Supplementary Notes for
"SPA_{SQR} applies smoothed quantile regression to detect heterogeneous genetic associations "

Contents

List of Figures

List of Tables

1 Additional Methods

1.1 Partial normal approximation of the cumulant generating function

The saddlepoint approximation (SPA) provides accurate tail probabilities for the score statistic $S = \sum_{i=1}^{n} R_i G_i$, where R_i are residuals from the null quantile regression model (smoothed or nonsmooth) and $G_i \in \{0, 1, 2\}$ is the genotype. The method relies on the moment generating function (MGF) of S. Under the null hypothesis, treating G_i as independent Binomial$(2, \mu)$ variables with μ the alternative allele frequency, the MGF is

$$M_S(t) = \mathbb{E}(e^{tS}) = \prod_{i=1}^{n} \left[1 - \mu + \mu e^{R_i t}\right]^2,$$

so the cumulant generating function (CGF) is

$$K_S(t) = \log M_S(t) = 2 \sum_{i=1}^{n} \log(1 - \mu + \mu e^{R_i t}).$$

Direct evaluation of K_S and its derivatives over all n individuals is computationally expensive because of the exponential and logarithmic operations. We therefore adopt the partial normal approximation idea of Dey et al. [1].

Specifically, we classify residuals into outliers and non-outliers using an IQR rule with clipping. Let $\alpha_{0.25}$ and $\alpha_{0.75}$ be the 25th and 75th percentiles of the residual vector $\boldsymbol{R}$, and $\text{IQR} = \alpha_{0.75} - \alpha_{0.25}$. A residual R_i is flagged as an outlier if

$$R_i \notin \left[\max(\alpha_{0.25} - 1.5\,\text{IQR}, -0.55),\ \min(\alpha_{0.75} + 1.5\,\text{IQR}, 0.55)\right].$$

For nonsmooth quantile regression, values such as $\pm 0.6, \pm 0.7, \pm 0.8, \pm 0.9$ are deemed outliers; for smoothed quantile regression, the proportion of outliers usually ranges from 0% to 25%, depending on the bandwidth h and quantile τ.

Without loss of generality, reorder the participants so that the first n_1 have outlier residuals and the remaining $n - n_1$ are non-outliers. The score statistic decomposes as

$$S = \sum_{i=1}^{n_1} R_i G_i + \sum_{i=n_1+1}^{n} R_i G_i =: S_{\text{out}} + S_{\text{non}}.$$

For the non-outlier component, we approximate its distribution by a normal distribution with mean and variance:

$$\mathbb{E}(S_{\text{non}}) = 2\mu\, \boldsymbol{R}_{\text{non}}^{\top} \mathbf{1}_{n-n_1}, \qquad \text{Var}(S_{\text{non}}) = 2\mu(1-\mu)\, \boldsymbol{R}_{\text{non}}^{\top} \boldsymbol{R}_{\text{non}},$$

where $\boldsymbol{R}_{\text{non}}$ is the subvector of residuals for the non-outliers. Consequently, the CGF of S_{non} is

$$K_{S_{\text{non}}}(t) = \mathbb{E}(S_{\text{non}})\, t + \tfrac{1}{2} \text{Var}(S_{\text{non}})\, t^2,$$

and the overall CGF is approximated by

$$K_S(t) = 2 \sum_{i=1}^{n_1} \log(1 - \mu + \mu e^{R_i t}) + \mathbb{E}(S_{\text{non}})\, t + \tfrac{1}{2} \text{Var}(S_{\text{non}})\, t^2. \tag{1}$$

We emphasise that $\mathbb{E}(S_{\text{non}})$ is generally nonzero, even when the null model includes an intercept.

With this approximate CGF, the first two derivatives of K_S are

$$K_S'(t) = 2 \sum_{i=1}^{n_1} \frac{\mu R_i e^{R_i t}}{1 - \mu + \mu e^{R_i t}} + \mathbb{E}(S_{\text{non}}) + \text{Var}(S_{\text{non}})\, t, \tag{2}$$

$$K_S''(t) = 2 \sum_{i=1}^{n_1} \frac{\mu(1-\mu) R_i^2 e^{R_i t}}{(1 - \mu + \mu e^{R_i t})^2} + \text{Var}(S_{\text{non}}). \tag{3}$$

which are used during the process of saddle point approximation.

1.2 The SPA$_{\text{SQR}}$ algorithm

Algorithm 1 gives the complete recipe of SPA$_{\text{SQR}}$ applied to a single phenotype Y. In what follows, note that the residuals R_i are specific to chromosome (c) and quantile (τ), but we omit subscripts for notational simplicity.

Algorithm 1 SPA$_{\text{SQR}}$ — Smoothed Quantile Regression Association Testing

Input: Phenotype $\boldsymbol{Y} \in \mathbb{R}^n$; covariate matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ (including intercept); genotype matrix $\boldsymbol{G} \in \mathbb{R}^{n\times m}$; quantile levels $\{\tau_1, \ldots, \tau_d\}$.

Output: Per-quantile p-values $\{p_{\tau_k}\}$ and combined p-value p_{cct} for each variant.

Step 0a: LOCO polygenic score construction

Apply rank-based inverse normal transformation to $\boldsymbol{Y}$ to obtain $\tilde{\boldsymbol{Y}}$, then compute LOCO polygenic scores $\hat{\boldsymbol{Y}}_{-c}$ for each chromosome $c \in \{1, 2, \ldots, 22\}$ using LDAK-KVIK Step 1.

Step 0b (optional): Sparse GRM construction

Compute genetic relationship matrix $\boldsymbol{\Phi}$ using PLINK 2, retaining pairs with genetic correlation coefficient above a chosen threshold (e.g. 0.05); if sparse GRM is not provided set $\boldsymbol{\Phi} \leftarrow \boldsymbol{I}_n$.

Step 1: Fit null smoothed quantile regression models

for each chromosome c and quantile level τ **do**

Compute bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$.

Fit the null model by minimizing $f_{h,\tau}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \ell_{h,\tau}(\tilde{Y}_i - \hat{Y}_{-c,i} - \boldsymbol{X}_i^\top \boldsymbol{\beta})$, null model coefficient is denoted as $\tilde{\boldsymbol{\beta}}_{h,\tau}$.

Obtain smoothed quantile regression residuals $R_i = \tau - \mathcal{K}\Big(-\frac{e_i(\tilde{\boldsymbol{\beta}}_{h,\tau})}{h}\Big)$.

Classify each R_i as outlier or non-outlier via the IQR rule (see Methods).

Step 2: Per-variant association testing

for each variant $\boldsymbol{G}_j$ (column j of $\boldsymbol{G}$) on chromosome c **do**

for each quantile level τ **do**

Score statistic $s = \sum_{i=1}^{n} R_i\, G_{ij}$.

Null variance: $\widehat{\sigma}_g^2(\boldsymbol{G}_j) = \frac{1}{n-1}\sum_{i=1}^{n}(G_{ij} - \bar{G}_j)^2$, $\quad \text{Var}(S) = \widehat{\sigma}_g^2(\boldsymbol{G}_j)\, \boldsymbol{R}^\top \boldsymbol{\Phi}\, \boldsymbol{R}$.

if $|s| < r\sqrt{\text{Var}(S)}$ with $r = 2$ **then**

Normal approximation: $p_\tau = 2\,\mathcal{K}(-|s|/\sqrt{\text{Var}(S)})$.

else

Form the CGF via partial normal approximation:

$$K_S(t) \;=\; 2\sum_{i=1}^{n_1} \log(1 - \mu + \mu e^{R_i t}) + \mathbb{E}[S_{\text{non}}]\, t + \tfrac{1}{2}\,\text{Var}(S_{\text{non}})\, t^2,$$

Variance-adjusted score $s_{\text{adj}} = s\sqrt{\text{Var}_{\text{CGF}}(S)/\,\text{Var}(S)}$, where $\text{Var}_{\text{CGF}}(S) = K_S''(0)$.

Solve $K_S'(\zeta) = s_{\text{adj}}$ for ζ via Newton's method.

$\omega = \text{sgn}(\zeta)\sqrt{2(\zeta s_{\text{adj}} - K_S(\zeta))}$, $\quad \nu = \zeta\sqrt{K_S''(\zeta)}$.

Tail probability $\Pr(S < s_{\text{adj}}) \approx \mathcal{K}\Big(\omega + \frac{1}{\omega}\log(\nu/\omega)\Big)$.

Two-sided p-value: $p_\tau = \Pr(S < -|s_{\text{adj}}|) + (1 - \Pr(S < |s_{\text{adj}}|))$.

Cauchy combination: $p_{\text{cct}} = \frac{1}{2} - \frac{1}{\pi}\arctan\Big(\frac{1}{d}\sum_{k=1}^{d}\tan\big((0.5 - p_{\tau_k})\pi\big)\Big)$.

return $\{p_{\tau_k}\}$ and p_{cct} for each variant.

2 Solving the null smoothed quantile regression model

This section gives the algorithmic details behind line 5 of Algorithm 1, which fits the null smoothed quantile regression model [2, 3] on the trait after LOCO PGS offset $\boldsymbol{Y}^{(c)} := \tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c}$.

2.1 Objective function

Let $\rho_\tau(u) = u(\tau - 1\{u < 0\})$ denote the standard check loss at quantile $\tau \in (0, 1)$, and let $K_h(u) = h^{-1}\phi(u/h)$ be a Gaussian kernel of bandwidth $h > 0$, where ϕ is the standard normal density. The convolution-smoothed quantile regression loss is

$$\ell_{h,\tau}(u) \;=\; (\rho_\tau \star K_h)(u) \;=\; \int_{-\infty}^{\infty} \rho_\tau(u - v)\, K_h(v)\, \mathrm{d}v. \tag{4}$$

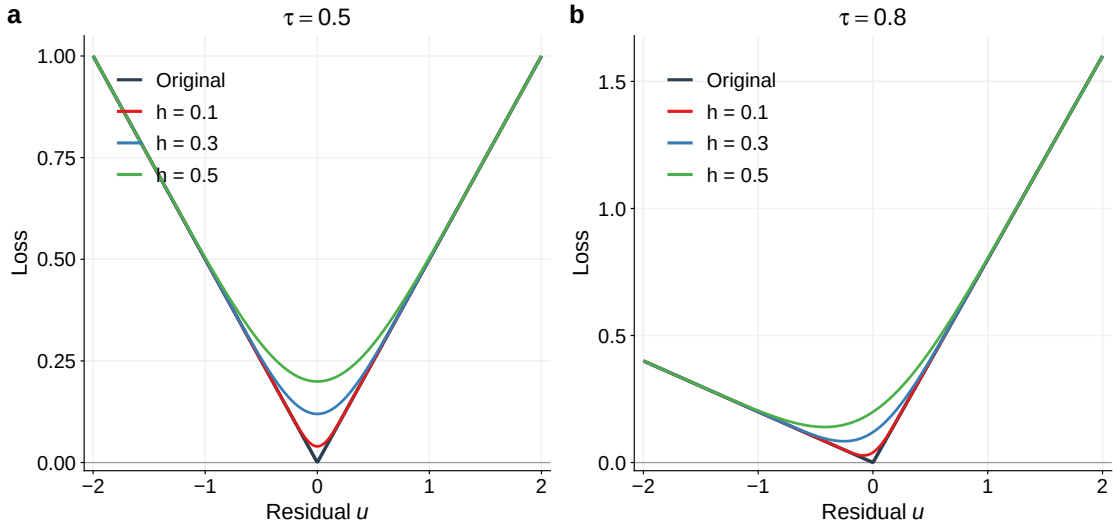
For Gaussian kernel this admits the closed form

$$\ell_{h,\tau}(u) \;=\; (\tau - \tfrac{1}{2})\, u \;+\; \tfrac{u}{2}\,[2\mathcal{K}(u/h) - 1] \;+\; h\,\phi(u/h), \tag{5}$$

where $\mathcal{K}$ is the standard-normal CDF. The optimization objective is

$$f_{h,\tau}(\boldsymbol{\beta}) \;=\; \frac{1}{n}\sum_{i=1}^{n} \ell_{h,\tau}(Y_i^{(c)} - \boldsymbol{X}_i^\top \boldsymbol{\beta}). \tag{6}$$

$f_{h,\tau}$ is twice continuously differentiable, convex, and has a Lipschitz-continuous gradient [3]. As $h \downarrow 0$, $\ell_{h,\tau}(u) \to \rho_\tau(u)$ pointwise, so the minimiser of (6) approaches the classic nonsmooth quantile regression estimate.



Supplementary Figure 1: Convolution smoothing of the quantile check loss. The original quantile check loss ρ_τ (dark) and its Gaussian-kernel convolution-smoothed counterpart $\ell_{h,\tau}$ from Equation (5), evaluated at bandwidths $h \in \{0.1, 0.3, 0.5\}$ and shown at **a** $\tau = 0.5$ and **b** $\tau = 0.8$.

Differentiating (6) gives the closed-form gradient

$$\nabla_\beta f_{h,\tau}(\boldsymbol{\beta}) \;=\; -\frac{1}{n}\,\boldsymbol{X}^\top \boldsymbol{d}(\boldsymbol{\beta}), \qquad d_i(\boldsymbol{\beta}) \;=\; \tau \;-\; \mathcal{K}\!\left(-\frac{Y_i^{(c)} - \boldsymbol{X}_i^\top \boldsymbol{\beta}}{h}\right). \tag{7}$$

The smoothed quantile regression residuals R_i in line 6 of Algorithm 1 are exactly the entries $d_i(\boldsymbol{\beta})$ at the optimal solution $\tilde{\boldsymbol{\beta}}_{h,\tau}$ to problem (6).

2.2 Quadratic majorization-minimization with extrapolation

We minimize (6) via the Quadratic Majorization-Minimization with Extrapolation (QMME) algorithm, developed in our prior works [4, 5]. Compared with the Barzilai-Borwein gradient descent algorithm of He et al. [3], QMME has rigorous convergence guarantees and exhibits stable, monotone descent behavior in practice, whereas Barzilai-Borwein iterates often oscillate and may even diverge. QMME is at least as fast as Barzilai-Borwein on well-conditioned problems and substantially faster on ill-conditioned ones. It additionally recycles the Gram matrix and its Cholesky decomposition across quantile levels, further reducing computational cost.

The Gaussian-smoothed quantile loss has bounded second derivative, $\ell''_{h,\tau}(u) = K_h(u) \leq 1/(\sqrt{2\pi}\,h)$, so that the Hessian matrix is $\nabla^2 f_{h,\tau}(\boldsymbol{\beta}) = (1/n)\,\boldsymbol{X}^\top \boldsymbol{W}(\boldsymbol{\beta})\,\boldsymbol{X}$ with $W_{ii}(\boldsymbol{\beta}) = K_h\big(Y_i^{(c)} - \boldsymbol{X}_i^\top \boldsymbol{\beta}\big)$. $\nabla^2 f_{h,\tau}(\boldsymbol{\beta})$ is upper bounded, uniformly in $\boldsymbol{\beta}$, by the positive definite matrix

$$\boldsymbol{H} \;=\; \frac{1}{n\sqrt{2\pi}\,h}\,\boldsymbol{X}^\top \boldsymbol{X} \;+\; \delta\,\boldsymbol{I}_p, \tag{8}$$

where $\delta = 10^{-6}$ is a small ridge term introduced to improve numerical stability. This yields a global quadratic upper bound

$$f_{h,\tau}(\boldsymbol{\beta}) \;\leq\; f_{h,\tau}(\boldsymbol{\beta}') + \nabla f_{h,\tau}(\boldsymbol{\beta}')^\top(\boldsymbol{\beta} - \boldsymbol{\beta}') + \tfrac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}')^\top \boldsymbol{H}\,(\boldsymbol{\beta} - \boldsymbol{\beta}'), \quad \forall\,\boldsymbol{\beta}, \boldsymbol{\beta}' \in \mathbb{R}^p. \tag{9}$$

Let $\ell_k \in \mathbb{N}$ denote a momentum counter, initialised at $\ell_1 = 1$. Each QMME iteration extrapolates the current iterate based on the previous iterate and then minimizes the majorant:

$$\boldsymbol{\eta}^{(k)} \;=\; \boldsymbol{\beta}^{(k)} \;+\; \frac{\ell_k}{\ell_k + 2}\,(\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}^{(k-1)}), \tag{10}$$

$$\boldsymbol{\beta}^{(k+1)} \;=\; \boldsymbol{\eta}^{(k)} \;-\; \boldsymbol{H}^{-1}\,\nabla f_{h,\tau}(\boldsymbol{\eta}^{(k)}). \tag{11}$$

After each step the counter is updated as $\ell_{k+1} = \ell_k + 1$. We reset $\ell_{k+1} = 1$ in two cases. When ℓ_k reaches the maximum restart period P (we use $P = 100$), the momentum is reset and the extrapolated iterate $\boldsymbol{\beta}^{(k+1)}$ is kept as is. Additionally, when the descent test $f_{h,\tau}(\boldsymbol{\beta}^{(k+1)}) \leq f_{h,\tau}(\boldsymbol{\beta}^{(k)})$ fails, the momentum is reset and the extrapolated iterate is replaced by a naive MM step anchored at $\boldsymbol{\beta}^{(k)}$,

$$\boldsymbol{\beta}^{(k+1)} \;\leftarrow\; \boldsymbol{\beta}^{(k)} \;-\; \boldsymbol{H}^{-1}\nabla f_{h,\tau}(\boldsymbol{\beta}^{(k)}). \tag{12}$$

The previous iterate $\boldsymbol{\beta}^{(k)}$ itself is unchanged, but $\boldsymbol{\beta}^{(k+1)}$ is overwritten. By (9) the replacement satisfies $f_{h,\tau}(\boldsymbol{\beta}^{(k+1)}) \leq f_{h,\tau}(\boldsymbol{\beta}^{(k)})$, so $\{f_{h,\tau}(\boldsymbol{\beta}^{(k)})\}_{k\geq 0}$ is monotone non-increasing. We declare convergence when $\|\nabla f_{h,\tau}(\boldsymbol{\eta}^{(k)})\|_\infty \leq \varepsilon_{\text{tol}}$ (default $\varepsilon_{\text{tol}} = 10^{-9}$), and cap the maximum number of iterations at 5000.

Because $\boldsymbol{H}$ in (8) does not depend on $\boldsymbol{\beta}$, its Cholesky factor $\boldsymbol{L}$ satisfying $\boldsymbol{H} = \boldsymbol{L}\boldsymbol{L}^\top$ can be computed once and reused across iterations: each QMME step then requires only a gradient evaluation in $\mathcal{O}(np)$ via (7) and a forward substitution against $\boldsymbol{L}$ followed by a back substitution against $\boldsymbol{L}^\top$, both are of $\mathcal{O}(p^2)$ complexity. A second level of reuse arises in the LOCO loop of Algorithm 1: $\boldsymbol{X}^\top \boldsymbol{X}$ depends only on the covariates and is computed once per trait, while the bandwidth h varies across chromosomes but not across the d quantile levels. Thus for each chromosome the Cholesky factor of $\boldsymbol{H}$ is constructed once and reused for all d quantiles.

