Supplementary Information for "Contrasting regional architectures of schizophrenia and other complex diseases using fast variance components analysis"

Po-Ru Loh, Gaurav Bhatia, Alexander Gusev, Hilary K Finucane, Brendan K Bulik-Sullivan, Samuela J Pollack, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Teresa R de Candia, Sang Hong Lee, Naomi R Wray, Kenneth S Kendler, Michael C O'Donovan, Benjamin M Neale, Nick Patterson, Alkes L Price

Contents

1 Literature on REML algorithms

We begin by overviewing the existing literature on restricted maximum likelihood (REML) [24] estimation for linear mixed models (LMMs)—i.e., variance components analysis—focusing on methods that have been most commonly applied in genetics and methods that apply Monte Carlo approximation and thus are germane to this work.

1.1 Exact methods

By "exact methods" we mean methods that locally converge to the REML optimum (vs. methods that approximate the REML optimum, which we discuss next). All REML procedures are iterative and compute analytically exact solutions only in the limit of infinite iterations; additionally, all iterative optimizers are at least somewhat sensitive to their starting points and are not guaranteed to find the global optimum from an arbitrary starting point.

First derivative methods. These methods perform an analog of gradient descent and thus are most robust—as they monotonically increase the likelihood and never leave the valid parameter domain—but converge slowly.

- **Expectation maximization (EM).** Dempster et al. [66] is the canonical reference; see Sec. 8.3 of Searle's text [67] for an in-depth discussion and derivation of update equations. The popular GCTA software [2] offers EM as a non-default option.
- **EM with parameter expansion (PX-EM).** Liu et al. [68] first proposed PX-EM as a general approach to speed up EM convergence; Foulley and van Dyk [69] provide a comprehensive treatment of PX-EM for LMMs.

Second derivative methods. These methods use exact or approximate second derivative information about the log likelihood surface and therefore converge much faster than EM when near the optimum but are less robust when far from the optimum. Consequently, many implementations start by taking a few EM steps (GCTA) or PX-EM steps (GEMMA [70], PX-AI [71]). Second derivative methods can be classified by the way in which the Hessian is computed or approximated.

- **Newton-Raphson (NR)** (exact Hessian). Newton's method is the standard derivative-based optimization procedure; in the context of REML, Zhou and Stephens [70] provided an important computational advance for NR REML by developing an efficient eigendecomposition-based algorithm (implemented in the GEMMA software) to compute the gradient and Hessian of the log likelihood.

- **Method of scoring, a.k.a. Fisher scoring** (approximate Hessian). This approach replaces the negative Hessian, which is the *observed* Fisher information, with its expected value, which is the information matrix (*independent* of the observation y) and is sometimes easier to compute than the Hessian [67]. GCTA offers Fisher scoring as a non-default option.

- **Average information (AI)** (approximate Hessian). Gilmour et al. [26] proposed using a matrix equal to the average (appropriately defined) of the negative Hessian (used in Newton-Raphson) and the Fisher information (used in the method of scoring), terming this matrix the "average information" as it is an average of the observed and expected information. This matrix is computationally convenient because it eliminates a trace term involving an inverse matrix; however, AI REML still requires $O(N^3)$-time matrix operations. GCTA uses AI REML as its default option.

- **Broyden's method** (approximate Hessian, quasi-Newton). The basic idea is to avoid direct computation of the Hessian by approximating it using a multi-dimensional analog of the secant method. Groeneveld [72] applies Broyden's method to REML in genetics.

1.2 Monte Carlo methods

Unlike exact methods, Monte Carlo REML methods use random sampling to approximate derivatives of the log likelihood, with the goal of trading off a small amount of accuracy in favor of greater computational speed.

Existing aproaches.

- **MC EM.** Wei and Tanner [73] first proposed Monte Carlo EM for general EM parameter estimation; Garcia-Cortes et al. [25] applied MC EM to LMMs. Matilainen et al. [74] investigate MC EM REML in detail; see the introduction of their paper for further references.

- **MC "NR"/AI/Broyden.** To our knowledge, Matilainen et al. [23] were the first to use Monte Carlo to directly approximate second derivative-based methods—which offer far faster convergence than EM—in the context of LMM REML, where the computational challenge is problem size rather than model complexity. They investigated three algorithms that all use Monte Carlo sampling to approximate the gradient of the log likelihood and differ in their Hessian approximation methods:

    1. MC "NR" REML: Uses Monte Carlo to approximate the Fisher information matrix (not the exact Hessian, as in true Newton-Raphson).
    2. MC AI REML: Uses the AI matrix, which does not require Monte Carlo, only requiring solution of the mixed model equations.

3. MC BM REML: Uses Broyden's method to estimate the Hessian from Monte Carlo gradients.

Matilainen et al.'s study [23] is a proof of concept and compares the three algorithms above with MC EM REML on a set of simulations.

As an aside, the term "Monte Carlo Newton-Raphson" is unfortunately confusing because it refers to several methods. Many authors have investigated Monte Carlo approaches within routines for solving general maximum likelihood problems for which analytic solutions do not exist and hence Newton-Raphson iteration is used (within EM). In these cases, MC NR refers to applying MCMC or Gibbs sampling from conditional distributions [75, 76].

1.3 Contribution of BOLT-REML

The algorithm we present here, BOLT-REML, uses a Monte Carlo AI REML algorithm that broadly speaking applies the same framework outlined in Matilainen et al. [23], thus reducing per-iteration running time to $\approx O(MN^{1.5})$. However, we make the following contributions that are essential to making the method robust and practical:

- **Trust region optimization using approximate change in log likelihood.** We achieve robust convergence by applying a trust region method to detect and prevent large, detrimental steps when far from the optimum. (Trust region methods are among the most commonly used numerical methods for robust multi-parameter optimization [77].) The basic idea is to identify where the local quadratic model of the log likelihood (based on its approximate gradient and Hessian) starts breaking down by comparing model-predicted vs. computed values of the log likelihood. Here, computing the log likelihood is infeasible, so instead we apply a gradient-based approximation of the change in log likelihood.

- **Simple convergence criterion via reuse of Monte Carlo samples.** By reusing pseudo-random Monte Carlo sample components across iterations, the *approximate, Monte Carlo* gradient we compute varies smoothly with variance parameter estimates. (The AI matrix is exactly computed and does not use Monte Carlo, hence is always a smooth function of the variance parameters.) Consequently, our MC AI REML iteration truly converges rather than jumping around near the optimum (as is typically the case for Monte Carlo parameter optimization methods), allowing easy assessment of convergence. Garcia-Cortes et al. [25] did take such an approach in their MC EM REML algorithm, but Matilainen et al. [23] did not, instead suggesting more complex convergence criteria that have been developed for general MC EM methods.

- **Efficient software implementation.** We implement our MC AI REML algorithm in a well-optimized, open source software package. In particular, BOLT-REML uses a computational

framework requiring no $O(MN)$ or $O(N^2)$ memory allocations beyond genotype data storage and performs all rate-limiting computations using multithreaded BLAS Level 3 matrix operations.

When specialized to the single-trait, single-variance component case, BOLT-REML is very similar to the heritability estimation procedure implemented in BOLT-LMM [20]; we expect that our use of secant iteration in BOLT-LMM vs. the trust region approach in BOLT-REML is unlikely to reduce robustness in the univariate case.

2 Mathematics of BOLT-REML

2.1 Single-trait case

We begin by developing the mathematics of Monte Carlo REML for one phenotype. We will later generalize to multiple phenotypes, but the multi-trait notation is more cumbersome, so we start with the single-phenotype case to convey the intuition behind the approach.

2.1.1 Model and log likelihood

We consider the model [21]

$$y = \sigma_0 u_0 + \sigma_1 Z_1 u_1 + \cdots + \sigma_K Z_K u_K,$$

in which y denotes phenotype, $\sigma_1 Z_1 u_1, \ldots, \sigma_K Z_K u_K$ are K random genetic effects, and $\sigma_0 u_0$ is a random environmental (noise) effect, where $\sigma_0, \ldots, \sigma_K$ are variance parameters to be estimated, the Z_k contain normalized SNPs corresponding to the k-th non-identity variance component, and all entries of all the u_k are iid standard normal. (Note that the u_k are vectors of different lengths: u_k has length equal to the number of SNPs in the k-th variance component. Also note that for the case of a single non-identity variance component containing M SNPs with normalized genotypes X, we may write $Z = X/\sqrt{M}$ and $u = \sqrt{M}\beta$, so that $\beta \sim N(0, \sigma_g^2/M) \iff u \sim N(0, \sigma_g^2)$, $X\beta = Zu$, and $XX'/M = ZZ'$ is the GRM.) We have not included fixed effect covariates in the model for notational convenience, but covariates can be incorporated by projecting them out of all components and adjusting dimensionality accordingly (the REML approach [24]).

