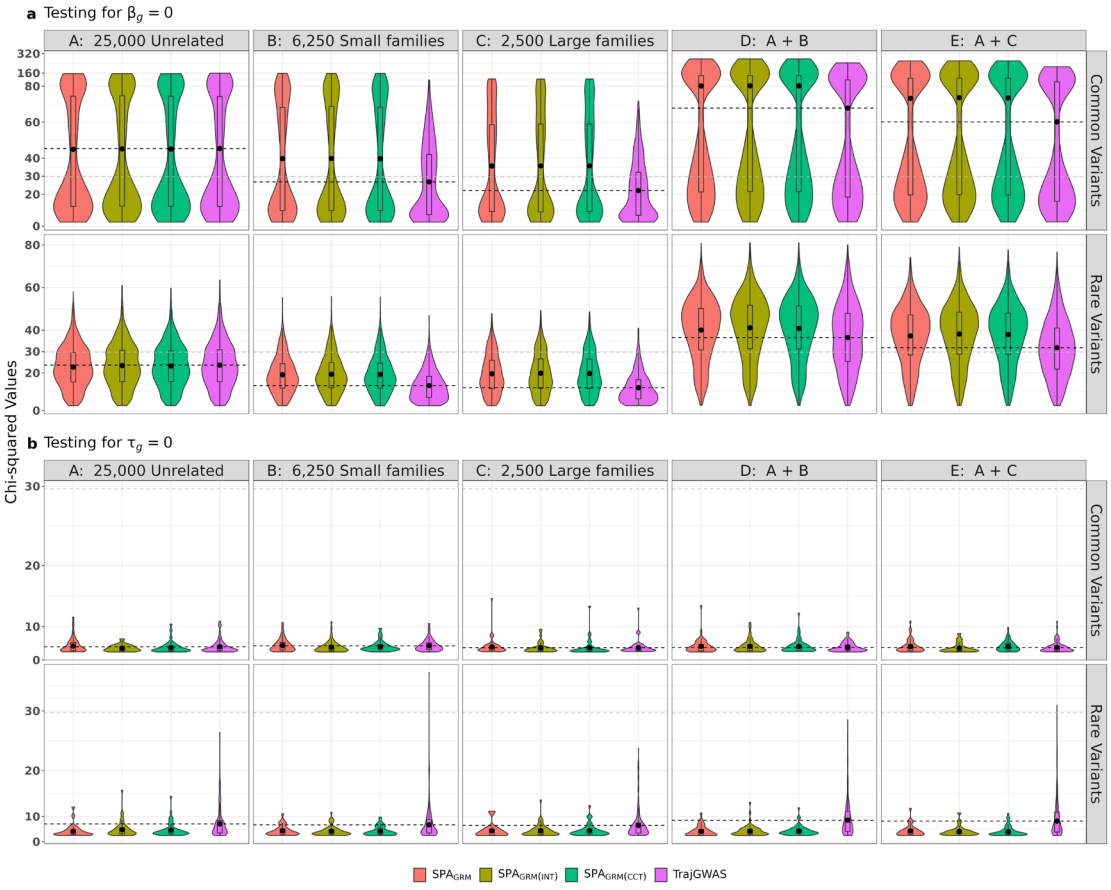
Supplementary Figure 12. Scatter plots of GWAS results for fasting glucose and BMI when the analysis is restricted to unrelated White British descents. From top to bottom, (**a**) is the result of analyzing fasting glucose, and (**b**) is the result of analyzing BMI. From left to right are results for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. Two methods TrajGWAS (x-axis) and SPA$_{GRM}$ (y-axis) are compared. The red dashed line represents the genome-wide significance ($P = 5 \times 10^{-8}$). Genetic variants are grouped based on minor allele frequency (MAF). Common variants with MAF = (0.05, 0.5); low-frequency variants with MAF = (0.01, 0.05); rare variants with MAF = (0.002, 0.01); and ultra-rare variants with MAF = (2e-4, 0.002).



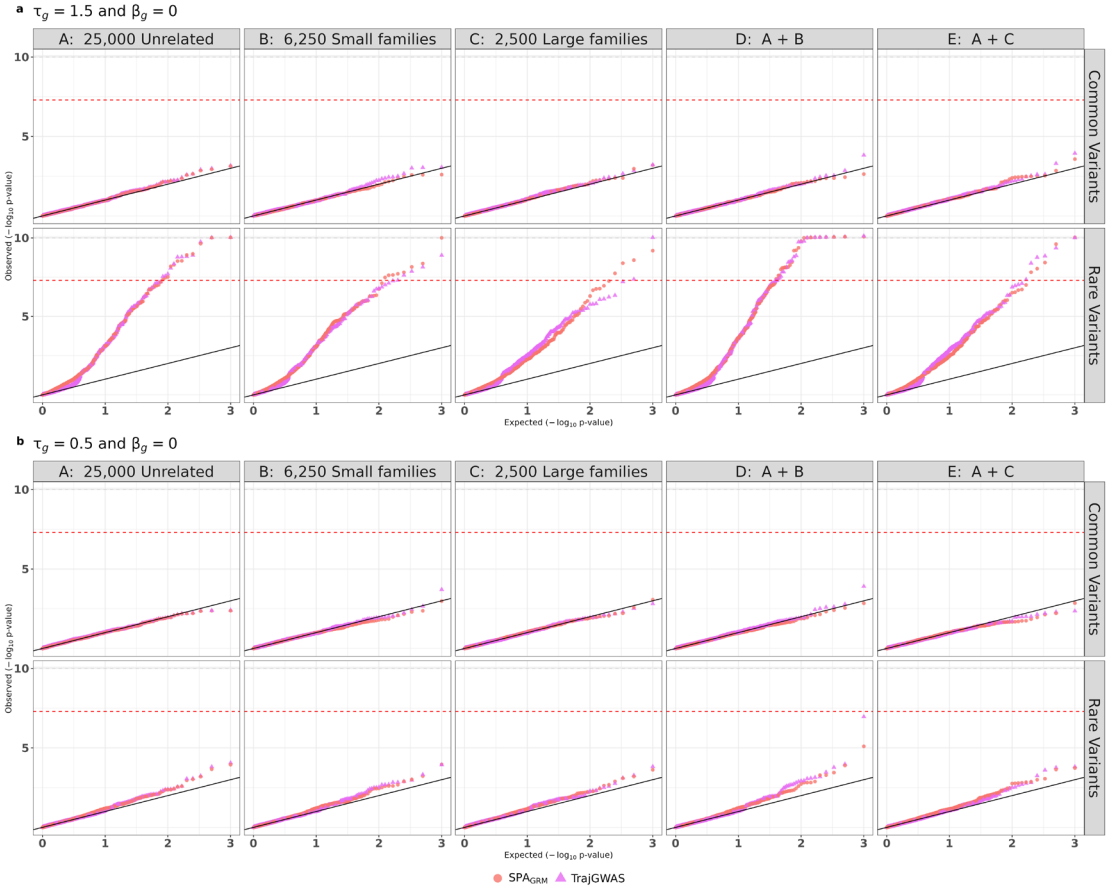
Supplementary Figure 13. Distribution of chi-square statistics of SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, and TrajGWAS methods in scenario 3 (i.e., $\boldsymbol{\beta_g = 0}$ and $\boldsymbol{\tau_g \neq 0}$). (**a**) and (**b**) are for empirical power for testing $\beta_g = 0$ and $\tau_g = 0$ at common variants and rare variants, respectively. From left to right, the plots considered five relatedness scenarios of Dataset A: 25,000 unrelated subjects (all were unrelated and used in TrajGWAS analysis); Dataset B: 25,000 related subjects from 4-member families (50% were unrelated and used in TrajGWAS analysis); Dataset C: 25,000 related subjects from 10-member families (40% were unrelated and used in TrajGWAS analysis); Dataset D: a combination of Dataset A and B (75% were unrelated and used in TrajGWAS analysis); Dataset E: a combination of Dataset A and C (70% were unrelated and used in TrajGWAS analysis). Genetic effect sizes were set to $-\log10(\text{MAF}) \times 0.1$ for common variants and $-\log10(\text{MAF}) \times 0.02$ for rare variants, with an additional multiplier of 0 for β_g and 1.5 for τ_g. The chi-square statistics less than 3.84 (i.e. p values > 0.05) were filtered out. The black dots in the box plot represent the mean, while the box indicates the interquartile range (IQR). The whiskers extend to 1.5 times the IQR from the quartiles. The black dashed line represents the mean chi-squared values of TrajGWAS analysis results, and the grey dashed line represents the chi-squared statistic corresponding to the genome-wide significance ($P = 5 \times 10^{-8}$). Common variants with MAF = (0.05, 0.5); and rare variants with MAF = (2e-4, 0.05). MAF, minor allele frequency. 80 and 30 are breakpoints for the different scales of the y-axis in subplots (**a**) and (**b**), respectively. Mean chi-square statistics are displayed in **Supplementary Table 1**, and empirical power estimates are displayed in **Supplementary Table 2**.


a Testing for $\beta_g = 0$
A: 25,000 Unrelated
B: 6,250 Small families
C: 2,500 Large families
D: A + B
E: A + C
Common Variants
Rare Variants
Chi-squared Values
b Testing for $\tau_g = 0$
SPA_{GRM}
$SPA_{GRM(INT)}$
$SPA_{GRM(CCT)}$
TrajGWAS

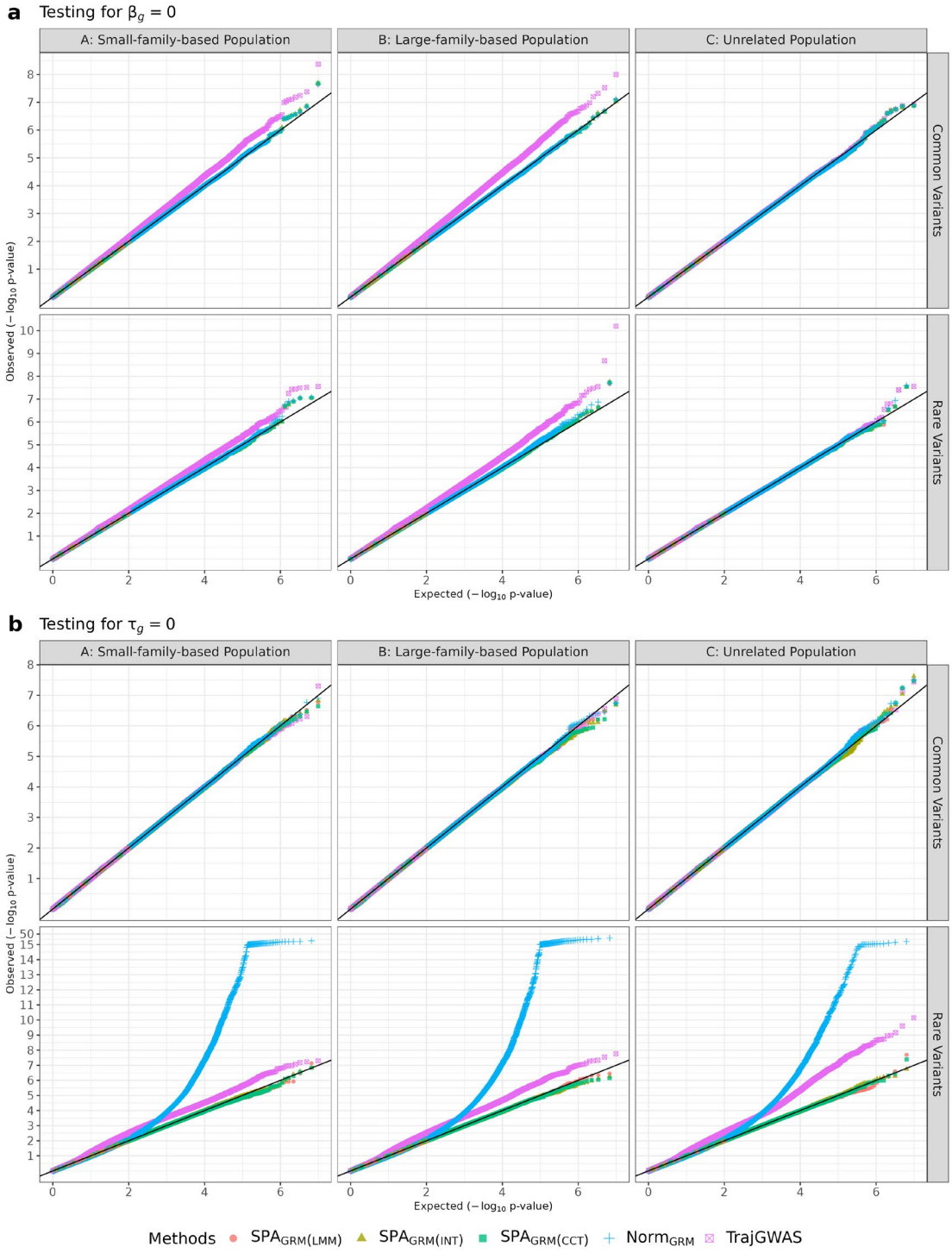
Supplementary Figure 14. Distribution of chi-square statistics of SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, and TrajGWAS methods in scenario 4 (i.e., $\boldsymbol{\beta_g \neq 0}$ and $\boldsymbol{\tau_g = 0}$). (**a**) and (**b**) are for empirical power for testing $\beta_g = 0$ and $\tau_g = 0$ at common variants and rare variants, respectively. From left to right, the plots considered five relatedness scenarios of Dataset A: 25,000 unrelated subjects (all were unrelated and used in TrajGWAS analysis); Dataset B: 25,000 related subjects from 4-member families (50% were unrelated and used in TrajGWAS analysis); Dataset C: 25,000 related subjects from 10-member families (40% were unrelated and used in TrajGWAS analysis); Dataset D: a combination of Dataset A and B (75% were unrelated and used in TrajGWAS analysis); Dataset E: a combination of Dataset A and C (70% were unrelated and used in TrajGWAS analysis). Genetic effect sizes were set to $-\log10(\text{MAF}) \times 0.1$ for common variants and $-\log10(\text{MAF}) \times 0.02$ for rare variants, with an additional multiplier of 1 for β_g and 0 for τ_g. The chi-square statistics less than 3.84 (i.e. p values > 0.05) were filtered out. The black dots in the box plot represent the mean, while the box indicates the interquartile range (IQR). The whiskers extend to 1.5 times the IQR from the quartiles. The black dashed line represents the mean chi-squared values of TrajGWAS analysis results, and the grey dashed line represents the chi-squared statistic corresponding to the genome-wide significance ($P = 5 \times 10^{-8}$). Common variants with MAF = (0.05, 0.5); and rare variants with MAF = (2e-4, 0.05). MAF, minor allele frequency. 80 and 30 are breakpoints for the different scales of the y-axis in subplots (**a**) and (**b**), respectively. Mean chi-square statistics are displayed in **Supplementary Table 1**, and empirical power estimates are displayed in **Supplementary Table 2**.


a Testing for $\beta_g = 0$
A: 25,000 Unrelated
B: 6,250 Small families
C: 2,500 Large families
D: A + B
E: A + C
Common Variants
Rare Variants
Chi-squared Values
b Testing for $\tau_g = 0$
SPA$_{GRM}$
SPA$_{GRM(INT)}$
SPA$_{GRM(CCT)}$
TrajGWAS

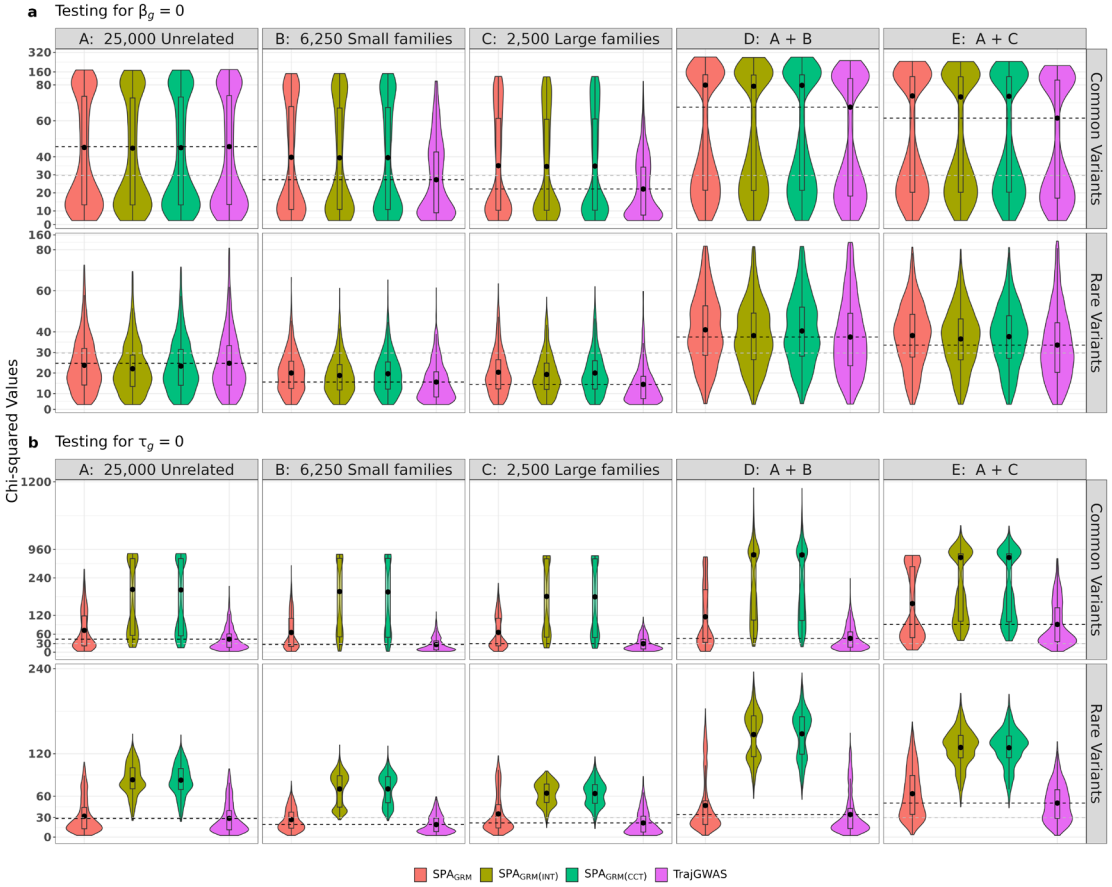
Supplementary Figure 15. Quantile-quantile (QQ) plots of p values from TrajGWAS and SPA$_{GRM}$ methods in scenario 3 (i.e., $\boldsymbol{\beta_g = 0}$ and $\boldsymbol{\tau_g \neq 0}$) when the true effect size $\boldsymbol{\tau_g}$ is smaller. Subplots a and b present QQ plots of (**a**) the true $\tau_g = 1.5$ & $\beta_g = 0$ and (**b**) the true $\tau_g = 0.5$ & $\beta_g = 0$. From left to right, the plots considered five relatedness scenarios of Dataset A: 25,000 unrelated subjects (all were unrelated and used in TrajGWAS analysis); Dataset B: 25,000 related subjects from 4-member families (50% were unrelated and used in TrajGWAS analysis); Dataset C: 25,000 related subjects from 10-member families (40% were unrelated and used in TrajGWAS analysis); Dataset D: a combination of Dataset A and B (75% were unrelated and used in TrajGWAS analysis); Dataset E: a combination of Dataset A and C (70% were unrelated and used in TrajGWAS analysis). Common variants with MAF = (0.05, 0.5); and rare variants with MAF = (2e-4, 0.05). MAF, minor allele frequency.