2.3 Invariance of the score statistic to covariate adjustment

We now establish a useful property of the score statistic S defined in line 10 of Algorithm 1: it is invariant under projection of the genotype onto the null space of the covariates $\boldsymbol{X}$. In other words, $S \;=\; \tilde{\boldsymbol{G}}^\top \boldsymbol{R} \;=\; \boldsymbol{G}^\top \boldsymbol{R}$.

Proof. By the definition of $\tilde{\boldsymbol{G}}$,

$$\tilde{\boldsymbol{G}}^\top \boldsymbol{R} \;=\; \boldsymbol{G}^\top \boldsymbol{R} \;-\; \boldsymbol{G}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{R}.$$

It therefore suffices to show that $\boldsymbol{X}^\top \boldsymbol{R} = \mathbf{0}$. The convolution-smoothed quantile loss can be written as

$$f_{h,\tau}(\boldsymbol{\beta}) \;=\; \tfrac{1}{n}\, \mathbf{1}^\top \ell_{h,\tau}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}),$$

where $\ell_{h,\tau}$ is applied element-wise. Differentiating with respect to $\boldsymbol{\beta}$ gives

$$\nabla_\beta\, f_{h,\tau}(\boldsymbol{\beta}) \;=\; -\tfrac{1}{n}\, \boldsymbol{X}^\top \Big[\tau\, \mathbf{1}_n - \mathcal{K}\Big(-\tfrac{\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}}{h}\Big)\Big],$$

where $\mathcal{K}(\cdot)$ is applied element-wise. Setting this gradient to zero at $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}_{h,\tau}$ yields

$$\boldsymbol{X}^\top \left[\tau\, \mathbf{1}_n - \mathcal{K}\left(-\frac{\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_{h,\tau}}{h}\right)\right] \;=\; \mathbf{0},$$

which is precisely $\boldsymbol{X}^\top \boldsymbol{R} = \mathbf{0}$. □

3 Effect-size estimation

The score test in Algorithm 1 reports a p-value for each variant but does not estimate the size of the genetic effect. This is deliberate: the score test is fast because it fits the null model *once* and then sweeps through millions of variants without re-fitting the quantile regression model. For most genome-wide screens, a p-value is all that is needed at this stage.

For follow-up analyses on a small number of variants of interest — for example, comparing the effect of a top SNP across different quantiles of the phenotype distribution to characterize effect heterogeneity — the size of the effect itself becomes the primary quantity of interest. To support this, SPA_{SQR} provides a complementary *Wald mode* that, for each selected variant, refits the smoothed quantile regression model with the genotype $\boldsymbol{G}$ included alongside the covariates and reports the estimated effect size $\hat{\gamma}(\tau)$ together with its standard error. This Wald mode is intended for short, curated variant lists rather than genome-wide scans; on each variant it pays the cost of a full model refit, but in exchange yields directly interpretable effect sizes and standard errors.

3.1 Refitting the model with the genotype included

For a single variant $\boldsymbol{G} \in \mathbb{R}^n$ and quantile level $\tau \in (0,1)$, SPA_{SQR} fits smoothed quantile regression using both covariates and the genotype as the design matrix. Stacking the covariates and genotype into one design vector $\boldsymbol{Z}_i = (\boldsymbol{X}_i^\top, G_i)^\top$, and writing $\boldsymbol{\theta}(\tau) = (\,\boldsymbol{\beta}(\tau)^\top, \gamma(\tau))^\top$, $\gamma(\tau)$ is the per-allele effect on the τ-th quantile of $\boldsymbol{Y}$. The estimator of $\boldsymbol{\theta}(\tau)$ is obtained by solving

$$\hat{\boldsymbol{\theta}}(\tau) \;=\; \underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}}\; \frac{1}{n} \sum_{i=1}^{n} \ell_{h,\tau}(Y_i^{(c)} - \boldsymbol{Z}_i^\top \boldsymbol{\theta}), \tag{13}$$

where $\ell_{h,\tau}$ is the smoothed quantile regression loss (5) and $Y_i^{(c)}$ is the phenotype (after LOCO PGS offset). Problem (13) is solved using the same QMME algorithm as null models. Notice that in score testing we typically use a relatively large value of h, such as $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. A large h promotes statistical power in association testing but may introduce statistical bias in the coefficients. Here, we use a relatively conservative value of h, namely $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/5$.

3.2 Standard error of the effect estimate

To attach a standard error to $\hat{\gamma}(\tau)$ we use the sandwich formula for M-estimators [6]. Concretely,

$$\hat{\boldsymbol{A}}(\tau) \;=\; \frac{1}{n}\sum_{i=1}^{n} K_h(-\hat{e}_i)\,\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top}, \qquad \hat{\boldsymbol{B}}(\tau) \;=\; \frac{1}{n}\sum_{i=1}^{n} R_i^2\,\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top}, \tag{14}$$

where $\hat{e}_i = Y_i^{(c)} - \boldsymbol{Z}_i^{\top}\hat{\boldsymbol{\theta}}(\tau)$ are the fitted residuals, K_h is the Gaussian kernel of bandwidth h, and $R_i = \tau - \mathcal{K}(-\hat{e}_i/h)$ are the smoothed quantile regression residuals evaluated at $\hat{\boldsymbol{\theta}}(\tau)$. The estimated covariance of $\hat{\boldsymbol{\theta}}(\tau)$ is then the "sandwich"

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}(\tau)) \;=\; \frac{1}{n}\,\hat{\boldsymbol{A}}(\tau)^{-1}\hat{\boldsymbol{B}}(\tau)\,\hat{\boldsymbol{A}}(\tau)^{-1}. \tag{15}$$

The per-allele effect size estimate and its standard error are simply the last entry of $\hat{\boldsymbol{\theta}}(\tau)$ and the square root of the corresponding diagonal entry of $\mathrm{Var}(\hat{\boldsymbol{\theta}}(\tau))$:

$$\hat{\gamma}(\tau) \;=\; \hat{\theta}_{p+1}(\tau), \qquad \mathrm{SE}(\hat{\gamma}(\tau)) \;=\; \sqrt{\big[\mathrm{Var}(\hat{\boldsymbol{\theta}}(\tau))\big]_{p+1,\,p+1}}. \tag{16}$$

A two-sided Wald test of $H_0\colon \gamma(\tau) = 0$ uses the z-statistic $\hat{\gamma}(\tau)/\,\mathrm{SE}(\hat{\gamma}(\tau))$; an approximate 95% confidence interval for the estimated effect size is $\hat{\gamma}(\tau) \pm 1.96\ \mathrm{SE}(\hat{\gamma}(\tau))$.

4 Command-line pipeline for SPA$_{\mathbf{SQR}}$

This section gives an example of the command-line workflow for running SPA$_{\mathrm{SQR}}$. The pipeline consists of (i) Obtain LOCO polygenic scores via LDAK-KVIK or REGENIE, (ii) sparse GRM construction with PLINK 2, and (iii) SPA$_{\mathrm{SQR}}$ association testing through the `GRAB` command-line interface. Throughout we assume the genotype data are stored as a PLINK 1 fileset `simu_geno.{bed,bim,fam}` and that we wish to test two quantitative traits `Quantitative1` and `Quantitative2`, adjusting for covariates `MALE, PC1, PC2, PC3, PC4`. The phenotypes and covariates are stored in `simu_geno.pheno`. The same workflow generalizes to any number of traits or covariates.

Throughout the rest of this tutorial we assume a working directory `tutorial/` containing the genotype file and the phenotype/covariate file:

```
tutorial/
simu_geno.{bed,bim,fam}    # PLINK 1 genotype fileset
simu_geno.pheno            # phenotype and covariate columns
simu_geno_wald_extract     # variant IDs for Step 3
grab2                      # GRAB binary
ldak6.2.linux              # LDAK binary
regenie                    # REGENIE binary
plink2                     # PLINK 2 binary
```

The phenotype file `simu_geno.pheno` is in PLINK format, with FID/IID in the first two columns and phenotype/covariate data in the remaining columns:

```
FID  IID     MALE  PC1         ...  Quantitative1  Quantitative2
0    S00001  1     0.00965326  ...  0.09583754    -0.58042778
0    S00002  0     0.01326490  ...  0.74086200     0.80441222
```

Each phenotype file or covariate file must start with a `FID IID` pair (as above) or a single `IID` key column, and phenotype and covariate columns may share a single file or live in two files specified via `--pheno` and `--covar`. We recommend applying a rank-based inverse normal transformation (INT) to the non-missing values of each trait prior to LOCO PGS construction:

```
grab2 --int-pheno --pheno simu_geno.pheno \
  --pheno-name Quantitative1,Quantitative2 --out simu_geno_int
```

This writes `simu_geno_int.txt`, retaining the `FID IID` key columns and replacing each requested trait column with its INT-transformed version:

```
FID  IID     Quantitative1  Quantitative2
0    S00001  -1.26311218   -0.59183920
0    S00002   0.69147316    0.81234832
...
```

`--int-pheno` only writes the requested trait columns in `--pheno`.

Step 0a: Compute LOCO polygenic scores

SPA_{SQR} accepts LOCO PGS produced by either LDAK or REGENIE; GRAB auto-detects the LOCO file format from each LOCO file's header. We illustrate both pipelines below.

Option A: LDAK-KVIK Step 1. LDAK-KVIK encodes missing values as `NA` (not the PLINK convention of `-9`) in the covariates file, so we need to convert any `-9` or blank cells in the input table to `NA` before running LDAK. LDAK selects phenotypes via `--mpheno` and covariates via `--covar-names`. For instance, we may compute LDAK LOCO PGS for both `Quantitative1, Quantitative2` using `MALE,PC1,PC2,PC3,PC4` as covariates via

```
./ldak6.2.linux --kvik-step1 ldak_step1 --bfile simu_geno \
          --pheno simu_geno_int.txt --mpheno ALL \
          --covar simu_geno.pheno   --covar-names MALE,PC1,PC2,PC3,PC4 \
          --max-threads 8
```

This writes one LOCO PGS file per phenotype: `ldak_step1.step1.pheno1.loco.prs` for `Quantitative1` and `ldak_step1.step1.pheno2.loco.prs` for `Quantitative2`. For `GRAB` to be able to leverage these LOCO PGS as offsets, we must manually create a prediction list file

```
cat > simu_geno_ldak_pred.list <<EOF
Quantitative1   $(pwd)/ldak_step1.step1.pheno1.loco.prs
Quantitative2   $(pwd)/ldak_step1.step1.pheno2.loco.prs
EOF
```

which supplies `GRAB` information on which phenotype is paired with which LOCO PGS file. The file `simu_geno_ldak_pred.list` will be passed to `GRAB` using the argument `--pred-list`.

Option B: REGENIE Step 1. REGENIE selects phenotypes via `--phenoColList` and covariates via `--covarColList`:

```
./regenie --step 1 --bed simu_geno \
    --phenoFile simu_geno_int.txt --phenoColList Quantitative1,Quantitative2 \
    --covarFile simu_geno.pheno   --covarColList MALE,PC1,PC2,PC3,PC4 \
    --bsize 1000 --threads 8 \
    --out simu_geno_regenie
```

This generates two LOCO PGS files: `simu_geno_regenie_1.loco`, `simu_geno_regenie_2.loco`, plus an auto-generated prediction list `simu_geno_regenie_pred.list`, which can be directly passed to `GRAB --pred-list`.

Step 0b: Construct a sparse GRM

SPA_{SQR} uses a sparse genetic relationship matrix $\boldsymbol{\Phi}$ to calibrate the null score variance under relatedness, $\text{Var}(S) = \widehat{\sigma}_g^2(\boldsymbol{G}_j)\, \boldsymbol{R}^\top \boldsymbol{\Phi}\, \boldsymbol{R}$. Historically the sparse GRM is typically computed with GCTA [7], but this can be very time-consuming at the scale of hundreds of thousands of participants. Since late 2025, PLINK 2 [8] also supports sparse GRM construction and is substantially faster. We may construct a sparse GRM via

```
plink2 --bfile simu_geno --maf 0.01 --make-grm-sparse 0.05 \
   --threads 8 --out simu_geno
```

This restricts GRM computation to variants with MAF ≥ 0.01 and retains genetic correlation coefficients above 0.05. The output is `simu_geno.grm.sp` (a text file with three columns: 0-based sample i index, 0-based sample j index, genetic correlation coefficient between sample-i and sample-j) plus a companion `simu_geno.grm.id` listing the sample IDs in the same order as `simu_geno.fam`.

Steps 1–2: Run SPA_{SQR} via the `GRAB` CLI

The `grab2` binary performs Steps 1–2 of Algorithm 1 in a single command: it reads the phenotype, subtracts the chromosome-specific LOCO PGS as offsets, fits the null SQR model with bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/k$, and tests every variant on the chromosome subject to quality-control filters.

For example, we may perform SPA_{SQR} association testing using LDAK-KVIK LOCO PGS as offsets while incorporating a sparse GRM via

```
./grab2  --bfile simu_geno \
  --pheno simu_geno_int.txt --pheno-name Quantitative1,Quantitative2 \
  --covar simu_geno.pheno   --covar-name MALE,PC1,PC2,PC3,PC4 \
  --sp-grm-plink2 simu_geno.grm.sp \
  --pred-list simu_geno_ldak_pred.list \
  --spasqr-h-scale 3 \
  --spasqr-taus 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9 \
  --threads 8 \
  --pheno-transform int \
  --out spasqr_results
```

Key flags:

- `--pheno-name` accepts a comma-separated list of trait column names; tests for all listed traits are run jointly with shared genotype I/O.
- `--spasqr-h-scale` sets the bandwidth divisor k in $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/k$. The score-mode default is $k = 3$, matching our main analyses.
- `--spasqr-taus` accepts a comma-separated list of quantile levels.
- `--sp-grm-plink2` reads the PLINK 2 `.grm.sp` file (the companion `.grm.id` is detected automatically) and enables GRM-aware variance correction $\text{Var}(S) = \widehat{\sigma}_g^2(\boldsymbol{G})\, \boldsymbol{R}^\top \boldsymbol{\Phi}\, \boldsymbol{R}$. Omitting `--sp-grm-plink2` falls back to the unrelated-samples variance $\widehat{\sigma}_g^2(\boldsymbol{G})\, \boldsymbol{R}^\top \boldsymbol{R}$. We recommend supplying a sparse GRM when the cohort exhibits substantial relatedness. The GRM feature is not recommended for genotypes spanning multiple ancestries, since the GRM itself typically becomes highly inaccurate in that setting.

- `--pheno-transform` applies a transformation internally to the non-missing trait values before subtracting the LOCO PGS and fitting the null SQR model. Options are `int` (rank-based inverse normal) and `standardize` (center and scale to unit variance). The transformation must match the one used during PGS construction: pass `int` when the PGS was trained on INT-transformed traits, and `standardize` when it was trained on raw traits (LDAK-KVIK and REGENIE both standardize traits before PGS construction).

`GRAB` produces one tab-delimited file per phenotype: `spasqr_results.Quantitative1.SPAsqr` and `spasqr_results.Quantitative2.SPAsqr`. For each phenotype the output reports per-quantile p-values (`P_tau0.1, ..., P_tau0.9`), the Cauchy-combined p-value `P_CCT` (line 21 of Algorithm 1), the corresponding Z-scores (`Z_tau0.1, ..., Z_tau0.9`), plus several variant-level QC fields:

```
CHROM  POS  ID  REF  ALT  MISS_RATE  ALT_FREQ  MAC  HWE_P  P_CCT
P_tau0.1 ... P_tau0.9   Z_tau0.1 ... Z_tau0.9
```

Step 3 (optional): Estimating effect sizes

After the genome-wide scan identifies a small number of variants of interest, the same `grab2` binary can re-fit the smoothed quantile regression models on each selected variant (Section 3) and report effect size estimate $\hat{\gamma}$ together with its standard error. We also compute Wald-test p-values at each requested τ. The user prepares a plain-text file (here `simu_geno_wald_extract`) listing one variant ID per line — typically the genome-wide-significant hits from the score-mode output of Steps 1–2, or a curated list from prior literature. Here suppose that `simu_geno_wald_extract` contains the following 8 variants:

```
$ cat simu_geno_wald_extract
SNP_1031
SNP_1040
SNP_1428
SNP_187
SNP_2170
SNP_2287
SNP_3240
SNP_4380
```

We estimate the effect sizes for those variants at different quantiles on both traits via:

```
grab2 --method SPAsqr --spasqr-mode wald \
  --bfile simu_geno \
  --pheno simu_geno_int.txt --pheno-name Quantitative1,Quantitative2 \
  --covar simu_geno.pheno   --covar-name MALE,PC1,PC2,PC3,PC4 \
  --pred-list simu_geno_ldak_pred.list --extract simu_geno_wald_extract \
  --spasqr-taus 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9 \
  --pheno-transform int \
  --threads 8 \
  --out spasqr_effect
```

The outputs `spasqr_effect.Quantitative1.SPAsqr`, `spasqr_effect.Quantitative2.SPAsqr` have the following header:

```
CHROM  POS  ID  REF  ALT  MISS_RATE  ALT_FREQ  MAC  HWE_P  P_CCT
P_tau0.1    ...  P_tau0.9
Z_tau0.1    ...  Z_tau0.9
BETA_tau0.1 ...  BETA_tau0.9
SE_tau0.1   ...  SE_tau0.9
```

For each quantile τ, the column `BETA_tau` reports the smoothed-QR effect size estimate $\hat{\gamma}_\tau$; `SE_tau` is the standard error obtained via the sandwich formula; `Z_tau` and `P_tau` are the Wald z scores and two-sided p-values. Plotting `BETA_tau` against τ with ± 1.96 `SE_tau` as error bars visualizes whether the genetic effect varies across different quantiles of the phenotype distribution. Wald mode defaults to a smaller smoothing bandwidth ($k = 5$, vs. $k = 3$ in score mode); override via `--spasqr-h-scale <k>` when desired.

Multithreading strategy of `GRAB`

`GRAB` parallelizes association testing using only the C++17 standard library (`std::thread`, `std::atomic`, `std::mutex`, `std::condition_variable`), with no external parallel runtime such as OpenMP or MKL; this keeps `GRAB` a single self-contained binary. The number of worker threads T is set by `--threads`.

The unit of parallel work is a *chunk* of markers. Variants are partitioned into chromosome-aligned chunks of at most `--chunk-size` markers (default 8192); a chunk never crosses a chromosome boundary. Work is distributed by a single lock-free atomic counter: each worker repeatedly claims the next available chunk and processes it to completion, so faster threads automatically absorb more chunks (greedy load balancing) and no central scheduler is needed.

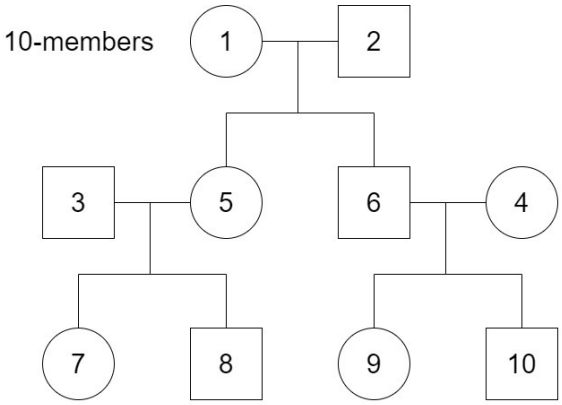
Each worker is fully self-contained: it owns a deep copy of the null model and its own genotype reader, and shares no working memory with the other workers. Within a chunk, the genotype matrix is decoded a *single* time and reused to test all phenotypes and quantile levels at once, so the cost of reading and decoding genotypes is paid only once no matter how many traits are analyzed. Output is serialized by one dedicated *writer* thread: workers deposit each finished chunk's formatted results and signal the writer, which emits chunks strictly in their original order. This makes the output byte-for-byte deterministic regardless of thread timing; it is the only point at which the threads synchronize.