We wish to find the values of the variance parameters $\sigma_0^2, \sigma_1^2, \ldots, \sigma_K^2$ that maximize the likelihood of observing the phenotype y (a vector of observations over N individuals) under the above model. We can write the model equivalently as $y \sim N(0, V)$, where

$$V = \sigma_0^2 I_N + \sigma_1^2 Z_1 Z_1' + \cdots + \sigma_K^2 Z_K Z_K'.$$

The pdf of this multivariate normal distribution is given by

$$f(y) = \frac{1}{\sqrt{(2\pi)^N \det V}} e^{-\frac{1}{2} y' V^{-1} y},$$

where N is the number of samples (which changes to $N - C$ in REML with C independent covariates projected out). Thus, up to a constant that we ignore, the log likelihood of observing y assuming parameters $\sigma_0^2, \sigma_1^2, \ldots, \sigma_K^2$ is given by

$$\ell = \ell(\sigma_0^2, \sigma_1^2, \ldots, \sigma_K^2) = -\frac{1}{2}(\log \det V + y' V^{-1} y).$$

2.1.2 Monte Carlo approximation of the gradient of the log likelihood; MC REML

We may compute the gradient and Hessian of the log likelihood wrt σ_k^2 by using the matrix calculus formulas

$$d \log \det V = \text{tr } V^{-1} dV, \qquad dV^{-1} = -V^{-1} dV V^{-1},$$

giving

$$\frac{\partial \ell}{\partial \sigma_k^2} = -\frac{1}{2} \left(\text{tr } V^{-1} Z_k Z_k' - y' V^{-1} Z_k Z_k' V^{-1} y \right).$$

For large sample sizes, building and inverting V is computationally challenging, leading us to rewrite the trace term in terms of an expectation that does not involve matrix inverses, following equation (14.12) of ref. [78]:

$$\begin{aligned}
\frac{\partial \ell}{\partial \sigma_k^2} &= -\frac{1}{2} \left(\text{tr } V^{-1} Z_k Z_k' V^{-1} V - y' V^{-1} Z_k Z_k' V^{-1} y \right) \\
&= -\frac{1}{2} \left(\text{tr } V^{-1} Z_k Z_k' V^{-1} E[y_V y_V'] - y' V^{-1} Z_k Z_k' V^{-1} y \right) \\
&= -\frac{1}{2} \left(E[\text{tr } V^{-1} Z_k Z_k' V^{-1} y_V y_V'] - y' V^{-1} Z_k Z_k' V^{-1} y \right) \\
&= -\frac{1}{2} \left(E[y_V' V^{-1} Z_k Z_k' V^{-1} y_V] - y' V^{-1} Z_k Z_k' V^{-1} y \right), \qquad (1)
\end{aligned}$$

where y_V is a random vector such that

$$E[y_V y_V'] = V.$$

For fixed σ_k^2, sampling such a vector y_V is easily done: we build

$$y_V = \sqrt{\sigma_0^2} u_0 + \sqrt{\sigma_1^2} Z_1 u_1 + \cdots + \sqrt{\sigma_K^2} Z_K u_K, \qquad (2)$$

where the components of all the u_k vectors are iid standard normal. We note in passing that were V a linear combination of variance components with some negative coefficients, we could still apply the above trick by writing $V = E[y_{V^+} y'_{V^+}] - E[y_{V^-} y'_{V^-}]$ with y_{V^+} and y_{V^-} built from terms corresponding to the positive and negative variance components, respectively.

The first-order conditions for optimality of the variance parameters are the vanishing of the gradient of the log likelihood, i.e., $\frac{\partial \ell}{\partial \sigma_k^2} = 0$. As noted in ref. [78], $\sigma_k^2 Z'_k V^{-1} y$ is the BLUP estimate $\widehat{u_k}$ of u_k, so the terms appearing in equation (1) are expected and observed squared norms of BLUP estimates, and REML optimization amounts to equating these quantities. We previously presented the rudimentary heritability estimation algorithm used within the BOLT-LMM mixed model association software [20] in this way. Similarly, the parameter updates of EM REML are scaled differences between observed and expected squared norms of BLUP estimates (Sec. 4); comparing equations (1) and (13–14) show that EM REML is very closely related to gradient descent.

In practice, we estimate the gradient of the log likelihood by replacing the expectation in equation (1) with an empirical average over a finite number S of Monte Carlo samples of y_V:

$$\overline{\frac{\partial \ell}{\partial \sigma_k^2}} = -\frac{1}{2}\left(\overline{y'_V V^{-1} Z_k Z'_k V^{-1} y_V} - y' V^{-1} Z_k Z'_k V^{-1} y\right), \tag{3}$$

where the overline denotes a Monte Carlo average over S samples. For large N, typically only a few samples are required to obtain precise enough estimates of these derivatives to use within an optimization routine. Computing the quantities on the right only requires solving the mixed model equations, which can be done efficiently with conjugate gradient iteration [18, 19]. Moreover, by reusing the same S sets of pseudorandom components $u_0, \ldots, u_K$ used in equation (2) to build the Monte Carlo samples y_V [25] for all parameter values σ_k^2, the Monte Carlo gradient $\overline{\nabla \ell} = \left(\overline{\frac{\partial \ell}{\partial \sigma_k^2}}\right)$ becomes a smooth function of the variance parameters σ_k^2, which is convenient for numerical optimization. We refer to the approach of approximately maximizing the log likelihood by solving $\overline{\nabla \ell} = 0$ as MC REML.

We note that while we have an efficient Monte Carlo method for estimating the gradient of the log likelihood, this method does not estimate the log likelihood itself, which involves a log determinant. We are not aware of an efficient way to estimate the log determinant that applies here; ref. [79] proposes an approximation based on the expansion

$$-\log\det(I - A) = -\text{tr}\ \log(I - A) = \text{tr}\,(A + \frac{1}{2}A^2 + \frac{1}{3}A^3 + \cdots) = \sum E\left[\frac{1}{k} \cdot \frac{x' A^k x}{x' x}\right],$$

where $x \sim N(0, I)$ are Monte Carlo samples, but this method requires the eigenvalues of A to be between -1 and 1 and appears unsuitable here.

2.1.3 Average information (AI) approximation of the Hessian

We may similarly express the second-order partials (i.e., the entries of the Hessian) as

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \sigma_k^2 \partial \sigma_j^2} &= -\frac{1}{2}\left(-\text{tr}\, V^{-1} Z_k Z_k' V^{-1} Z_j Z_j' + 2y' V^{-1} Z_k Z_k' V^{-1} Z_j Z_j' V^{-1} y\right) \\ &= -\frac{1}{2}\left(-E[y_V' V^{-1} Z_k Z_k' V^{-1} Z_j Z_j' V^{-1} y_V] + 2y' V^{-1} Z_k Z_k' V^{-1} Z_j Z_j' V^{-1} y\right). \quad (4)\end{aligned}$$

Instead of approximating this matrix using Monte Carlo sampling, we use the average information (AI) approximation of ref. [26]. This approximation amounts to replacing the random y_V in the expectation term of equation (4) with the observed phenotype y, giving

$$\frac{\partial^2 \ell}{\partial \sigma_k^2 \partial \sigma_j^2} \approx -\frac{1}{2} y' V^{-1} Z_k Z_k' V^{-1} Z_j Z_j' V^{-1} y = \mathscr{I}_A.$$

2.2 Multiple traits

We now generalize to multiple traits, following ref. [22, 70, 80]. To avoid notational clutter, we assume a model with only one genetic effect term, but all of the following easily generalizes to multiple random genetic effects.

2.2.1 Model

Consider D correlated traits, each measured at the same N individuals. We denote the "stacked" DN-vector of trait values by $\mathbf{y}$ and denote the N-element subvectors corresponding to individual traits by $\mathbf{y}_d$, $d = 1, \ldots, D$, so that

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_D \end{pmatrix}.$$

We assume the linear random effects model

$$\mathbf{y} = (I_D \otimes Z)\mathbf{u}_g + \mathbf{u}_e, \quad (5)$$

where Z is an $N \times M$ coefficient matrix (e.g., $Z = X/\sqrt{M}$ under our usual SNP normalization), $\mathbf{u}_g$ is a stacked DM-vector of effect sizes, and $\mathbf{u}_e$ is a stacked DN-vector of noise residuals such that

$$\mathbf{u}_g \sim N(\mathbf{0}, V_g \otimes I_M), \quad \mathbf{u}_e \sim N(\mathbf{0}, V_e \otimes I_N).$$

Thus,

$$\mathbf{y} \sim N(\mathbf{0}, \mathbf{V}),$$

where

$$\mathbf{V} = (V_g \otimes ZZ') + (V_e \otimes I_N).$$

The distributional assumption on $\mathbf{u}_g$ says that for every effect index m, the d-vector of effect sizes for the d traits is distributed as $N(0, V_g)$, and across different m, effect sizes are independent of each other. An analogous statement holds for the environmental terms $\mathbf{u}_e$. Marginalizing across individual traits, each trait satisfies the usual LMM:

$$\mathbf{y}_d = Z\mathbf{u}_{g,d} + \mathbf{u}_{e,d}.$$

We treat the genetic and environmental $D \times D$ covariance matrices V_g and V_e, each of which have $D(D+1)/2$ independent coordinates (due to symmetry), as unknowns to be estimated. In the general case of K random genetic effects, each random effect as well as the environmental term has its own set of $D(D+1)/2$ covariance parameters, so that the model has a total of $P = (1+K)D(D+1)/2$ parameters.

We note that fixed effect covariates need to be included independently for every trait. In particular, for D traits, D "all-1s" covariate vectors need to included, namely, the indicator vectors corresponding to the coordinates of each of the D traits. In REML analysis, doing so is equivalent to projecting covariates out from the genotypes and each phenotype independently (and adjusting dimensionality accordingly).