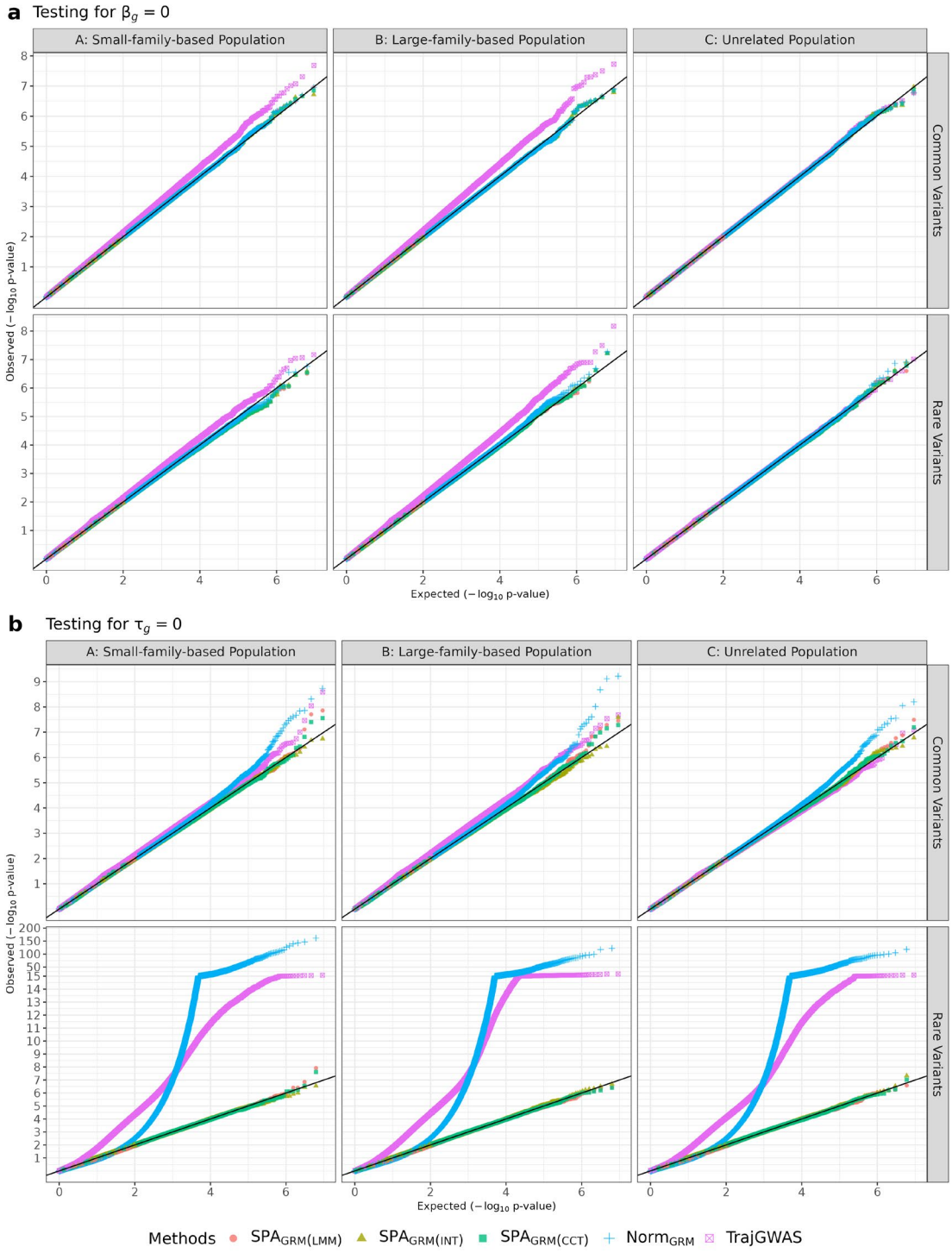
Supplementary Figure 16. Quantile-quantile (QQ) plots of p values from SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, Norm$_{GRM}$, and TrajGWAS methods in scenario 1 (i.e., $\beta_g = 0$ and $\tau_g = 0$) with no random effects on the WS variability. Top and bottom plots represent p values for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. From left to right, the plots consider three family relatedness of small-family-based dataset; large-family-based dataset; and unrelated dataset. Longitudinal traits were simulated 100 times, and association tests were conducted on 100,000 common and rare variants, respectively. Totally 1×10^7 replications were conducted in each scenario.



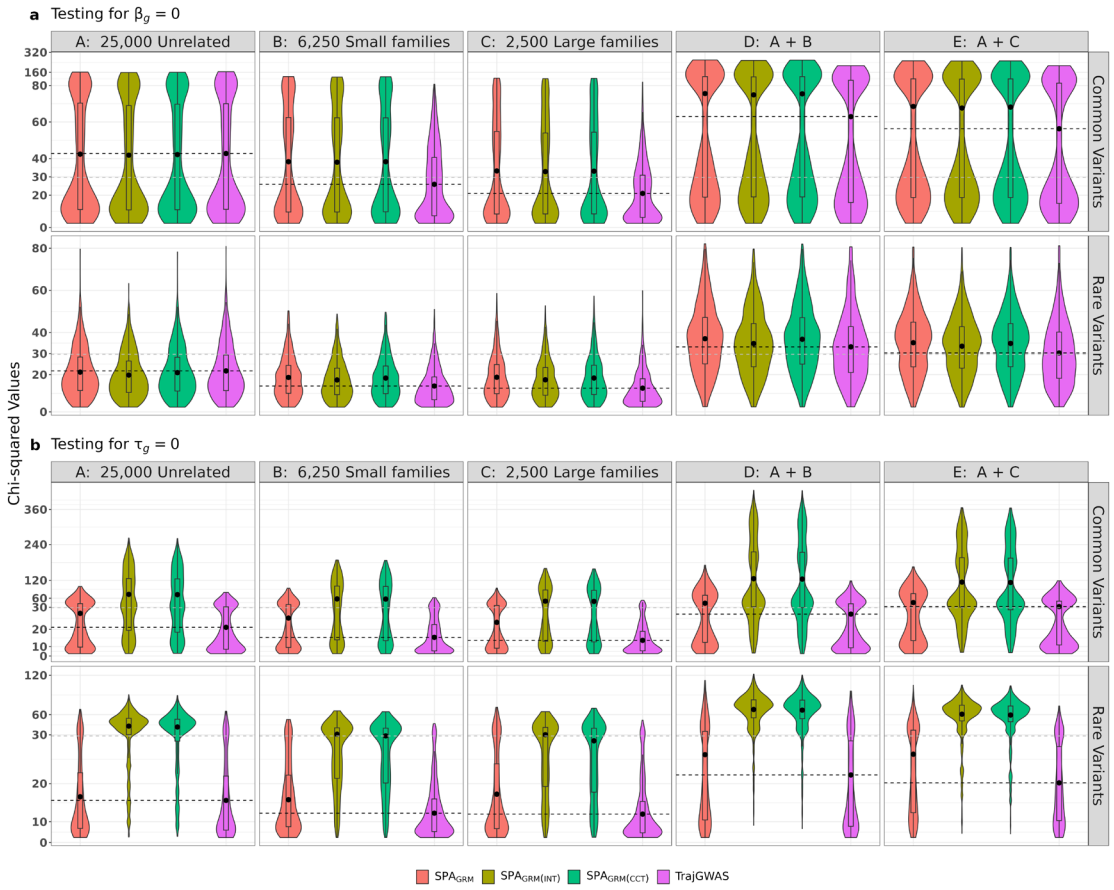
Supplementary Figure 17. Distribution of chi-square statistics of SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, and TrajGWAS methods in scenario 2 (i.e., $\beta_g \neq 0$ and $\tau_g \neq 0$) with no random effects on the WS variability. Top and bottom plots are for empirical power for testing $\beta_g = 0$ and $\tau_g = 0$ at common variants and rare variants, respectively. From left to right, the plots considered five relatedness scenarios of Dataset A: 25,000 unrelated subjects (all were unrelated and used in TrajGWAS analysis); Dataset B: 25,000 related subjects from 4-member families (50% were unrelated and used in TrajGWAS analysis); Dataset C: 25,000 related subjects from 10-member families (40% were unrelated and used in TrajGWAS analysis); Dataset D: a combination of Dataset A and B (75% were unrelated and used in TrajGWAS analysis); Dataset E: a combination of Dataset A and C (70% were unrelated and used in TrajGWAS analysis). Genetic effect sizes were set to $-\log10(\text{MAF}) \times 0.1$ for common variants and $-\log10(\text{MAF}) \times 0.02$ for rare variants, with an additional multiplier of 1 for β_g and 1.5 for τ_g. The chi-square statistics less than 3.84 (i.e. p values > 0.05) were filtered out. The black dots in the box plot represent the mean, while the box indicates the interquartile range (IQR). The whiskers extend to 1.5 times the IQR from the quartiles. The black dashed line represents the mean chi-squared values of TrajGWAS analysis results, and the grey dashed line represents the chi-squared statistic corresponding to the genome-wide significance $(P = 5 \times 10^{-8})$. Common variants with MAF = (0.05, 0.5); and rare variants with MAF = (2e-4, 0.05). MAF, minor allele frequency.



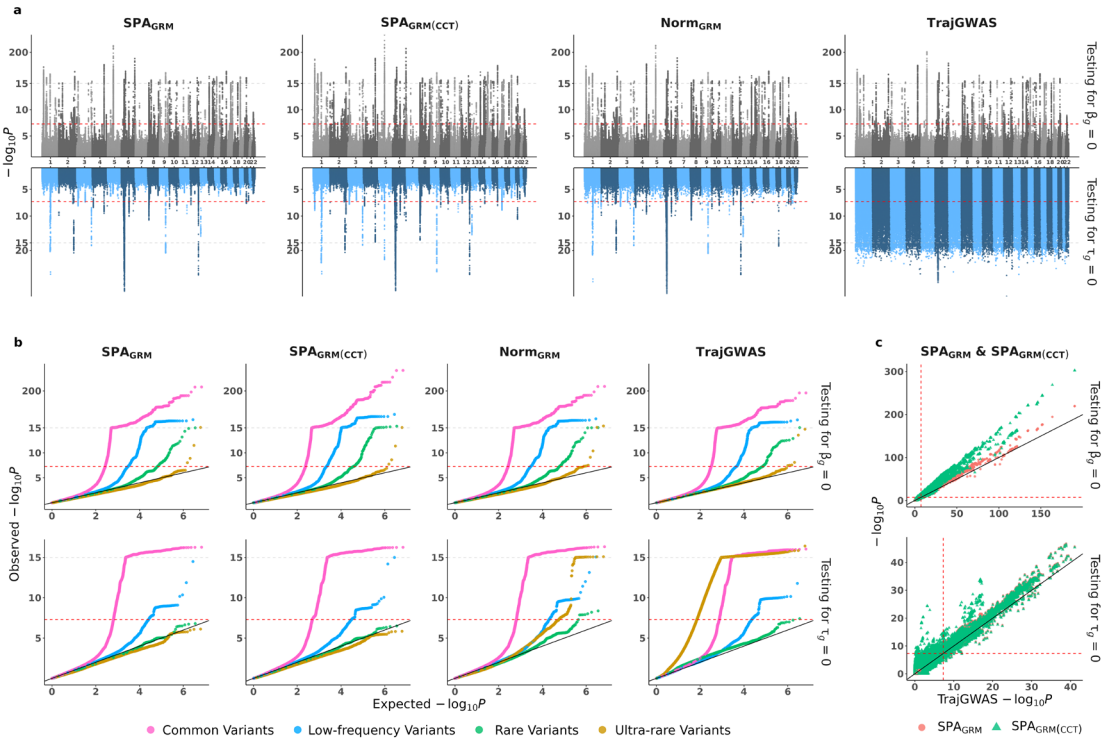
Supplementary Figure 18. Quantile-quantile (QQ) plots of p values from SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, Norm$_{GRM}$, and TrajGWAS methods in scenario 1 (i.e., $\beta_g = 0$ and $\tau_g = 0$) when use the inverse-gamma model to simulate within-subject (WS) random effects. Top and bottom plots represent p values for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. From left to right, the plots consider three family relatedness of small-family-based dataset; large-family-based dataset; and unrelated dataset. Longitudinal traits were simulated 100 times, and association tests were conducted on 100,000 common and rare variants, respectively. Totally 1×10^7 replications were conducted in each scenario.



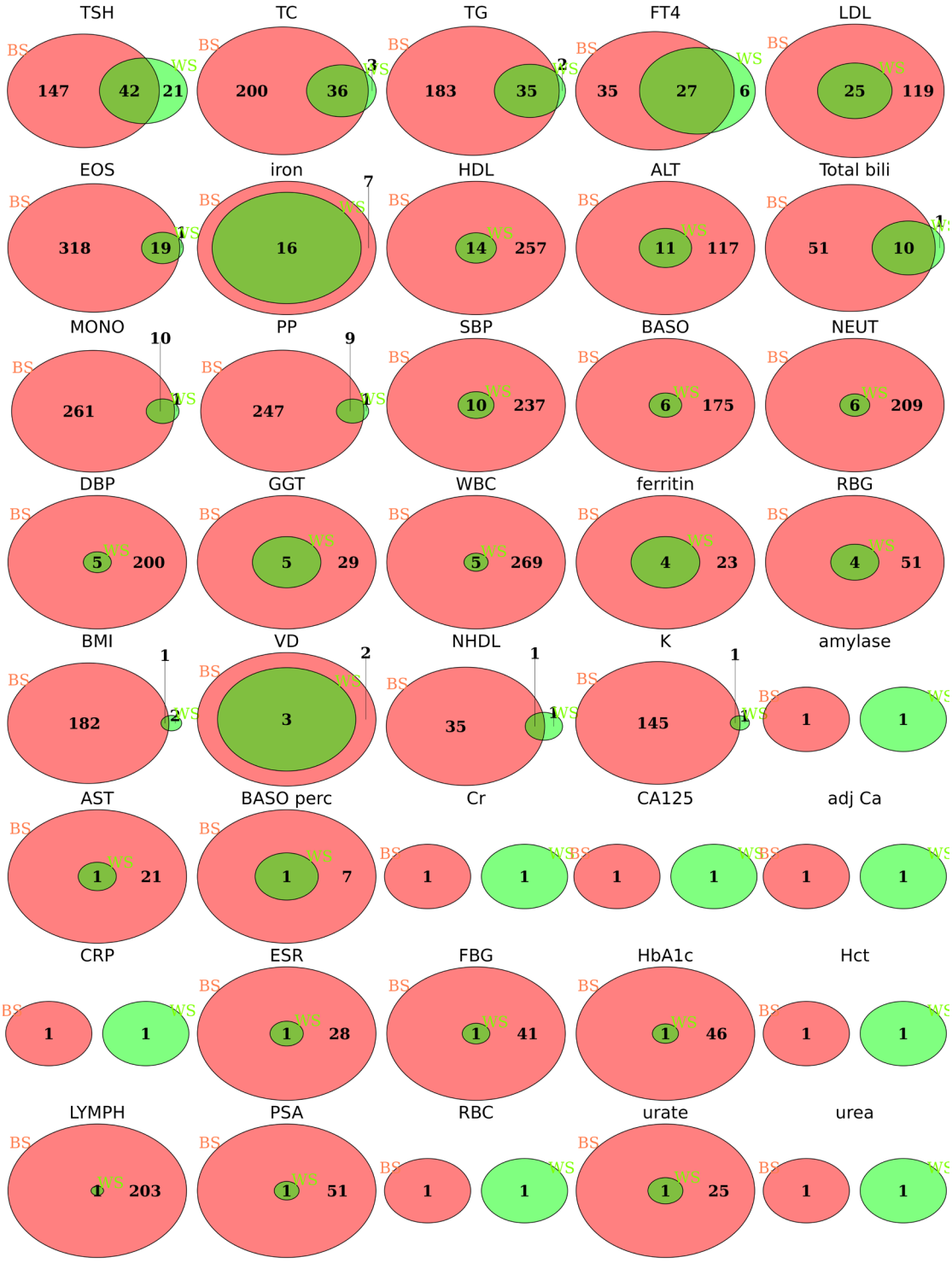
Supplementary Figure 19. Distribution of chi-square statistics of SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, and TrajGWAS methods in scenario 2 (i.e., $\beta_g \neq 0$ and $\tau_g \neq 0$) when use the inverse-gamma model to simulate within-subject (WS) random effects. Top and bottom plots are for empirical power for testing $\beta_g = 0$ and $\tau_g = 0$ at common variants and rare variants, respectively. From left to right, the plots considered five relatedness scenarios of Dataset A: 25,000 unrelated subjects (all were unrelated and used in TrajGWAS analysis); Dataset B: 25,000 related subjects from 4-member families (50% were unrelated and used in TrajGWAS analysis); Dataset C: 25,000 related subjects from 10-member families (40% were unrelated and used in TrajGWAS analysis); Dataset D: a combination of Dataset A and B (75% were unrelated and used in TrajGWAS analysis); Dataset E: a combination of Dataset A and C (70% were unrelated and used in TrajGWAS analysis). Genetic effect sizes were set to $-\log10(\text{MAF}) \times 0.1$ for common variants and $-\log10(\text{MAF}) \times 0.02$ for rare variants, with an additional multiplier of 1 for β_g and 1.5 for τ_g. The chi-square statistics less than 3.84 (i.e. p values > 0.05) were filtered out. The black dots in the box plot represent the mean, while the box indicates the interquartile range (IQR). The whiskers extend to 1.5 times the IQR from the quartiles. The black dashed line represents the mean chi-squared values of TrajGWAS analysis results, and the grey dashed line represents the chi-squared statistic corresponding to the genome-wide significance ($P = 5 \times 10^{-8}$). Common variants with MAF = (0.05, 0.5); and rare variants with MAF = (2e-4, 0.05). MAF, minor allele frequency.



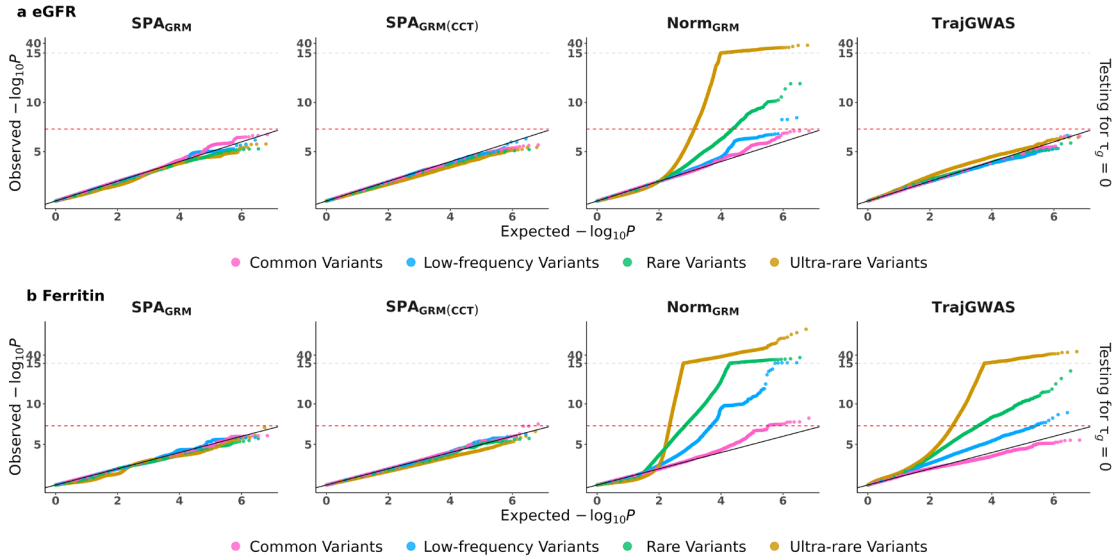
Supplementary Figure 20. Manhattan plots and quantile-quantile (QQ) plots of GWAS results for serum thyroid stimulating hormone. a) mirror Manhattan plots of GWAS results of SPA_{GRM}, $SPA_{GRM(CCT)}$, $Norm_{GRM}$, and TrajGWAS; **b**) QQ plots in which genetic variants are grouped based on minor allele frequency (MAF): common variants with MAF in (0.05, 0.5), low-frequency variants with MAF in (0.01, 0.05), rare variants with MAF in (0.002, 0.01), and ultra-rare variants with MAF in (2e-4, 0.002); **c**) scatterplots comparing SPA_{GRM} and $SPA_{GRM(CCT)}$ with TrajGWAS on common variants (i.e. MAF > 0.05). Across all subplots, upper and lower are results for testing mean profile ($\beta_g = 0$) and WS variability ($\tau_g = 0$), respectively. The red dashed line represents p value of $(P = 5 \times 10^{-8})$, and the grey dashed line represents the breakpoint for the different scales of the y-axis $(-log_{10}P = 15)$. For thyroid stimulating hormone (TSH), 145,519 subjects were used for $Norm_{GRM}$, SPA_{GRM}, and $SPA_{GRM(CCT)}$ analysis, of which 121,916 (83.8%) were used for TrajGWAS. An average of 4.99 TSH records were measured per subject. P value is calculated using two-sided score tests.