For leave-one-chromosome-out (LOCO) analyses, the autosomes are processed one at a time—each requires its own chromosome-specific offset—while the chunk-level scheme above parallelizes the markers within each chromosome; the per-chromosome null model is built once and shared, read-only, across all workers. Because parallelism arises from many independent single-threaded workers operating on disjoint chunks, runtime scales nearly linearly in T up to the available physical cores and I/O constraints.

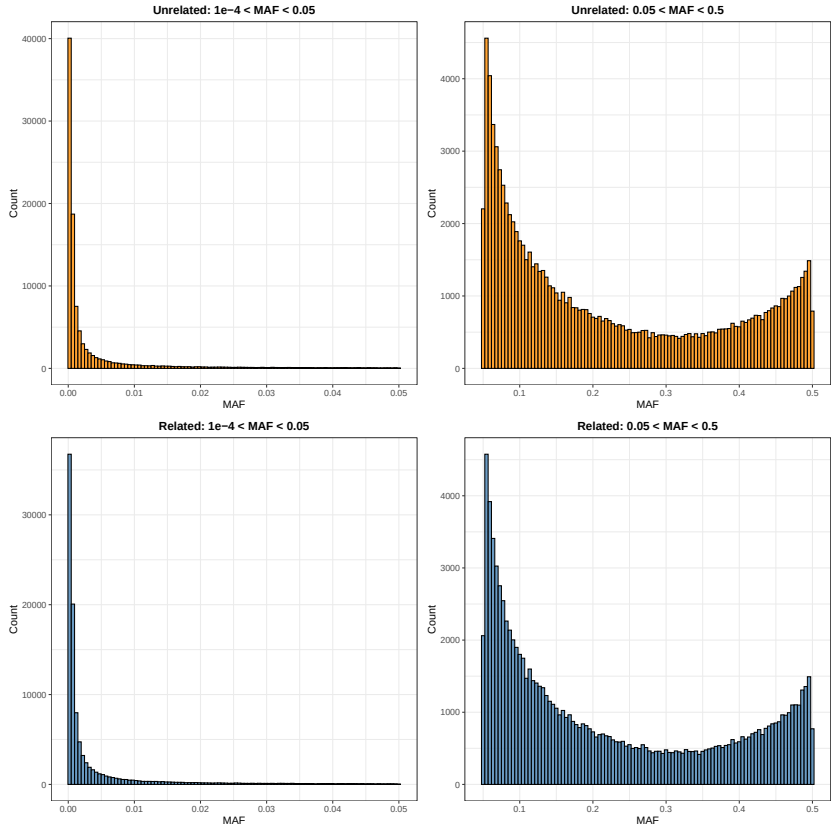
5 Genotype generation

We largely follow the protocol used in our prior work SPA_{GRM} [9] to simulate genotypes. To mimic real genotypes, we used haplotype blocks from unrelated White British participants in the UK Biobank as founder haplotypes. For each family, founder haplotypes were used as input and gene-dropping simulation [10] was performed through the pedigree shown in Supplementary Figure 2 to generate descendant genotypes.

We first generated an unrelated dataset of 50,000 participants, in which each individual constitutes its own family. We then generated a study cohort of 50,000 subjects comprising 25,000 unrelated subjects and 25,000 related subjects from 2,500 families of 10 members each (Supplementary Figure 2). Each family followed a three-generation pedigree consisting of founders (grandparents), their offspring (parents), and grandchildren. This design captures a broad range of kinship relationships, including both first- and second-degree relatives. Common variants (MAF > 0.05) were generated using haplotype blocks from genotype calls (UKB field ID: 22418), while rare variants ($10^{-4} <$ MAF < 0.05) were generated using sequence data (field ID: 23155). We subsequently filtered to 100,000 common variants and 100,000 rare variants. The resulting MAF distributions of common and rare variants are shown in Supplementary Figure 3.



Supplementary Figure 2: Family structure in the related dataset. We simulated related samples using the family structure shown above. Members #1–#4 are the subjects assigned founder haplotypes (the maximal set of unrelated subjects in a family); members #5–#10 are their descendants generated by gene-dropping. We generated 100,000 common variants and 100,000 rare variants by performing gene-dropping simulation using real genotypes as founder haplotypes that were propagated through the pedigree.



Supplementary Figure 3: Minor allele frequency distribution for the unrelated and related datasets. Each dataset contains 100,000 rare variants ($10^{-4} < \text{MAF} < 0.05$) and 100,000 common variants ($0.05 < \text{MAF} < 0.5$). Histograms are constructed using 100 bins.

6 Additional simulation results

Supplementary Table 1: Type I error inflation ratio at significance level 5×10^{-5} for REGENIE, LDAK-KVIK, REGENIE.QRS, $\text{Normal}_{\text{QR}}$, $\text{Normal}_{\text{SQR}}$, SPA_{QR}, and SPA_{SQR} based on 2×10^7 simulations, using LDAK-KVIK LOCO PGS as offsets for QR GWAS methods. The inflation ratio is the empirical type I error divided by the nominal significance level. SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. We consider two datasets: one consisting of 50,000 unrelated participants and a second consisting of 25,000 unrelated participants plus 25,000 related participants. Common variants have MAF in the range $(0.05, 0.5)$; rare variants have MAF in the range $(10^{-4}, 0.05)$. **Bold** entries indicate uncontrolled type I error rates (inflation ratio greater than 2).

			Scenarios			
Cohort	**Variant**	**Method**	**Homo-geneous**	**Domin-ance**	**Hetero-skedastic**	**Hetero-geneous**
Unrelated	Common	REGENIE	1.01	1.01	0.98	1.02
		LDAK-KVIK	1.02	1.02	1.03	1.03
		REGENIE.QRS	1.07	1.07	1.09	1.04
		$\text{Normal}_{\text{QR}}$	0.99	0.99	0.98	1.00
		$\text{Normal}_{\text{SQR}}$	1.04	1.04	1.08	1.09
		SPA_{QR}	0.98	0.98	0.96	0.98
		SPA_{SQR}	1.07	1.07	1.08	1.08
	Rare	REGENIE	1.02	1.06	1.05	1.13
		LDAK-KVIK	1.00	1.05	1.11	1.13
		REGENIE.QRS	**3.67**	**3.55**	**3.58**	**3.67**
		$\text{Normal}_{\text{QR}}$	**3.84**	**3.72**	**3.83**	**3.87**
		$\text{Normal}_{\text{SQR}}$	**3.70**	**3.77**	**3.99**	**3.76**
		SPA_{QR}	0.77	0.70	0.74	0.73
		SPA_{SQR}	0.86	0.74	0.81	0.79
Related	Common	REGENIE	**4.97**	**3.33**	**3.81**	**2.95**
		LDAK-KVIK	0.79	0.92	0.82	0.88
		REGENIE.QRS	0.78	0.91	0.95	1.00
		$\text{Normal}_{\text{QR}}$	0.94	0.92	0.91	0.92
		$\text{Normal}_{\text{SQR}}$	1.04	1.06	1.05	1.09
		SPA_{QR}	0.93	0.91	0.90	0.91
		SPA_{SQR}	1.07	1.05	1.05	1.08
	Rare	REGENIE	**5.14**	**3.62**	**4.12**	**3.39**
		LDAK-KVIK	0.81	1.02	0.99	1.02
		REGENIE.QRS	**2.80**	**3.12**	**3.35**	**3.71**
		$\text{Normal}_{\text{QR}}$	**3.26**	**3.44**	**3.34**	**3.44**
		$\text{Normal}_{\text{SQR}}$	**3.14**	**3.57**	**3.76**	**3.77**
		SPA_{QR}	0.80	0.84	0.78	0.78
		SPA_{SQR}	0.86	0.87	0.84	0.88

Supplementary Table 2: Type I error inflation ratio at significance level 5×10^{-7} for REGENIE.QRS, $\text{Normal}_{\text{QR}}$, $\text{Normal}_{\text{SQR}}$, SPA_{QR}, and SPA_{SQR} based on 2×10^7 simulations, using REGENIE LOCO PGS as offsets for QR GWAS methods. The inflation ratio is the empirical type I error divided by the nominal significance level. SPA_{SQR} uses bandwidth $h = \text{IQR}(\check{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$.

Cohort	Variant	Method	Scenarios			
			Homo-geneous	**Domin-ance**	**Hetero-skedastic**	**Hetero-geneous**
Unrelated	Common	REGENIE.QRS	1.61	1.13	0.72	1.29
		$\text{Normal}_{\text{QR}}$	1.13	0.64	1.01	1.29
		$\text{Normal}_{\text{SQR}}$	1.13	0.56	1.53	1.21
		SPA_{QR}	0.88	0.64	0.64	1.21
		SPA_{SQR}	1.05	0.40	1.29	1.21
	Rare	REGENIE.QRS	**20.86**	**21.34**	**19.83**	**17.75**
		$\text{Normal}_{\text{QR}}$	**22.81**	**24.42**	**22.62**	**20.72**
		$\text{Normal}_{\text{SQR}}$	**20.68**	**21.71**	**22.19**	**20.77**
		SPA_{QR}	1.42	1.13	0.85	0.66
		SPA_{SQR}	0.94	0.76	0.47	0.47
Related	Common	REGENIE.QRS	**6.83**	**4.02**	**6.18**	**3.69**
		$\text{Normal}_{\text{QR}}$	1.37	1.29	0.72	0.80
		$\text{Normal}_{\text{SQR}}$	1.69	1.29	1.37	0.72
		SPA_{QR}	1.37	1.20	0.72	0.80
		SPA_{SQR}	1.69	1.12	1.29	0.72
	Rare	REGENIE.QRS	**65.81**	**46.04**	**54.13**	**44.72**
		$\text{Normal}_{\text{QR}}$	**30.61**	**28.25**	**29.19**	**24.64**
		$\text{Normal}_{\text{SQR}}$	**37.19**	**31.73**	**32.76**	**31.07**
		SPA_{QR}	0.85	0.85	1.04	0.75
		SPA_{SQR}	1.32	0.94	0.94	1.04

Supplementary Table 3: Type I error inflation ratio at significance level 5×10^{-5} for REGENIE.QRS, $\text{Normal}_{\text{QR}}$, $\text{Normal}_{\text{SQR}}$, SPA_{QR}, and SPA_{SQR} based on 2×10^7 simulations, using REGENIE LOCO PGS as offsets for QR GWAS methods. The inflation ratio is the empirical type I error divided by the nominal significance level. SPA_{SQR} uses bandwidth $h = \text{IQR}(\check{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$.

Cohort	Variant	Method	Scenarios			
			Homo-geneous	Domin-ance	Hetero-skedastic	Hetero-geneous
Unrelated	Common	REGENIE.QRS	1.06	1.05	1.02	1.06
		$\text{Normal}_{\text{QR}}$	0.99	1.01	1.01	1.00
		$\text{Normal}_{\text{SQR}}$	1.04	1.07	1.04	1.06
		SPA_{QR}	0.96	0.99	0.99	0.97
		SPA_{SQR}	1.03	1.06	1.04	1.06
	Rare	REGENIE.QRS	**3.64**	**3.73**	**3.67**	**3.67**
		$\text{Normal}_{\text{QR}}$	**3.91**	**3.86**	**3.82**	**3.83**
		$\text{Normal}_{\text{SQR}}$	**3.79**	**3.74**	**3.81**	**3.77**
		SPA_{QR}	0.80	0.78	0.73	0.72
		SPA_{SQR}	0.77	0.80	0.73	0.73
Related	Common	REGENIE.QRS	**3.38**	**2.55**	**2.89**	**2.49**
		$\text{Normal}_{\text{QR}}$	0.99	0.91	0.90	0.96
		$\text{Normal}_{\text{SQR}}$	1.07	1.05	1.00	1.00
		SPA_{QR}	0.98	0.90	0.89	0.95
		SPA_{SQR}	1.07	1.04	1.00	0.99
	Rare	REGENIE.QRS	**9.21**	**7.30**	**8.16**	**7.07**
		$\text{Normal}_{\text{QR}}$	**4.73**	**4.33**	**4.50**	**4.24**
		$\text{Normal}_{\text{SQR}}$	**4.86**	**4.46**	**4.58**	**4.40**
		SPA_{QR}	0.80	0.94	0.95	0.88
		SPA_{SQR}	1.13	1.01	1.06	0.98

Supplementary Table 4: Type I error inflation ratio at significance level 5×10^{-7} for $\mathbf{SPA_{SQR}}$ based on 2×10^7 simulations, across bandwidths $h = \mathrm{IQR}(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/2$ and $h = \mathrm{IQR}(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/5$, using either REGENIE or LDAK-KVIK LOCO PGS as offsets. The inflation ratio is the empirical type I error divided by the nominal significance level.

				Scenarios			
Cohort	**Variant**	**Bandwidth**	**Offset**	**Homo-geneous**	**Domin-ance**	**Hetero-skedastic**	**Hetero-geneous**
Unrelated	Common	IQR /2	REGENIE	1.45	0.64	1.53	1.53
			LDAK-KVIK	1.13	1.13	0.96	1.37
		IQR /5	REGENIE	1.05	0.80	1.21	1.21
			LDAK-KVIK	1.53	0.80	0.64	1.37
	Rare	IQR /2	REGENIE	1.51	1.13	0.38	0.47
			LDAK-KVIK	0.66	0.66	0.28	0.57
		IQR /5	REGENIE	0.76	0.85	0.38	0.38
			LDAK-KVIK	0.47	0.47	0.47	0.47
Related	Common	IQR /2	REGENIE	1.45	1.20	1.37	0.96
			LDAK-KVIK	0.48	1.12	0.80	1.29
		IQR /5	REGENIE	1.05	1.37	1.12	0.88
			LDAK-KVIK	1.06	1.04	0.72	0.80
	Rare	IQR /2	REGENIE	1.51	1.04	1.32	0.94
			LDAK-KVIK	0.47	0.75	0.66	0.66
		IQR /5	REGENIE	0.94	1.22	0.75	1.22
			LDAK-KVIK	0.47	0.85	0.28	0.75

Supplementary Table 5: Type I error inflation ratio at significance level 5×10^{-5} for $\mathbf{SPA_{SQR}}$ based on 2×10^7 simulations, across bandwidths $h = \mathrm{IQR}(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/2$ and $h = \mathrm{IQR}(\hat{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/5$, using either REGENIE or LDAK-KVIK LOCO PGS as offsets. The inflation ratio is the empirical type I error divided by the nominal significance level.

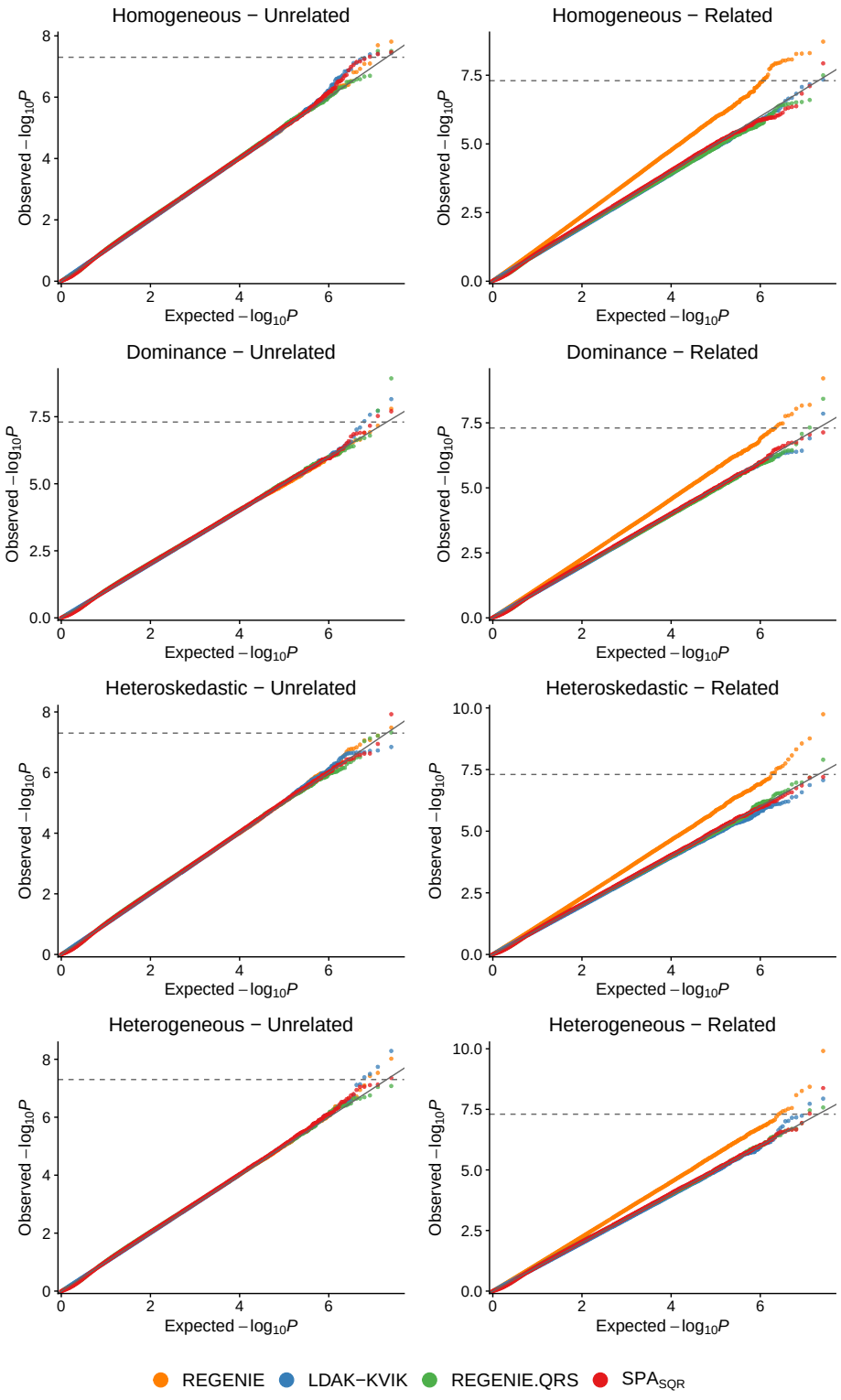
				Scenarios			
Cohort	**Variant**	**Bandwidth**	**Offset**	**Homo-geneous**	**Domin-ance**	**Hetero-skedastic**	**Hetero-geneous**
Unrelated	Common	IQR /2	REGENIE	1.06	1.08	1.03	1.04
			LDAK-KVIK	1.09	1.07	1.08	1.06
		IQR /5	REGENIE	1.05	1.01	1.04	1.03
			LDAK-KVIK	1.08	1.06	1.06	1.06
	Rare	IQR /2	REGENIE	1.22	0.87	0.83	0.81
			LDAK-KVIK	0.92	0.84	0.88	0.89
		IQR /5	REGENIE	0.70	0.73	0.67	0.65
			LDAK-KVIK	0.82	0.66	0.74	0.71
Related	Common	IQR /2	REGENIE	1.06	1.04	1.01	0.98
			LDAK-KVIK	1.09	1.05	1.02	1.07
		IQR /5	REGENIE	1.07	1.01	1.00	1.04
			LDAK-KVIK	1.08	1.06	1.04	1.05
	Rare	IQR /2	REGENIE	1.22	1.09	1.14	1.07
			LDAK-KVIK	0.92	0.89	0.91	0.95
		IQR /5	REGENIE	1.08	0.94	1.01	0.98
			LDAK-KVIK	0.82	0.84	0.80	0.83

Supplementary Table 6: Per-quantile type I error inflation ratio at significance level 5×10^{-7} for REGENIE.QRS, $\text{Normal}_{\text{SQR}}$, and SPA_{SQR} on *rare* variants (MAF in $(10^{-4}, 0.05)$), using LDAK-KVIK LOCO PGS as offsets. The inflation ratio is the empirical type I error at quantile level τ divided by the nominal significance level; each cell pools $\approx 4 \times 10^7$ null tests across the four phenotype scenarios (homogeneous, dominance, heteroskedastic, heterogeneous). SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. **Bold** entries indicate uncontrolled type I error (inflation ratio greater than 2). The normal-approximation methods REGENIE.QRS and $\text{Normal}_{\text{SQR}}$ are severely inflated at the extreme quantiles ($\tau = 0.1, 0.9$) yet controlled near the median, whereas SPA_{SQR} is calibrated across all quantiles.