2.2.2 Monte Carlo gradient approximation

We denote the log likelihood of $\mathbf{y}$ assuming variance parameters V_g, V_e by $\ell = \ell(V_g, V_e)$ as before, and we combine the distinct entries of the covariance matrices V_g, V_e ($D(D+1)/2$ entries each) into the parameter vector $\theta = (\theta_p)$, where $p = 1, \ldots, P$ indexes parameters. Thus,

$$\begin{aligned}
\ell &= -\frac{1}{2}\left(\log\det \mathbf{V} + \mathbf{y}'\mathbf{V}^{-1}\mathbf{y}\right) \\
\frac{\partial \ell}{\partial \theta_p} &= -\frac{1}{2}\left(\text{tr}\, \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p} - \mathbf{y}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\mathbf{y}\right) \\
&= -\frac{1}{2}\left(E\left[\mathbf{y}'_{\mathbf{V}}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\mathbf{y}_{\mathbf{V}}\right] - \mathbf{y}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\mathbf{y}\right).
\end{aligned}$$

To compute a Monte Carlo estimate $\overline{\nabla \ell}$ of the gradient $\nabla \ell = \left(\frac{\partial \ell}{\partial \theta_p}\right)$, we replace the expectation with an empirical average over S Monte Carlo samples of $\mathbf{y}_{\mathbf{V}} \sim N(\mathbf{0}, \mathbf{V})$. Explicitly, we generate $\mathbf{y}_{\mathbf{V}}$ via

$$\mathbf{y}_{\mathbf{V}} = (I_D \otimes Z)\mathbf{u}_{g,\text{rand}} + \mathbf{u}_{e,\text{rand}}$$

where

$$\mathbf{u}_{g,\text{rand}} = \texttt{vec}(\texttt{randn}(M, D) \cdot \texttt{chol}(V_g)'), \quad \mathbf{u}_{e,\text{rand}} = \texttt{vec}(\texttt{randn}(N, D) \cdot \texttt{chol}(V_e)'),$$

so that

$$\mathbf{y}_\mathbf{V} = \texttt{vec}(Z \cdot \texttt{randn}(M, D) \cdot \texttt{chol}(V_g)' + \texttt{randn}(N, D) \cdot \texttt{chol}(V_e)').$$

To make the Monte Carlo gradient $\overline{\nabla \ell}$ a smooth function of θ, we generate S fixed sets of matrices $\texttt{randn}(M, D)$ and $\texttt{randn}(N, D)$, one per Monte Carlo trial, and reuse these matrices for all parameter values θ, as described in Sec. 2.1.2.

2.2.3 Average information (AI) approximation of the Hessian

As in equation (4), we write

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \theta_p \partial \theta_q} &= -\frac{1}{2}\left(-\text{tr}\, \mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_q} + 2\mathbf{y}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_q}\mathbf{V}^{-1}\mathbf{y}\right) \quad (6) \\ &\approx -\frac{1}{2}\mathbf{y}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_q}\mathbf{V}^{-1}\mathbf{y} = -\mathscr{I}_A. \quad (7) \end{aligned}$$

to approximate the Hessian using the negative AI matrix. (Note that the second-order partials of $\mathbf{V}$ do not appear in equation (6) because under our parameterization, $\partial \mathbf{V}/\partial \theta_p$ is a constant matrix for each θ_p, so that second-order partials conveniently vanish.)

The AI approximation has two significant advantages over Monte Carlo estimation of the Hessian. First, it is guaranteed to be positive semidefinite, leading to more robust numerical optimization of the log likelihood, especially when some covariances are close to singular. Positive semidefiniteness can be seen by horizontally concatenating the P vectors $\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\mathbf{y}$ into a $DN \times P$ matrix

$$\mathbf{W} = \left[\frac{\partial \mathbf{V}}{\partial \theta_1}\mathbf{V}^{-1}\mathbf{y} \;\; \cdots \;\; \frac{\partial \mathbf{V}}{\partial \theta_P}\mathbf{V}^{-1}\mathbf{y}\right],$$

in which case we may write

$$\mathscr{I}_A = \frac{1}{2}\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}.$$

Because $\mathbf{V}$ is psd (i.e., has a Cholesky factorization), it follows that the same holds for $\mathscr{I}_A$.

Second, computing $\mathscr{I}_A$ is cheaper than using Monte Carlo: evaluating the expressions in equation (7) only requires solving the mixed model equations once per variance parameter θ_p.

2.3 Standard errors

In general, standard errors of maximum likelihood parameter estimates can be obtained by approximating the covariance matrix of the estimates with the inverse of the negative Hessian. (This approximation amounts to fitting a quadratic form to the log likelihood, noting the vanishing of its gradient at the MLE. The negative Hessian is the observed Fisher information and measures the curvature of the log likelihood.) Approximating the negative Hessian with the AI matrix, we have

$$\text{Cov}\left(\widehat{\theta_p}\right) \approx \mathscr{I}_A^{-1}. \tag{8}$$

The usual (non-Monte Carlo) AI REML estimated standard error of $\widehat{\theta_p}$ is the square root of the p-th diagonal entry of the above covariance. For MC AI REML, we multiply the standard error by a factor of $\sqrt{1 + 1/S}$ to take into account the additional variance introduced because MC AI REML essentially fits the observed data to a reference set of S simulated data sets.

2.3.1 Standard errors under transformed coordinates

It is most convenient (for exposition and implementation) to parameterize variance components using entries of the covariance matrices as we have done above: as noted earlier, the (θ_p) parameterization results in simple formulas for $\frac{\partial \mathbf{V}}{\partial \theta_p}$ and the vanishing of second-order partials of $\mathbf{V}$. However, the quantities we typically wish to report are heritabilities (i.e., fractions of total variance contributed by random effects) or genetic correlations (i.e., ratios of off-diagonal covariance parameters to square roots of products of diagonal entries). Computing point estimates of these quantities is trivial once maximum likelihood variance parameter estimates have been found, but estimating their standard errors requires transforming the Hessian of the log likelihood according to the desired reparameterization, which requires a little more work. We derive the transformation formulas below; once we have the Hessian of the log likelihood with respect to transformed coordinates, we can then estimate error covariances by the usual approach of inverting the negative Hessian.

Consider a scalar function $f(x)$ (where x is vector-valued); denote its Hessian by H_x. Suppose we wish to reparameterize in terms of coordinates X via a transformation $\tau : X \mapsto x$ with Jacobian $J = \left(\frac{\partial x}{\partial X}\right)$. Letting $F(X) = f(\tau(X)) = f(x)$, we have

$$\nabla_X F = J' \nabla_x f,$$

i.e., $(\nabla_X F)' = (\nabla_x f)' J$, so that

$$H_X = \nabla_X (\nabla_X F)' = J' \nabla_x ((\nabla_x f)' J).$$

In general, applying the product rule to $\nabla_x((\nabla_x f)'J)$ gives both a Hessian term $\nabla_x(\nabla_x f)'J = H_x J$ and a term involving $\nabla_x f'$ and derivatives of J. However, at a critical point of f (such as the MLE if f is a log likelihood), the gradient vanishes, leaving the simple formula

$$H_X = J' H_x J \qquad \text{if } \nabla_x f = 0. \tag{9}$$

We also note that the AI matrix defined in equation (7) obeys the same transformation formula under change of coordinates. We can derive the transformation by observing that for a reparameterization $\phi \mapsto \theta$, the partial derivatives of $\mathbf{V}$ transform according to

$$\frac{\partial \mathbf{V}}{\partial \phi_q} = \sum_p \frac{\partial \theta_p}{\partial \phi_q} \frac{\partial \mathbf{V}}{\partial \theta_p}.$$

The partials of θ_p with respect to the new parameters ϕ_q are entries of the Jacobian, and expanding gives the transformation $\mathscr{I}_A \mapsto J' \mathscr{I}_A J$.

In BOLT-REML, we implement the following two sets of transformations to reparameterize from covariance parameters θ_p (i.e., entries of V_g and V_e) to genetic correlations and heritabilities. We carry out the first set of transformations independently across all variance components (environmental and genetic), after which we carry out the second set independently across all traits.

Transformation from covariances to correlations. For each variance component, we have $D(D+1)/2$ independent variables that parameterize its covariance matrix. We wish to transform the $D(D-1)/2$ off-diagonal covariance parameters into correlations: that is, we start with $D(D+1)/2$ parameters σ_{ij}, $1 \le i \le j \le D$ (ignoring the index of the variance component under consideration for ease of notation), and we wish to reparameterize using $D(D-1)/2$ parameters $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$, $1 \le i < j \le D$, along with the D unchanged diagonal parameters σ_{ii}, $1 \le i \le D$. Writing

$$\sigma_{ij} = \rho_{ij}\sqrt{\sigma_{ii}\sigma_{jj}},$$

the nonzero entries of the Jacobian are given by

$$\begin{aligned}
J_{ii}^{ii} &= \frac{\partial \sigma_{ii}}{\partial \sigma_{ii}} = 1 \\
J_{ii}^{ij} &= \frac{\partial \sigma_{ij}}{\partial \sigma_{ii}} = \frac{\rho_{ij}\sqrt{\sigma_{jj}}}{2\sqrt{\sigma_{ii}}} \\
J_{ij}^{ij} &= \frac{\partial \sigma_{ij}}{\partial \rho_{ij}} = \sqrt{\sigma_{ii}\sigma_{jj}} \\
J_{jj}^{ij} &= \frac{\partial \sigma_{ij}}{\partial \sigma_{jj}} = \frac{\rho_{ij}\sqrt{\sigma_{ii}}}{2\sqrt{\sigma_{jj}}}
\end{aligned}$$

where J_c^r denotes the r, c entry of the Jacobian.