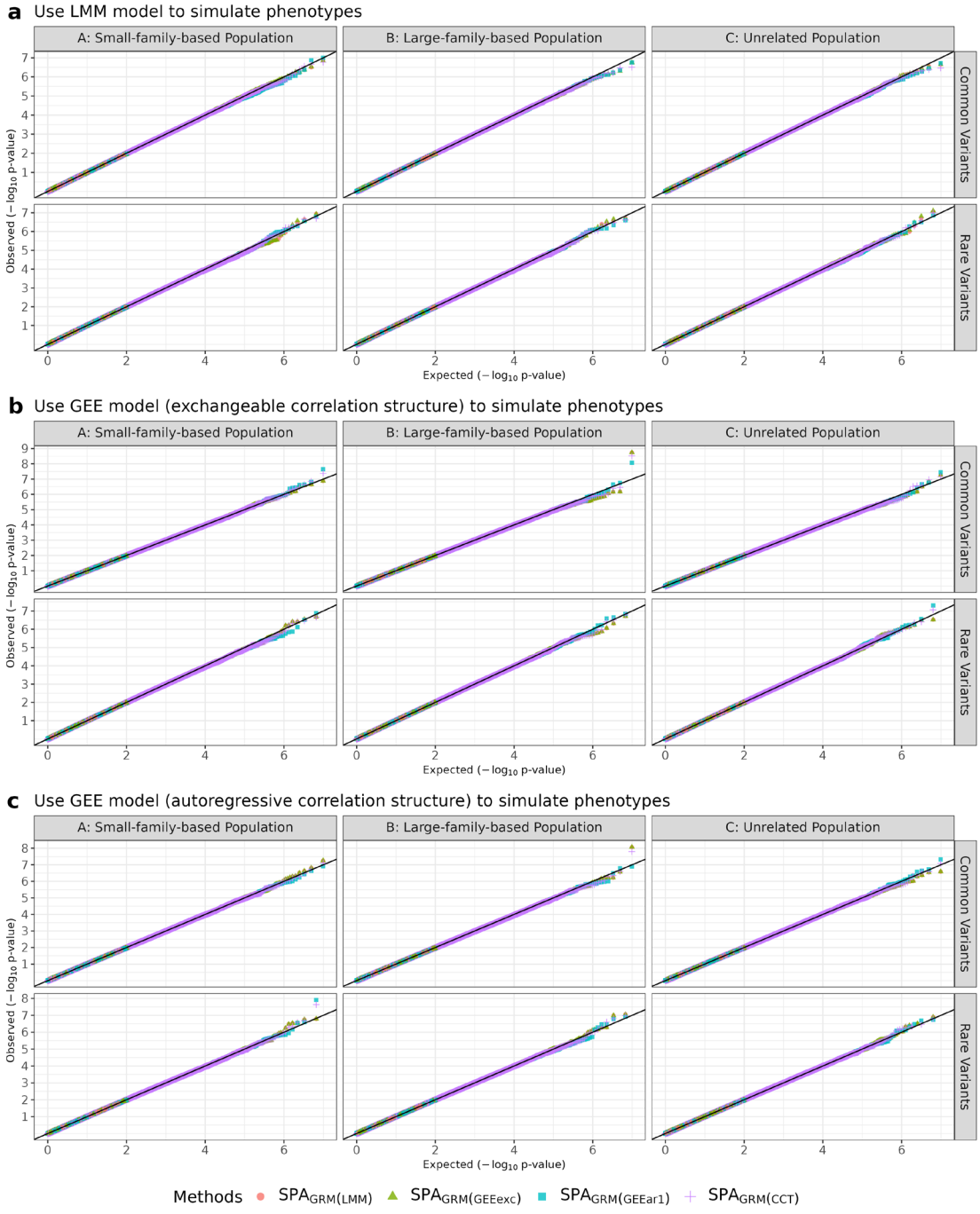
Supplementary Figure 21. Venn diagrams of genes affecting the mean level and WS variability of 40 longitudinal traits. "BS" is the abbreviation for between-subject variability, which represents the mean level; "WS" is the abbreviation for within-subject variability. From top to bottom, from left to right, 40 longitudinal traits are sorted by the number of loci associated with WS variability of these traits: TSH, thyroid stimulating hormone; TC, serum cholesterol; TG, serum triglyceride; FT4, free thyroxine; LDL, Low density lipoprotein cholesterol; EOS, eosinophil count; iron, serum iron; HDL, high density lipoprotein cholesterol; ALT, alanine aminotransferase; Total bili, total bilirubin; MONO, monocyte count; PP, pulse pressure; SBP, systolic blood pressure; BASO, basophil count; NEUT, neutrophill count; DBP, diastolic blood pressure; GGT, gamma-glutamyl transferase; WBC, white blood cell count; ferritin, ferritin; RBG, random blood glucose; BMI, body mass index; VD, Vitamin D; NHDL, non-high density lipoprotein cholesterol; K, serum potassium; amylase, amylase; AST, aspartate aminotransferase; BASO perc, basophil percentage; Cr, blood creatinine; CA125, carbohydrate antigen 125; adj Ca, adjusted calcium; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FBG, fasting blood glucose; HbA1c, glycated haemoglobin; Hct, haematocrit percentage; LYMPH, lymphocyte count; PSA, prostate-specific antigen; RBC, red blood cell count; urate, serum urate; urea, serum urea. There is additionally one locus affect the WS variability of FT3 (free triiodothyronine), but not affect its mean level.


TSH
BS
WS
147
42
21
TC
BS
WS
200
36
3
TG
BS
WS
183
35
2
FT4
BS
WS
35
27
6
LDL
BS
WS
25
119
EOS
BS
WS
318
19
1
iron
BS
WS
16
7
HDL
BS
WS
14
257
ALT
BS
WS
11
117
Total bili
BS
WS
51
10
1
MONO
BS
WS
261
1
10
PP
BS
WS
247
1
9
SBP
BS
WS
10
237
BASO
BS
WS
6
175
NEUT
BS
WS
6
209
DBP
BS
WS
5
200
GGT
BS
WS
5
29
WBC
BS
WS
5
269
ferritin
BS
WS
4
23
RBG
BS
WS
4
51
BMI
BS
WS
182
2
1
VD
BS
WS
3
2
NHDL
BS
WS
35
1
1
K
BS
WS
145
1
1
amylase
BS
WS
1
1
AST
BS
WS
1
21
BASO perc
BS
WS
1
7
Cr
BS
WS
1
1
CA125
BS
WS
1
1
adj Ca
BS
WS
1
1
CRP
BS
WS
1
1
ESR
BS
WS
1
28
FBG
BS
WS
1
41
HbA1c
BS
WS
1
46
Hct
BS
WS
1
1
LYMPH
BS
WS
1
203
PSA
BS
WS
1
51
RBC
BS
WS
1
1
urate
BS
WS
1
25
urea
BS
WS
1
1

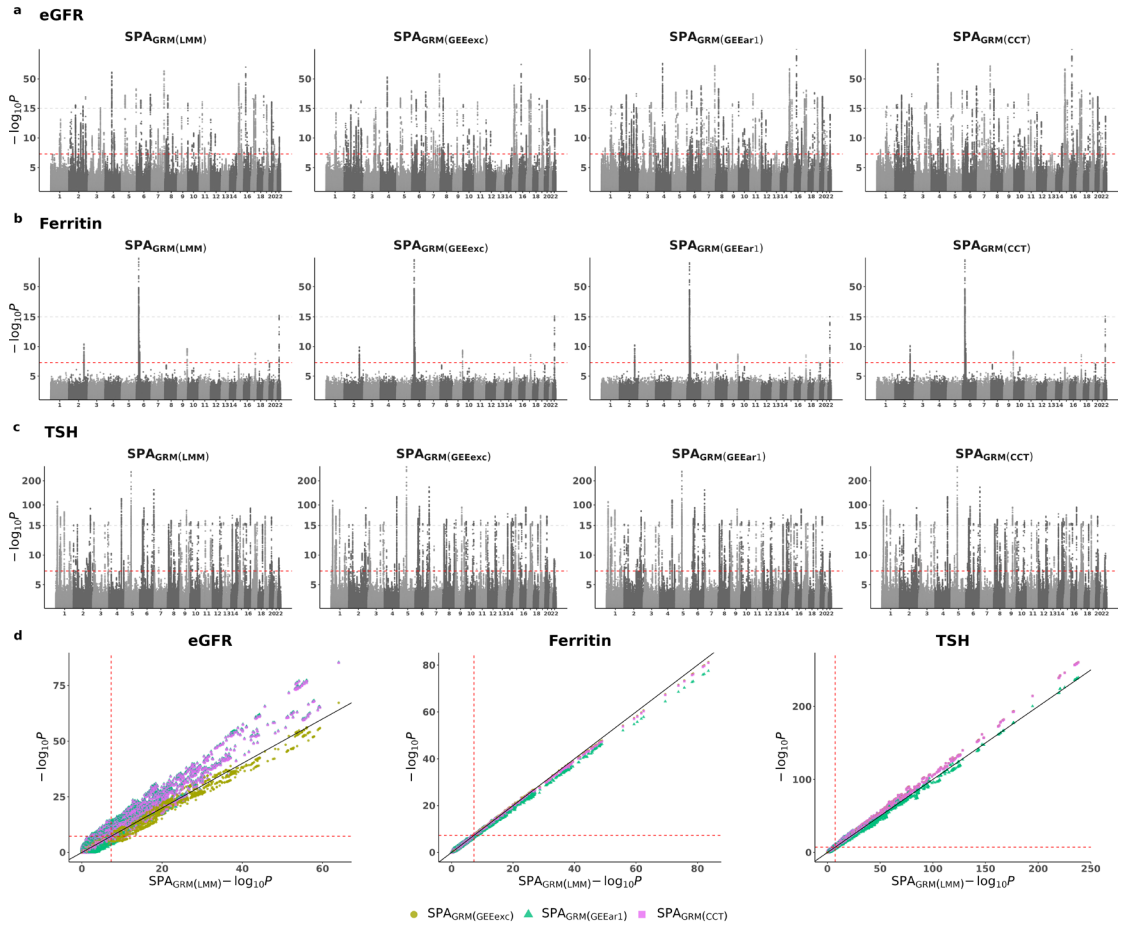
Supplementary Figure 22. Quantile-quantile (QQ) plots of p values from SPA$_{GRM}$, SPA$_{GRM(CCT)}$, Norm$_{GRM}$, and TrajGWAS methods when we shuffled the order of the residuals from fitted null models of eGFR and serum ferritin. Subplots **a** and **b** present results of scenarios we shuffled the order of the residuals for testing $\tau_g = 0$ from fitted null models of eGFR and ferritin, respectively. Genetic variants are grouped based on minor allele frequency (MAF): common variants with MAF in (0.05, 0.5), low-frequency variants with MAF in (0.01, 0.05), rare variants with MAF in (0.002, 0.01), and ultra-rare variants with MAF in (2e-4, 0.002).



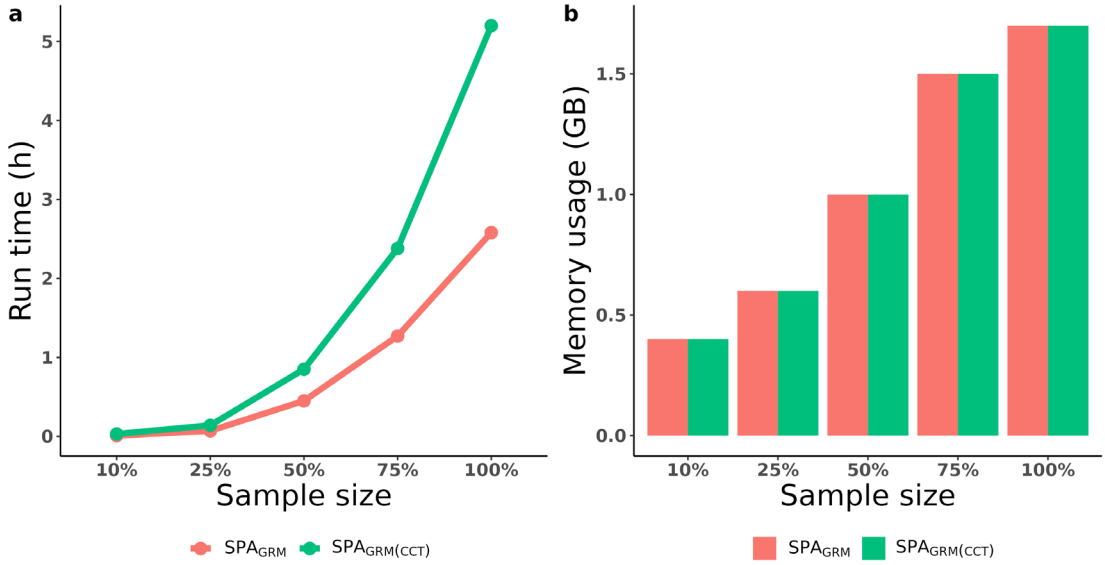
Supplementary Figure 23. Quantile-quantile (QQ) plots of p values of $SPA_{GRM(LMM)}$, $SPA_{GRM(GEEexc)}$, $SPA_{GRM(GEEar1)}$, and $SPA_{GRM(CCT)}$ methods for longitudinal traits using different generative models. From top to bottom plots, longitudinal traits were simulated using (**a**) linear mixed model (LMM); (**b**) generalized estimation equations (GEE) model with exchangeable working correlation structure and (**c**) GEE model with autoregressive working correlation structure. From left to right, the plots consider three family relatedness of small-family-based dataset; large-family-based dataset; and unrelated dataset. In each scenario, phenotypes were replicated 100 times and association tests were conducted on 100,000 common variants and 100,000 rare variants, respectively. Totally, $1 \times 10^7 = 100 \times 10^5$ tests were conducted in each scenario.



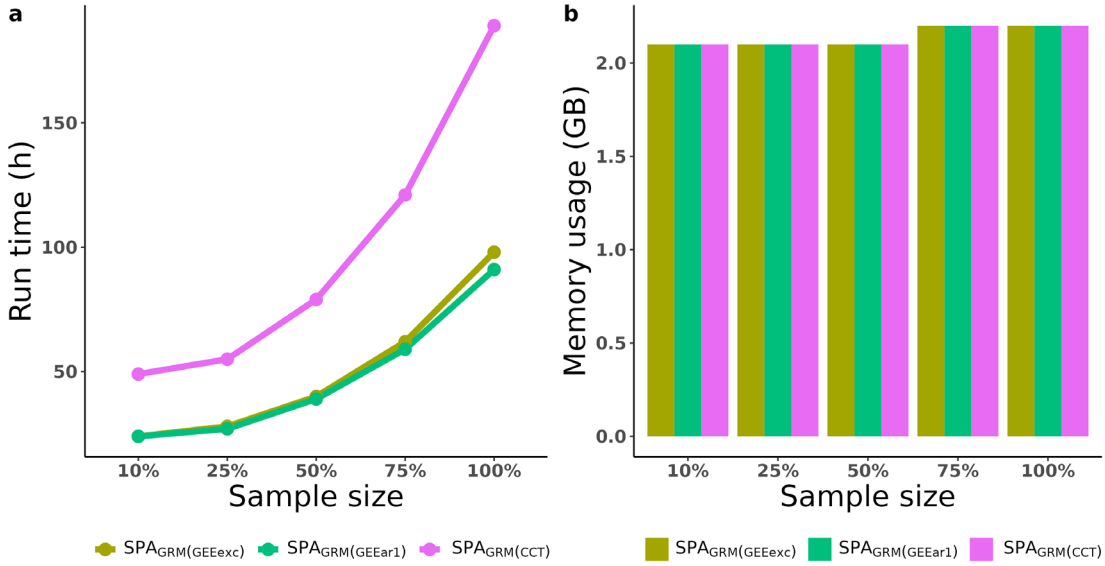
Supplementary Figure 24. Manhattan plots and scatter plots of GWAS results from $SPA_{GRM(LMM)}$, $SPA_{GRM(GEEexc)}$, $SPA_{GRM(GEEar1)}$, and $SPA_{GRM(CCT)}$ methods. (**a**)-(**c**) are Manhattan plots of GWAS results of three longitudinal traits: eGFR, ferritin, and TSH, respectively. (**d**) is scatter plots of GWAS results from $SPA_{GRM(LMM)}$, $SPA_{GRM(GEEexc)}$, $SPA_{GRM(GEEar1)}$, and $SPA_{GRM(CCT)}$ methods. In subplot d, x-axis represents $-log_{10}P$ of $SPA_{GRM(LMM)}$ results, and y-axis represents $-log_{10}P$ of $SPA_{GRM(GEEexc)}$, $SPA_{GRM(GEEar1)}$, and $SPA_{GRM(CCT)}$ results. The red dashed line in subplot a-d represents the genome-wide significance ($P = 5 \times 10^{-8}$), and the grey dashed line in (**a**)-(**c**) represents the breakpoint for the different scales of the y-axis ($-log_{10}P = 15$). Dots in (**d**) can be overlapped because $SPA_{GRM(CCT)}$ performs very similarly to $SPA_{GRM(GEEexc)}$ or $SPA_{GRM(GEEar1)}$.



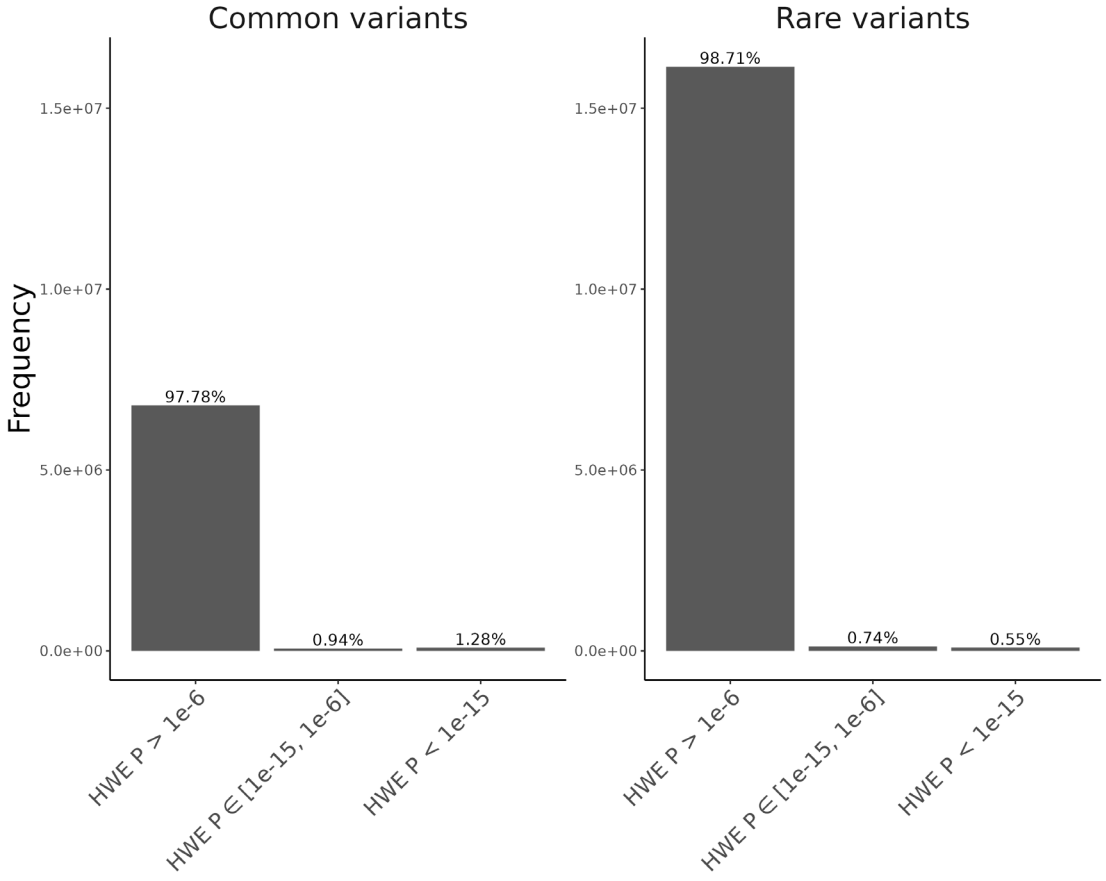
Supplementary Figure 25. Run time and memory usage of SPA_{GRM}, and $SPA_{GRM(CCT)}$ methods in step 0. (**a**) Runtime. The x axis represents the sample size and the y axis represents the run time in hourly units. (**b**) Memory usage. The x axis represents the sample size and the y axis represents the memory usage in GB units. $SPA_{GRM(CCT)}$ is the combination of SPA_{GRM} and $SPA_{GRM(INT)}$. Analysis was performed on 10%, 25%, 50%, 75%, and 100% of 175,443 individuals with longitudinal BMI measurements. All analyses were conduct on a CPU model of Intel(R) Xeon(R) Gold 6342 CPU 2.80GHz.