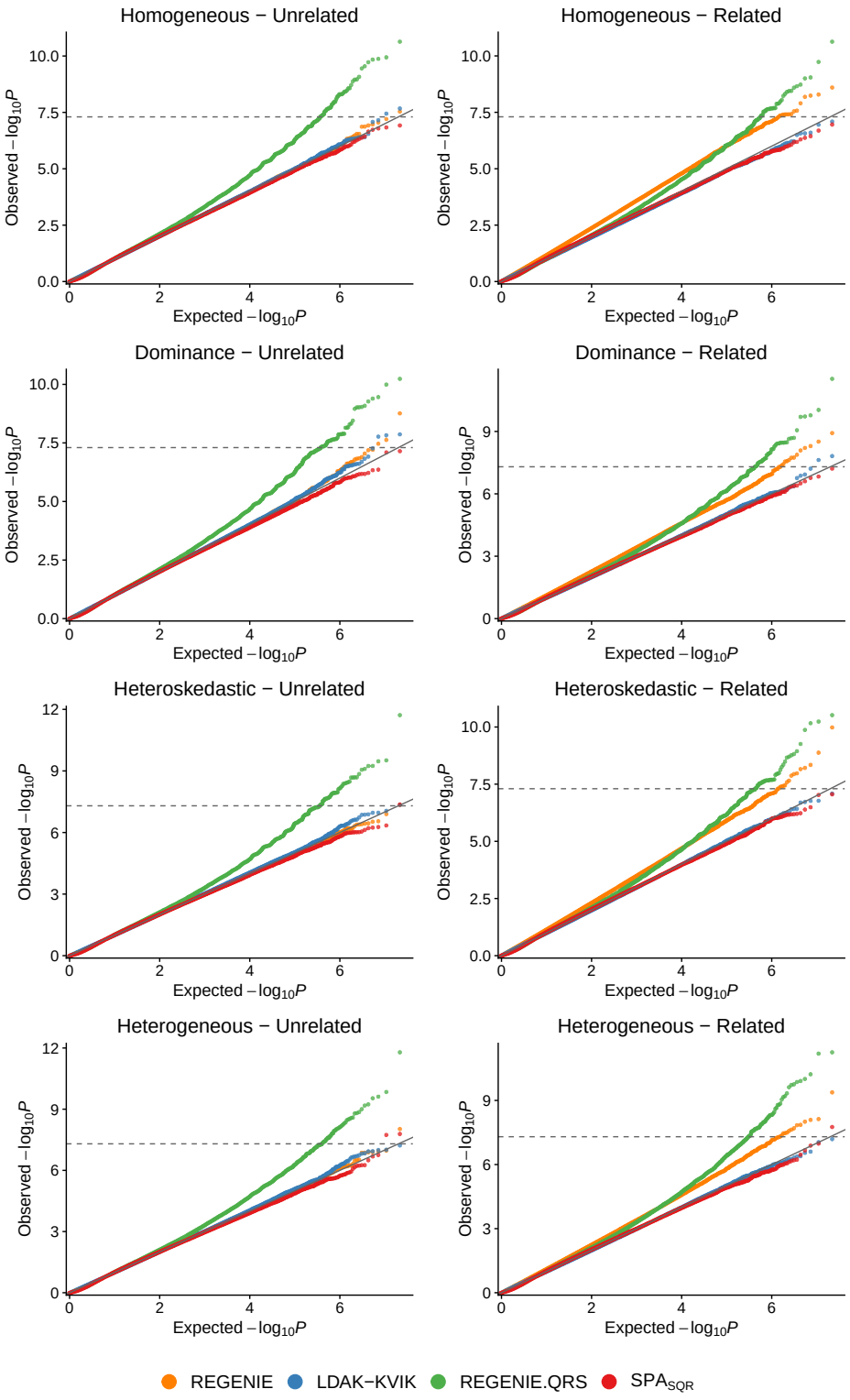
		Quantile level τ								
Cohort	**Method**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Unrelated	REGENIE.QRS	**28.81**	**6.10**	**2.46**	0.98	0.15	0.64	1.82	**5.36**	**29.75**
	$\text{Normal}_{\text{SQR}}$	**35.94**	**8.16**	**2.75**	1.13	0.49	0.54	1.77	**7.43**	**38.50**
	SPA_{SQR}	0.64	0.69	0.64	0.49	0.49	0.39	0.25	0.44	0.34
Related	REGENIE.QRS	**27.56**	**4.66**	1.08	0.49	0.39	0.49	1.42	**5.30**	**26.23**
	$\text{Normal}_{\text{SQR}}$	**32.02**	**5.44**	**2.11**	1.03	0.54	0.93	**2.26**	**7.21**	**32.75**
	SPA_{SQR}	0.93	0.83	0.69	0.39	0.54	0.54	0.64	0.88	1.32

Supplementary Table 7: Per-quantile type I error inflation ratio at significance level 5×10^{-5} for REGENIE.QRS, $\text{Normal}_{\text{SQR}}$, and SPA_{SQR} on *rare* variants (MAF in $(10^{-4}, 0.05)$), using LDAK-KVIK LOCO PGS as offsets.

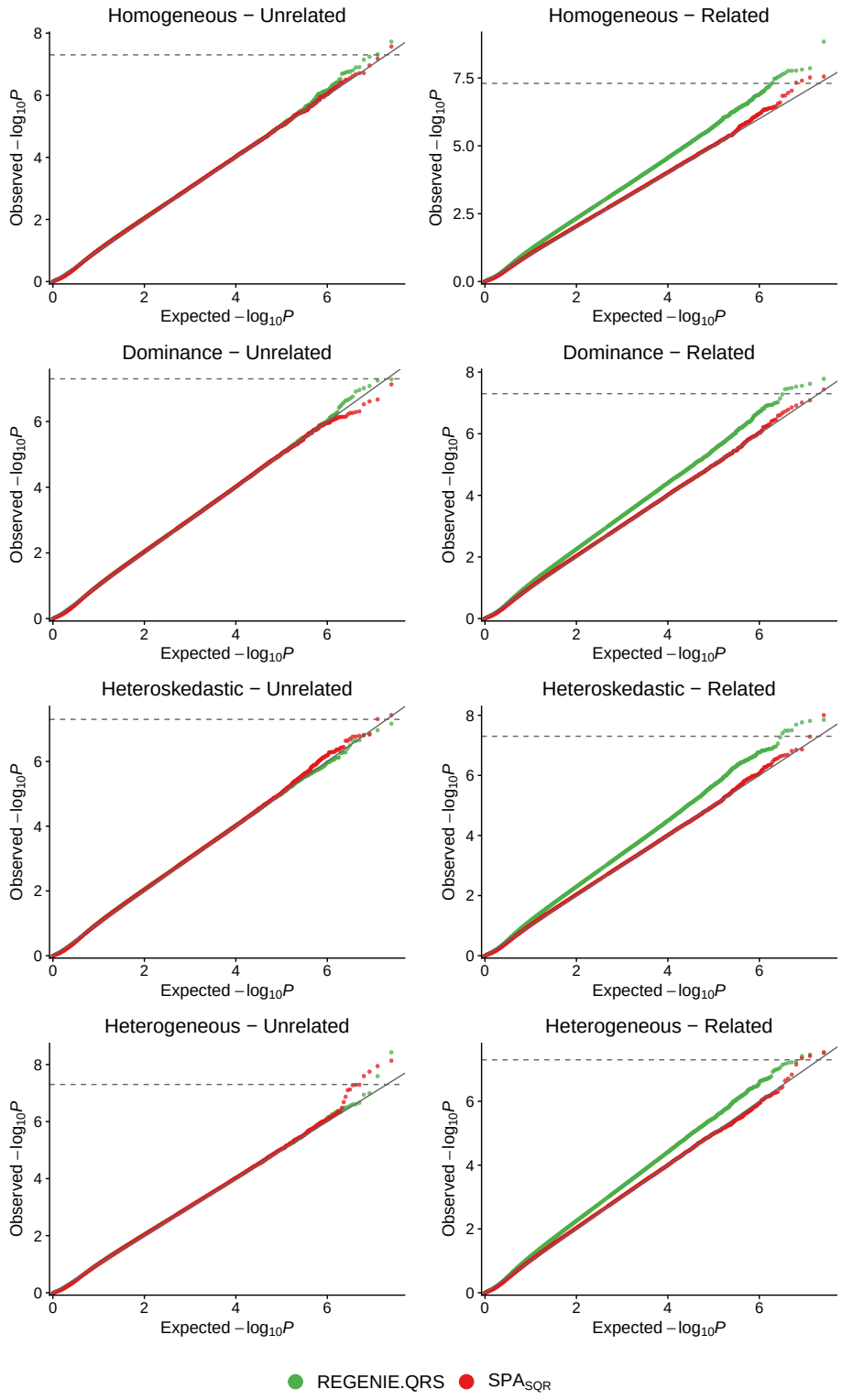
		Quantile level τ								
Cohort	**Method**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Unrelated	REGENIE.QRS	**4.93**	**2.06**	1.24	0.80	0.68	0.82	1.17	**2.08**	**4.94**
	$\text{Normal}_{\text{SQR}}$	**5.30**	**2.24**	1.30	0.92	0.80	0.90	1.32	**2.28**	**5.22**
	SPA_{SQR}	0.78	0.88	0.81	0.76	0.80	0.71	0.76	0.85	0.84
Related	REGENIE.QRS	**4.45**	1.83	0.97	0.67	0.58	0.67	0.98	1.82	**4.62**
	$\text{Normal}_{\text{SQR}}$	**4.82**	**2.14**	1.26	0.89	0.81	0.92	1.32	**2.17**	**4.89**
	SPA_{SQR}	0.99	0.90	0.82	0.75	0.81	0.76	0.82	0.90	1.03



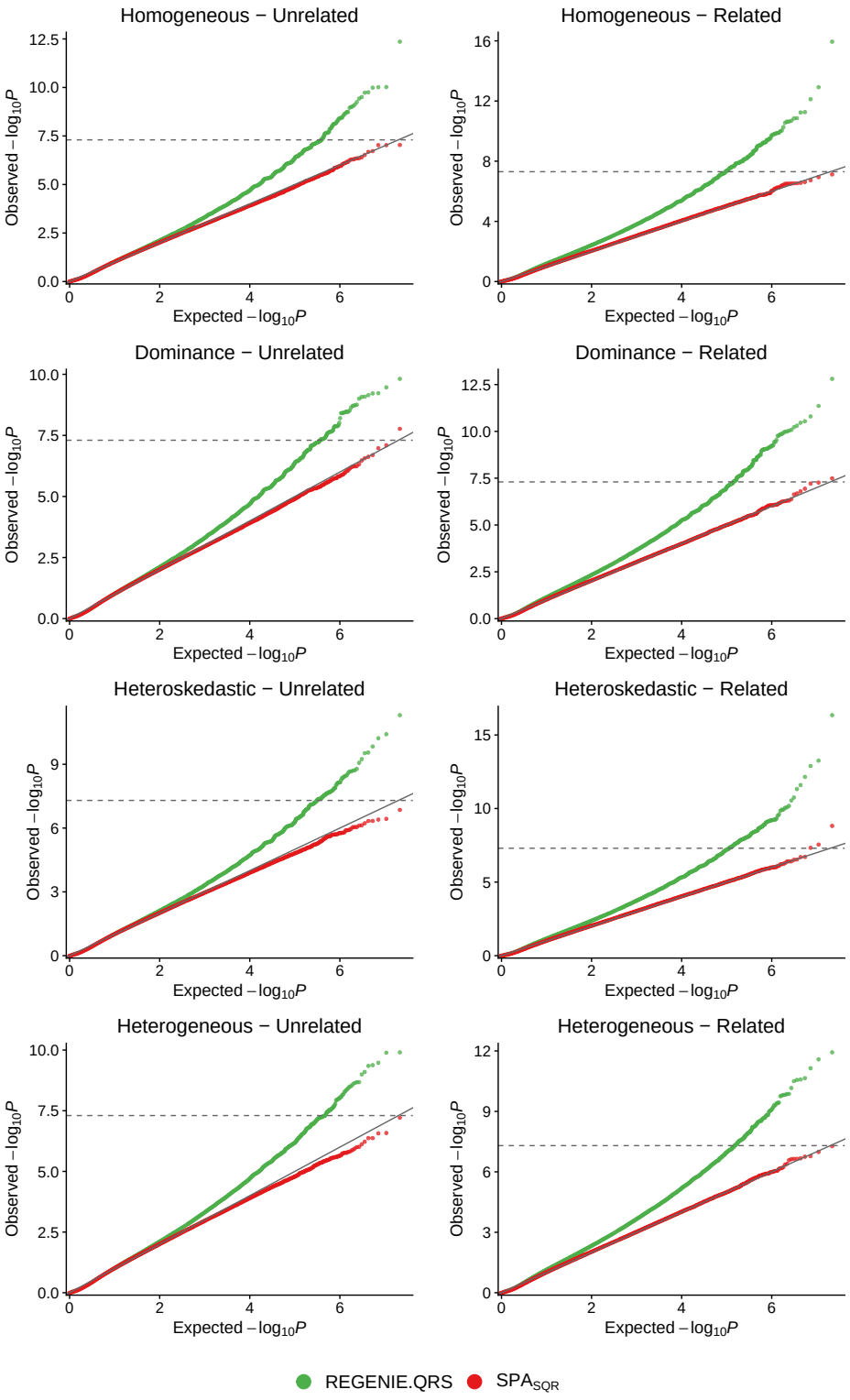
Supplementary Figure 4: QQ plots of null p-values for common variants, using LDAK-KVIK LOCO PGS as offsets. Quantile–quantile plots of $-\log_{10}(p)$ for REGENIE, LDAK-KVIK, REGENIE.QRS, and SPA_{SQR} under four scenarios (rows: homogeneous, dominance, heteroskedastic, heterogeneous) and two cohorts (columns: unrelated, related), for common variants with MAF in $(0.05, 0.5)$. REGENIE.QRS and SPA_{SQR} use the LDAK-KVIK LOCO PGS as offsets; SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. The dashed grey line marks the genome-wide significance threshold 5×10^{-8}.



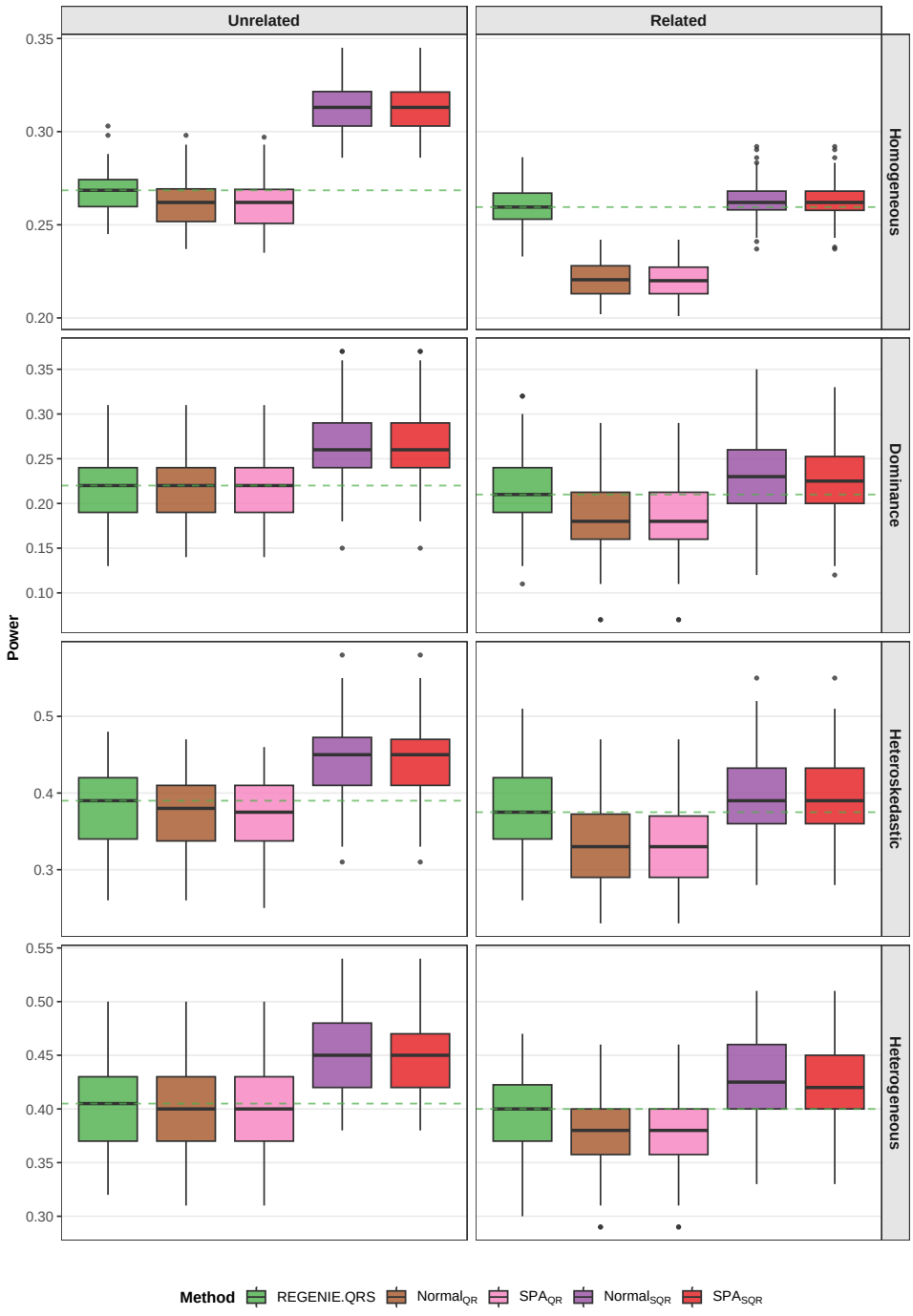
Supplementary Figure 5: QQ plots of null p-values for rare variants, using LDAK-KVIK LOCO PGS as offsets. Quantile–quantile plots of $-\log_{10}(p)$ for REGENIE, LDAK-KVIK, REGENIE.QRS, and SPA_{SQR} under four scenarios (rows: homogeneous, dominance, heteroskedastic, heterogeneous) and two cohorts (columns: unrelated, related), for rare variants with MAF in $(10^{-4}, 0.05)$. REGENIE.QRS and SPA_{SQR} use the LDAK-KVIK LOCO PGS as offsets; SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. The dashed grey line marks the genome-wide significance threshold 5×10^{-8}.