Transformation from variances to heritabilities. For each trait, we have $1 + K$ independent variables $\sigma_0^2, \sigma_1^2, \ldots, \sigma_K^2$ that parameterize the variances of the environmental and genetic variance components for that trait (ignoring the index of the trait under consideration for ease of notation). We wish to reparameterize with 1 overall variance parameter σ^2 and K heritability fractions h_k^2, $1 \leq k \leq K$, such that

$$\sigma_0^2 = \sigma^2 \left(1 - \sum_{k=1}^{K} h_k^2\right) \quad \text{and} \quad \sigma_k^2 = \sigma^2 h_k^2, \quad 1 \leq k \leq K.$$

The Jacobian is given by

$$J = \begin{pmatrix} \frac{\partial(\sigma_0^2)}{\partial(\sigma^2)} = 1 - \sum_k h_k^2 & \frac{\partial(\sigma_0^2)}{\partial(h_1^2)} = -\sigma^2 & \cdots & \frac{\partial(\sigma_0^2)}{\partial(h_K^2)} = -\sigma^2 \\ \frac{\partial(\sigma_1^2)}{\partial(\sigma^2)} = h_1^2 & \frac{\partial(\sigma_1^2)}{\partial(h_1^2)} = \sigma^2 & \cdots & \frac{\partial(\sigma_1^2)}{\partial(h_K^2)} = 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(\sigma_K^2)}{\partial(\sigma^2)} = h_K^2 & \frac{\partial(\sigma_K^2)}{\partial(h_1^2)} = 0 & \cdots & \frac{\partial(\sigma_K^2)}{\partial(h_K^2)} = \sigma^2 \end{pmatrix}.$$

2.3.2 Breakdown of SE estimates at parameter space bounds, e.g., zero heritability

In the above discussion of standard errors and transformed SEs, we need to be careful that if the MLE variance parameter vector is located on the boundary of the parameter space due to domain constraints, the gradient of the log likelihood may not equal zero. In the single-trait case, this situation typically arises when a variance parameter would have been estimated to be negative if negative variances were allowed, but is instead estimated to be zero because of the nonnegativity constraint. An analogous statement holds in the multi-trait case with covariances required to be positive semidefinite.

If the gradient does not vanish at the MLE due to bound constraints, then the quadratic form approximation of the log likelihood, which underlies the analytic standard error from equation (8), breaks down. (Similarly, the transformation formula (9) for change of coordinates no longer holds.) Therefore, analytic standard errors should only be trusted for variance components not at boundary constraints, and we further recommend that in situations with zero-heritability variance components, REML should be rerun with those variance components removed.

2.3.3 Distinction between SE and likelihood ratio test for nonzero variance

We also note that the analytic standard error of a heritability parameter provides only an approximate test of whether the corresponding variance component explains nonzero heritability (by computing $z = \widehat{\sigma^2}/\text{s.e.}(\widehat{\sigma^2})$): such a test assumes a quadratic approximation to the log likelihood, i.e.,

a normal distribution of the posterior density of the heritability, which in reality is only locally valid. The proper way to test for nonzero heritability is to perform a likelihood ratio test; however, a limitation of MC REML is that it does not efficiently compute likelihood ratio test statistics (as it achieves speedup by circumventing likelihood computations).

In particular, for variance components of very low rank (e.g., modeling a small subset of SNPs), we have observed that computing a naive z-score based on the analytic standard error produces an extremely underpowered test for nonzero contribution to variance: even if the component explains a very large fraction of variance (reflected in an appropriately large REML parameter estimate), the analytic standard error is also very large, so that the z-score achieved is only $\approx$2. The reason is that the posterior distribution on the variance parameter is highly non-normal and asymmetric. For example, consider the case of a variance component containing just one SNP estimated to explain 1% of variance. Despite this huge effect, REML cannot be sure whether the true variance parameter is actually 2% (and the SNP just happened to get a smaller effect size under the random effects model). Thus, REML will report a large standard error (reflecting the uncertainty in the upper tail)—even though a closer examination of the likelihood actually would show that at the lower tail, the variance explained is convincingly $>$0% (as the likelihood ratio test would show).

For the goal of computing efficient likelihood ratio tests for whether low-rank variance components contribute nonzero variance—which is distinct from our primary goal of estimating variance parameters—we recommend the spectral decomposition approach of FaST-LMM-Set [81], which performs fast set association tests using two random-effects components, one for the SNP set being tested and the other to model potential genome-wide confounders. (In order to achieve computational efficiency, the latter component can only use a small subset of SNPs, so precautions such as including principal components as fixed effect covariates may be necessary [82]. We further note that for case-control ascertained data sets—in which ascertainment induces correlations among causal SNPs—modeling only a small subset of SNPs should help to avoid loss of power when testing SNP sets for associations [27]; in contrast, when estimating heritability contributed by a SNP set, all SNPs need to be modeled in order to avoid inflation of estimates due to the same phenomenon.)

3 BOLT-REML algorithm

We now describe our algorithm to compute approximate REML estimates of variance parameters. The overall structure of the algorithm is derivative-based iterative optimization using a trust region method for robust convergence. We begin with an initial guess $\theta = (\theta_1, \ldots, \theta_P)$. In each iteration, we propose an update step $\mathbf{p} = (\Delta\theta_1, \ldots, \Delta\theta_P)$ such that $\theta + \mathbf{p}$ achieves the largest predicted increase in log likelihood based on a local quadratic model of the log likelihood (built from the Monte Carlo approximate gradient and AI approximate negative Hessian), subject to parameter

domain and trust region constraints. We evaluate the approximate gradient at the tentative new parameter vector $\theta + \mathbf{p}$ and use its value to approximate the actual change in log likelihood. We compare the approximate actual change in log likelihood to the predicted change based the local quadratic model, and based on this information, we decide whether to accept the step and/or modify the trust region radius. We declare convergence of the algorithm when the approximate change in log likelihood falls below a specified threshold (corresponding to the distance of the current iterate from the optimum in units of standard errors).

3.1 Initial guess

3.1.1 Single trait

Because of the possibility of local optima, a good initial guess is very important to the robustness of an optimization routine. In the single-trait case with K non-identity variance components, we can use the sample variance of the phenotype as a rough guess of the overall variance parameter σ^2, leaving K heritability fractions to create guesses for as follows:

1. Estimate the total heritability h^2 explained by all variance components by combining them into a single component and applying the secant iteration implemented in the original BOLT-LMM software [20].

2. For each variance component k, estimate the total heritability h^2_{-k} explained by all variance components except k using one step of secant iteration (as in the original BOLT-LMM re-estimation of heritability for LOCO reps [20]; the only difference here is that we leave each variance component—rather than each chromosome—out in turn).

3. Estimate the heritability explained by each variance component as

$$h^2_k = \max\left\{0, \left(1 - \frac{(n-1)h^2_{-k}}{\sum_{l=1}^{n} h^2_{-l}}\right) h^2\right\}.$$

3.1.2 Multiple traits

To obtain guesses of the parameters along the diagonal terms of the covariances V_g, V_e, we apply the single-trait initial guess procedure above to each trait in turn. To obtain guesses of the off-diagonal covariance terms, we compute for each pair of traits the phenotypic correlation between the two traits. We then assume that the environmental correlation and genetic correlation(s) (plural in the case of multiple genetic variance components) all equal the estimated phenotypic correlation.

3.2 Trust region iterative optimization

After making an initial parameter vector guess, we proceed to take steps in parameter space designed to iteratively optimize the log likelihood. At each iteration, we need to take into account two important considerations: (1) optimizing the step to try to increase the log likelihood as much as possible (according to the local quadratic model), to achieve fast convergence; but also (2) being careful not to take too ambitious of a step, to ensure robustness. We balance these considerations by using a trust region approach that enforces an adaptive bound Δ on the maximum allowed size of the step (with coordinates scaled to improve conditioning).

3.2.1 Step computation

In each iteration, we begin with a current parameter vector estimate θ and wish to compute a step $\mathbf{p} = (\Delta\theta_p)$ to an updated estimate $\theta + \mathbf{p}$. We choose the step to maximize the predicted change in log likelihood according to the local quadratic model

$$\text{predicted } \Delta\ell = \overline{\nabla\ell}'\mathbf{p} - \frac{1}{2}\mathbf{p}'\mathscr{I}_A\mathbf{p} \tag{10}$$

(where $\overline{\nabla\ell}$ is the Monte Carlo gradient and $\mathscr{I}_A$ is the AI negative Hessian) over steps $\mathbf{p}$ satisfying the following constraints:

1. REML parameter domain: Positive semidefiniteness of the covariance matrices (one per variance component) corresponding to the new parameter vector $\theta + \mathbf{p}$.

2. Trust region bound on the scaled step norm: $||\texttt{diag}(\mathscr{I}_A)\mathbf{p}|| < \Delta$.

The value of Δ bounds the region in which the local quadratic approximation of ℓ is trusted. Note that the objective function in equation (10) is a simple quadratic, so that if there were no constraints, the optimal $\mathbf{p}$ would be the Newton step. On the opposite extreme, if Δ is very small, the optimal $\mathbf{p}$ is approximately parallel to $\overline{\nabla\ell}$ as in gradient descent.