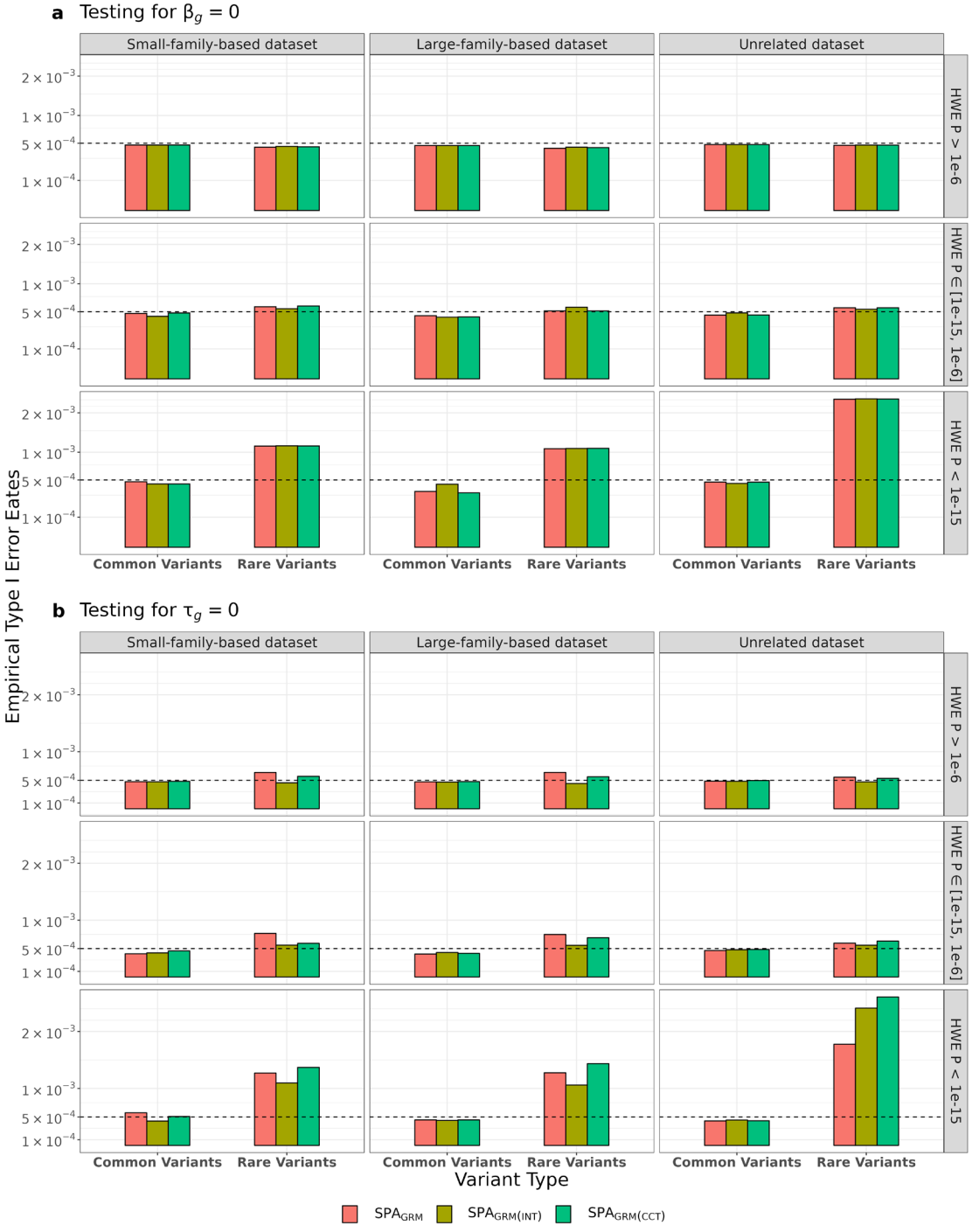
Supplementary Figure 26. Run time and memory usage of $SPA_{GRM(GEEexc)}$, $SPA_{GRM(GEEar1)}$, and $SPA_{GRM(CCT)}$ methods in step 2. (**a**) Runtime. The x axis represents the sample size and the y axis represents the run time in hourly units. (**b**) Memory usage. The x axis represents the sample size and the y axis represents the memory usage in GB units. $SPA_{GRM(CCT)}$ is the combination of $SPA_{GRM(GEEexc)}$ and $SPA_{GRM(GEEar1)}$. Analysis was performed on 10%, 25%, 50%, 75%, and 100% of 175,443 individuals with longitudinal BMI measurements. The association tests were conducted on 23 million imputed variants. All analyses were conduct on a CPU model of Intel(R) Xeon(R) Gold 6342 CPU 2.80GHz.



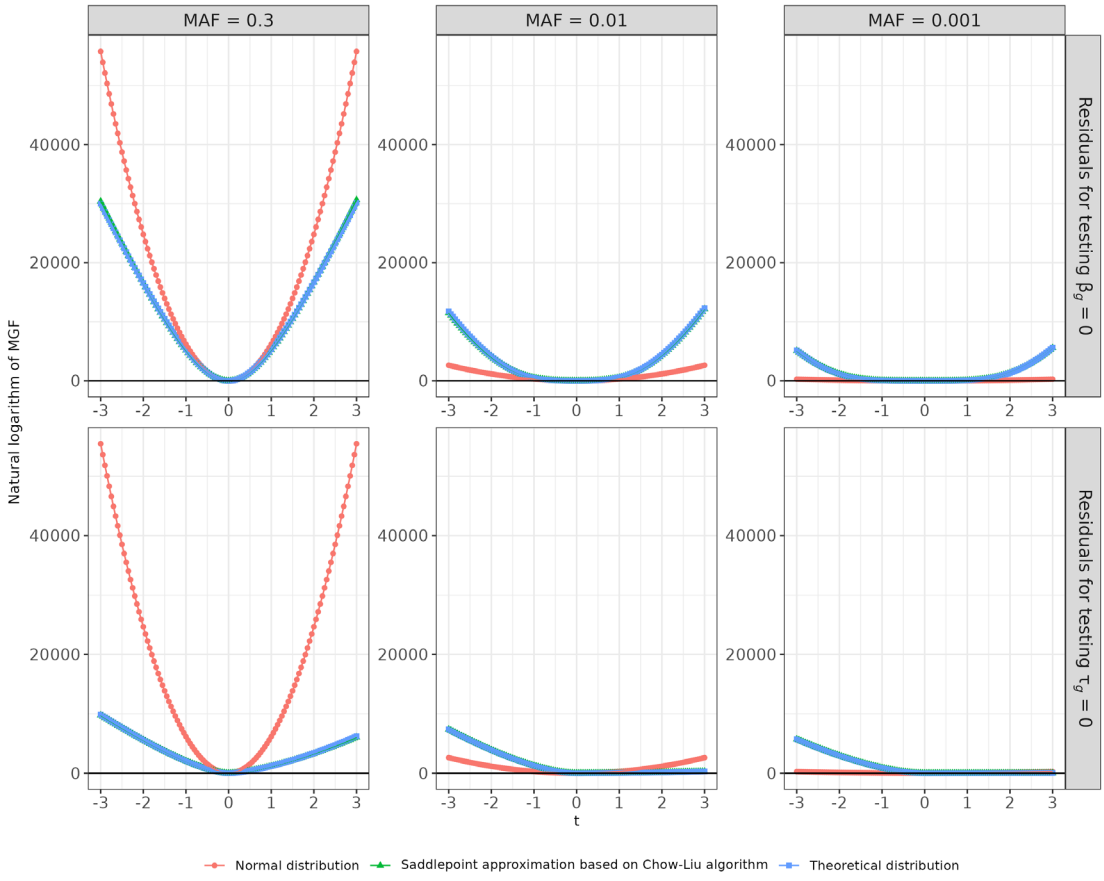
Supplementary Figure 27. Histograms of Hardy-Weinberg Equilibrium (HWE) p values for 23 million imputed variants. These 23 million variants were grouped as 6,941,361 common variants (MAF > 0.05) and 16,360,256 rare variants (0.05 > MAF > 2e-4). We divided common and rare variants into three groups according to their HWE p values: HWE P > 1e-6; 1e-6 > HWE P > 1e-15; HWE P < 1e-15, respectively. Each bar represents the proportion of this type of variants. MAF, minor allele frequency.



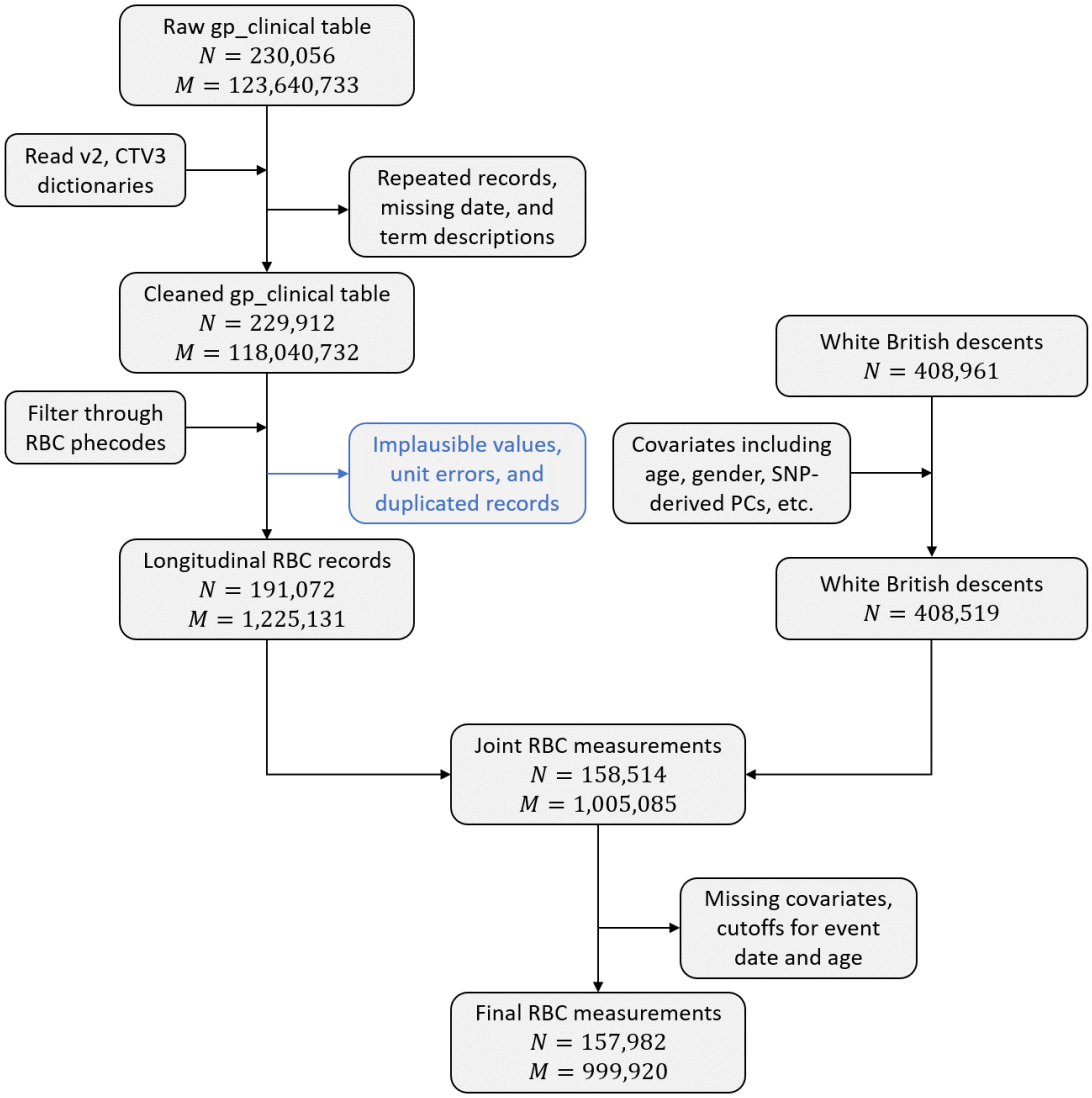
Supplementary Figure 28. Empirical type I error rates of SPA$_{GRM}$-based methods based on 10^9 replications. Top and bottom plots represent empirical type I error rates for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. From left to right, the plots consider three family relatedness of small-family-based dataset; large-family-based dataset; and unrelated dataset. We divided 100,000 common variants and 100,000 rare variants into three groups according to their HWE p values: HWE P > 1e-6; 1e-6 > HWE P > 1e-15; HWE P < 1e-15, respectively. We used the significance level of 5×10^{-4} to make sure a sufficient number of tests violate the null hypothesis to make the empirical type I error rates more accurate. No variants with HWE p value $\in [1e-15, 1e-6]$ violate the null hypothesis at the significance level of 5×10^{-8}.



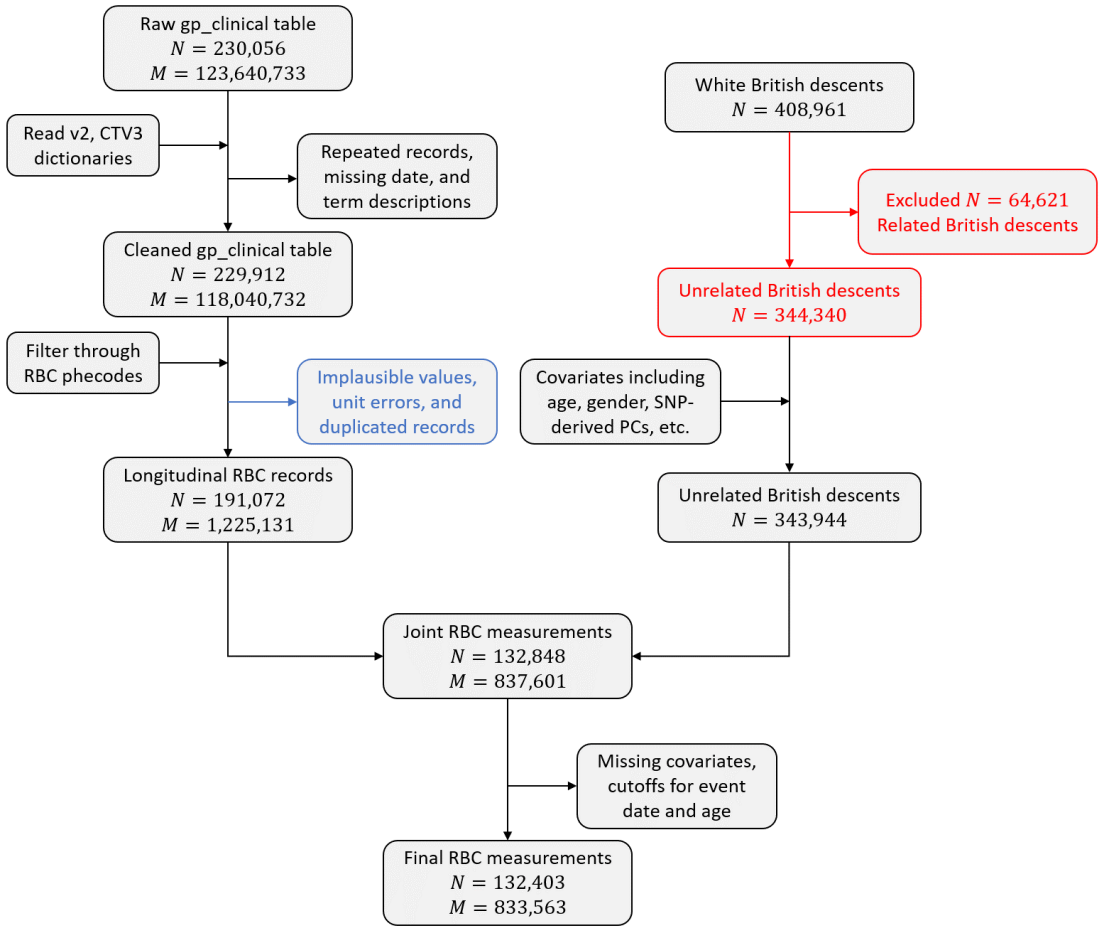
Supplementary Figure 29. Natural logarithm of MGFs using normal distribution approximation, Chow-Liu algorithm approximation, and theoretical distribution. We used pedigree data consisting of 6250 4-member families as showed in **Supplementary Fig. 1**. Genotypes with MAFs of 0.3, 0.01, and 0.001 were considered to represent common and rare variants. We considered residuals for testing $\beta_g = 0$ and $\tau_g = 0$, to mimic balanced and unbalanced phenotype distributions. In each scenario, we compared the natural logarithm of MGFs using these three methods.



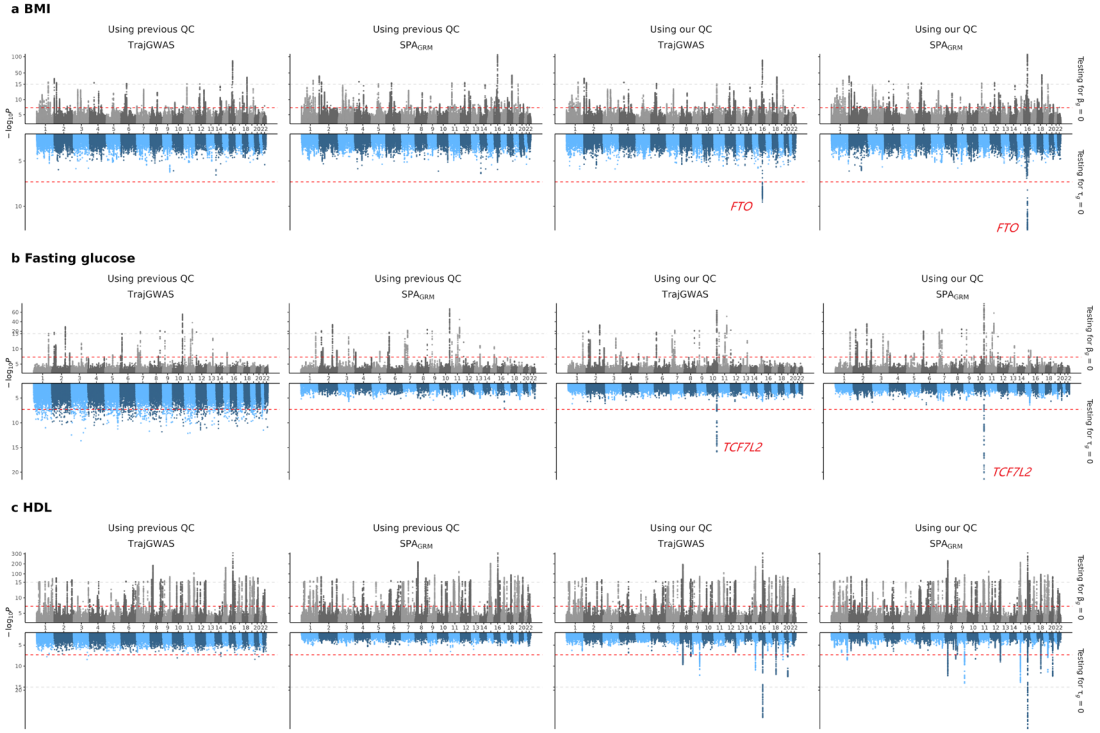
Supplementary Figure 30. Cohort curation for red blood cell in SPAGRM analyses from UK Biobank primary care data



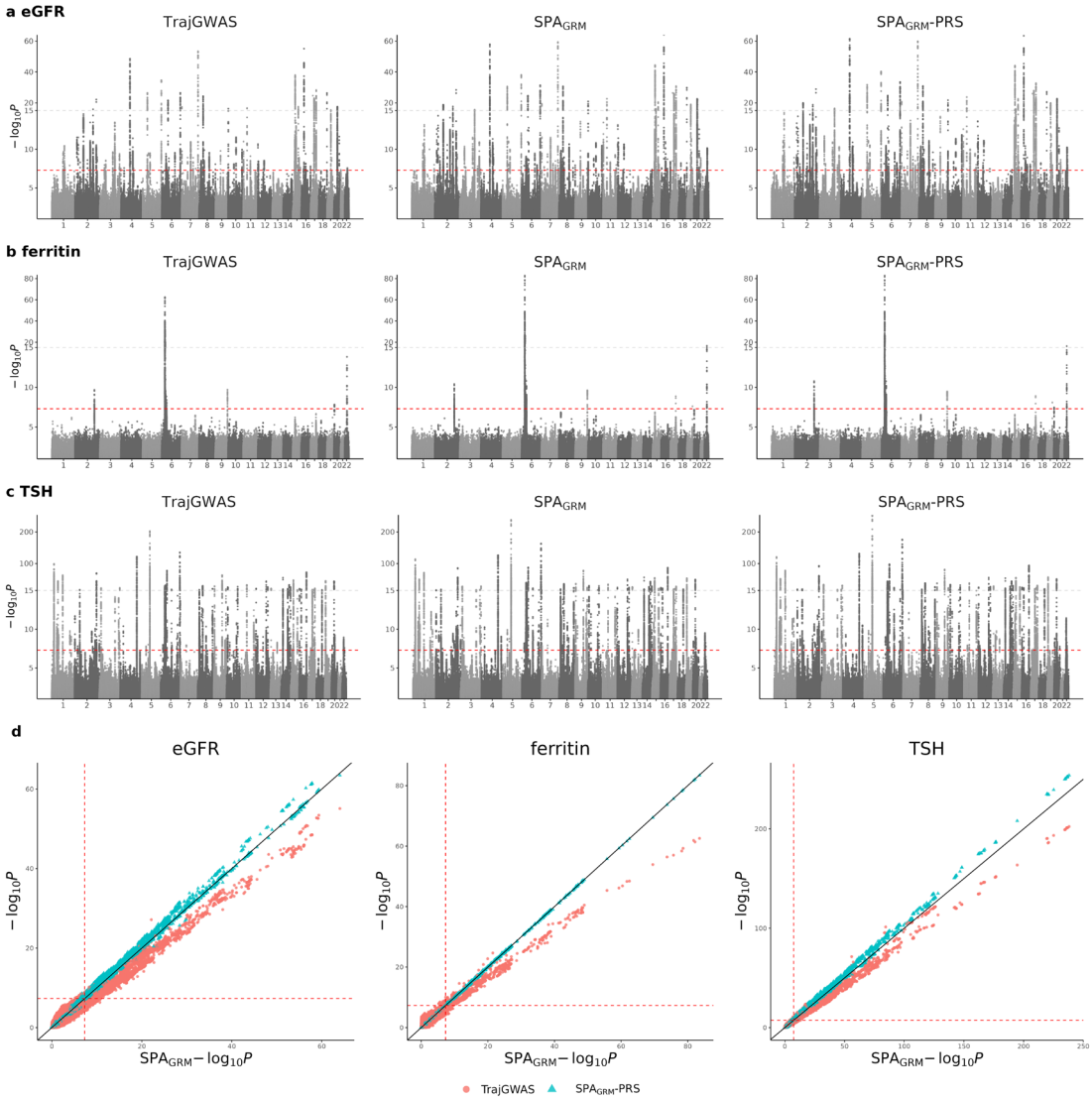
Supplementary Figure 31. Cohort curation for red blood cell in TrajGWAS analyses from UK Biobank primary care data