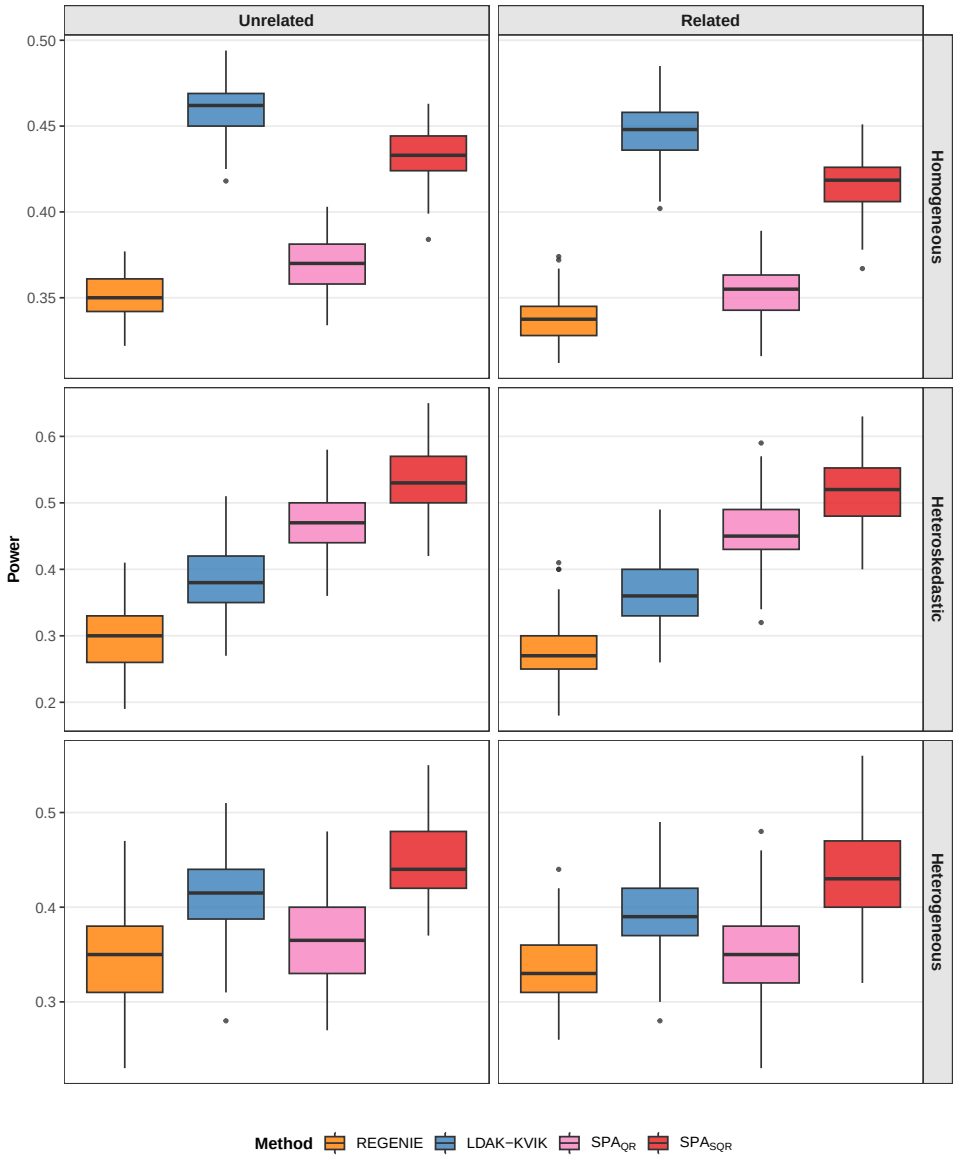
Supplementary Figure 6: QQ plots of null p-values for common variants, using REGENIE LOCO PGS as offsets. Quantile–quantile plots of $-\log_{10}(p)$ for REGENIE.QRS and SPA_{SQR} under four scenarios (rows: homogeneous, dominance, heteroskedastic, heterogeneous) and two cohorts (columns: unrelated, related), for common variants with MAF in $(0.05, 0.5)$. REGENIE.QRS and SPA_{SQR} use the REGENIE LOCO PGS as offsets; SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. The dashed grey line marks the genome-wide significance threshold 5×10^{-8}.



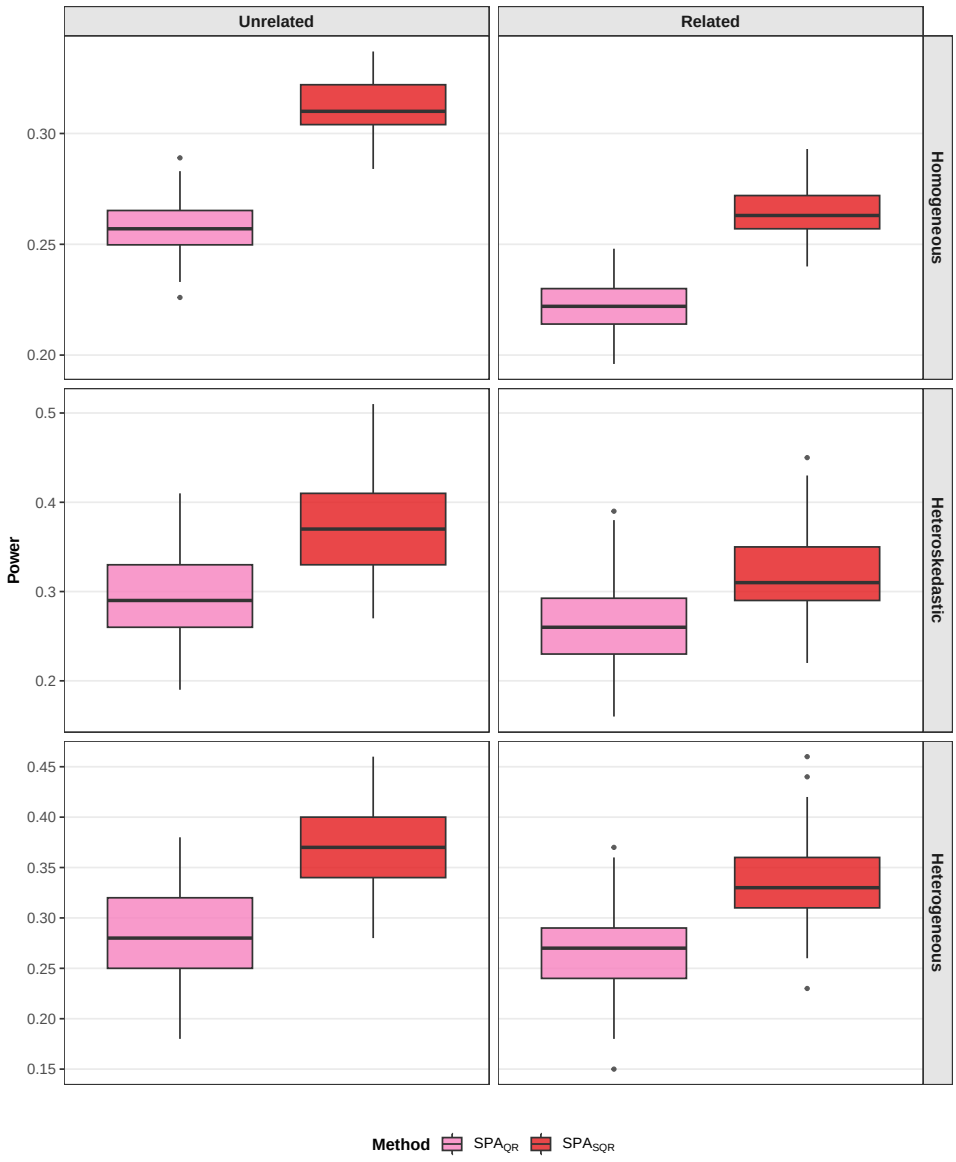
Supplementary Figure 7: QQ plots of null p-values for rare variants, using REGENIE LOCO PGS as offsets. Quantile–quantile plots of $-\log_{10}(p)$ for REGENIE.QRS and SPA_{SQR} under four scenarios (rows: homogeneous, dominance, heteroskedastic, heterogeneous) and two cohorts (columns: unrelated, related), for rare variants with MAF in $(10^{-4}, 0.05)$. REGENIE.QRS and SPA_{SQR} use the REGENIE LOCO PGS as offsets; SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. The dashed grey line marks the genome-wide significance threshold 5×10^{-8}.



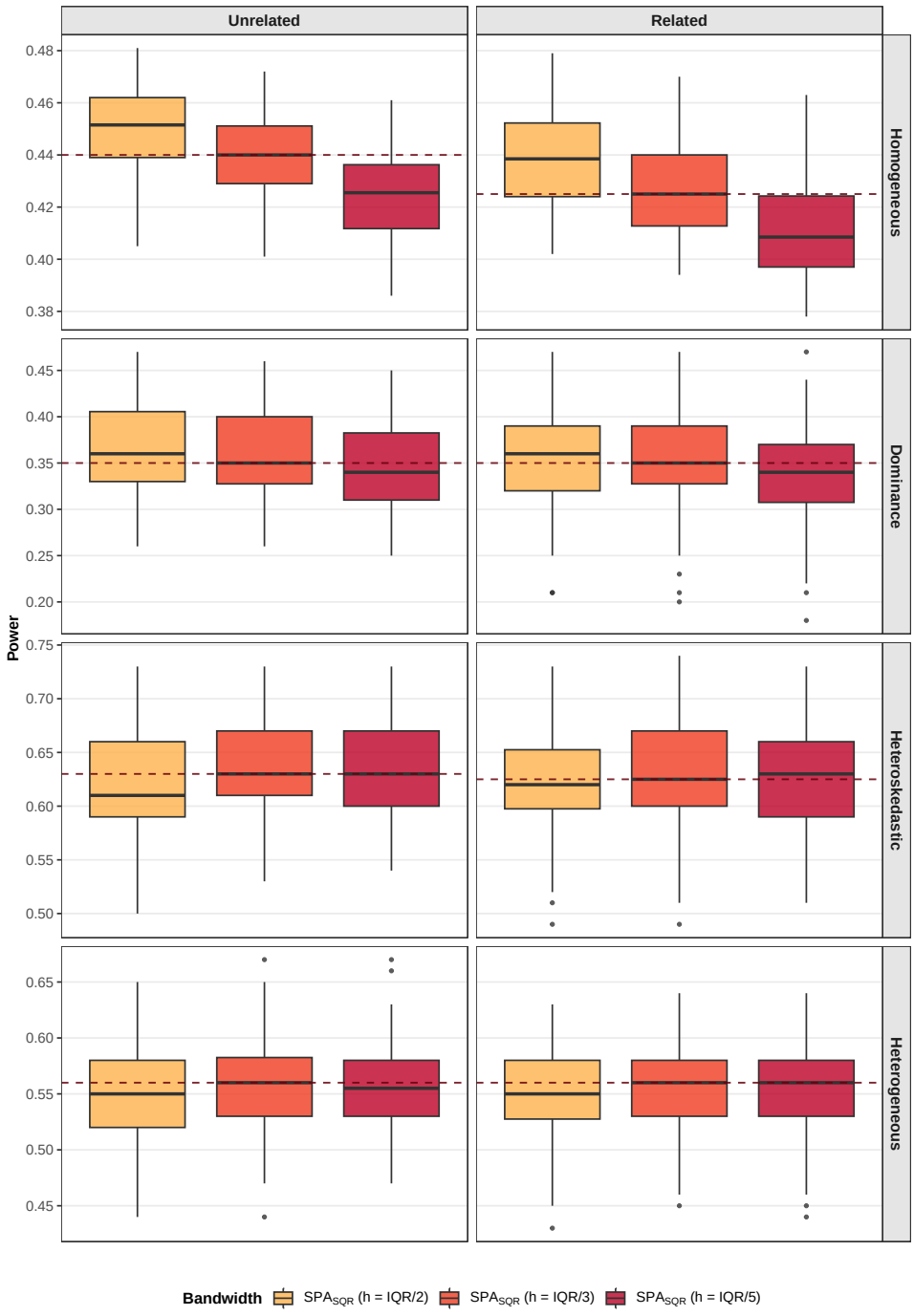
Supplementary Figure 8: Power box plots for common variants using REGENIE LOCO PGS as offsets. Power at significance level 5×10^{-8} for REGENIE.QRS, $\text{Normal}_{\text{QR}}$, SPA_{QR}, $\text{Normal}_{\text{SQR}}$, and SPA_{SQR} on common variants (MAF > 0.05), across four scenarios (Homogeneous, Dominance, Heteroskedastic, Heterogeneous) on unrelated (left column) and related (right column) samples, based on 100 simulations. Per simulation, 1,000 causal common variants were tested under Homogeneous and 100 under each of the other three scenarios. All QR GWAS methods use the REGENIE LOCO PGS as offsets in the null fit; SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$. The dashed green line marks the per-panel median power of REGENIE.QRS for visual reference.



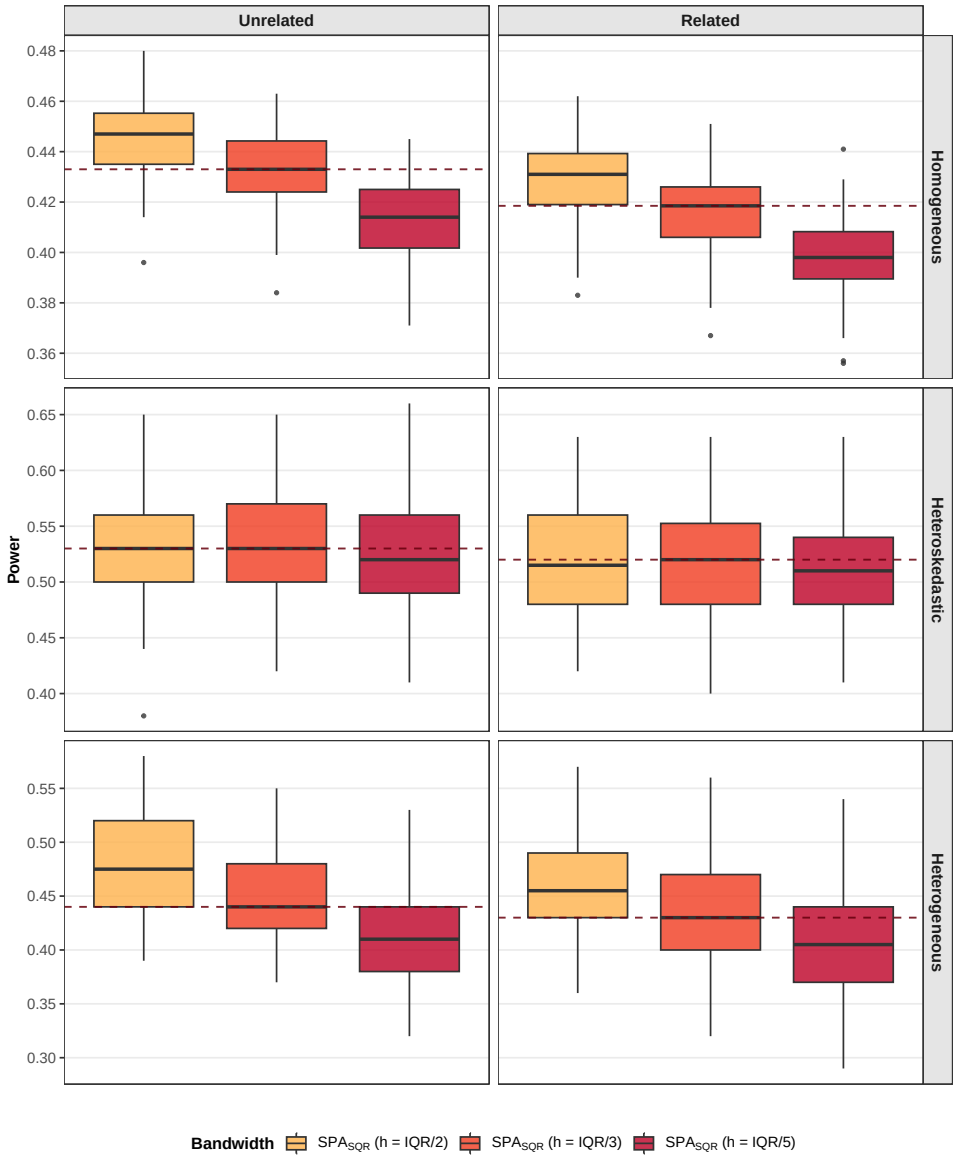
Supplementary Figure 9: Power box plots for rare variants using LDAK-KVIK LOCO PGS as offsets. Power at significance level 5×10^{-8} for REGENIE, LDAK-KVIK, SPA_{QR}, and SPA_{SQR} on rare variants ($10^{-4} < \text{MAF} < 0.05$), across three scenarios (Homogeneous, Heteroskedastic, Heterogeneous; Dominance omitted due to insufficient homozygous minor genotypes), on unrelated (left) and related (right) samples, based on 100 simulations. Per simulation, 1,000 causal rare variants were tested under Homogeneous and 100 under each of the other two scenarios. QR GWAS methods use the LDAK-KVIK LOCO PGS as offsets; SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$.



Supplementary Figure 10: Power box plots for rare variants using REGENIE LOCO PGS as offsets. Power at significance level 5×10^{-8} for SPA_{QR} and SPA_{SQR} on rare variants ($10^{-4} < \text{MAF} \leq 0.05$), across three scenarios (Homogeneous, Heteroskedastic, Heterogeneous) on unrelated (left) and related (right) samples, based on 100 simulations. Per simulation, 1,000 causal rare variants were tested under Homogeneous and 100 under each of the other two scenarios. Both methods use the REGENIE LOCO PGS as offsets. SPA_{SQR} uses bandwidth $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$.



Supplementary Figure 11: Sensitivity of $\mathbf{SPA_{SQR}}$ to the smoothing bandwidth on common variants. Power at significance level 5×10^{-8} for SPA_{SQR} run with three bandwidths, $h \in \{\text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/2,\ \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3,\ \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/5\}$ on common variants (MAF > 0.05), across four scenarios (Homogeneous, Dominance, Heteroskedastic, Heterogeneous) on unrelated (left) and related (right) samples, based on 100 simulations. Per simulation, 1,000 causal common variants were tested under Homogeneous and 100 under each of the other three scenarios. LDAK-KVIK LOCO PGS were used as offsets for all three runs. The dashed gray line marks the per-panel median power of the default $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$ run.