Finding the step $\mathbf{p}$ that maximizes the predicted change in log likelihood given in equation (10) subject to the above bounds is itself an optimization problem, but because the objective function is quadratic, we can apply standard numerical optimization methods to solve it at trivial computational cost. To increase robustness, we run three optimization routines implemented in NLopt [83]—MMA [84], CCSAQ [84], and SLSQP [85]—and use the best result.

3.2.2 Step acceptance criterion and trust region radius update

After computing a potential step $\mathbf{p}$ using the above procedure, we need to decide whether to accept the step and also whether and how to update the trust region radius Δ. Usually, trust region methods

do so by comparing the predicted change in the objective function (here, the log likelihood) to the actual change, but a challenge that our method faces is that it can only estimate derivatives of the log likelihood: we are unable to compute or estimate the log likelihood itself. Instead, we use the following derivative-based proxy for the change in the objective function:

$$\Delta\ell \approx \mathbf{p}' \cdot \frac{1}{2}\left(\overline{\nabla\ell}(\theta) + \overline{\nabla\ell}(\theta+\mathbf{p})\right). \tag{11}$$

(Note that this quantity is different from the *predicted* change in log likelihood given in equation (10), which was what we optimized to compute the step.) The rationale is that the change in log likelihood equals the dot product of the step vector $\mathbf{p}$ with the mean value of the gradient along the step from θ to $\theta+\mathbf{p}$. We can approximate the mean gradient with the average of the (approximate) gradients at the beginning and end of the step, i.e., at θ and $\theta+\mathbf{p}$, and this approximation is precisely equation (11). The approximation is an accurate estimate of the change in the log likelihood as long as the approximate gradient is reasonably close to linear over the step from θ to $\theta+\mathbf{p}$, which in particular holds if the step size is not too big.

We update the trust region radius Δ according to how well the change in log likelihood predicted by the quadratic model (10) matches the approximate actual change (11). We use the procedure and parameters recommended by ref. [77] with the additional safeguard that we reject steps that more than double the norm of the gradient (suggesting that the local quadratic model may be invalid). Details are provided in the pseudocode (Sec. 3.4).

3.3 Convergence criterion

We declare convergence of the iterative optimization when the predicted change in log likelihood drops below a set threshold (10^{-4} by default). Note that suboptimality of the achieved vs. optimal log likelihood has a simple interpretation in terms of the accuracy of the final parameter estimates. If our converged parameter estimates θ_{final} give log likelihood ℓ_{final} whereas optimal REML estimates θ_{opt} give log likelihood ℓ_{opt}, then

$$(\theta_{p,\text{opt}} - \theta_{p,\text{final}})' \mathscr{I}_A^{-1} (\theta_{p,\text{opt}} - \theta_{p,\text{final}}) \approx \ell_{\text{opt}} - \ell_{\text{final}}$$

if we assume that the log likelihood is locally well-approximated by the quadratic $-\mathscr{I}_A$ (which is the assumption underlying the analytic standard error estimates discussed in Section 2.3). If we further assume near-independence of coordinates, we have

$$\sum_{p=1}^{P} \left(\frac{\theta_{p,\text{opt}} - \theta_{p,\text{final}}}{\text{SE}(\theta_p)}\right)^2 \approx \ell_{\text{opt}} - \ell_{\text{final}}.$$

3.4 Pseudocode for main trust region iteration

```
const eta1 = 0.0001, eta2 = 0.99;
const alpha1 = 0.25, alpha2 = 3.5;
Delta = Inf; // initialize trust region radius
x = initial variance parameter guesses
compute gradMC(x) using fixed pseudorandom vectors
do
  // compute step
  compute HessAI(x)
  compute constrained Newton step p:
    optimize quadratic model: p' * gradMC - 0.5 * p' * HessAI(x) * p
    subject to: ||diag(HessAI(x))*p|| < Delta, covariance matrices psd
  dLLpred = p' * gradMC(x) - 0.5 * p' * HessAI(x) * p

  // check convergence of algorithm
  if dLLpred < tolLL
    break

  // assess step quality
  compute gradMC(x+p) using fixed pseudorandom vectors
  dLLapprox = p' * 0.5 * (gradMC(x) + gradMC(x+p))
  if ||gradMC(x+p)|| > 2*||gradMC(x)||
    rho = -1; // dangerous model deviation: reject step; reduce trust radius
  else
    rho = dLLapprox / dLLpred

  // decide step acceptance
  if rho > eta1
    x = x+p

  // update trust region radius
  if rho < eta1
    Delta = alpha1 * ||diag(HessAI(x))*p||;
  else if rho < eta2
    ;
  else
    Delta = max(Delta, alpha2 * ||diag(HessAI(x))*p||);
done
```

3.5 Number of Monte Carlo samples and speedup via two-phase optimization

In choosing the number of Monte Carlo samples S to use in estimating the gradient of the log likelihood, we need to balance two considerations: (1) accuracy, which increases with S (because the variance of our estimates scales roughly with $1 + 1/S$, as discussed in Sec. 2.3); and (2) computational cost, which increases linearly with S plus a constant (Sec. 3.6). In our BOLT-REML implementation, we finesse this trade-off by applying a two-phase optimization using different values of S. First, we run a "coarse" optimization using $S = 15$ samples and stopping the iteration once the predicted change in log likelihood drops below 0.01. Then, starting from the final iterate of the coarse optimization, which should be close to the true optimum, we run a "fine" optimization using $S = 100$ samples and iterating to a log likelihood tolerance of 10^{-4}.

3.6 Computational cost

The cost of evaluating the approximate gradient in equation (3) is dominated by computation of the terms $\mathbf{V}^{-1}\mathbf{y}_\mathbf{V}$ for each of S Monte Carlo simulated phenotype vectors $\mathbf{y}_\mathbf{V}$ and the term $\mathbf{V}^{-1}\mathbf{y}$ (for the real phenotype vector $\mathbf{y}$). This computation amounts to solving a batch of $S + 1$ mixed model equation systems, which we do using conjugate gradient iteration [18, 19]. To compute the AI matrix entries given in equation (7), we need to further compute $\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_p}\mathbf{V}^{-1}\mathbf{y}$ for each parameter θ_p, requiring $P = (1 + K)D(D + 1)/2$ additional solutions of the mixed model equations.

Thus, the total cost of the algorithm per AI iteration is $O(S + KD^2)$ conjugate gradient runs, each of which has cost scaling roughly with $MN^{1.5}D^{1.5}$ according to ref. [20]. (As defined above, S is the number of Monte Carlo samples, K is the number of variance components, D is the number of traits, M is the total number of SNPs across all variance components, and N is the number of individuals.) In our experience, fewer than 10 AI iterations are typically needed for convergence.

3.6.1 Comparison to other approaches

For context, we now provide a detailed discussion of the computational requirements of BOLT-REML vs. existing approaches. To streamline the exposition, we restrict to the single-trait case $D = 1$, which is most frequently used in practice, and we focus on the scaling of computational costs with N and M. We primarily compare BOLT-REML to GCTA [2], which is the most widely used and most versatile among publicly available REML software for genetics; we point out other methods that contribute algorithms of particular interest.

Computation of genetic relationship matrices (GRMs). For variance component models built from genetic random-effect terms, most REML algorithms (e.g., GCTA) begin by computing the genetic relationship matrices corresponding to the random effects. (BOLT-REML and the FaST-LMM [86] family of spectral decomposition-based methods are exceptions.) Computing the GRMs takes $O(MN^2)$ time, which may appear to be the dominant term in typical settings with $M > N$; however, this computation can easily be parallelized across a computing cluster (e.g., with PLINK2 [60]), so in practice, it is not typically a computational barrier.

Computation per REML iteration. For large N, the running time of standard Newton-type REML methods is dominated by the costs (per REML iteration) of inverting the estimated phenotypic covariance (i.e., computing $\mathbf{V}^{-1}$) and performing a few $N \times N$ matrix multiplications per entry of the Hessian. Both inversion and matrix multiplication are $O(N^3)$ operations, and the inversion is particularly difficult to parallelize across multiple machines. BOLT-REML iterations have cost scaling roughly with $MN^{1.5}$ (which is also difficult to parallelize across machines), so for typical $M < 10^6$ in genetics, BOLT-REML achieves a speedup once N exceeds several thousand individuals, with the break-even point increasing with M.

We note that in the special case of only one non-identity variance component, spectral decomposition techniques drastically reduce computational cost per-REML iteration [70, 87]. We also note that for situations in which all or all but one of the variance components have low rank, two other approaches exist that can (for certain ranges of M and N) achieve speedup beyond BOLT-REML. Listgarten et al. [81] observed that if the total number of SNPs in the model is smaller than the sample size (i.e., $M < N$), then spectral decomposition can be used to reduce computational cost to $O(M^2N)$ (for one SVD) per REML iteration. FaST-LMM-Set [81] implements this approach for mixed model set association tests. Speed et al. [88] observed that if only one variance component has full rank (as in our estimation of per-Mb schizophrenia SNP-heritability), then its GRM only needs to be computed and inverted once at $O(MN^2 + N^3)$ cost, after which the Woodbury matrix identity can be used to efficiently perform REML iterations at $O(M_{\text{other}}N^2)$ cost, where M_{other} is the total number of SNPs in the low-rank components. MultiBLUP [88] implements this approach for mixed model prediction.