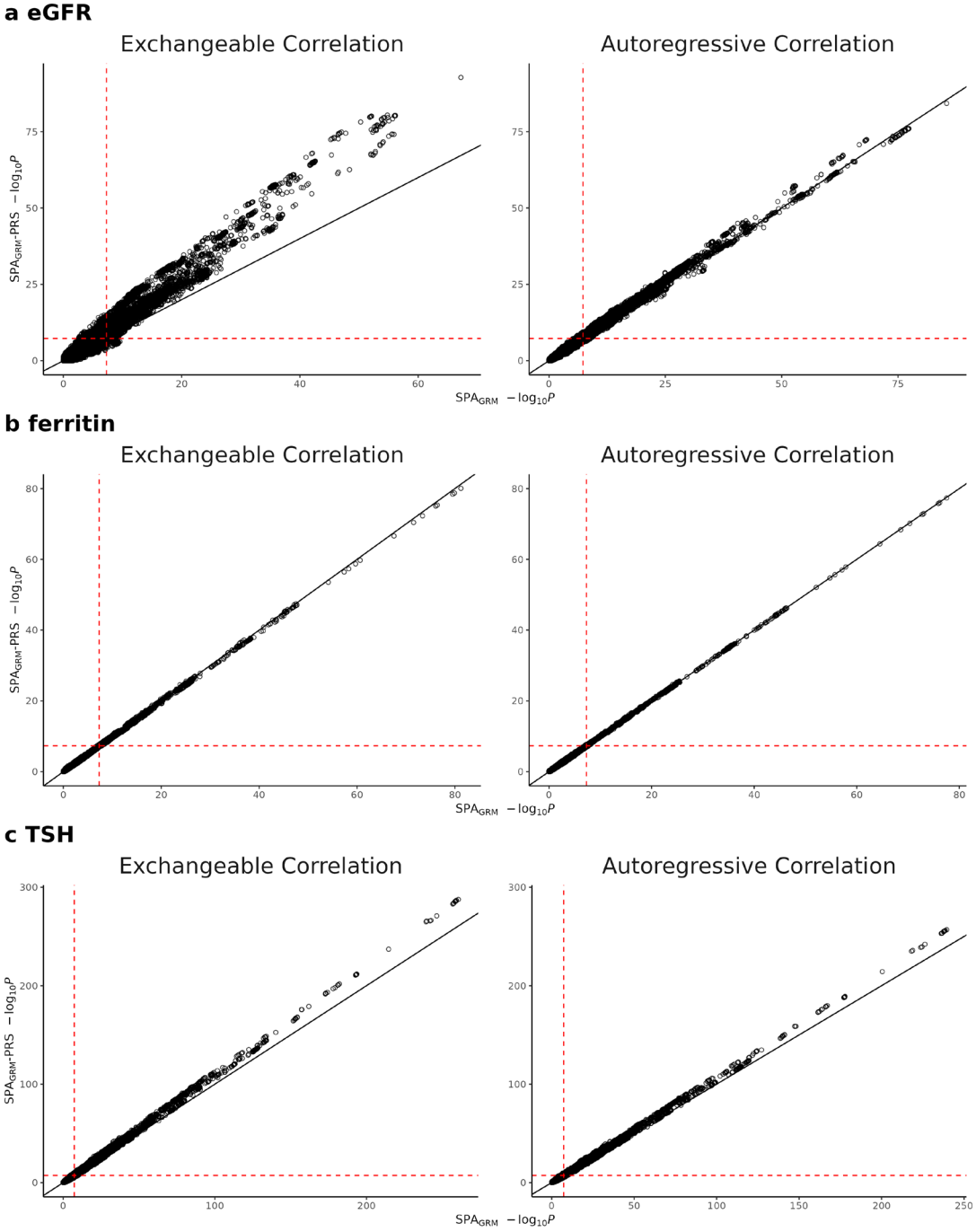
Supplementary Figure 32. Manhattan plots of GWAS results for longitudinal body mass index, fasting glucose, and high-density lipoprotein cholesterol, whether to conduct precise quality control. (**a**)-(**c**) are mirror Manhattan plots of GWAS results of body mass index (BMI), fasting glucose, and high-density lipoprotein cholesterol (HDL-C). Across all subplots, upper and lower are results for testing $\beta_g = 0$ and $\tau_g = 0$, respectively. From left to right, we compared whether to conduct precise quality control (QC) and two methods: TrajGWAS and SPA_{GRM}. The left two are "Using previous QC", which means no QC on WS variability; and the right two are "Using our QC", which means QC on WS variability. The red dashed line represents the genome-wide significance ($P = 5 \times 10^{-8}$), and the grey dashed line represents the breakpoint for the different scales of the y-axis ($-log_{10}P = 15$).



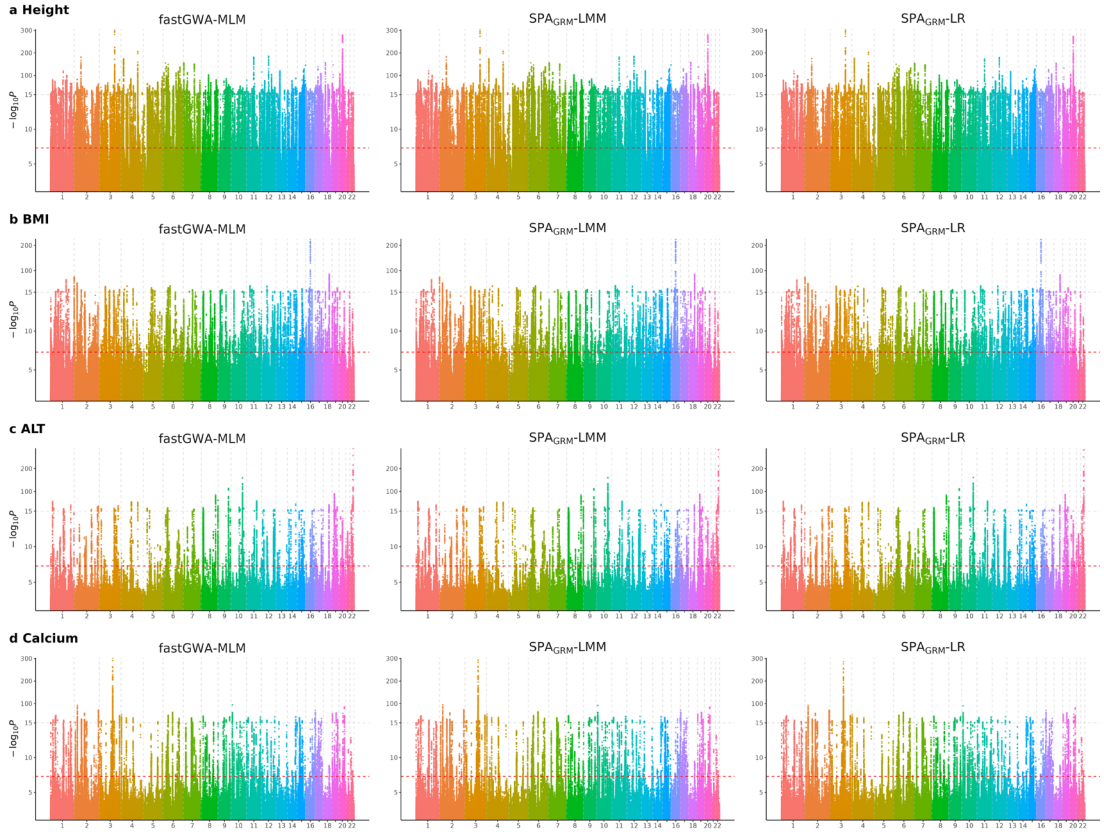
Supplementary Figure 33. Manhattan plots and quantile-quantile (QQ) plots of GWAS results from TrajGWAS, SPA$_{GRM}$, and SPA$_{GRM}$-PGS methods. (**a**)-(**c**) are Manhattan plots of GWAS results of three longitudinal traits: eGFR, ferritin, and TSH, respectively. (**d**) is scatterplots of GWAS results from TrajGWAS, SPA$_{GRM}$, and SPA$_{GRM}$-PGS methods. In (**d**), x-axis represents $-log_{10}P$ of SPA$_{GRM}$ results, and y-axis represents $-log_{10}P$ of TrajGWAS and SPA$_{GRM}$-PGS results. The red dashed line in subplot a-d represents the genome-wide significance ($P = 5 \times 10^{-8}$), and the grey dashed line in subplot a-c represents the breakpoint for the different scales of the y-axis ($-log_{10}P = 15$).



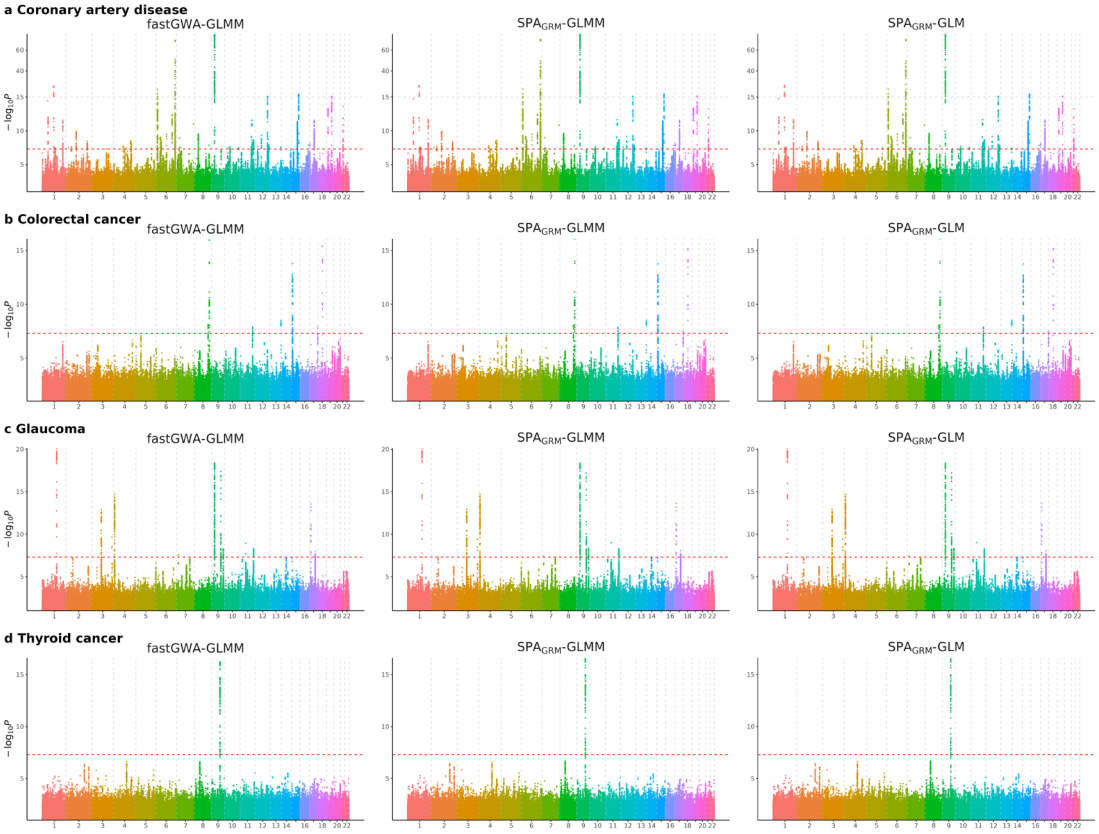
Supplementary Figure 34. Scatter plots of GWAS results from SPA$_{GRM}$ and SPA$_{GRM}$-PGS methods using different GEE models for different longitudinal traits. (**a**)-(**c**) are GWAS results of three longitudinal traits: eGFR, ferritin, and TSH, respectively. In each subplot, x-axis represents $-log_{10}P$ of SPA$_{GRM}$ results, and y-axis represents $-log_{10}P$ of SPA$_{GRM}$-PGS results. From left to right, we considered using exchangeable correlation structure and autoregressive correlation structure to fir the null model. The red dashed line in (**a**)-(**c**) represents the genome-wide significance ($P = 5 \times 10^{-8}$).



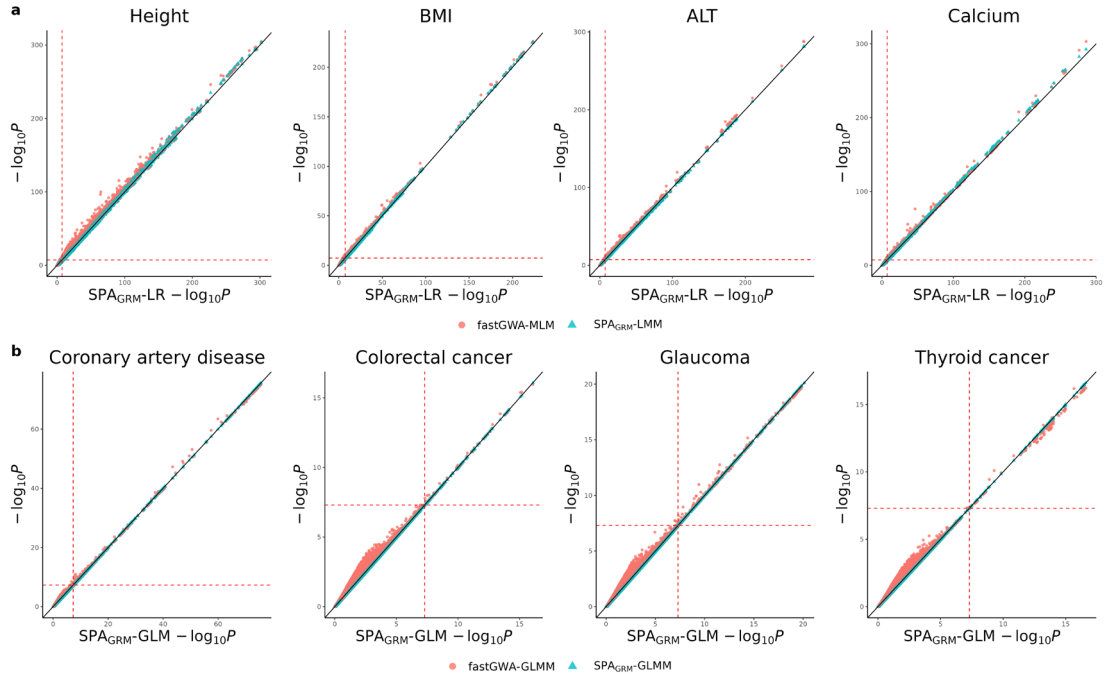
Supplementary Figure 35. Comparison of methods on four quantitative traits from the UK Biobank. (**a**)-(**d**) are Manhattan plots of using fastGWA-MLM, SPA_{GRM}-LMM, and SPA_{GRM}-LR for the analysis of height ($N = 407{,}443$), body mass index (BMI, $N = 405{,}187$), alanine transaminase (ALT, $N = 382{,}371$), and calcium ($N = 356{,}043$). The tests were performed on 9.4 million imputed SNPs with minor allele frequency > 0.01. the bottom red dashed horizontal line represents the genome-wide significance ($P = 5 \times 10^{-8}$) and the top grey dashed horizontal line represents the breakpoint for the different scaling of the y axis ($-log_{10}P = 15$). The dashed vertical lines separate the 22 chromosomes.



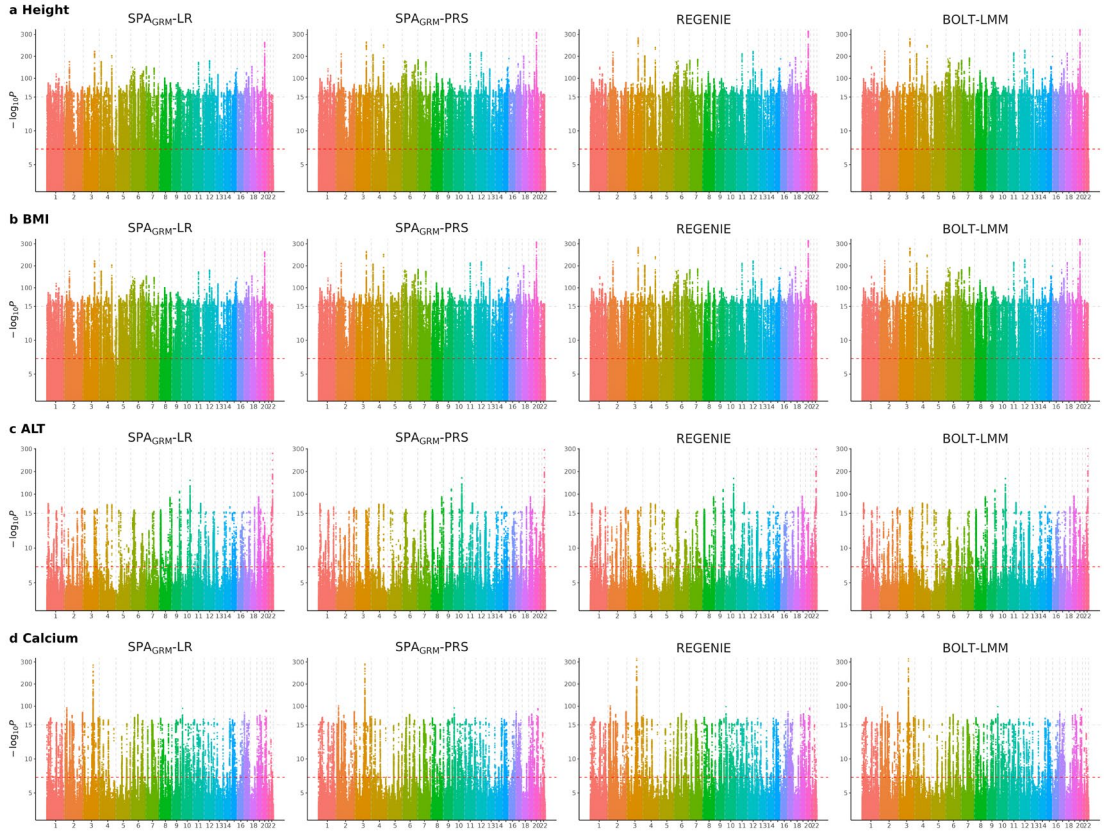
Supplementary Figure 36. Comparison of methods on four binary traits from the UK Biobank. (**a**)-(**d**) are Manhattan plots of using fastGWA-GLMM, SPA$_{GRM}$-GLMM, and SPA$_{GRM}$-GLM for the analysis of coronary artery disease (case-control ratio is 1:10), colorectal cancer (1:64), glaucoma (1:67), and thyroid cancer (1:710) with the same sample size $N = 408{,}510$. The tests were performed on 9.4 million imputed SNPs with minor allele frequency > 0.01. the bottom red dashed horizontal line represents the genome-wide significance ($P = 5 \times 10^{-8}$) and the top grey dashed horizontal line represents the breakpoint for the different scaling of the y axis ($-log_{10}P = 15$). The dashed vertical lines separate the 22 chromosomes.



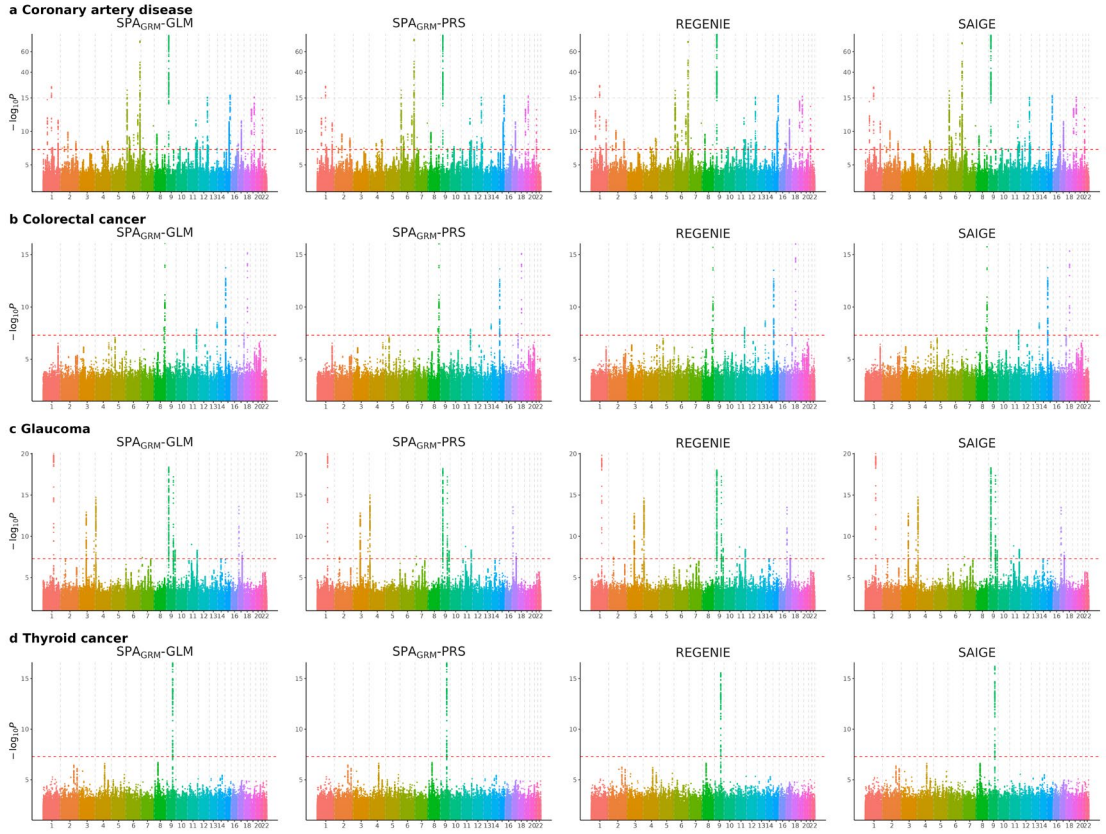
Supplementary Figure 37. Quantile-quantile (QQ) plots comparing methods on four quantitative and binary traits from the UK Biobank. (**a**) displays $-log_{10}P$ comparison results of using SPA$_{GRM}$-LR (x-axis), fastGWA-MLM (y-axis), and SPA$_{GRM}$-LMM (y-axis) for the analysis of height ($N = 407{,}443$), body mass index (BMI, $N = 405{,}187$), alanine transaminase (ALT, $N = 382{,}371$), and calcium ($N = 356{,}043$). (**b**) displays $-log_{10}P$ comparison results of using SPA$_{GRM}$-GLM (x-axis), fastGWA-GLMM (y-axis), and SPA$_{GRM}$-GLMM (y-axis) for the analysis of coronary artery disease (case-control ratio is 1:10), colorectal cancer (1:64), glaucoma (1:67), and thyroid cancer (1:710) with the same sample size $N = 408{,}510$. The tests were performed on 9.4 million imputed SNPs with minor allele frequency > 0.01. The red dashed line represents the genome-wide significance ($P = 5 \times 10^{-8}$).