Supplementary Figure 12: Sensitivity of $\mathbf{SPA_{SQR}}$ to the kernel bandwidth on rare variants. Power at significance level 5×10^{-8} for SPA_{SQR} run with three bandwidths, $h \in \{\text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/2,\ \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3,\ \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/5\}$ (middle is the default), on rare variants ($10^{-4} < \text{MAF} \leq 0.05$), across three scenarios (Homogeneous, Heteroskedastic, Heterogeneous; Dominance omitted) on unrelated (left) and related (right) samples, 100 simulations. Per simulation, 1,000 causal rare variants were tested under Homogeneous and 100 under each of the other two scenarios. LDAK-KVIK LOCO PGS were used as offsets. The dashed gray line marks the per-panel median power of the default $h = \text{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$ run.

Supplementary Table 8: 66 quantitative traits from the UK Biobank. Field ID refers to the UK Biobank data-field identifier. N_1 and N_2 are the numbers of participants with non-missing phenotype values in the discovery and replication cohorts, respectively.

Field	N_1	N_2	Abbrev.	Description	Category
47	406,829	49,973	*handgrip*	Hand grip strength (right)	Anthropometry
48	407,800	50,067	*wc*	Waist circumference	Anthropometry
49	407,755	50,065	*hc*	Hip circumference	Anthropometry
50	407,608	50,048	*height*	Standing height	Anthropometry
102	381,588	46,893	*hr*	Pulse rate, automated reading	Cardiovascular
3062	372,341	45,737	*fvc*	Forced vital capacity	Lung function
3063	372,341	45,737	*fev1*	Forced expiratory volume in 1 s	Lung function
3064	372,341	45,737	*pef*	Peak expiratory flow	Lung function
4079	381,588	46,893	*dbp*	Diastolic blood pressure, automated reading	Cardiovascular
4080	381,580	46,889	*sbp*	Systolic blood pressure, automated reading	Cardiovascular
20015	407,250	49,994	*height_sit*	Sitting height	Anthropometry
21001	407,179	49,994	*bmi*	Body mass index	Anthropometry
21002	407,324	50,014	*weight*	Weight	Anthropometry
23099	401,137	49,300	*bf_pct*	Body fat percentage	Anthropometry
23100	400,714	49,236	*fatmass*	Whole body fat mass	Anthropometry
23105	401,351	49,325	*bmr*	Basal metabolic rate	Anthropometry
30000	396,334	48,625	*wbc*	White blood cell (leukocyte) count	White blood cell
30010	396,338	48,626	*rbc*	Red blood cell (erythrocyte) count	Red blood cell
30020	396,338	48,626	*hgb*	Haemoglobin concentration	Red blood cell
30030	396,338	48,626	*hct*	Haematocrit percentage	Red blood cell
30040	396,336	48,626	*mcv*	Mean corpuscular volume	Red blood cell
30050	396,335	48,626	*mch*	Mean corpuscular haemoglobin	Red blood cell
30070	396,336	48,626	*rdw*	Red blood cell distribution width	Red blood cell
30080	396,335	48,626	*plt*	Platelet count	Platelet
30090	396,331	48,626	*pct*	Platelet crit	Platelet
30100	396,330	48,626	*mpv*	Mean platelet (thrombocyte) volume	Platelet
30110	396,330	48,626	*pdw*	Platelet distribution width	Platelet
30180	395,647	48,527	*lymph_pct*	Lymphocyte percentage	White blood cell
30190	395,647	48,527	*mono_pct*	Monocyte percentage	White blood cell
30200	395,647	48,527	*neut_pct*	Neutrophill percentage	White blood cell
30210	395,647	48,527	*eos_pct*	Eosinophill percentage	White blood cell
30220	395,647	48,527	*baso_pct*	Basophill percentage	White blood cell
30240	389,999	47,685	*retic_pct*	Reticulocyte percentage	Reticulocyte
30260	390,000	47,684	*mrv*	Mean reticulocyte volume	Reticulocyte
30270	390,000	47,685	*mscv*	Mean sphered cell volume	Reticulocyte
30280	390,000	47,684	*irf*	Immature reticulocyte fraction	Reticulocyte
30290	390,000	47,685	*hlr_pct*	High light scatter reticulocyte percentage	Reticulocyte
30510	396,808	48,677	*crea_urine*	Creatinine (enzymatic) in urine	Urine
30520	395,932	48,594	*uk*	Potassium in urine	Urine
30530	395,968	48,558	*una*	Sodium in urine	Urine
30600	356,634	43,797	*alb*	Albumin	Liver function
30610	389,490	47,847	*alp*	Alkaline phosphatase	Liver function
30620	389,338	47,827	*alt*	Alanine aminotransferase	Liver function
30630	354,503	43,504	*apoA*	Apolipoprotein A	Lipids
30640	387,584	47,605	*apoB*	Apolipoprotein B	Lipids
30650	388,045	47,673	*ast*	Aspartate aminotransferase	Liver function
30670	389,203	47,818	*urea*	Urea	Kidney function
30680	356,502	43,783	*ca*	Calcium	Metabolic
30690	389,469	47,850	*chol*	Cholesterol	Lipids
30700	389,275	47,827	*crea*	Creatinine	Kidney function
30710	388,631	47,750	*crp*	C-reactive protein	Metabolic

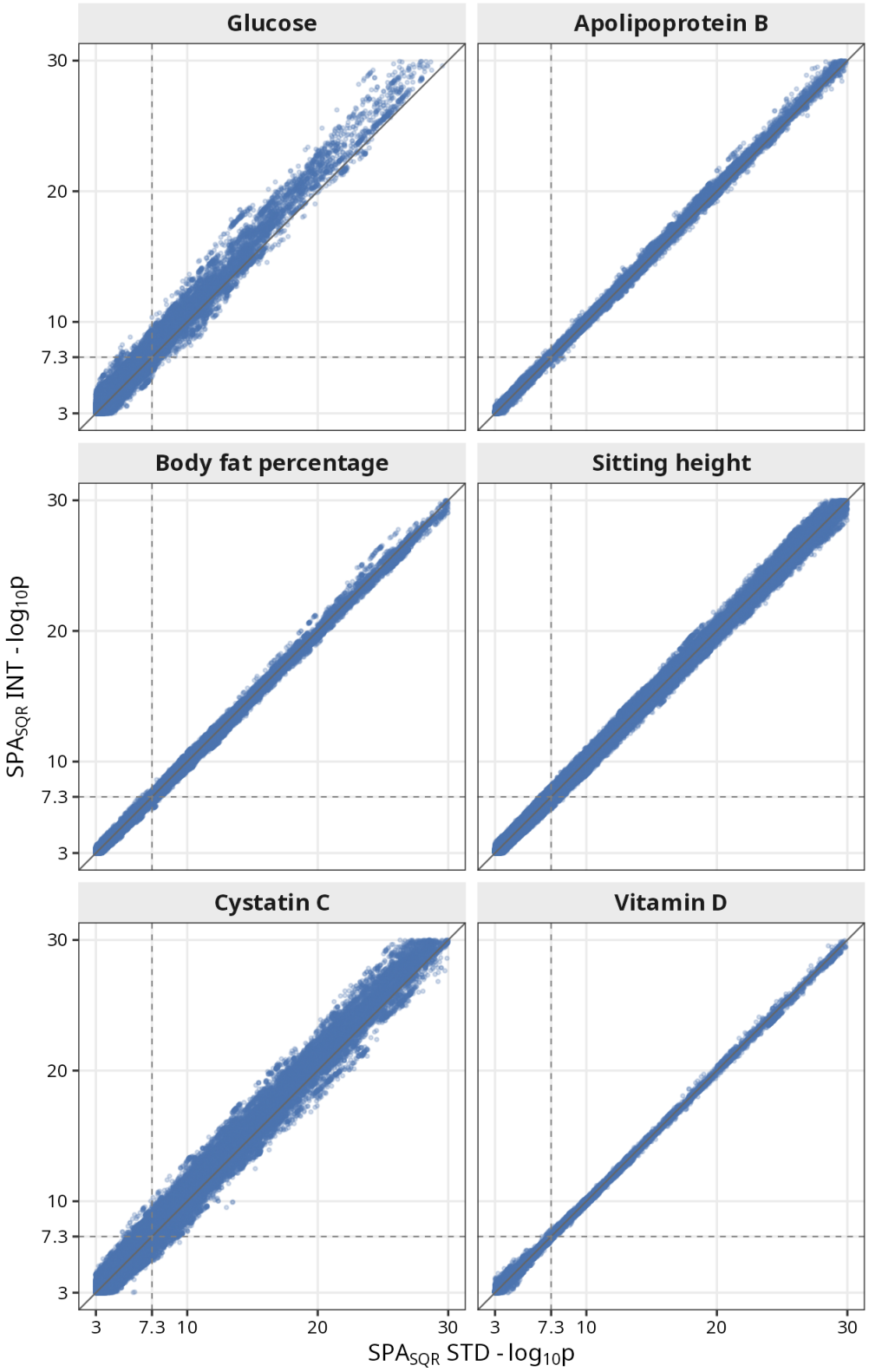
continued on next page

(continued from previous page)

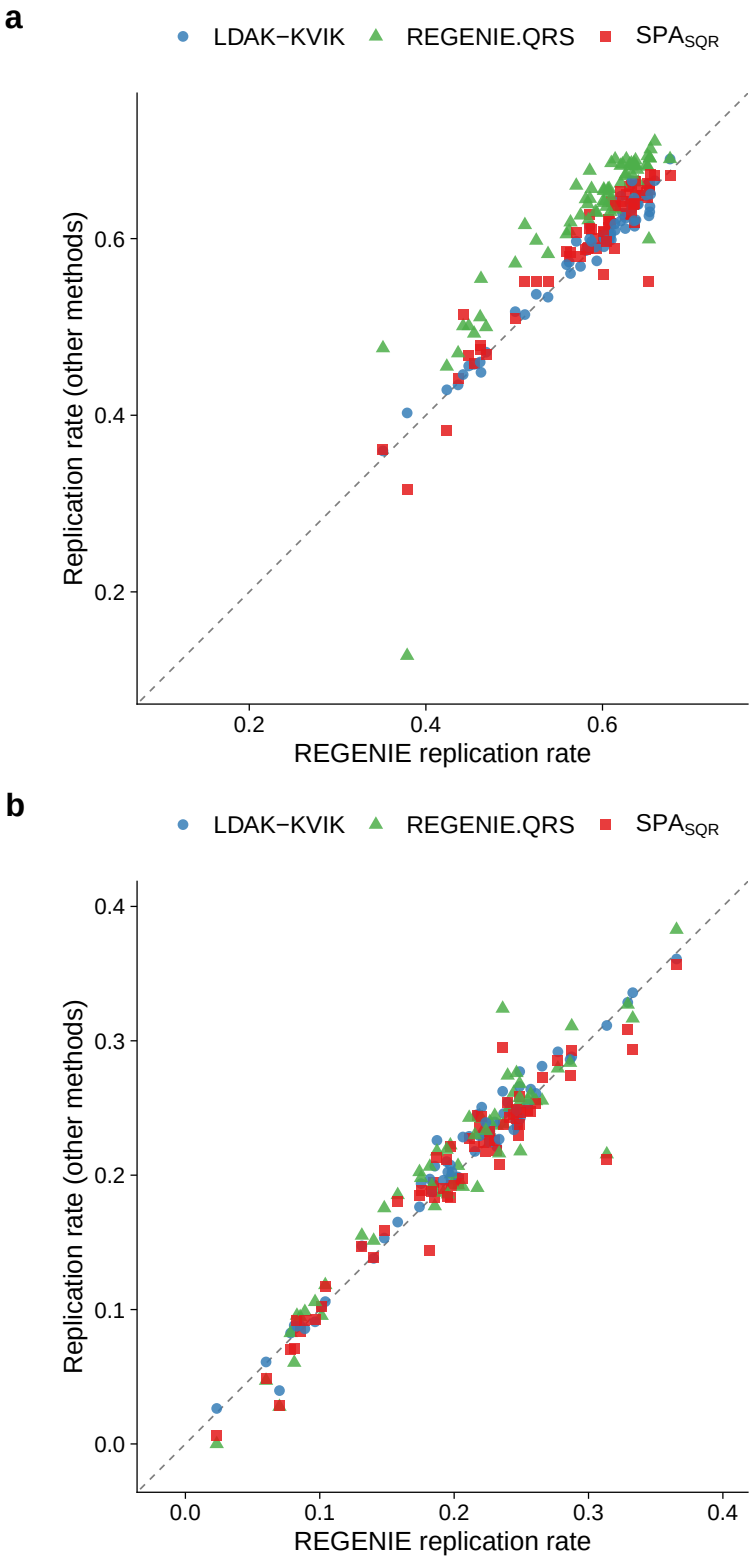
Field	N_1	N_2	Abbrev.	Description	Category
30720	389,438	47,849	*cysc*	Cystatin C	Kidney function
30730	389,271	47,821	*ggt*	Gamma glutamyltransferase	Liver function
30740	356,232	43,721	*glu*	Glucose	Glucose/HbA1c
30750	389,369	47,829	*hba1c*	Glycated haemoglobin	Glucose/HbA1c
30760	356,475	43,773	*hdl*	HDL cholesterol	Lipids
30770	387,363	47,599	*igf1*	IGF-1	Endocrine
30780	388,753	47,758	*ldl*	LDL direct	Lipids
30810	355,959	43,693	*phos*	Phosphate	Metabolic
30830	353,216	43,324	*shbg*	SHBG	Endocrine
30840	387,850	47,648	*tbil*	Total bilirubin	Liver function
30850	352,829	43,162	*testo*	Testosterone	Endocrine
30860	356,225	43,751	*tprot*	Total protein	Liver function
30870	389,162	47,812	*tg*	Triglycerides	Lipids
30880	388,996	47,799	*urate*	Urate	Metabolic
30890	372,308	45,843	*vitd*	Vitamin D	Endocrine

Supplementary Table 9: Number of genome-wide significant variants ($p < 5 \times 10^{-8}$) identified by each method summed over all 66 traits, stratified by minor allele frequency (MAF).

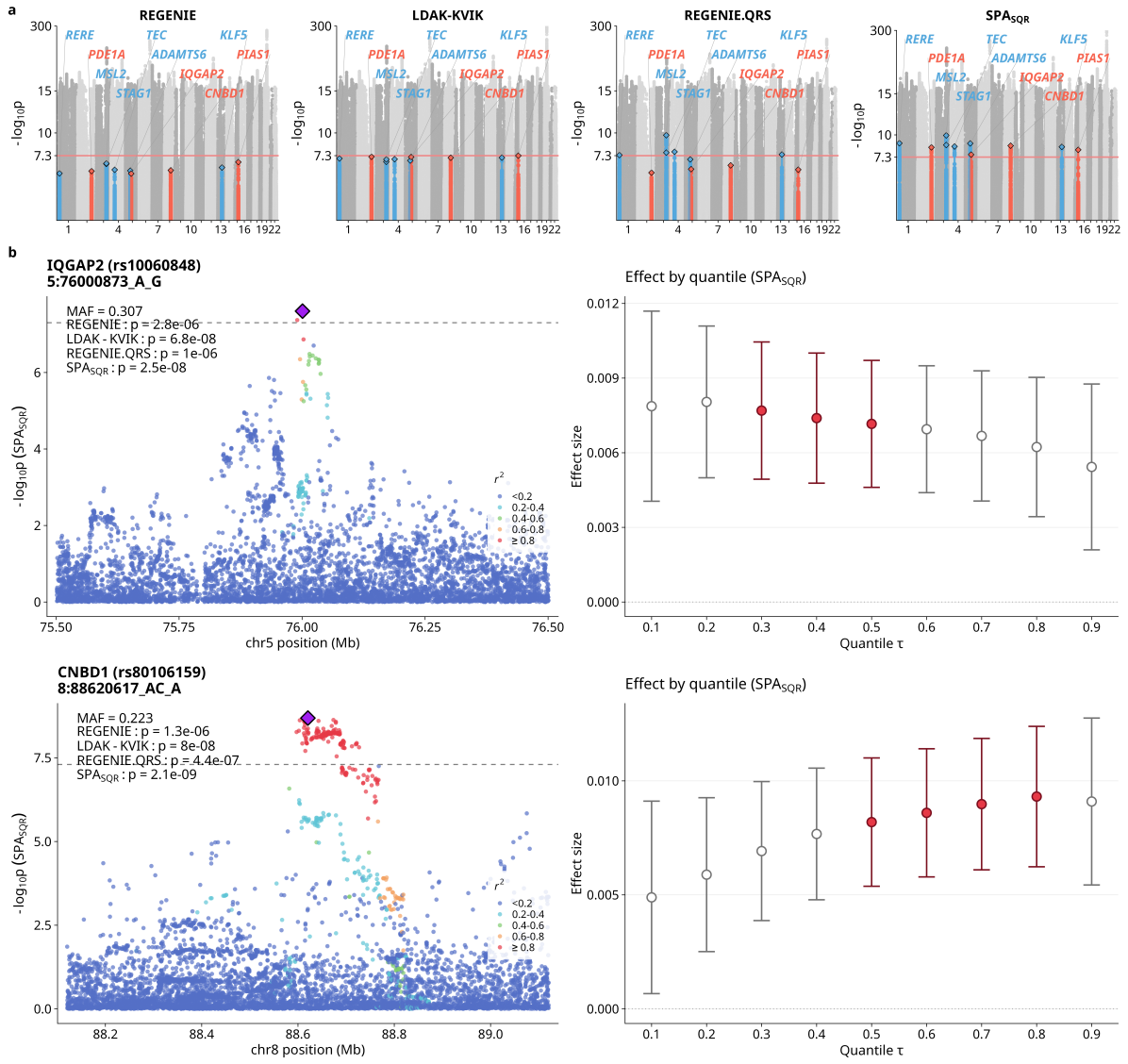
Method	0.0001–0.001	0.001–0.01	0.01–0.05	$\geq$ 0.05
Discovery cohort (∼408,000)				
SPA_{SQR}	4,549	46,058	256,369	4,016,734
$\text{Normal}_{\text{SQR}}$	5,335	46,716	256,777	4,017,104
REGENIE.QRS	4,092	34,305	194,813	3,157,688
LDAK-KVIK	5,032	48,191	262,093	4,140,499
REGENIE	4,701	43,618	242,500	3,864,019
Replication cohort (∼50,000)				
SPA_{SQR}	73	1,681	8,376	201,525
$\text{Normal}_{\text{SQR}}$	2,299	2,082	8,513	201,690
REGENIE.QRS	1,746	1,611	6,411	142,479
LDAK-KVIK	127	1,852	9,435	215,758
REGENIE	105	1,684	9,061	193,681