Memory use. Standard REML methods compute and work with one $N \times N$ GRM per variance component; GCTA currently stores all of these matrices in memory, so that in practice, memory constraints become a computational barrier before running time. For models with many variance components, the memory footprint could be reduced by rereading each GRM from disk when necessary; however, $O(N^2)$ memory would still be needed to work with at least a few $N \times N$ matrices. In contrast, BOLT-REML (like BOLT-LMM [20]) stores only raw genotypes in memory, requiring $\approx MN/4$ bytes of memory (regardless of the number of variance components). We have found this

memory requirement to be mild in practice, but it could potentially be reduced by storing genotypes in compressed form and decompressing on-the-fly in each REML iteration (which would especially help for rare SNPs) or by rereading genotypes from disk in each REML iteration; the latter approach would incur a substantial performance penalty, however.

4 Appendix: EM REML

Here we derive the BLUP estimates (assuming known variance parameters) and EM REML update equations (for variance parameter estimation) for the multi-trait mixed model given in equation (5), following ref. [67]. We derive these equations primarily to point out the relationship between Monte Carlo gradient approximation and the EM REML update formulas. EM REML can also be efficiently approximated with Monte Carlo [25], but the first derivative-based MC EM algorithm converges much more slowly than Newton-type MC REML methods [23] such as our MC AI REML algorithm.

4.1 Mixed model equations: BLUP estimates

To solve the mixed model equations and obtain BLUP estimates of random effects, we write out the joint distribution of $\mathbf{y}$, $\mathbf{u}_g$, and $\mathbf{u}_e$. From the model equations (5), we have

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{u}_g \\ \mathbf{u}_e \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & V_g \otimes Z & V_e \otimes I_N \\ V_g \otimes Z' & V_g \otimes I_M & 0 \\ V_e \otimes I_N & 0 & V_e \otimes I_N \end{pmatrix}\right).$$

Conditioning on $\mathbf{y}$, we have (by Appendix S.3 of ref. [67])

$$\begin{aligned} \begin{pmatrix} \mathbf{u}_g \\ \mathbf{u}_e \end{pmatrix} \Bigg| \mathbf{y} \;\sim\; N\Bigg(&\begin{pmatrix} V_g \otimes Z' \\ V_e \otimes I_N \end{pmatrix} \mathbf{V}^{-1}\mathbf{y}, \\ &\begin{pmatrix} V_g \otimes I_M & 0 \\ 0 & V_e \otimes I_N \end{pmatrix} - \begin{pmatrix} V_g \otimes Z' \\ V_e \otimes I_N \end{pmatrix} \mathbf{V}^{-1} \begin{pmatrix} V_g \otimes Z & V_e \otimes I_N \end{pmatrix}\Bigg). \end{aligned} \tag{12}$$

Therefore, the BLUP random effect estimates given observed $\mathbf{y}$ are

$$\hat{\mathbf{u}}_g = (V_g \otimes Z')\mathbf{V}^{-1}\mathbf{y}, \quad \hat{\mathbf{u}}_e = (V_e \otimes I_N)\mathbf{V}^{-1}\mathbf{y}.$$

4.2 EM REML update equations

Following the derivation of ref. [67] Sec. 8.3 (noting in particular equation (14) on p. 300), the EM equations for estimating V_g and V_e amount to the successive updates

$$V_{g,pq} \leftarrow V_{g,pq} + \frac{1}{M}\left(\hat{\mathbf{u}}'_{g,p}\hat{\mathbf{u}}_{g,q} - E[\hat{\mathbf{u}}'_{g,p}\hat{\mathbf{u}}_{g,q}]\right) \tag{13}$$

$$V_{e,pq} \leftarrow V_{e,pq} + \frac{1}{N}\left(\hat{\mathbf{u}}'_{e,p}\hat{\mathbf{u}}_{e,q} - E[\hat{\mathbf{u}}'_{e,p}\hat{\mathbf{u}}_{e,q}]\right), \tag{14}$$

for $p, q = 1, \ldots, D$, where the expectations are taken over random effect estimates based on randomly generated $\mathbf{y}$ according to the assumed model parameters V_g and V_e. Note that at convergence, the pairwise dot products of estimated random effects (resp. residuals) across all pairs of traits equal their expectations.

Detailed derivation of EM update equations. For $p, q = 1, \ldots, D$, we wish to update $V_{g,pq}$ and $V_{e,pq}$ according to the sufficient statistics

$$V_{g,pq} \leftarrow \frac{1}{M}E[\mathbf{u}'_{g,p}\mathbf{u}_{g,q} \mid \mathbf{y}], \quad V_{e,pq} \leftarrow \frac{1}{N}E[\mathbf{u}'_{e,p}\mathbf{u}_{e,q} \mid \mathbf{y}].$$

We will carry out the derivation for the genetic terms; the residual terms behave analogously. Write

$$\begin{aligned} E[\mathbf{u}'_{g,p}\mathbf{u}_{g,q} \mid \mathbf{y}] &= E[\mathbf{u}'_{g,p} \mid \mathbf{y}]'E[\mathbf{u}_{g,q} \mid \mathbf{y}] + \text{tr Cov}(\mathbf{u}_{g,p}, \mathbf{u}_{g,q} \mid \mathbf{y}) \\ &= \hat{\mathbf{u}}'_{g,p}\hat{\mathbf{u}}_{g,q} + \text{tr}\left(V_{g,pq}I_M - (V'_{g,p} \otimes Z')\mathbf{V}^{-1}(V_{g,q} \otimes Z)\right) \end{aligned}$$

using equation (12) for the covariance conditional on $\mathbf{y}$. Now

$$\text{tr } V_{g,pq}I_M = M \cdot V_{g,pq}$$

and

$$\begin{aligned} \text{tr}\,(V'_{g,p} \otimes Z')\mathbf{V}^{-1}(V_{g,q} \otimes Z) &= \text{tr } \mathbf{V}^{-1}(V_{g,q} \otimes Z)(V'_{g,p} \otimes Z') \\ &= \text{tr } \mathbf{V}^{-1}(V_{g,q} \otimes Z)(V'_{g,p} \otimes Z')\mathbf{V}^{-1}\mathbf{V} \\ &= \text{tr } \mathbf{V}^{-1}(V_{g,q} \otimes Z)(V'_{g,p} \otimes Z')\mathbf{V}^{-1}E[\mathbf{y}_\mathbf{V}\mathbf{y}'_\mathbf{V}] \\ &= E[\mathbf{y}'_\mathbf{V}\mathbf{V}^{-1}(V_{g,q} \otimes Z)(V'_{g,p} \otimes Z')\mathbf{V}^{-1}\mathbf{y}_\mathbf{V}] \\ &= E[\hat{\mathbf{u}}'_{g,q}\hat{\mathbf{u}}_{g,p}], \end{aligned}$$

and substituting gives the desired update equation.

References

1. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565–569 (2010).

2. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82 (2011).

3. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *American Journal of Human Genetics* **88**, 294–305 (2011).

4. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519–525 (2011).

5. Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genetics* **44**, 247–250 (2012).

6. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *American Journal of Human Genetics* **95**, 535–552 (2014).

7. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).

8. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* (2013).

9. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

10. Ripke, S. *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics* **43**, 969 (2011).

11. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics* **45**, 1150–1159 (2013).

12. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

13. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nature Genetics* (2013).

14. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics* **46**, 234–244 (2014).

15. Visscher, P. M. *et al.* Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLOS Genetics* **10**, e1004269 (2014).

16. Visscher, P. M. & Goddard, M. E. A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics* (2015).

17. Golan, D., Lander, E. S. & Rosset, S. Measuring missing heritability: Inferring the contribution of common variants. *Proceedings of the National Academy of Sciences* **111**, E5272–E5281 (2014).

18. Legarra, A. & Misztal, I. Computing strategies in genome-wide selection. *Journal of Dairy Science* **91**, 360–366 (2008).

19. VanRaden, P. Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414–4423 (2008).

20. Loh, P.-R. *et al.* Efficient Bayesian mixed model analysis increases association power in large cohorts. *Nature Genetics* (2015).

21. Henderson, C. *Application of Linear Models in Animal Breeding* (University of Guelph, 1984).

22. Henderson, C. & Quaas, R. Multiple trait evaluation using relatives' records. *Journal of Animal Science* (1976).

23. Matilainen, K., Mäntysaari, E. A., Lidauer, M. H., Strandén, I. & Thompson, R. Employing a Monte Carlo Algorithm in Newton-Type Methods for Restricted Maximum Likelihood Estimation of Genetic Parameters. *PLOS ONE* **8**, e80821 (2013).

24. Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).

25. García-Cortés, L. A., Moreno, C., Varona, L. & Altarriba, J. Variance component estimation by resampling. *Journal of Animal Breeding and Genetics* **109**, 358–363 (1992).

26. Gilmour, A. R., Thompson, R. & Cullis, B. R. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 1440–1450 (1995).

27. 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. *Nature Genetics* **46**, 100–106 (2014).

28. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics* **91**, 1011–1021 (2012).

29. Lee, S. H. *et al.* Estimation of SNP heritability from dense genotype data. *American Journal of Human Genetics* **93**, 1151–1155 (2013).

30. Gusev, A. *et al.* Quantifying missing heritability at known GWAS loci. *PLOS Genetics* **9**, e1003993 (2013).

31. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genetics* **44**, 483–489 (2012).

32. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46**, 1173–1186 (2014).