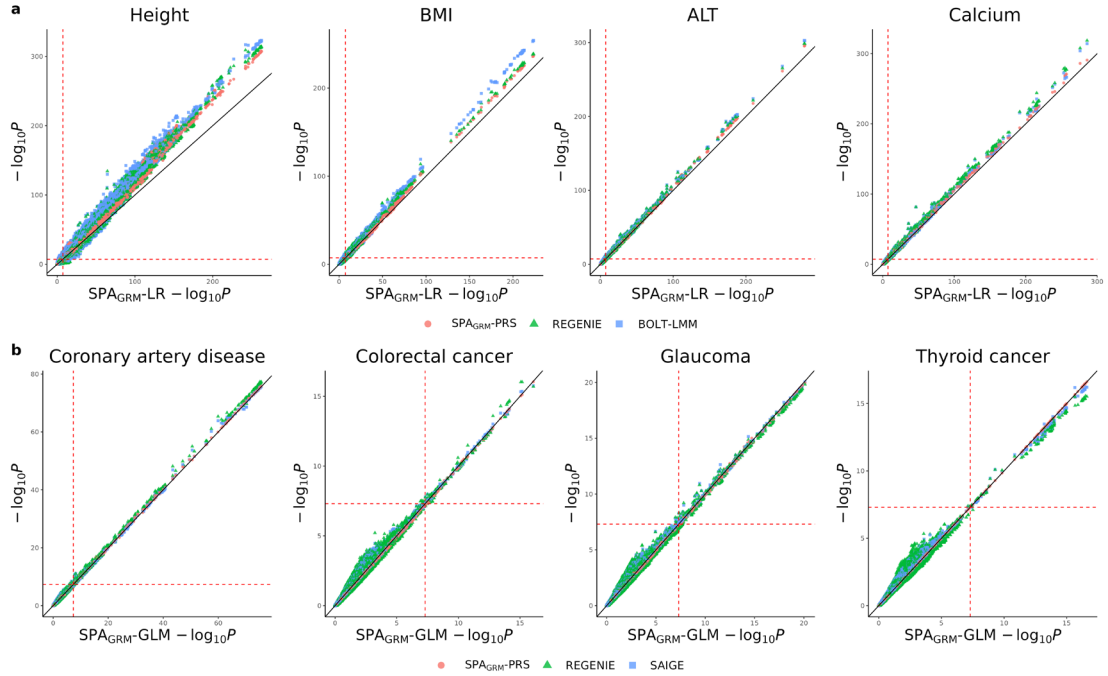
Supplementary Figure 38. Comparison of methods on four quantitative traits from the UK Biobank. (**a**)-(**d**) are Manhattan plots of using SPA$_{GRM}$-LR, SPA$_{GRM}$-PGS, REGENIE, and BOLT-LMM for the analysis of height ($N = 407{,}443$), body mass index (BMI, $N = 405{,}187$), alanine transaminase (ALT, $N = 382{,}371$), and calcium ($N = 356{,}043$). The tests were performed on 9.4 million imputed SNPs with minor allele frequency > 0.01. the bottom red dashed horizontal line represents the genome-wide significance ($P = 5 \times 10^{-8}$) and the top grey dashed horizontal line represents the breakpoint for the different scaling of the y axis ($-log_{10}P = 15$). The dashed vertical lines separate the 22 chromosomes.



Supplementary Figure 39. Comparison of methods on four binary traits from the UK Biobank. (**a**)-(**d**) are Manhattan plots of using SPA$_{GRM}$-GLM, SPA$_{GRM}$-PGS, REGENIE, and SAIGE for the analysis of coronary artery disease (case-control ratio is 1:10), colorectal cancer (1:64), glaucoma (1:67), and thyroid cancer (1:710) with the same sample size $N = 408{,}510$. The tests were performed on 9.4 million imputed SNPs with minor allele frequency > 0.01. the bottom red dashed horizontal line represents the genome-wide significance ($P = 5 \times 10^{-8}$) and the top grey dashed horizontal line represents the breakpoint for the different scaling of the y axis ($-log_{10}P = 15$). The dashed vertical lines separate the 22 chromosomes.



Supplementary Figure 40. Quantile-quantile (QQ) plots comparing methods on four quantitative and binary traits from the UK Biobank. (**a**) displays $-log_{10}P$ comparison results of using SPA$_{GRM}$-LR (x-axis), SPA$_{GRM}$-PGS (y-axis), REGENIE (y-axis), and BOLT-LMM (y-axis) for the analysis of height ($N = 407{,}443$), body mass index (BMI, $N = 405{,}187$), alanine transaminase (ALT, $N = 382{,}371$), and calcium ($N = 356{,}043$). (**b**) displays $-log_{10}P$ comparison results of using SPA$_{GRM}$-GLM (x-axis), SPA$_{GRM}$-PGS (y-axis), REGENIE (y-axis), and SAIGE (y-axis) for the analysis of coronary artery disease (case-control ratio is 1:10), colorectal cancer (1:64), glaucoma (1:67), and thyroid cancer (1:710) with the same sample size $N = 408{,}510$. The tests were performed on 9.4 million imputed SNPs with minor allele frequency > 0.01. The red dashed line represents the genome-wide significance ($P = 5 \times 10^{-8}$).



Supplementary Table 1. Mean chi-square statistics of SPA$_{GRM}$, SPA$_{GRM(INT)}$, and SPA$_{GRM(CCT)}$, and TrajGWAS methods in scenarios 2-4.

Simulation conditions			Mean chi-square statistics					
			Scenario 2		Scenario 3		Scenario 4	
Family relatedness	Variant type	Methods	$\beta_g \neq 0$	$\tau_g \neq 0$	$\beta_g = 0$	$\tau_g \neq 0$	$\beta_g \neq 0$	$\tau_g = 0$
A: 25,000 Unrelated	Common variants	SPA$_{GRM}$	44.6	**16.6**	5.7	**16.1**	44.9	5.6
		SPA$_{GRM(INT)}$	44.7	**96.2**	5.5	**96.6**	45.2	5
		SPA$_{GRM(CCT)}$	44.9	**95**	5.7	**95.6**	45.1	5.2
		TrajGWAS	44.7	11.7	5.8	11.7	45.4	5.4
	Rare variants	SPA$_{GRM}$	20.4	18.8	11.8	18.1	22.8	5.5
		SPA$_{GRM(INT)}$	20.4	**58.3**	9.3	**57.4**	23.4	6.1
		SPA$_{GRM(CCT)}$	21.4	**57**	11.6	**56.1**	23.2	5.9
		TrajGWAS	20.9	17.4	12.2	17.1	23.6	7.9
B: 6,250 Small families	Common variants	SPA$_{GRM}$	**39.8**	**15.4**	5.5	**16.1**	**39.8**	5.7
		SPA$_{GRM(INT)}$	**39.7**	**83.3**	5.4	**83.9**	**39.8**	5.3
		SPA$_{GRM(CCT)}$	**39.9**	**82.6**	5.5	**83.3**	**39.7**	5.4
		TrajGWAS	27	10.1	5.8	10.6	26.9	5.6
	Rare variants	SPA$_{GRM}$	**18.8**	**15.9**	11.2	**15.5**	**19.1**	5.7
		SPA$_{GRM(INT)}$	**17.9**	**43.7**	8.3	**43.2**	**19.4**	5.6
		SPA$_{GRM(CCT)}$	**19**	**42.8**	10.9	**42.4**	**19.3**	5.5
		TrajGWAS	15.3	13.4	10.9	13.6	14.1	7.6
C: 2,500 Large families	Common variants	SPA$_{GRM}$	**34.6**	**16.2**	5.5	**16.8**	**35.7**	5.3
		SPA$_{GRM(INT)}$	**34.3**	**73.9**	5.7	**73.5**	**35.8**	5.1
		SPA$_{GRM(CCT)}$	**34.6**	**72.9**	5.6	**72.6**	**35.8**	5.1
		TrajGWAS	22.1	9.9	5.5	9.8	22.1	5.2
	Rare variants	SPA$_{GRM}$	**19**	**19**	9.4	**19.4**	**19.6**	5.7
		SPA$_{GRM(INT)}$	**17.4**	**42.1**	7.7	**42.3**	**19.9**	5.7
		SPA$_{GRM(CCT)}$	**18.9**	**40.9**	9.3	**41**	**19.7**	5.8
		TrajGWAS	14.1	13.8	9.6	14.9	13.1	7.5
D: A + B	Common variants	SPA$_{GRM}$	**78.2**	**17.2**	5.7	**18.2**	**80.5**	5.5
		SPA$_{GRM(INT)}$	**78.5**	**143.2**	5.6	**140.9**	**81.2**	5.4
		SPA$_{GRM(CCT)}$	**78.7**	**143.6**	5.6	**140.2**	**80.8**	5.5
		TrajGWAS	64.7	10.9	5.8	12	67.8	5.2
	Rare variants	SPA$_{GRM}$	**33**	**23.2**	14	**23.7**	**40.1**	5.5
		SPA$_{GRM(INT)}$	**33.8**	**94.9**	9.8	**94.3**	**41.1**	5.6
		SPA$_{GRM(CCT)}$	**35.3**	**93.7**	13.5	**93**	**40.8**	5.7
		TrajGWAS	28.8	20.4	14.8	20.7	36.6	9
E: A + C	Common variants	SPA$_{GRM}$	**71.7**	**29**	6	**29.6**	**73.1**	5.4
		SPA$_{GRM(INT)}$	**71.2**	**152.7**	5.7	**155**	**73.6**	5
		SPA$_{GRM(CCT)}$	**71.6**	**151.3**	5.8	**154.1**	**73.4**	5.4
		TrajGWAS	59.7	19	5.7	19.4	60.1	5.2
	Rare variants	SPA$_{GRM}$	**33.7**	**24.7**	9.7	**25.8**	**37.3**	5.6
		SPA$_{GRM(INT)}$	**32.1**	**85.5**	7.6	**85.9**	**38.2**	5.5
		SPA$_{GRM(CCT)}$	**33.9**	**84.1**	9.6	**84.5**	**37.9**	5.4
		TrajGWAS	28.8	19.4	10.5	20.2	31.8	8.7

Mean chi-square statistics of SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, and TrajGWAS methods in scenarios 2-4. Refer to **Figs. 3** and **4**, and **Supplementary Fig. 20** for the table legend. Bold numbers represent superior performance in statistical power compared to TrajGWAS. P value is calculated using two-sided score tests. Exact p values are provided in **Supplementary Data 4**.

Supplementary Table 2. Empirical power of SPA$_{GRM}$, SPA$_{GRM(INT)}$, and SPA$_{GRM(CCT)}$, and TrajGWAS methods at the significant level of 5×10^{-8} in scenarios 2-4.

Simulation conditions			Empirical power					
			Scenario 2		Scenario 3		Scenario 4	
Family relatedness	Variant type	Methods	$\beta_g \neq 0$	$\tau_g \neq 0$	$\beta_g = 0$	$\tau_g \neq 0$	$\beta_g \neq 0$	$\tau_g = 0$
A: 25,000 Unrelated	Common variants	SPA$_{GRM}$	48.2%	**14.5%**	0	**12.5%**	46.5%	0
		SPA$_{GRM(INT)}$	48.1%	**67.8%**	0	**69.4%**	46.4%	0
		SPA$_{GRM(CCT)}$	48.1%	**66.1%**	0	**68.1%**	46.4%	0
		TrajGWAS	48.4%	5.0%	0	5.5%	46.7%	0
	Rare variants	SPA$_{GRM}$	19.0%	22.4%	4.7%	19.2%	24.4%	0
		SPA$_{GRM(INT)}$	19.9%	**86.8%**	1.7%	**86.8%**	28.5%	0
		SPA$_{GRM(CCT)}$	22.2%	**85.7%**	3.7%	**84.9%**	27.2%	0
		TrajGWAS	20.3%	20.3%	4.8%	17.7%	27.9%	0
B: 6,250 Small families	Common variants	SPA$_{GRM}$	**45.9%**	**13.4%**	0	**14.9%**	**45.3%**	0
		SPA$_{GRM(INT)}$	**46.1%**	**62.1%**	0	**61.8%**	**45.1%**	0
		SPA$_{GRM(CCT)}$	**46.0%**	**60.5%**	0	**59.4%**	**45.1%**	0
		TrajGWAS	41.3%	1.8%	0	2.9%	41.0%	0
	Rare variants	SPA$_{GRM}$	**15.3%**	**14.8%**	3.0%	**12.2%**	**12.4%**	0
		SPA$_{GRM(INT)}$	**12.0%**	**76.3%**	0	**76.7%**	**13.4%**	0
		SPA$_{GRM(CCT)}$	**15.8%**	**75.8%**	2.8%	**74.0%**	**13.1%**	0
		TrajGWAS	8.5%	5.6%	1.8%	8.9%	2.7%	0.5%
C: 2,500 Large families	Common variants	SPA$_{GRM}$	**43.6%**	**13.5%**	0	**14.0%**	**45.5%**	0
		SPA$_{GRM(INT)}$	**43.8%**	**57.9%**	0	**57.1%**	**45.3%**	0
		SPA$_{GRM(CCT)}$	**43.7%**	**56.6%**	0	**55.2%**	**45.4%**	0
		TrajGWAS	29.9%	1.8%	0	0.8%	29.1%	0
	Rare variants	SPA$_{GRM}$	**16.4%**	**21.9%**	1.8%	**22.3%**	**16.4%**	0
		SPA$_{GRM(INT)}$	**11.9%**	**75.8%**	0	**75.3%**	**18.2%**	0
		SPA$_{GRM(CCT)}$	**16.1%**	**73.1%**	2.0%	**73.0%**	**17.8%**	0
		TrajGWAS	7.4%	7.5%	0.8%	10.8%	2.9%	0
D: A + B	Common variants	SPA$_{GRM}$	**61.3%**	**16.7%**	0	**18.7%**	**62.9%**	0
		SPA$_{GRM(INT)}$	**61.8%**	**83.9%**	0	**83.7%**	**63.0%**	0
		SPA$_{GRM(CCT)}$	**61.4%**	**82.7%**	0	**82.2%**	**62.9%**	0
		TrajGWAS	55.9%	3.7%	0	5.5%	56.6%	0
	Rare variants	SPA$_{GRM}$	**51.8%**	**25.2%**	9.5%	**25.3%**	**76.7%**	0
		SPA$_{GRM(INT)}$	**55.9%**	**98.5%**	4.0%	**98.0%**	**77.6%**	0
		SPA$_{GRM(CCT)}$	**58.6%**	**98.0%**	9.3%	**97.6%**	**77.4%**	0
		TrajGWAS	40.3%	23.9%	9.7%	24.0%	64.7%	0
E: A + C	Common variants	SPA$_{GRM}$	**59.5%**	**35.3%**	0	**37.1%**	**58.8%**	0
		SPA$_{GRM(INT)}$	**59.4%**	**87.5%**	0	**88.9%**	**59.0%**	0
		SPA$_{GRM(CCT)}$	**59.6%**	**85.6%**	0	**87.7%**	**58.8%**	0
		TrajGWAS	51.8%	20.1%	0	22.9%	50.6%	0
	Rare variants	SPA$_{GRM}$	**56.3%**	**31.1%**	2.0%	**34.1%**	**72.0%**	0
		SPA$_{GRM(INT)}$	**53.8%**	**97.4%**	0.6%	**96.7%**	**73.0%**	0
		SPA$_{GRM(CCT)}$	**57.6%**	**96.4%**	2.1%	**96.1%**	**72.7%**	0
		TrajGWAS	43.4%	21.8%	2.3%	24.5%	54.3%	0.6%

Empirical power of SPA$_{GRM}$, SPA$_{GRM(INT)}$, SPA$_{GRM(CCT)}$, and TrajGWAS methods at significant level of 5×10^{-8} in scenarios 2-4. Refer to **Figs. 3** and **4**, and **Supplementary Fig. 20** for the table legend. Bold numbers represent superior performance in statistical power compared to TrajGWAS. P value is calculated using two-sided score tests. Exact p values are provided in **Supplementary Data 4**.

Supplementary Table 3. Clinical terms to extract longitudinal traits from UKB primary care data.

Trait	Acronyms	Unit	Terminology	Code terms
activated partial thromboplastin time	aPTT	seconds	Read v2/CTV3	42jG, 42Q6, XS9U4
adjusted calcium	adj Ca	mmol/L	Read v2/CTV3	44h[479], 44I[8C], 4Q721, X771W, Xabp[kr], XaDvd, XaIR[kn], XaZyY, XE2q3
alanine aminotransferase	ALT	U/L	Read v2/CTV3	44G[.3AB], X771f, XaIRi, XaLJx
alkaline phosphatase	ALP	U/L	Read v2/CTV3	44CU, 44F[.123Z], XaIRj, XE2px
amylase	amylase	U/L	Read v2/CTV3	44C[4NT], XaIRh, XM14L
aspartate aminotransferase	AST	U/L	Read v2/CTV3	44H[5BC], X771i, XaES6, XE25Q
basophil count	BASO	10^9 /L	Read v2/CTV3	42L.
basophil percentage	BASO perc	%	Read v2/CTV3	42b3, XE2mS
bicarbonate	HCO_3	mmol/L	Read v2/CTV3	44h3, 44i0, 44I7, XaDvc, XaItM, XE2q2
blood albumin	Alb	g/L	Read v2/CTV3	44M[4I], XaIRc, XE2eA
blood creatinine	Cr	umol/L	Read v2/CTV3	44J3[.0123z], 44J[CF], XE2q5, XaETQ, 4Q40, X771Q
blood globulin	Glob	g/L	Read v2/CTV3	44M5., 44MP, XaItX, XE2eB
body mass index	BMI	kg/m^2	Read v2/CTV3	XaCDR, XaJJH, XaJqk, XaZcl, 22K, 229, 22A, 162[23], X76CG, XE1h4, XM01G, Xa7wI
calcium	Ca	mmol/L	Read v2/CTV3	44h[479], 44I[8C], 4Q721, X771W, Xabp[kr], XaDvd, XaIR[kn], XaZyY, XE2q3
carbohydrate antigen 125	CA125	U/mL	Read v2/CTV3	XaJNf
chloride	Cl	mmol/L	Read v2/CTV3	44h2, 44i1, 44I6, XaDvb, XaItN, XE2q1
C-reactive protein	CRP	mg/L	Read v2/CTV3	44C[CS], XaINL, XE2dy
creatine kinase	CK	U/L	Read v2/CTV3	44H[4EG], XaES[5B]
diastolic blood pressure	DBP	mmHg	Read v2/CTV3	246[.abcdefgABEFGJQRTVWXY12345679], XaF4[abFKLOSZ], XaJ2[EF], XaKF[xw], XaKj[FG], 662L, R1y2, G20
eosinophil count	EOS	10^9 /L	Read v2/CTV3	42K.
eosinophil percentage	EOS perc	%	Read v2/CTV3	42b9, XaCJj
erythrocyte sedimentation rate	ESR	mm/h	Read v2/CTV3	42B6, XE2m7
estimated glomerular filtration rate	eGFR	$mL/min/1.73m^2$	Read v2/CTV3	44J3[.0123z], 44J[CF], XE2q5, XaETQ, 4Q40, X771Q
fasting blood glucose	FBG	mmol/L	Read v2/CTV3	44[fg]1, 44T[2K], XE2mq
ferritin	ferritin	μg/L	Read v2/CTV3	42d4, 42R4, XaItW, XE24r
folate	folate	μg/L	Read v2/CTV3	42U[.1235EZ], X76tC
Forced expiratory volume in 1-second	FEV1	L	Read v2/CTV3	33972, 339[abef], 339O., X77Qu, XaIx[QRUV]
Forced expired volume in 1-second/forced vital capacity ratio	FEV1/FVC	%	Read v2/CTV3	X77Ra, XaEFy
forced vital capacity	FVC	L	Read v2/CTV3	3396[.013], 339[hs], XaJ3K, XaPpI
Framingham coronary heart disease 10 year risk score	CHD score	%	Read v2/CTV3	38DP, XaFsZ
free thyroxine	FT4	pmol/L	Read v2/CTV3	442[7cV], XaER[rs]
free triiodothyronine	FT3	pmol/L	Read v2/CTV3	442[45U], XaERq