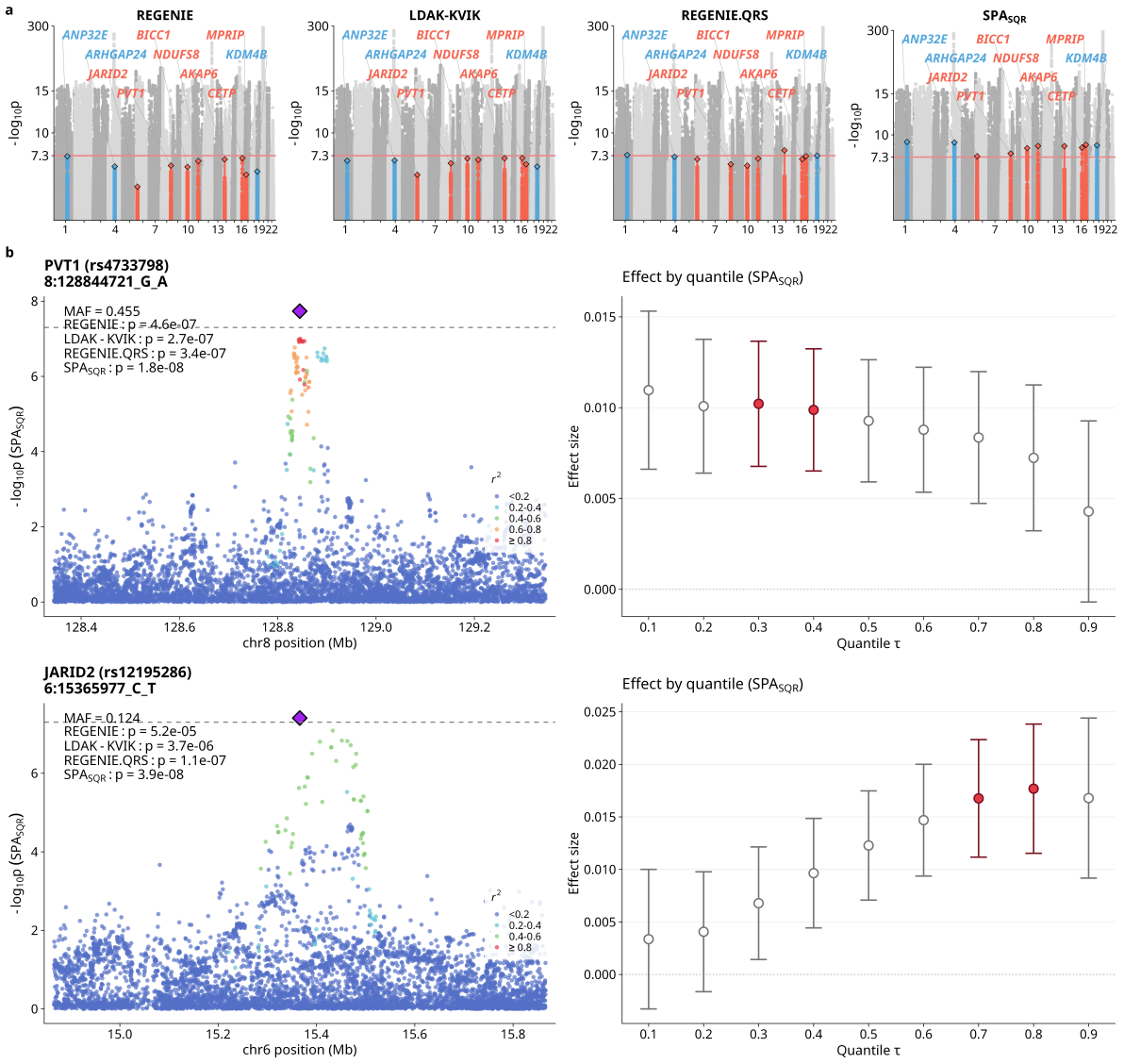
Supplementary Figure 13: Per-variant $-\log_{10} p$ from SPA$_{\mathbf{SQR}}$ ($h = \mathrm{IQR}(\tilde{\boldsymbol{Y}} - \hat{\boldsymbol{Y}}_{-c})/3$) rank-based inverse normal transformation (INT, y-axis) versus under standardization (STD, x-axis) , for the six highlighted traits (discovery cohort, ∼408,000 white British participants). Only variants with $3 < -\log_{10} p < 30$ on both axes are shown. The diagonal is the line $y = x$ and dashed lines mark genome-wide significance ($-\log_{10} p = 7.3$).



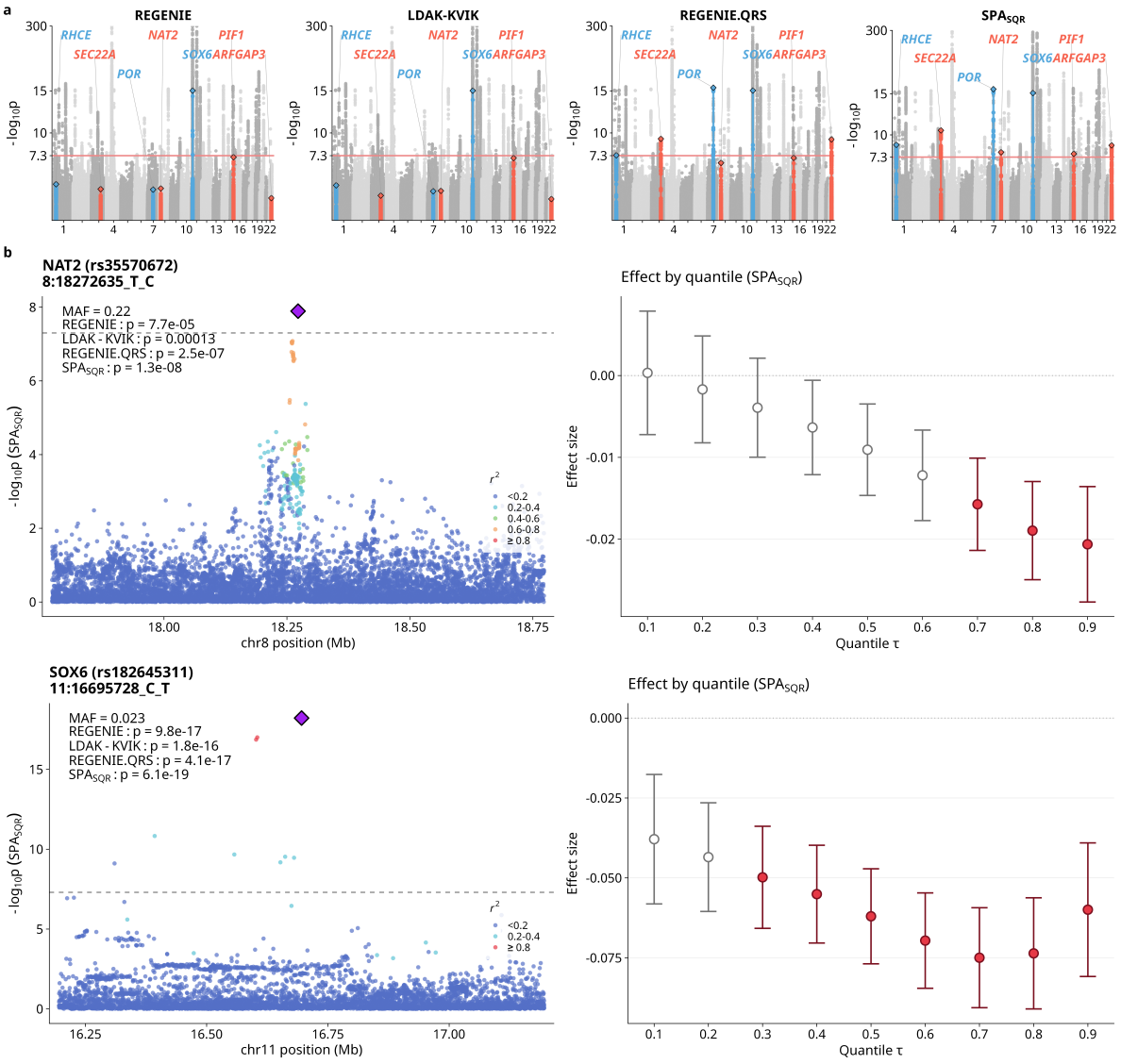
Supplementary Figure 14: Replication rates of 66 quantitative traits in the UK Biobank. In each panel the horizontal axis shows the replication rate of REGENIE and the vertical axis shows the replication rate of each of the other three methods—LDAK-KVIK (blue), REGENIE.QRS (green) and SPA_{SQR} (red); the dashed line is $y = x$. **a** Replication rate using a significance threshold of 0.05. **b** Replication rate using the more stringent threshold of 5×10^{-4}.



Supplementary Figure 15: GWAS results for sitting height (UKB field 20015) using approximately 408,000 UK Biobank white British individuals. a. Genome-wide Manhattan plots for REGENIE, LDAK-KVIK, REGENIE.QRS and SPA$_{\mathrm{SQR}}$. The red horizontal line marks the genome-wide significance threshold ($p = 5 \times 10^{-8}$). In each panel, we highlight up to 10 loci that were identified by SPA$_{\mathrm{SQR}}$ but missed by the baseline methods. Loci discovered by SPA$_{\mathrm{SQR}}$ but missed by both linear GWAS methods are highlighted in blue, while those missed by all three competing methods are highlighted in orange. **b.** Two GWAS loci with heterogeneous effect patterns. For each locus, the left panel shows a 1 Mb regional zoom plot, with each variant colored by its LD r^2 relative to the lead variant. The upper-left corner annotates its minor allele frequency and p-values from all four GWAS methods in the discovery cohort. The right panel displays the quantile-specific effect size $\hat{\gamma}(\tau)$ estimated at $\tau = 0.1, 0.2, \ldots, 0.9$, with error bars indicating 95% confidence intervals. Quantiles with genome-wide significant Wald test p-values are highlighted as filled red circles.



Supplementary Figure 16: GWAS results for cystatin C (UKB field 30720) using approximately 408,000 UK Biobank white British participants.



Supplementary Figure 17: GWAS results for vitamin D (UKB field 30890) using approximately 408,000 UK Biobank white British participants.

Supplementary Table 10: Novel SPA_{SQR}-discovered loci for Glucose highlighted in the Manhattan plot. These comprise loci missed by both linear methods (REGENIE and LDAK-KVIK) and loci additionally missed by REGENIE.QRS. Discovery p-values are from ∼408,000 white British and replication p-values from ∼50,000 non-British European participants.

SNP	Ref/Alt	MAF	Gene (location)	Stage	p-value			
					REGENIE	LDAK-KVIK	REGENIE.QRS	SPA_{SQR}
10:94466439	A/G	0.435	*HHEX* (11 kb downstream)	Disc.	1.6×10^{-6}	4.5×10^{-6}	8.3×10^{-20}	5.3×10^{-23}
				Repl.	0.0985	0.165	0.00324	0.0199
6:32626302	T/G	0.304	*HLA-DQB1* (1 kb downstream)	Disc.	2.8×10^{-4}	2.8×10^{-4}	1.5×10^{-11}	1.5×10^{-13}
				Repl.	0.0284	0.0256	0.0284	0.00723
2:227117778	C/G	0.353	*IRS1* (478 kb downstream)	Disc.	0.00173	0.00236	6.4×10^{-10}	1.2×10^{-11}
				Repl.	0.347	0.393	0.0149	0.0388
11:55982676	C/T	0.076	*OR5T2* (16 kb downstream)	Disc.	6.2×10^{-8}	5.6×10^{-8}	1.2×10^{-8}	6.5×10^{-10}
				Repl.	0.00583	0.00908	0.0291	0.0288
6:137256327	A/G	0.417	*SLC35D3* (9 kb downstream)	Disc.	1.4×10^{-7}	1.3×10^{-7}	2.9×10^{-9}	1.8×10^{-9}
				Repl.	0.148	0.15	0.00524	0.0185
3:12396913	G/A	0.120	*PPARG* (inside)	Disc.	1.8×10^{-5}	1.9×10^{-5}	1.2×10^{-7}	2.3×10^{-9}
				Repl.	0.0168	0.0196	0.0417	0.0469
10:126103639	G/GCGAGACTGTGT	0.182	*OAT* (inside)	Disc.	1.8×10^{-5}	6.6×10^{-5}	8.9×10^{-11}	5.1×10^{-9}
				Repl.	0.0863	0.0389	0.014	0.0162
11:55473187	C/G	0.075	*OR4C6* (39 kb downstream)	Disc.	1.4×10^{-6}	1.4×10^{-6}	1.3×10^{-7}	1.7×10^{-8}
				Repl.	0.00193	0.00322	0.021	0.0172
20:42996110	G/C	0.097	*HNF4A* (inside)	Disc.	8.8×10^{-4}	3.7×10^{-4}	1.9×10^{-7}	2.0×10^{-8}
				Repl.	0.0845	0.119	0.0682	0.0483
15:99279846	C/T	0.364	*IGF1R* (inside)	Disc.	1.6×10^{-5}	8.0×10^{-6}	8.8×10^{-7}	3.6×10^{-8}
				Repl.	0.0352	0.0285	0.0332	0.0411

Supplementary Table 11: Novel SPA_{SQR}-discovered loci for Apolipoprotein B highlighted in the Manhattan plot.

SNP	Ref/Alt	MAF	Gene (location)	Stage	p-value			
					REGENIE	LDAK-KVIK	REGENIE.QRS	SPA_{SQR}
16:81537760	C/G	0.255	*CMIP* (inside)	Disc.	2.7×10^{-6}	2.3×10^{-6}	4.6×10^{-9}	1.2×10^{-12}
				Repl.	0.125	0.0932	0.0539	0.041
2:136707982	T/C	0.241	*DARS* (inside)	Disc.	4.3×10^{-7}	3.8×10^{-7}	2.5×10^{-7}	4.7×10^{-10}
				Repl.	0.00878	0.0122	0.00922	0.0452
15:44027885	T/C	0.025	*PDIA3* (10 kb upstream)	Disc.	3.0×10^{-4}	1.9×10^{-4}	2.2×10^{-8}	8.0×10^{-10}
				Repl.	0.0486	0.0144	0.00137	0.00347
11:64004723	G/C	0.043	*VEGFB* (inside)	Disc.	5.2×10^{-5}	4.0×10^{-5}	1.3×10^{-7}	1.3×10^{-9}
				Repl.	7.5×10^{-5}	5.9×10^{-5}	1.7×10^{-4}	3.2×10^{-5}
4:155489608	C/T	0.005	*FGB* (inside)	Disc.	8.4×10^{-7}	7.7×10^{-7}	1.9×10^{-7}	2.0×10^{-9}
				Repl.	0.263	0.179	0.0194	0.033
17:66413965	A/G	0.241	*ARSG* (inside)	Disc.	2.3×10^{-7}	5.6×10^{-8}	1.7×10^{-7}	7.7×10^{-9}
				Repl.	0.0429	0.0296	0.0447	0.0294
17:47244412	C/T	0.349	*B4GALNT2* (inside)	Disc.	9.9×10^{-6}	8.6×10^{-7}	3.3×10^{-7}	7.8×10^{-9}
				Repl.	0.0522	0.0514	0.0537	0.0408
15:62270765	G/A	0.353	*VPS13C* (inside)	Disc.	7.5×10^{-7}	6.1×10^{-7}	8.2×10^{-8}	1.6×10^{-8}
				Repl.	0.0531	0.0224	0.046	0.0339
16:31074787	GA/G	0.382	*ZNF668* (inside)	Disc.	1.1×10^{-6}	3.7×10^{-7}	3.3×10^{-7}	2.4×10^{-8}
				Repl.	0.00601	0.0122	0.0507	0.0239
7:128418372	T/C	2.8×10^{-4}	*OPN1SW* (2 kb upstream)	Disc.	8.5×10^{-7}	6.7×10^{-7}	3.2×10^{-8}	3.3×10^{-8}
				Repl.	0.00847	0.00469	0.0235	0.011

Supplementary Table 12: Novel SPA_{SQR}-discovered loci for Body fat % highlighted in the Manhattan plot.

SNP	Ref/Alt	MAF	Gene (location)	Stage	*p*-value			
					REGENIE	LDAK-KVIK	REGENIE.QRS	SPA_{SQR}
3:156795525	A/G	0.399	*LEKR1* (31 kb downstream)	Disc.	2.9×10^{-6}	1.2×10^{-6}	2.6×10^{-13}	5.3×10^{-17}
				Repl.	0.152	0.189	0.0351	0.0472
12:57526646	G/A	0.392	*LRP1* (inside)	Disc.	3.1×10^{-8}	1.0×10^{-8}	2.0×10^{-8}	3.9×10^{-10}
				Repl.	0.0778	0.062	0.0344	0.0358
6:15381161	T/TC	0.487	*JARID2* (inside)	Disc.	1.2×10^{-7}	5.0×10^{-8}	1.2×10^{-8}	1.2×10^{-9}
				Repl.	0.078	0.0908	0.0241	0.0273
4:100019089	A/C	0.010	*ADH5* (9 kb upstream)	Disc.	3.0×10^{-7}	1.1×10^{-7}	8.7×10^{-9}	1.4×10^{-9}
				Repl.	0.0143	0.00266	0.00224	0.00542
3:25141824	A/C	0.390	*RARB* (73 kb upstream)	Disc.	6.2×10^{-7}	4.4×10^{-7}	1.5×10^{-6}	4.6×10^{-9}
				Repl.	0.00267	0.00581	0.0228	0.0071
8:106336304	T/A	0.473	*ZFPM2* (inside)	Disc.	3.8×10^{-7}	1.7×10^{-7}	1.7×10^{-7}	5.0×10^{-9}
				Repl.	0.143	0.0754	0.0813	0.0388
3:70897294	G/A	0.286	*FOXP1* (106 kb downstream)	Disc.	7.2×10^{-7}	3.0×10^{-7}	4.2×10^{-7}	6.3×10^{-9}
				Repl.	0.0273	0.0516	0.0294	0.0227
1:32175918	A/G	0.118	*COL16A1* (5 kb upstream)	Disc.	5.1×10^{-7}	1.8×10^{-7}	8.1×10^{-8}	7.6×10^{-9}
				Repl.	0.144	0.0997	0.0333	0.0421
13:98026620	C/G	0.305	*MBNL2* (inside)	Disc.	5.5×10^{-8}	1.8×10^{-7}	1.8×10^{-9}	9.2×10^{-9}
				Repl.	0.01	0.00989	0.00589	0.00705
1:84603487	G/A	0.241	*PRKACB* (inside)	Disc.	7.5×10^{-7}	9.6×10^{-8}	3.5×10^{-7}	1.3×10^{-8}
				Repl.	0.0983	0.0879	0.00237	0.00155

Supplementary Table 13: Novel SPA_{SQR}-discovered loci for Sitting height highlighted in the Manhattan plot.