33. Koren, A. *et al.* Differential relationship of DNA replication timing to different forms of human mutation and variation. *American Journal of Human Genetics* **91**, 1033–1040 (2012).

34. The International HapMap Consortium, K. A., Frazer *et al.* A second generation human haplotype map of over 3.1 million snps. *Nature* **449**, 851–861 (2007).

35. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLOS Genetics* **5**, e1000471 (2009).

36. Banovich, N. E. *et al.* Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLOS Genetics* **10**, e1004663 (2014).

37. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proceedings of the National Academy of Sciences* **111**, E455–E464 (2014).

38. Goldstein, D. B. Common genetic variation and human traits. *New England Journal of Medicine* **360**, 1696 (2009).

39. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

40. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications* **6** (2015).

41. Gottesman, I. I. & Shields, J. A polygenic theory of schizophrenia. *Proceedings of the National Academy of Sciences* **58**, 199–205 (1967).

42. Sullivan, P. F. Puzzling over schizophrenia: schizophrenia as a pathway disease. *Nature Medicine* **18**, 210–211 (2012).

43. Gelfman, S., Cohen, N., Yearim, A. & Ast, G. DNA-methylation effect on cotranscriptional splicing is dependent on gc architecture of the exon–intron structure. *Genome Research* **23**, 789–799 (2013).

44. Gibson, G. Rare and common variants: twenty arguments. *Nature Reviews Genetics* **13**, 135–145 (2012).

45. Lohmueller, K. E. The impact of population demography and selection on the genetic architecture of complex traits. *PLOS Genetics* **10**, e1004379 (2014).

46. Ferreira, M. A. *et al.* Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006–1014 (2011).

47. Bønnelykke, K. *et al.* Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nature Genetics* **45**, 902–906 (2013).

48. Hinds, D. A. *et al.* A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nature Genetics* **45**, 907–911 (2013).

49. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *bioRxiv* 014498 (2015).

50. Vattikuti, S., Guo, J. & Chow, C. C. Heritability and genetic correlations explained by common snps for metabolic syndrome traits. *PLOS Genetics* **8**, e1002637 (2012).

51. Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *American Journal of Human Genetics* (2015).

52. Cheverud, J. M. A comparison of genetic and phenotypic correlations. *Evolution* 958–968 (1988).

53. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLOS Genetics* **9**, e1003264 (2013).

54. Haseman, J. & Elston, R. The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* **2**, 3–19 (1972).

55. Finucane, H. K. *et al.* Partitioning heritability by functional category using GWAS summary statistics. *bioRxiv* 014241 (2015).

56. Chen, C.-Y. *et al.* Improved ancestry inference using weights from external reference panels. *Bioinformatics* btt144 (2013).

57. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

58. Manichaikul, A. *et al.* Population structure of Hispanics in the United States: the multiethnic study of atherosclerosis. *PLOS Genetics* **8**, e1002640 (2012).

59. Hoffmann, T. J. *et al.* Next generation genome-wide association tool: Design and coverage of a high-throughput European-optimized SNP array. *Genomics* **98**, 79–89 (2011).

60. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* (2015).

61. Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics* **29**, 51–76 (1965).

62. Galinsky, K. J. *et al.* Fast principal components analysis reveals independent evolution of ADH1B gene in Europe and East Asia. *bioRxiv* 018143 (2015).

63. Tange, O. GNU Parallel - The Command-Line Power Tool. *The USENIX Magazine* **36**, 42–47 (2011). URL `http://www.gnu.org/s/parallel`.

64. Kostem, E. & Eskin, E. Improving the accuracy and efficiency of partitioning heritability into the contributions of genomic regions. *American Journal of Human Genetics* **92**, 558–564 (2013).

65. Bengio, Y. & Grandvalet, Y. No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research* **5**, 1089–1105 (2004).

66. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1–38 (1977).

67. Searle, S. R., Casella, G. & McCulloch, C. E. *Variance components* (John Wiley & Sons, 2006).

68. Liu, J. S. & Wu, Y. N. Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94**, 1264–1274 (1999).

69. Foulley, J.-L. & Van Dyk, D. A. The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genetics Selection Evolution* **32**, 1–21 (2000).

70. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods* **11**, 407–409 (2014).

71. Meyer, K. *et al.* PX $\times$ AI: Algorithmics for better convergence in restricted maximum likelihood estimation. In *8th World Congress on Genetics Applied to Livestock Production* (2006).

72. Groeneveld, E. A reparameterization to improve numerical optimization in multivariate REML (co)variance component estimation. *Genetics Selection Evolution* **26**, 537–545 (1994).

73. Wei, G. C. & Tanner, M. A. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699–704 (1990).

74. Matilainen, K., Mäntysaari, E. A., Lidauer, M. H., Strandén, I. & Thompson, R. Employing a Monte Carlo algorithm in expectation maximization restricted maximum likelihood estimation of the linear mixed model. *Journal of Animal Breeding and Genetics* **129**, 457–468 (2012).

75. Kuk, A. Y. & Cheng, Y. W. The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical Computation and Simulation* **59**, 233–250 (1997).

76. McCulloch, C. E. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170 (1997).

77. Gould, N. I., Orban, D., Sartenaer, A. & Toint, P. L. Sensitivity of trust-region algorithms to their parameters. *4OR* **3**, 227–241 (2005).

78. McCulloch, C., Searle, S. & Neuhaus, J. *Generalized, linear, and mixed models* (Wiley, 2008), 2nd edn.

79. Barry, R. P. & Kelley Pace, R. Monte Carlo estimates of the log determinant of large sparse matrices. *Linear Algebra and its Applications* **289**, 41–54 (1999).

80. Korte, A. *et al.* A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics* **44**, 1066–1071 (2012).

81. Listgarten, J. *et al.* A powerful and efficient set test for genetic markers that handles confounders. *Bioinformatics* **29**, 1526–1533 (2013).

82. Tucker, G., Price, A. L. & Berger, B. A. Improving the power of GWAS and avoiding confounding from population stratification with PC-Select. *Genetics* (2014).

83. Johnson, S. G. The NLopt nonlinear-optimization package. URL `http://ab-initio.mit.edu/nlopt`.

84. Svanberg, K. A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM Journal on Optimization* **12**, 555–573 (2002).

85. Kraft, D. Algorithm 733: TOMP–Fortran modules for optimal control calculations. *ACM Transactions on Mathematical Software (TOMS)* **20**, 262–281 (1994).

86. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods* **8**, 833–835 (2011).

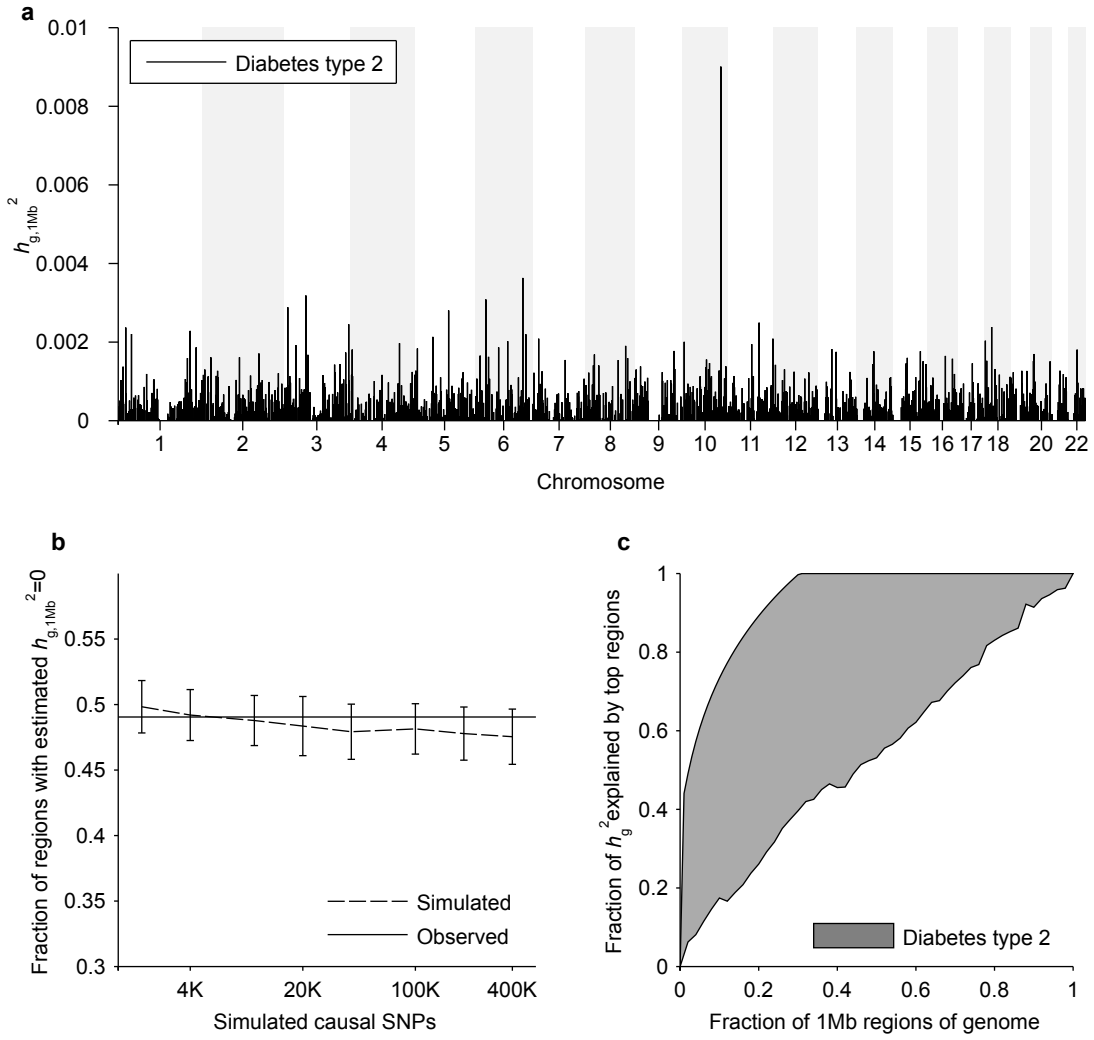
87. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).