gamma-glutamyl transferase	GGT	U/L	Read v2/CTV3	44G[479], XaES[34]
glycated haemoglobin	HbA1c	mmol/mol	Read v2/CTV3	XaPbt, XaERp, X772q, 42W[.12345Z], 44TB.
haematocrit percentage	Hct	%	Read v2/CTV3	425., X76t[bc], XE2Zq
haemoglobin	Hb	g/dL	Read v2/CTV3	423[.123456789ABZ], Xa96v, XE2m6, XM1Vu
high density lipoprotein cholesterol	HDL	mmol/L	Read v2/CTV3	44d[23A], X772M, 44P[5BC], XaEVr
Low density lipoprotein cholesterol	LDL	mmol/L	Read v2/CTV3	44d[45B], X772N, 44P[6DEI], XaEVs, XaIp4
lymphocyte count	LYMPH	10^9 /L	Read v2/CTV3	42M[.14Z]
lymphocyte percentage	LYMPH perc	%	Read v2/CTV3	42b1, XE2mQ
mean corpuscular haemoglobin	MCH	pg	Read v2/CTV3	428., XE2pb
mean corpuscular haemoglobin concentration	MCHC	g/dL	Read v2/CTV3	429
mean corpuscular volume	MCV	fL	Read v2/CTV3	42A[.12345Z]
mean platelet volume	MPV	fL	Read v2/CTV3	42Z5
monocyte count	MONO	10^9 /L	Read v2/CTV3	42N[.125Z]
monocyte percentage	MONO perc	%	Read v2/CTV3	42b2, XE2mR
neutrophill count	NEUT	10^9 /L	Read v2/CTV3	42J[.15Z]
neutrophill percentage	NEUT perc	%	Read v2/CTV3	42b0, XE2mP
non-high density lipoprotein cholesterol	NHDL	mmol/L	Read v2/CTV3	44PL, XabE1, XaN3z
oxygen saturation at periphery	SpO_2	%	Read v2/CTV3	X770D
peak expiratory flow	PEF	L/min	Read v2/CTV3	339[345Acgno], 66YX, X77RW, XaEFM, XaEHe, XaIx[DPS], XaJEg, XaK6I, XE2wr, XE2xQ, XS7q5
plasma viscosity	PV	mPas	Read v2/CTV3	42B[.1Z], X76xg, XE2pd
platelet count	Plt	10^9 /L	Read v2/CTV3	42P[.1234Z], XE24o
prostate-specific antigen	PSA	ng/mL	Read v2/CTV3	43Z[2F], X80QD, XabAM, XaPqN, XE25C
prothrombin time	PT	1	Read v2/CTV3	41C5|^42jr, 42QE[.01], 9k21, XaPJW
pulse pressure	PP	mmHg	Read v2/CTV3	246[.abcdefgABEFGJQRTVWXY12345679], XaF4[abFKLOSZ], XaJ2[EF], XaKF[xw], XaKj[FG], 662L, R1y2, G20
pulse rate	PR	bpm	Read v2/CTV3	242, 243, X773s, XaIBo, XM02J
random blood glucose	RBG	mmol/L	Read v2/CTV3	44[fg][.0], 44T[1AJ], 44U., 8A17, X772z, XE2mp, XM0ly
red blood cell count	RBC	10^{12} /L	Read v2/CTV3	426[.123457Z]
red blood cell distribution width	RDW	%	Read v2/CTV3	42Z7, XE2mO
serum cholesterol	TC	mmol/L	Read v2/CTV3	44OE, 44P[.12349HJKZ], XE2eD, X772L, XSK14, XaFs9, XaIRd, XaJe9, XaLux
serum inorganic phosphate	Pi	mmol/L	Read v2/CTV3	44h5, 44i2, 44I9, XaDve, XaItO, XE2q4
serum iron	iron	umol/L	Read v2/CTV3	42d2, 42R[.1237Z], 4Q74, X76tH, X7733, XaIRe, XE24q
serum potassium	K	mmol/L	Read v2/CTV3	44h[08], 44I4, 4Q42, X771S, XaDvZ, XaIRl, XE2pz
serum sodium	Na	mmol/L	Read v2/CTV3	44h[16], 44I5, 4Q43, X771T, XaDva, XaIRf, XE2q0
serum triglyceride	TG	mmol/L	Read v2/CTV3	44l[2FGH], 44PF, XaEU[qr]
serum urate	urate	umol/L	Read v2/CTV3	44K, 4Q66, XaDvn, XE2e2, XM0ls, XM1Uq

serum urea	urea	mmol/L	Read v2/CTV3	44J[.1289AZ], X771P, XaDvl, XE25R, XM0lt
systolic blood pressure	SBP	mmHg	Read v2/CTV3	246[.abcdefgABEFGJQRTVWXY12345679], XaF4[abFKLOSZ], XaJ2[EF], XaKF[xw], XaKj[FG], 662L, R1y2, G20
thyroid stimulating hormone	TSH	mU/L	Read v2/CTV3	442[AWX], X80Gc, XaEL[VW], XE2wy
total bilirubin	Total bili	umol/L	Read v2/CTV3	44E[.1239CZ], XaERu, XaETf, XE2qu
total protein	TP	g/L	Read v2/CTV3	44M[.1237AVZ], 4Q3., XE2e[9C]
urine albumin	Ualb	mg/L	Read v2/CTV3	46N4, 46W[.1], XE2eI, XE2bw
urine albumin/creatinine ratio	UACR	mg/mmol	Read v2/CTV3	44J7, 46T[CD], XE2n3
urine creatinine	Ucr	mmol/L	Read v2/CTV3	46M7, XE2qO
vitamin B12	VB12	ng/L	Read v2/CTV3	42T[.12Z], 44L[ce], XaJ27, XE2pf
Vitamin D	VD	nmol/L	Read v2/CTV3	44LA, 4QB4[67], XaY6m, XE2e7
waist circumference	Waist	cm	Read v2/CTV3	22N0, Xa041
white blood cell count	WBC	10^9 /L	Read v2/CTV3	42H[.123578Z], XaId[YZ]

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

Supplementary Table 4. Summary statistics of 79 longitudinal traits extracted from UKB primary care data and the corresponding UK biobank field ID.

Trait	Acronyms	Unit	Sample size	Per obs.	Summary of trait	Male	Age	BMI	UKB field ID
			SPA_{GRM} (TrajGWAS, %)	Median (IQR)	Mean (SD)	%	Mean (SD)	Mean (SD)	
activated partial thromboplastin time	aPTT	seconds	7,949 (6,604 83.1%)	1 (1, 1)	30.0 (4.9)	40.0%	61.6 (8.8)	28.2 (5.3)	-
adjusted calcium	adj Ca	mmol/L	86,990 (72,889 83.8%)	2 (1, 3)	2.4 (0.1)	42.5%	62 (8.2)	27.8 (4.9)	30680
alanine aminotransferase	ALT	U/L	158,297 (132,663 83.8%)	4 (2, 8)	26.8 (14.6)	45.8%	62 (8.1)	27.7 (4.8)	30620
alkaline phosphatase	ALP	U/L	155,191 (129,973 83.8%)	5 (2, 9)	88.1 (44.2)	45.9%	61.5 (8.1)	27.7 (4.8)	30610
amylase	amylase	U/L	10,081 (8,397 83.3%)	1 (1, 1)	60.8 (25.0)	39.4%	60.6 (8.6)	27.9 (5)	-
aspartate aminotransferase	AST	U/L	54,017 (45,407 84.1%)	2 (1, 5)	25.8 (10.0)	46.1%	61.1 (8.4)	27.9 (4.9)	30650
basophil count	BASO	10^9 /L	155,885 (130,617 83.8%)	4 (2, 8)	0.04 (0.04)	44.7%	61.3 (8.7)	27.7 (4.8)	30160
basophil percentage	BASO perc	%	18,923 (16,137 85.3%)	3 (1, 6)	0.8 (0.4)	44.3%	60.6 (8.5)	27.6 (4.8)	30220
bicarbonate	HCO_3	mmol/L	28,648 (24,251 84.7%)	2 (1, 6)	27.1 (2.7)	46.0%	60.7 (8.3)	27.9 (4.9)	-
blood albumin	Alb	g/L	156,597 (131,205 83.8%)	5 (2, 9)	41.7 (3.8)	45.7%	61.6 (8.1)	27.7 (4.8)	30600
blood creatinine	Cr	umol/L	165,305 (138,634 83.9%)	6 (3, 11)	83.3 (21.0)	45.9%	62 (8.1)	27.6 (4.8)	30700
blood globulin	Glob	g/L	80,648 (67,555 83.8%)	4 (2, 8)	28.3 (4.3)	46.2%	61.6 (8)	27.8 (4.9)	
body mass index	BMI	kg/m^2	175,443 (147,314 84.0%)	5 (3, 9)	28.4 (5.7)	45.4%	58.3 (9.9)	-	21001
calcium	Ca	mmol/L	87,571 (73,514 83.9%)	2 (1, 4)	2.4 (0.11)	42.8%	61.7 (8.4)	27.7 (4.9)	30680
carbohydrate antigen 125	CA125	U/mL	6,163 (5,076 82.4%)	1 (1, 1)	13.7 (10.5)	0.0%	59.9 (8.3)	26.7 (4.8)	-
chloride	Cl	mmol/L	35,199 (30,111 85.5%)	3 (2, 7)	103.3 (3.1)	46.0%	62.2 (8.3)	27.8 (4.9)	-
C-reactive protein	CRP	mg/L	71,581 (59,965 83.8%)	1 (1, 2)	7.1 (9.6)	40.6%	62 (8.4)	27.8 (4.9)	30710
creatine kinase	CK	U/L	21,704 (18,026 83.1%)	1 (1, 2)	132.2 (97.4)	50.6%	61.8 (7.5)	28.5 (4.9)	-
diastolic blood pressure	DBP	mmHg	182,248 (153,061 84.0%)	12 (6, 24)	80.4 (10.5)	45.6%	59 (9.4)	27.5 (4.8)	4079
eosinophil count	EOS	10^9 /L	155,439 (130,230 83.8%)	4 (2, 7)	0.2 (0.1)	44.7%	61.2 (8.7)	27.7 (4.8)	30150
eosinophil percentage	EOS perc	%	18,964 (16,179 85.3%)	3 (1, 6)	3.1 (1.9)	44.3%	60.7 (8.5)	27.6 (4.8)	30210
erythrocyte sedimentation rate	ESR	mm/h	91,360 (76,232 83.4%)	2 (1, 3)	13.9 (13.8)	41.3%	60.3 (8.8)	27.7 (4.9)	-
estimated glomerular filtration rate	eGFR	$mL/min/1.73m^2$	165,305 (138,634 83.9%)	6 (3, 11)	77.4 (15.6)	45.9%	62 (8.1)	27.6 (4.8)	30700
fasting blood glucose	FBG	mmol/L	79,409 (66,518 83.8%)	2 (1, 4)	5.5 (1.3)	48.3%	61.4 (7.7)	28.1 (4.9)	30740
ferritin	ferritin	μg/L	66,729 (55,683 83.4%)	1 (1, 3)	96.9 (128.9)	36.9%	60.8 (9.3)	27.6 (5.1)	-
folate	folate	μg/L	51,828 (43,448 83.8%)	1 (1, 2)	9.6 (5.2)	39.8%	62.5 (8.7)	27.8 (5.1)	-
Forced expiratory volume in 1-second	FEV1	L	19,998 (16,515 82.6%)	1 (1, 3)	2.1 (0.8)	47.5%	63.7 (7.5)	28.4 (5.2)	3063
Forced expired volume in 1-second/forced vital capacity ratio	FEV1/FVC	%	16,541 (13,681 82.7%)	1 (1, 2)	68.6 (13.6)	47.5%	63.3 (7.5)	28.4 (5.2)	-
forced vital capacity	FVC	L	15,882 (13,133 82.7%)	1 (1, 2)	3.1 (1.0)	47.4%	63.3 (7.6)	28.4 (5.2)	3062
Framingham coronary heart disease 10 year risk score	CHD score	%	37,101 (31,113 83.9%)	1 (1, 3)	11.6 (7.5)	46.1%	59.5 (7.6)	27.8 (4.8)	-

free thyroxine	FT4	pmol/L	81,384 (68,544 84.2%)	2 (1, 5)	14.9 (3.4)	39.5%	60.3 (8.7)	27.9 (5)	-
free triiodothyronine	FT3	pmol/L	7,118 (5,886 82.7%)	2 (1, 4)	4.8 (1.2)	22.6%	59.4 (8.5)	28.1 (5.3)	-
gamma-glutamyl transferase	GGT	U/L	83,634 (69,826 83.5%)	3 (1, 6)	43.1 (46.2)	47.6%	60.6 (8.1)	27.9 (4.9)	-
glycated haemoglobin	HbA1c	mmol/mol	86,537 (72,668 84.0%)	2 (1, 4)	6.8 (1.4)	48.5%	63.3 (7.8)	28.4 (5.1)	30750
haematocrit percentage	Hct	%	157,896 (132,359 83.8%)	4 (2, 8)	41.2 (3.9)	44.8%	61 (8.8)	27.6 (4.8)	30030
haemoglobin	Hb	g/dL	159,098 (133,346 83.8%)	4 (2, 8)	13.7 (1.4)	44.8%	60.7 (9)	27.6 (4.8)	30020
high density lipoprotein cholesterol	HDL	mmol/L	157,928 (132,431 83.9%)	4 (2, 8)	1.5 (0.4)	46.6%	62 (7.7)	27.7 (4.8)	30760
Low density lipoprotein cholesterol	LDL	mmol/L	136,406 (114,475 83.9%)	3 (2, 6)	3.0 (1.0)	47.1%	61.9 (7.7)	27.8 (4.8)	30780
lymphocyte count	LYMPH	10^9 /L	156,961 (131,557 83.8%)	4 (2, 8)	2.0 (0.7)	44.8%	61.2 (8.7)	27.6 (4.8)	30120
lymphocyte percentage	LYMPH perc	%	19,300 (16,465 85.3%)	3 (2, 6)	30.9 (8.6)	44.2%	60.5 (8.5)	27.6 (4.8)	30180
mean corpuscular haemoglobin	MCH	pg	157,827 (132,277 83.8%)	4 (2, 8)	30.4 (2.0)	44.8%	60.9 (8.8)	27.6 (4.8)	30050
mean corpuscular haemoglobin concentration	MCHC	g/dL	120,754 (101,592 84.1%)	4 (2, 7)	33.4 (1.2)	44.6%	60.7 (8.8)	27.6 (4.8)	30060
mean corpuscular volume	MCV	fL	158,576 (132,911 83.8%)	4 (2, 8)	91.0 (5.2)	44.8%	60.8 (8.9)	27.6 (4.8)	30040
mean platelet volume	MPV	fL	16,667 (14,509 87.1%)	3 (2, 6)	9.8 (1.4)	44.1%	59.5 (8.6)	27.3 (5)	30100
monocyte count	MONO	10^9 /L	156,914 (131,509 83.8%)	4 (2, 8)	0.5 (0.2)	44.8%	61.2 (8.7)	27.6 (4.8)	30130
monocyte percentage	MONO perc	%	19,290 (16,454 85.3%)	3 (2, 6)	8.0 (2.3)	44.2%	60.6 (8.5)	27.6 (4.8)	30190
neutrophill count	NEUT	10^9 /L	157,085 (131,662 83.8%)	4 (2, 8)	3.8 (1.6)	44.8%	61.2 (8.7)	27.6 (4.8)	30140
neutrophill percentage	NEUT perc	%	19,265 (16,438 85.3%)	3 (2, 6)	57.3 (9.3)	44.3%	60.5 (8.6)	27.6 (4.8)	30200
non-high density lipoprotein cholesterol	NHDL	mmol/L	52,094 (43,179 82.9%)	1 (1, 2)	3.6 (1.1)	48.2%	65.7 (7.3)	28 (4.9)	-
oxygen saturation at periphery	SpO_2	%	10,659 (8,752 82.1%)	1 (1, 2)	97.0 (1.8)	46.0%	66.1 (7.2)	28.3 (5.2)	-
peak expiratory flow	PEF	L/min	34,855 (29,035 83.3%)	3 (1, 8)	396.4 (113.7)	43.5%	58.1 (9.5)	28.3 (5.2)	3064
plasma viscosity	PV	mPas	15,082 (12,597 83.5%)	1 (1, 3)	1.7 (0.1)	39.6%	58.2 (8.9)	27.5 (4.8)	-
platelet count	Plt	10^9 /L	158,721 (133,040 83.8%)	4 (2, 8)	258.2 (72.0)	44.8%	60.8 (8.9)	27.6 (4.8)	30080
prostate-specific antigen	PSA	ng/mL	38,961 (32,986 84.7%)	2 (1, 4)	2.3 (2.1)	100.0%	64 (6.9)	27.9 (4.2)	-
prothrombin time	PT	1	12,446 (10,440 83.9%)	2 (1, 32)	2.5 (0.8)	52.2%	66 (7.2)	28.7 (5.3)	-
pulse pressure	PP	mmHg	182,248 (153,061 84.0%)	12 (6, 24)	55.6 (13.9)	45.6%	59 (9.4)	27.5 (4.8)	-
pulse rate	PR	bpm	111,519 (92,954 83.4%)	2 (1, 4)	72.4 (11.8)	46.2%	63.4 (7.9)	27.8 (4.9)	102
random blood glucose	RBG	mmol/L	134,005 (112,103 83.7%)	3 (1, 5)	5.8 (2.1)	45.8%	60.4 (8.3)	27.8 (4.9)	30740
red blood cell count	RBC	10^{12} /L	157,982 (132,403 83.8%)	4 (2, 8)	4.5 (0.5)	44.8%	60.9 (8.9)	27.6 (4.8)	30010
red blood cell distribution width	RDW	%	90,355 (75,946 84.1%)	4 (2, 7)	13.5 (1.3)	44.7%	61.7 (8.6)	27.7 (4.8)	30070
serum cholesterol	TC	mmol/L	159,801 (133,931 83.8%)	5 (2, 10)	5.2 (1.2)	46.5%	61.1 (8)	27.7 (4.8)	30690
serum inorganic phosphate	Pi	mmol/L	63,106 (52,966 83.9%)	1 (1, 3)	1.1 (0.2)	41.5%	62.1 (8.3)	27.8 (5)	30810
serum iron	iron	umol/L	9,663 (8,203 84.9%)	1 (1, 2)	16.0 (7.2)	41.4%	61.3 (9.3)	27.9 (5.3)	-
serum potassium	K	mmol/L	164,066 (137,567 83.8%)	5 (3, 11)	4.4 (0.5)	45.8%	62 (8.1)	27.7 (4.8)	-
serum sodium	Na	mmol/L	164,285 (137,762 83.9%)	5 (3, 11)	140.1 (2.8)	45.8%	62 (8.1)	27.7 (4.8)	-
serum triglyceride	TG	mmol/L	151,901 (127,363 83.8%)	4 (2, 8)	1.6 (0.9)	46.8%	61.6 (7.8)	27.7 (4.8)	30870

serum urate	urate	umol/L	29,907 (24,917 83.3%)	1 (1, 2)	355.9 (104.2)	54.3%	60.3 (8.4)	28.7 (5)	30880
serum urea	urea	mmol/L	156,186 (130,830 83.8%)	5 (2, 10)	5.8 (2.0)	45.9%	61.8 (8.1)	27.7 (4.8)	30670
systolic blood pressure	SBP	mmHg	182,248 (153,061 84.0%)	12 (6, 24)	136.0 (17.7)	45.6%	59 (9.4)	27.5 (4.8)	4080
thyroid stimulating hormone	TSH	mU/L	145,519 (121,916 83.8%)	3 (2, 6)	2.1 (1.5)	43.2%	60.4 (8.7)	27.7 (4.9)	-
total bilirubin	Total bili	umol/L	160,595 (134,624 83.8%)	5 (2, 9)	10.9 (5.6)	45.8%	61.7 (8.1)	27.7 (4.8)	30840
total protein	TP	g/L	111,653 (93,447 83.7%)	4 (2, 8)	70.7 (4.3)	45.9%	61.5 (8)	27.7 (4.8)	30860
urine albumin	UAlb	mg/L	31,123 (26,014 83.6%)	3 (1, 6)	15.4 (28.4)	54.0%	64.8 (7.1)	29.9 (5.4)	30500
urine albumin/creatinine ratio	UACR	mg/mmol	31,123 (26,014 83.6%)	3 (1, 6)	1.9 (3.3)	54.0%	64.8 (7.1)	29.9 (5.4)	-
urine creatinine	UCr	mmol/L	31,123 (26,014 83.6%)	3 (1, 6)	8.9 (4.9)	54.0%	64.8 (7.1)	29.9 (5.4)	30510
vitamin B12	VB12	ng/L	58,599 (49,061 83.7%)	1 (1, 2)	379.9 (180.0)	39.7%	62.1 (8.8)	27.7 (5.1)	
Vitamin D	VD	nmol/L	8,976 (7,601 84.7%)	1 (1, 2)	58.2 (29.2)	28.4%	62.9 (7.9)	27.4 (5.4)	30890
waist circumference	Waist	cm	45,561 (37,619 82.6%)	1 (1, 2)	94.7 (15.7)	47.5%	61.6 (7.6)	27.8 (4.9)	48
white blood cell count	WBC	10^9 /L	158,727 (133,039 83.8%)	4 (2, 8)	6.6 (1.9)	44.8%	60.8 (8.9)	27.6 (4.8)	30000

6

Supplementary Table 5. Mean chi-square statistics of $SPA_{GRM(LMM)}$, $SPA_{GRM(GEEexc)}$, $SPA_{GRM(GEEar1)}$, and $SPA_{GRM(CCT)}$ methods for longitudinal traits using different generative models.