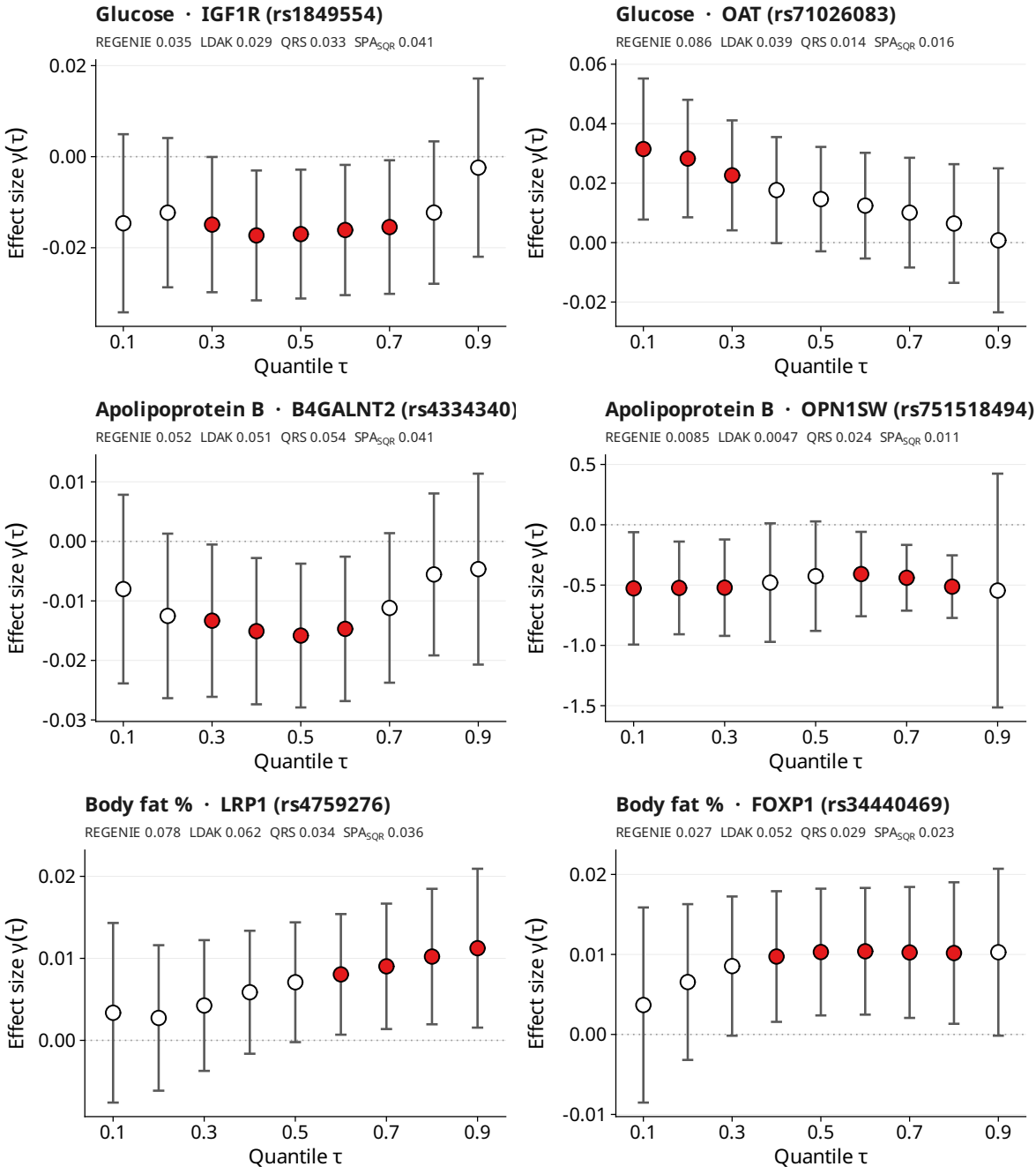
SNP	Ref/Alt	MAF	Gene (location)	Stage	*p*-value			
					REGENIE	LDAK-KVIK	REGENIE.QRS	SPA_{SQR}
3:135874930	T/G	0.268	*MSL2* (inside)	Disc.	3.1×10^{-7}	2.0×10^{-7}	1.9×10^{-10}	1.2×10^{-10}
				Repl.	0.00873	0.00112	1.3×10^{-4}	3.1×10^{-4}
1:8498326	T/TG	0.410	*RERE* (inside)	Disc.	2.6×10^{-6}	9.5×10^{-8}	4.3×10^{-8}	1.0×10^{-9}
				Repl.	0.0151	0.031	0.00353	0.019
5:64380395	G/A	0.329	*ADAMTS6* (64 kb downstream)	Disc.	1.4×10^{-6}	1.5×10^{-7}	1.2×10^{-7}	1.1×10^{-9}
				Repl.	8.5×10^{-4}	2.9×10^{-4}	0.00537	9.2×10^{-4}
3:136429906	T/A	0.218	*STAG1* (inside)	Disc.	2.9×10^{-7}	1.2×10^{-7}	2.2×10^{-8}	1.7×10^{-9}
				Repl.	0.00108	3.7×10^{-4}	2.6×10^{-5}	1.4×10^{-4}
8:88620617	AC/A	0.223	*CNBD1* (inside)	Disc.	1.3×10^{-6}	8.0×10^{-8}	4.4×10^{-7}	2.1×10^{-9}
				Repl.	0.0275	0.0414	0.0261	0.042
4:48260825	C/G	0.451	*TEC* (inside)	Disc.	1.2×10^{-6}	1.1×10^{-7}	1.7×10^{-8}	2.6×10^{-9}
				Repl.	0.00434	0.00113	1.3×10^{-4}	5.0×10^{-4}
13:73866451	TA/T	0.307	*KLF5* (214 kb downstream)	Disc.	7.2×10^{-7}	7.8×10^{-8}	3.7×10^{-8}	2.9×10^{-9}
				Repl.	0.154	0.157	0.967	0.035
2:183233310	T/C	0.491	*PDE1A* (inside)	Disc.	1.6×10^{-6}	6.5×10^{-8}	2.3×10^{-6}	3.3×10^{-9}
				Repl.	0.0511	0.0187	0.0844	0.0455
15:68279295	A/T	0.024	*PIAS1* (67 kb upstream)	Disc.	2.1×10^{-7}	5.0×10^{-8}	1.2×10^{-6}	6.5×10^{-9}
				Repl.	0.0603	0.0474	0.0629	0.0482
5:76000873	A/G	0.307	*IQGAP2* (inside)	Disc.	2.8×10^{-6}	6.8×10^{-8}	1.0×10^{-6}	2.5×10^{-8}
				Repl.	0.00854	0.0239	0.00577	0.0079

Supplementary Table 14: Novel SPA_{SQR}-discovered loci for Cystatin C highlighted in the Manhattan plot.

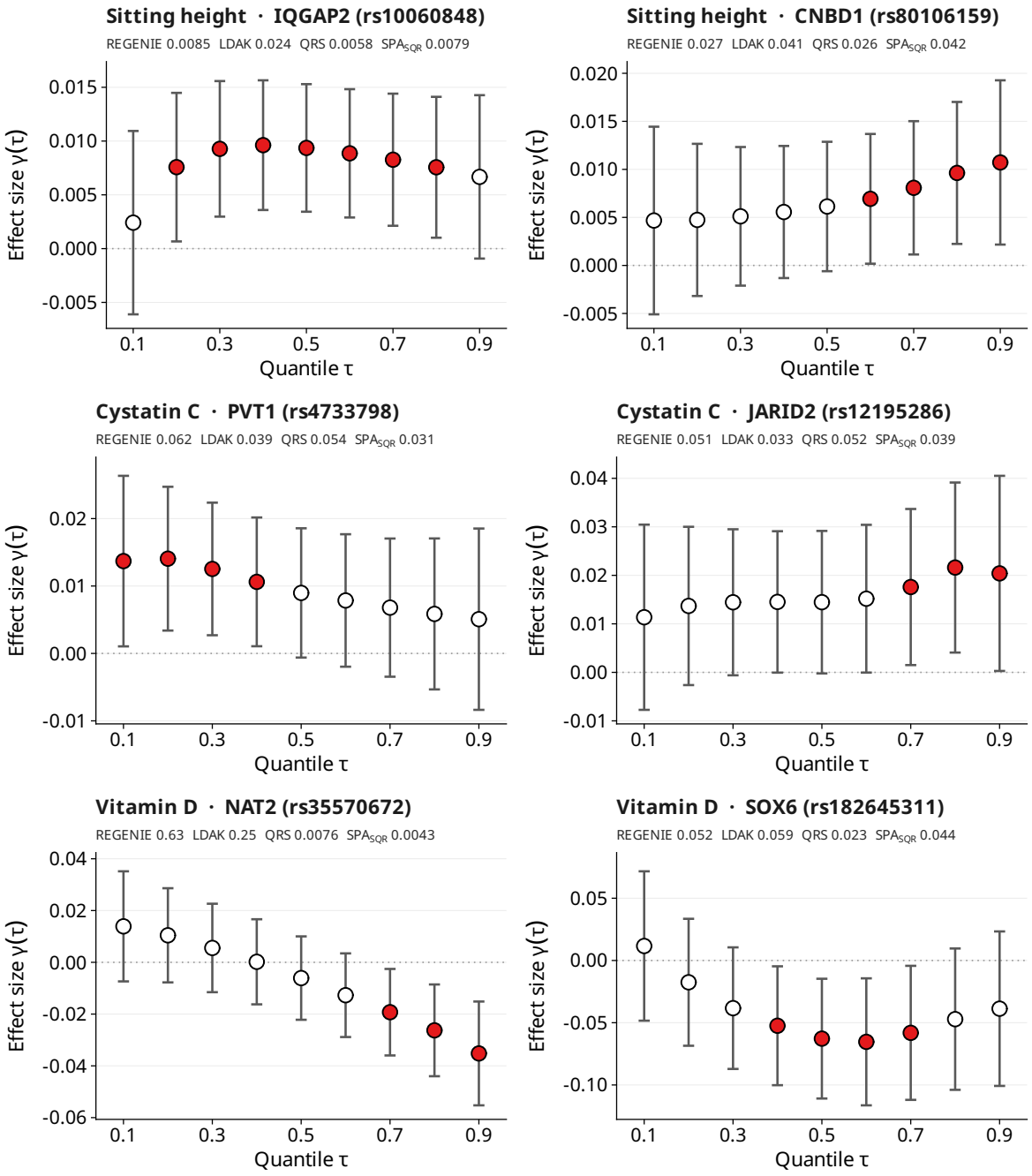
SNP	Ref/Alt	MAF	Gene (location)	Stage	*p*-value			
					REGENIE	LDAK-KVIK	REGENIE.QRS	SPA_{SQR}
1:150183322	G/T	0.134	*ANP32E* (7 kb downstream)	Disc.	5.7×10^{-8}	1.6×10^{-7}	4.1×10^{-8}	7.1×10^{-10}
				Repl.	0.157	0.168	0.014	0.0418
4:86857864	G/T	0.345	*ARHGAP24* (inside)	Disc.	5.6×10^{-7}	1.4×10^{-7}	6.5×10^{-8}	8.3×10^{-10}
				Repl.	0.0684	0.028	0.0444	0.0253
17:17075318	A/G	0.451	*MPRIP* (inside)	Disc.	3.5×10^{-6}	3.4×10^{-7}	5.6×10^{-8}	1.6×10^{-9}
				Repl.	0.00175	0.00167	0.00277	8.4×10^{-4}
19:4967739	C/T	0.207	*KDM4B* (1 kb upstream)	Disc.	1.7×10^{-6}	5.7×10^{-7}	5.0×10^{-8}	1.8×10^{-9}
				Repl.	0.00343	0.00483	7.9×10^{-4}	0.00232
11:67798381	G/A	0.093	*NDUFS8* (inside)	Disc.	1.7×10^{-7}	1.3×10^{-7}	9.6×10^{-8}	2.2×10^{-9}
				Repl.	0.0187	0.0639	0.0127	0.0361
14:33175822	G/A	0.423	*AKAP6* (inside)	Disc.	1.1×10^{-7}	8.9×10^{-8}	1.2×10^{-8}	2.3×10^{-9}
				Repl.	0.0179	0.00462	0.0601	0.0256
16:57000696	T/A	0.184	*CETP* (inside)	Disc.	8.7×10^{-8}	8.5×10^{-8}	1.1×10^{-7}	3.5×10^{-9}
				Repl.	0.0117	0.0349	0.062	0.0225
10:60371606	G/A	0.466	*BICC1* (inside)	Disc.	6.2×10^{-7}	9.4×10^{-8}	4.7×10^{-7}	4.1×10^{-9}
				Repl.	0.0149	0.00294	0.00315	0.00216
8:128844721	G/A	0.455	*PVT1* (inside)	Disc.	4.6×10^{-7}	2.7×10^{-7}	3.4×10^{-7}	1.8×10^{-8}
				Repl.	0.062	0.0391	0.0541	0.0312
6:15365977	C/T	0.124	*JARID2* (inside)	Disc.	5.2×10^{-5}	3.7×10^{-6}	1.1×10^{-7}	3.9×10^{-8}
				Repl.	0.051	0.0327	0.052	0.0386

Supplementary Table 15: Novel SPA_{SQR}-discovered loci for Vitamin D highlighted in the Manhattan plot.

SNP	Ref/Alt	MAF	Gene (location)	Stage	*p*-value			
					REGENIE	LDAK-KVIK	REGENIE.QRS	SPA_{SQR}
7:75612803	C/T	0.282	*POR* (inside)	Disc.	1.0×10^{-4}	1.4×10^{-4}	1.3×10^{-29}	2.7×10^{-36}
				Repl.	0.399	0.485	0.00282	0.00267
11:16695728	C/T	0.023	*SOX6* (inside)	Disc.	9.8×10^{-17}	1.9×10^{-16}	4.1×10^{-17}	6.1×10^{-19}
				Repl.	0.052	0.059	0.0227	0.044
3:122948722	A/T	0.024	*SEC22A* (inside)	Disc.	8.9×10^{-5}	3.8×10^{-4}	5.0×10^{-10}	2.8×10^{-11}
				Repl.	0.14	0.148	0.053	0.0435
1:25720248	C/G	0.442	*RHCE* (inside)	Disc.	3.0×10^{-5}	3.8×10^{-5}	4.5×10^{-8}	1.6×10^{-9}
				Repl.	0.0539	0.0533	0.0463	0.0492
22:43190817	T/C	0.447	*ARFGAP3* (1 kb downstream)	Disc.	6.7×10^{-4}	8.2×10^{-4}	5.8×10^{-10}	1.9×10^{-9}
				Repl.	0.653	0.779	0.0596	0.0375
8:18272635	T/C	0.220	*NAT2* (13 kb downstream)	Disc.	7.7×10^{-5}	1.3×10^{-4}	2.5×10^{-7}	1.3×10^{-8}
				Repl.	0.631	0.251	0.00761	0.00426
15:65095612	A/G	0.467	*PIF1* (12 kb downstream)	Disc.	7.0×10^{-8}	8.6×10^{-8}	8.3×10^{-8}	2.0×10^{-8}
				Repl.	0.0104	0.0108	0.013	0.0312



Supplementary Figure 18: The quantile-effect profiles of the highlighted $\mathrm{SPA_{SQR}}$ leads for glucose, apolipoprotein B and body fat percentage, in an independent cohort of ∼50,000 non-British European participants. Each panel shows the replication effect size $\hat{\gamma}(\tau) \pm 1.96\,\mathrm{SE}(\hat{\gamma}(\tau))$ estimated by $\mathrm{SPA_{SQR}}$ across quantile levels $\tau \in \{0.1, \ldots, 0.9\}$; a filled red point indicates nominal replication ($p < 0.05$). The four replication p-values above each panel are the combined p values across 9 quantiles.



Supplementary Figure 19: The quantile-effect profiles of the highlighted $\mathrm{SPA_{SQR}}$ leads for sitting height, cystatin C and vitamin D, in an independent cohort of ∼50,000 non-British European participants. Each panel shows the replication effect size $\hat{\gamma}(\tau) \pm 1.96\,\mathrm{SE}(\hat{\gamma}(\tau))$ estimated by $\mathrm{SPA_{SQR}}$ across quantile levels $\tau \in \{0.1, \ldots, 0.9\}$; a filled red point indicates nominal replication ($p < 0.05$). The four replication p-values above each panel are the combined p values across 9 quantiles.

Supplementary Table 16: Computational efficiency for the simultaneous analysis of multiple traits. LDAK-KVIK, REGENIE and SPA_{SQR} were each run as a single multi-trait command on the 10-million-variant simulated cohorts ($N = 400{,}000$ and $N = 50{,}000$), under two settings: 10 traits with 10 threads and 20 traits with 20 threads. SPA_{SQR} has no Step 1 and uses the LDAK-KVIK LOCO PGS for each trait. Wall-clock time and peak memory is captured by `/usr/bin/time -v`.

Experiment setting	Method	Stage	N	Wall (h)	Peak memory (GB)
10 traits 10 threads	LDAK-KVIK	Step 1	400,000	9.27	7.8
			50,000	1.72	3.5
		Step 2	400,000	7.08	8.1
			50,000	0.94	3.9
	REGENIE	Step 1	400,000	2.29	85.3
			50,000	0.78	10.9
		Step 2	400,000	5.27	5.5
			50,000	0.72	3.6
	SPA_{SQR}	Step 2	400,000	2.61	7.7
			50,000	0.34	4.7
20 traits 20 threads	LDAK-KVIK	Step 1	400,000	17.01	11.5
			50,000	3.00	5.6
		Step 2	400,000	12.09	9.7
			50,000	1.13	4.1
	REGENIE	Step 1	400,000	2.36	164.2
			50,000	0.77	21.0
		Step 2	400,000	4.42	5.7
			50,000	0.70	3.6
	SPA_{SQR}	Step 2	400,000	2.15	11.2
			50,000	0.26	5.5

Supplementary References

[1] Rounak Dey, Ellen M Schmidt, Goncalo R Abecasis, and Seunggeun Lee. A fast and accurate algorithm to test for binary phenotypes and its application to phewas. *The American Journal of Human Genetics*, 101(1):37–49, 2017.

[2] Marcelo Fernandes, Emmanuel Guerre, and Eduardo Horta. Smoothing quantile regressions. *Journal of Business & Economic Statistics*, 39(1):338–357, 2021.

[3] Xuming He, Xiaoou Pan, Kean Ming Tan, and Wen-Xin Zhou. Smoothed quantile regression with large-scale inference. *Journal of Econometrics*, 232(2):367–388, 2023.

[4] Qiang Heng, Hua Zhou, and Kenneth Lange. Tactics for improving least squares estimation. *Statistical Science*, 2026. Accepted.

[5] Qiang Heng and Caixing Wang. Inertial quadratic majorization minimization with application to kernel regularized learning. *arXiv preprint arXiv:2507.04247*, 2025.

[6] Leonard A Stefanski and Dennis D Boos. The calculus of M-estimation. *The American Statistician*, 56(1):29–38, 2002.

[7] Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1): 76–82, 2011.

[8] Christopher C. Chang, Carson C. Chow, Laurent C. A. M. Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1):7, 2015.

[9] He Xu, Yuzhuo Ma, Lin-Lin Xu, Yin Li, Yufei Liu, Ying Li, Xu-Jie Zhou, Wei Zhou, Seunggeun Lee, Peipei Zhang, Weihua Yue, and Wenjian Bi. SPA_{GRM}: effectively controlling for sample relatedness in large-scale genome-wide association studies of longitudinal traits. *Nature Communications*, 16(1):1413, 2025.

[10] Gonçalo R. Abecasis, Stacey S. Cherny, William O. Cookson, and Lon R. Cardon. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, 30(1): 97–101, 2002.