Simulation conditions			Mean chi-square statistics		
Family relatedness	Variant type	Methods	Scenario a	Scenario b	Scenario c
A: 25,000 Unrelated	Common variants	$SPA_{GRM(LMM)}$	**44.9**	94.9	85.6
		$SPA_{GRM(GEEexc)}$	44.1	**94.9**	85.6
		$SPA_{GRM(GEEar1)}$	40.6	88	**87.8**
		$SPA_{GRM(CCT)}$	43.5	94.2	87.5
	Rare variants	$SPA_{GRM(LMM)}$	**22.7**	47.5	42.9
		$SPA_{GRM(GEEexc)}$	22.3	**47.5**	42.9
		$SPA_{GRM(GEEar1)}$	20.1	44.4	**43.8**
		$SPA_{GRM(CCT)}$	21.8	46.9	43.7
B: 6,250 Small families	Common variants	$SPA_{GRM(LMM)}$	**39.8**	68	63.7
		$SPA_{GRM(GEEexc)}$	39.1	**68**	63.7
		$SPA_{GRM(GEEar1)}$	37.2	64.6	**64.9**
		$SPA_{GRM(CCT)}$	38.8	67.5	64.9
	Rare variants	$SPA_{GRM(LMM)}$	**19.1**	31.8	29.6
		$SPA_{GRM(GEEexc)}$	18.9	**31.8**	29.6
		$SPA_{GRM(GEEar1)}$	17.4	30.3	**29.8**
		$SPA_{GRM(CCT)}$	18.4	31.4	29.9
C: 2,500 Large families	Common variants	$SPA_{GRM(LMM)}$	**35.7**	58	53.9
		$SPA_{GRM(GEEexc)}$	35.2	**58**	53.9
		$SPA_{GRM(GEEar1)}$	33.2	55.6	54.7
		$SPA_{GRM(CCT)}$	34.9	57.5	**54.7**
	Rare variants	$SPA_{GRM(LMM)}$	**19.6**	30	29.3
		$SPA_{GRM(GEEexc)}$	19.3	**30**	29.3
		$SPA_{GRM(GEEar1)}$	17.9	29	29.5
		$SPA_{GRM(CCT)}$	18.9	29.8	29.5
D: A + B	Common variants	$SPA_{GRM(LMM)}$	**80.5**	156.8	144.8
		$SPA_{GRM(GEEexc)}$	78.9	**156.8**	144.8
		$SPA_{GRM(GEEar1)}$	73.6	147.4	**148**
		$SPA_{GRM(CCT)}$	78.4	156	147.7
	Rare variants	$SPA_{GRM(LMM)}$	**40.1**	72.6	68
		$SPA_{GRM(GEEexc)}$	39.5	**72.6**	68
		$SPA_{GRM(GEEar1)}$	35.9	69	**68.8**
		$SPA_{GRM(CCT)}$	38.8	71.9	68.8
E: A + C	Common variants	$SPA_{GRM(LMM)}$	**73.1**	140.5	129
		$SPA_{GRM(GEEexc)}$	71.5	**140.5**	129
		$SPA_{GRM(GEEar1)}$	66.9	132.5	**131.6**
		$SPA_{GRM(CCT)}$	70.9	139.7	131.5
	Rare variants	$SPA_{GRM(LMM)}$	**37.3**	63.6	60.3
		$SPA_{GRM(GEEexc)}$	36.9	**63.6**	60.3
		$SPA_{GRM(GEEar1)}$	33.6	60.7	**60.9**
		$SPA_{GRM(CCT)}$	36.2	62.9	60.9

Longitudinal traits were simulated using scenario (**a**) linear mixed model (LMM); scenario (**b**) generalized estimation equations (GEE) model with exchangeable working correlation structure and scenario (**c**) GEE model with autoregressive working correlation structure. Bold numbers represent superior performance in statistical power in each scenario. $SPA_{GRM(LMM)}$ and $SPA_{GRM(GEEexc)}$ performed very similarly in scenarios (**b**) and (**c**), as LMM with only a random intercept shares the same correlation structure as GEE with an exchangeable correlation structure. P value is calculated using two-sided score tests. Exact p values are provided in **Supplementary Data 5**.

Supplementary Table 6. Detailed covariates adjustment for each longitudinal trait.

Trait	Acronyms	Covariates
activated partial thromboplastin time	aPTT	Mean formula: aPTT ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: aPTT ~ 1 + sex + age + age2 + age*sex + BMI.
adjusted calcium	adj Ca	Mean formula: adj Ca ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: adj Ca ~ 1 + sex + age + age2 + age*sex + BMI.
alanine aminotransferase	ALT	Mean formula: ALP ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: ALP ~ 1 + sex + age + age2 + age*sex + BMI.
alkaline phosphatase	ALP	Mean formula: ALP ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: ALP ~ 1 + sex + age + age2 + age*sex + BMI.
amylase	amylase	Mean formula: amylase ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: amylase ~ 1 + sex + age + age2 + age*sex + BMI.
aspartate aminotransferase	AST	Mean formula: AST ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: AST ~ 1 + sex + age + age2 + age*sex + BMI.
basophil count	BASO	Mean formula: BASO ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: BASO ~ 1 + sex + age + age2 + age*sex + BMI.
basophil percentage	BASO perc	Mean formula: BASO perc ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: BASO perc ~ 1 + sex + age + age2 + age*sex + BMI.
bicarbonate	HCO_3	Mean formula: HCO3 ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: HCO3 ~ 1 + sex + age + age2 + age*sex + BMI.
blood albumin	Alb	Mean formula: Alb ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Alb ~ 1 + sex + age + age2 + age*sex + BMI.
blood creatinine	Cr	Mean formula: Cr ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Cr ~ 1 + sex + age + age2 + age*sex + BMI.
blood globulin	Glob	Mean formula: Glob ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Glob ~ 1 + sex + age + age2 + age*sex + BMI.
body mass index	BMI	Mean formula: BMI ~ 1 + sex + age + age2 + age*sex + 10PCs. WS formula: BMI ~ 1 + sex + age + age2 + age*sex.
calcium	Ca	Mean formula: Ca ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Ca ~ 1 + sex + age + age2 + age*sex + BMI.
carbohydrate antigen 125	CA125	Mean formula: CA125 ~ 1 + age + age2 + BMI + 10PCs. WS formula: CA125 ~ 1 + age + age2 + BMI.
chloride	Cl	Mean formula: Cl ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Cl ~ 1 + sex + age + age2 + age*sex + BMI.
C-reactive protein	CRP	Mean formula: CRP ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: CRP ~ 1 + sex + age + age2 + age*sex + BMI.
creatine kinase	CK	Mean formula: CK ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: CK ~ 1 + sex + age + age2 + age*sex + BMI.

diastolic blood pressure	DBP	Mean formula: DBP (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: DBP (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs.
eosinophil count	EOS	Mean formula: EOS ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: EOS ~ 1 + sex + age + age2 + age*sex + BMI.
eosinophil percentage	EOS perc	Mean formula: EOS perc ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: EOS perc ~ 1 + sex + age + age2 + age*sex + BMI.
erythrocyte sedimentation rate	ESR	Mean formula: ESR ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: ESR ~ 1 + sex + age + age2 + age*sex + BMI.
estimated glomerular filtration rate	eGFR	Mean formula: eGFR ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: eGFR ~ 1 + sex + age + age2 + age*sex + BMI.
fasting blood glucose	FBG	Mean formula: FBG ~ 1 + sex + age + age2 + age*sex + BMI + drug_status + 10PCs. WS formula: FBG ~ 1 + sex + age + age2 + age*sex + BMI + drug_status.
ferritin	ferritin	Mean formula: ferritin ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: ferritin ~ 1 + sex + age + age2 + age*sex + BMI.
folate	folate	Mean formula: folate ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: folate ~ 1 + sex + age + age2 + age*sex + BMI.
Forced expiratory volume in 1-second	FEV1	Mean formula: FEV1 ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status + 10PCs. WS formula: FEV1 ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status.
Forced expired volume in 1-second/forced vital capacity ratio	FEV1/FVC	Mean formula: FEV1/FVC ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status + 10PCs. WS formula: FEV1/FVC ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status.
forced vital capacity	FVC	Mean formula: FVC ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status + 10PCs. WS formula: FVC ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status.
Framingham coronary heart disease 10 year risk score	CHDscore	Mean formula: CHDscore ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: CHDscore ~ 1 + sex + age + age2 + age*sex + BMI.
free thyroxine	FT4	Mean formula: FT4 ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: FT4 ~ 1 + sex + age + age2 + age*sex + BMI.
free triiodothyronine	FT3	Mean formula: FT3 ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: FT3 ~ 1 + sex + age + age2 + age*sex + BMI.
gamma-glutamyl transferase	GGT	Mean formula: GGT ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: GGT ~ 1 + sex + age + age2 + age*sex + BMI.
glycated haemoglobin	HbA1c	Mean formula: HbA1c ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: HbA1c ~ 1 + sex + age + age2 + age*sex + BMI.
haematocrit percentage	Hct	Mean formula: Hct ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Hct ~ 1 + sex + age + age2 + age*sex + BMI.
haemoglobin	Hb	Mean formula: Hb ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Hb ~ 1 + sex + age + age2 + age*sex + BMI.

high density lipoprotein cholesterol	HDL	Mean formula: HDL (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: HDL (shifted) ~ 1 + sex + age + age2 + age*sex + BMI.
Low density lipoprotein cholesterol	LDL	Mean formula: LDL (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: LDL (shifted) ~ 1 + sex + age + age2 + age*sex + BMI.
lymphocyte count	LYMPH	Mean formula: LYMPH ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: LYMPH ~ 1 + sex + age + age2 + age*sex + BMI.
lymphocyte percentage	LYMPH perc	Mean formula: LYMPH perc ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: LYMPH perc ~ 1 + sex + age + age2 + age*sex + BMI.
mean corpuscular haemoglobin	MCH	Mean formula: MCH ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: MCH ~ 1 + sex + age + age2 + age*sex + BMI.
mean corpuscular haemoglobin concentration	MCHC	Mean formula: MCHC ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: MCHC ~ 1 + sex + age + age2 + age*sex + BMI.
mean corpuscular volume	MCV	Mean formula: MCV ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: MCV ~ 1 + sex + age + age2 + age*sex + BMI.
mean platelet volume	MPV	Mean formula: MPV ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: MPV ~ 1 + sex + age + age2 + age*sex + BMI.
monocyte count	MONO	Mean formula: MONO ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: MONO ~ 1 + sex + age + age2 + age*sex + BMI.
monocyte percentage	MONO perc	Mean formula: MONO perc ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: MONO perc ~ 1 + sex + age + age2 + age*sex + BMI.
neutrophill count	NEUT	Mean formula: NEUT ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: NEUT ~ 1 + sex + age + age2 + age*sex + BMI.
neutrophill percentage	NEUT perc	Mean formula: NEUT perc ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: NEUT perc ~ 1 + sex + age + age2 + age*sex + BMI.
non-high density lipoprotein cholesterol	NHDL	Mean formula: NHDL ~ 1 + sex + age + age2 + age*sex + BMI + drug_status + 10PCs. WS formula: NHDL ~ 1 + sex + age + age2 + age*sex + BMI + drug_status.
oxygen saturation at periphery	SpO_2	Mean formula: SpO2 ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status + 10PCs. WS formula: SpO2 ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status.
peak expiratory flow	PEF	Mean formula: PEF ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status + 10PCs. WS formula: PEF ~ 1 + sex + age + age2 + age*sex + BMI + genotyping_array + smoke_status.
plasma viscosity	PV	Mean formula: PV ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: PV ~ 1 + sex + age + age2 + age*sex + BMI.
platelet count	Plt	Mean formula: Plt ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Plt ~ 1 + sex + age + age2 + age*sex + BMI.
prostate-specific antigen	PSA	Mean formula: PSA ~ 1 + age + age2 + BMI + 10PCs. WS formula: PSA ~ 1 + age + age2 + BMI.
prothrombin time	PT	Mean formula: PT ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: PT ~ 1 + sex + age + age2 + age*sex + BMI.

pulse pressure	PP	Mean formula: PP (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: PP (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs.
pulse rate	PR	Mean formula: PR ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: PR ~ 1 + sex + age + age2 + age*sex + BMI.
random blood glucose	RBG	Mean formula: RBG ~ 1 + sex + age + age2 + age*sex + BMI + drug_status + 10PCs. WS formula: RBG ~ 1 + sex + age + age2 + age*sex + BMI + drug_status.
red blood cell count	RBC	Mean formula: RBC ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: RBC ~ 1 + sex + age + age2 + age*sex + BMI.
red blood cell distribution width	RDW	Mean formula: RDW ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: RDW ~ 1 + sex + age + age2 + age*sex + BMI.
serum cholesterol	TC	Mean formula: TC (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: TC (shifted) ~ 1 + sex + age + age2 + age*sex + BMI.
serum inorganic phosphate	Pi	Mean formula: Pi ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Pi ~ 1 + sex + age + age2 + age*sex + BMI.
serum iron	iron	Mean formula: iron ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: iron ~ 1 + sex + age + age2 + age*sex + BMI.
serum potassium	K	Mean formula: K ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: K ~ 1 + sex + age + age2 + age*sex + BMI.
serum sodium	Na	Mean formula: Na ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Na ~ 1 + sex + age + age2 + age*sex + BMI.
serum triglyceride	TG	Mean formula: TG (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: TG (shifted) ~ 1 + sex + age + age2 + age*sex + BMI.
serum urate	urate	Mean formula: urate ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: urate ~ 1 + sex + age + age2 + age*sex + BMI.
serum urea	urea	Mean formula: urea ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: urea ~ 1 + sex + age + age2 + age*sex + BMI.
systolic blood pressure	SBP	Mean formula: SBP (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: SBP (shifted) ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs.
thyroid stimulating hormone	TSH	Mean formula: TSH ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: TSH ~ 1 + sex + age + age2 + age*sex + BMI.
total bilirubin	Total bili	Mean formula: Total bili ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Total bili ~ 1 + sex + age + age2 + age*sex + BMI.
total protein	TP	Mean formula: TP ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: TP ~ 1 + sex + age + age2 + age*sex + BMI.
urine albumin	UAlb	Mean formula: UAlb ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: UAlb ~ 1 + sex + age + age2 + age*sex + BMI.
urine albumin/creatinine ratio	UACR	Mean formula: UACR ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: UACR ~ 1 + sex + age + age2 + age*sex + BMI.

urine creatinine	UCr	Mean formula: UCr ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: UCr ~ 1 + sex + age + age2 + age*sex + BMI.
vitamin B12	VB12	Mean formula: VB12 ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: VB12 ~ 1 + sex + age + age2 + age*sex + BMI.
Vitamin D	VD	Mean formula: VD ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: VD ~ 1 + sex + age + age2 + age*sex + BMI.
waist circumference	Waist	Mean formula: Waist ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: Waist ~ 1 + sex + age + age2 + age*sex + BMI.
white blood cell count	WBC	Mean formula: WBC ~ 1 + sex + age + age2 + age*sex + BMI + 10PCs. WS formula: WBC ~ 1 + sex + age + age2 + age*sex + BMI.

We adjusted for the first ten principal components (PCs) on the mean component. PCs were not adjusted for the WS variance because this makes no differences for the final results.

Supplemental References

1. Liang, K.-Y. & Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22 (1986).
2. Ballinger, G.A. Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods* **7**, 127-150 (2016).
3. Gardiner, J.C., Luo, Z. & Roman, L.A. Fixed effects, random effects and GEE: what are the differences? *Stat Med* **28**, 221-39 (2009).
4. Zorn, C.J.W. Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *American Journal of Political Science* **45**, 470 (2001).
5. Fitzmaurice, G.M. A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* **51**, 309-17 (1995).
6. Chow, C. & Liu, C. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory* **14**, 462-467 (1968).
7. Prim, R.C. Shortest Connection Networks And Some Generalizations. *Bell System Technical Journal* **36**, 1389-1401 (1957).
8. Wu, Z. *et al.* TSH-TSHR axis promotes tumor immune evasion. *J Immunother Cancer* **10**(2022).
9. Shi, R.L. *et al.* The Usefulness of Preoperative Thyroid-Stimulating Hormone for Predicting Differentiated Thyroid Microcarcinoma. *Otolaryngol Head Neck Surg* **154**, 256-62 (2016).
10. Zhou, W. *et al.* GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. *Nat Commun* **11**, 3981 (2020).
11. Williams, A.T. *et al.* Genome-wide association study of thyroid-stimulating hormone highlights new genes, pathways and associations with thyroid disease. *Nat Commun* **14**, 6713 (2023).
12. Liu, W. *et al.* Identification of BACH2 as a susceptibility gene for Graves' disease in the Chinese Han population based on a three-stage genome-wide association study. *Hum Genet* **133**, 661-71 (2014).
13. Gomes-Lima, C.J. *et al.* A Novel Risk Stratification System for Thyroid Nodules With Indeterminate Cytology-A Pilot Cohort Study. *Front Endocrinol (Lausanne)* **11**, 53 (2020).
14. Zhang, Y. *et al.* Genomic binding and regulation of gene expression by the thyroid carcinoma-associated PAX8-PPARG fusion protein. *Oncotarget* **6**, 40418-32 (2015).
15. Zhou, X., Curbo, S., Li, F., Krishnan, S. & Karlsson, A. Inhibition of glutamate oxaloacetate transaminase 1 in cancer cell lines results in altered metabolism with increased dependency of glucose. *BMC Cancer* **18**, 559 (2018).
16. Putlyaeva, L.V. *et al.* PTPN11 Knockdown Prevents Changes in the Expression of Genes Controlling Cell Cycle, Chemotherapy Resistance, and Oncogene-Induced Senescence in Human Thyroid Cells Overexpressing BRAF V600E Oncogenic Protein. *Biochemistry (Mosc)* **85**, 108-118 (2020).
17. Ko, S. *et al.* GWAS of longitudinal trajectories at biobank scale. *The American Journal of Human Genetics* (2022).
18. Yang, J. *et al.* FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267-72 (2012).
19. Ivarsdottir, E.V. *et al.* Effect of sequence variants on variance in glucose levels predicts type

2 diabetes risk and accounts for heritability. *Nat Genet* **49**, 1398-1402 (2017).
20. Hulse, A.M. & Cai, J.J. Genetic variants contribute to gene expression variability in humans. *Genetics* **193**, 95-108 (2013).
21. Wang, H. *et al.* Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci Adv* **5**, eaaw3538 (2019).
22. Jurgens, S.J. *et al.* Adjusting for common variant polygenic scores improves yield in rare variant association analyses. *Nat Genet* **55**, 544-548 (2023).
23. Campos, A.I. *et al.* Boosting the power of genome-wide association studies within and across ancestries by using polygenic scores. *Nat Genet* **55**, 1769-1776 (2023).
24. Bennett, D., O'Shea, D., Ferguson, J., Morris, D. & Seoighe, C. Controlling for background genetic effects using polygenic scores improves the power of genome-wide association studies. *Sci Rep* **11**, 19571 (2021).
25. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* **46**, 100-6 (2014).
26. Listgarten, J. *et al.* Improved linear mixed models for genome-wide association studies. *Nat Methods* **9**, 525-6 (2012).
27. Choi, S.W., Mak, T.S. & O'Reilly, P.F. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* **15**, 2759-2772 (2020).
28. German, C.A., Sinsheimer, J.S., Zhou, J. & Zhou, H. WiSER: Robust and scalable estimation and inference of within-subject variances from intensive longitudinal data. *Biometrics* (2021).
29. Dzubur, E. *et al.* 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 (2020).
30. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* **47**, 284-290 (2015).
31. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics* **50**, 1335-1341 (2018).
32. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool for biobank-scale data. *Nature Genetics* **53**, 1616-1621 (2021).
33. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A.P. & Price, A.L. Mixed-model association for biobank-scale datasets. *Nature genetics* **50**, 906-908 (2018).
34. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nature Genetics* **51**, 1749-1755 (2019).
35. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* **53**, 1097-1103 (2021).
36. Chen, C.-F. Score Tests for Regression Models. *Journal of the American Statistical Association* **78**, 158 (1983).