88. Speed, D. & Balding, D. J. MultiBLUP: improved SNP-based prediction for complex traits. *Genome Research* gr–169375 (2014).

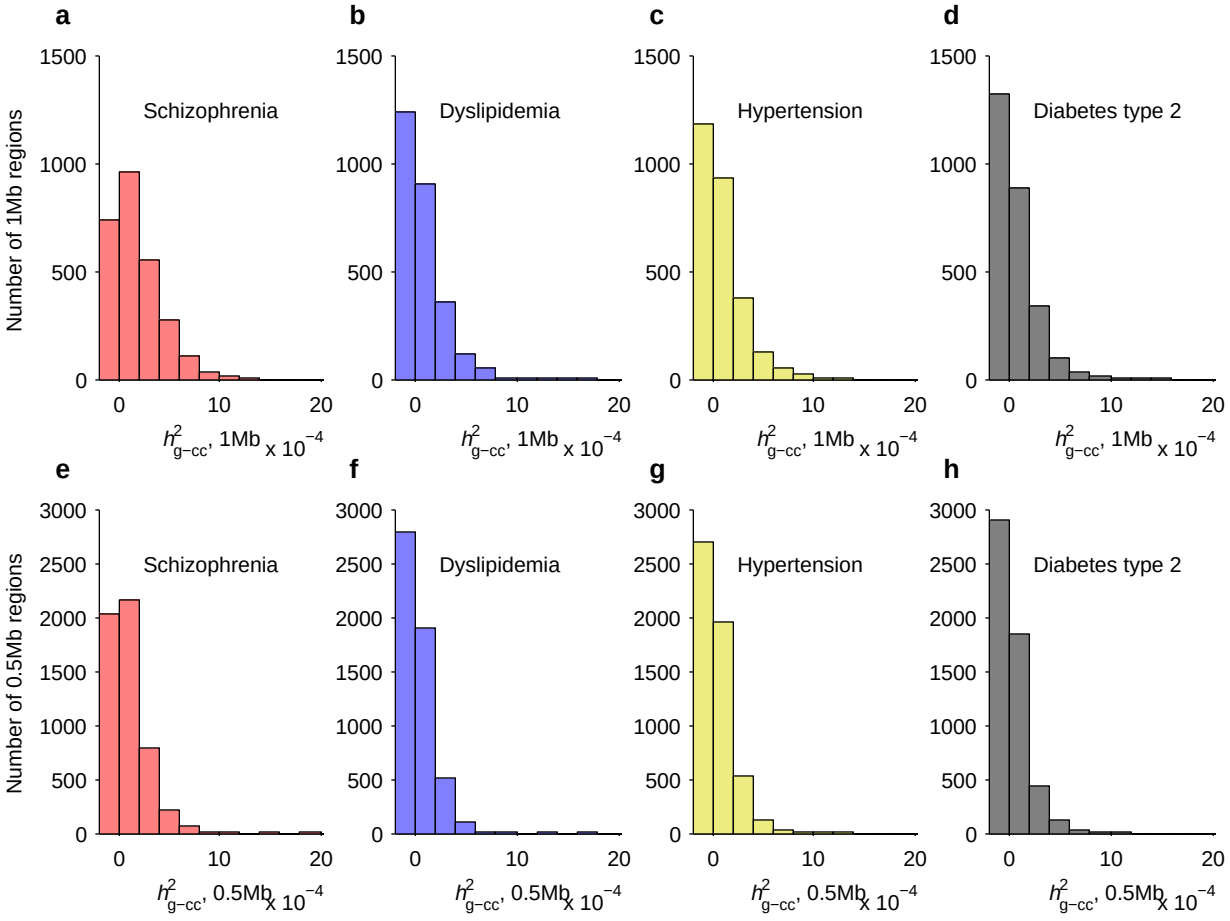
PGC-SCZ Members

The members of the Schizophrenia Working Group of the Psychiatric Genomics Consortium are Stephan Ripke, Benjamin M. Neale, Aiden Corvin, James T. R. Walters, Kai-How Farh, Peter A. Holmans, Phil Lee, Brendan Bulik-Sullivan, David A. Collier, Hailiang Huang, Tune H. Pers, Ingrid Agartz, Esben Agerbo, Margot Albus, Madeline Alexander, Farooq Amin, Silviu A. Bacanu, Martin Begemann, Richard A. Belliveau Jr, Judit Bene, Sarah E. Bergen, Elizabeth Bevilacqua, Tim B. Bigdeli, Donald W. Black, Anders D. Børglum, Richard Bruggeman, Nancy G. Buccola, Randy L. Buckner, William Byerley, Wiepke Cahn, Guiqing Cai, Dominique Campion, Rita M. Cantor, Vaughan J. Carr, Noa Carrera, Stanley V. Catts, Kimberly D. Chambert, Raymond C. K. Chan, Ronald Y. L. Chen, Eric Y. H. Chen, Wei Cheng, Eric F. C. Cheung, Siow Ann Chong, C. Robert Cloninger, David Cohen, Nadine Cohen, Paul Cormican, Nick Craddock, James J. Crowley, David Curtis, Michael Davidson, Kenneth L. Davis, Franziska Degenhardt, Jurgen Del Favero, Lynn E. DeLisi, Ditte Demontis, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Gary Donohoe, Elodie Drapeau, Jubao Duan, Frank Dudbridge, Naser Durmishi, Peter Eichhammer, Johan Eriksson, Valentina Escott-Price, Laurent Essioux, Ayman H. Fanous, Martilias S. Farrell, Josef Frank, Lude Franke, Robert Freedman, Nelson B. Freimer, Marion Friedl, Joseph I. Friedman, Menachem Fromer, Giulio Genovese, Lyudmila Georgieva, Elliot S. Gershon, Ina Giegling, Paola Giusti-Rodríguez, Stephanie Godard, Jacqueline I. Goldstein, Vera Golimbet, Srihari Gopal, Jacob Gratten, Jakob Grove, Lieuwe de Haan, Christian Hammer, Marian L. Hamshere, Mark Hansen, Thomas Hansen, Vahram Haroutunian, Annette M. Hartmann, Frans A. Henskens, Stefan Herms, Joel N. Hirschhorn, Per Hoffmann, Andrea Hofman, Mads V. Hollegaard, David M. Hougaard, Masashi Ikeda, Inge Joa, Antonio Julià, René S. Kahn, Luba Kalaydjieva, Sena Karachanak-Yankova, Juha Karjalainen, David Kavanagh, Matthew C. Keller, Brian J. Kelly, James L. Kennedy, Andrey Khrunin, Yunjung Kim, Janis Klovins, James A. Knowles, Bettina Konte, Vaidutis Kucinskas, Zita Ausrele Kucinskiene, Hana Kuzelova-Ptackova, Anna K. Kähler, Claudine Laurent, Jimmy Lee Chee Keong, S. Hong Lee, Sophie E. Legge, Bernard Lerer, Miaoxin Li, Tao Li, Kung-Yee Liang, Jeffrey Lieberman, Svetlana Limborska, Carmel M. Loughland, Jan Lubinski, Jouko Lönnqvist, Milan Macek Jr, Patrik K. E. Magnusson, Brion S. Maher, Wolfgang Maier, Jacques Mallet, Sara Marsal, Manuel Mattheisen, Morten Mattingsdal, Robert W. McCarley, Colm McDonald, Andrew M. McIntosh, Sandra Meier, Carin J. Meijer, Bela Melegh, Ingrid Melle, Raquelle I. Mesholam-Gately, Andres Metspalu, Patricia T. Michie, Lili Milani, Vihra Milanova, Younes Mokrab, Derek W. Morris, Ole Mors, Preben B. Mortensen, Kieran C. Murphy, Robin M. Murray, Inez Myin-Germeys, Bertram Müller-Myhsok, Mari Nelis, Igor Nenadic, Deborah A. Nertney, Gerald Nestadt, Kristin K. Nicodemus, Liene Nikitina-Zake, Laura Nisenbaum, Annelie Nordin, Eadbhard O'Callaghan, Colm O'Dushlaine, F. Anthony O'Neill, Sang-Yun Oh, Ann Olincy, Line Olsen, Jim Van Os, Psychosis Endophenotypes International Consortium, Christos Pantelis, George N. Papadimitriou, Sergi Papiol, Elena Parkhomenko, Michele T. Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Diana O. Perkins, Olli Pietiläinen, Jonathan Pimm, Andrew J. Pocklington, John Powell, Alkes Price, Ann E. Pulver, Shaun M. Purcell, Digby Quested, Henrik B. Rasmussen, Abraham Reichenberg, Mark A. Reimers, Alexander L. Richards, Joshua L.

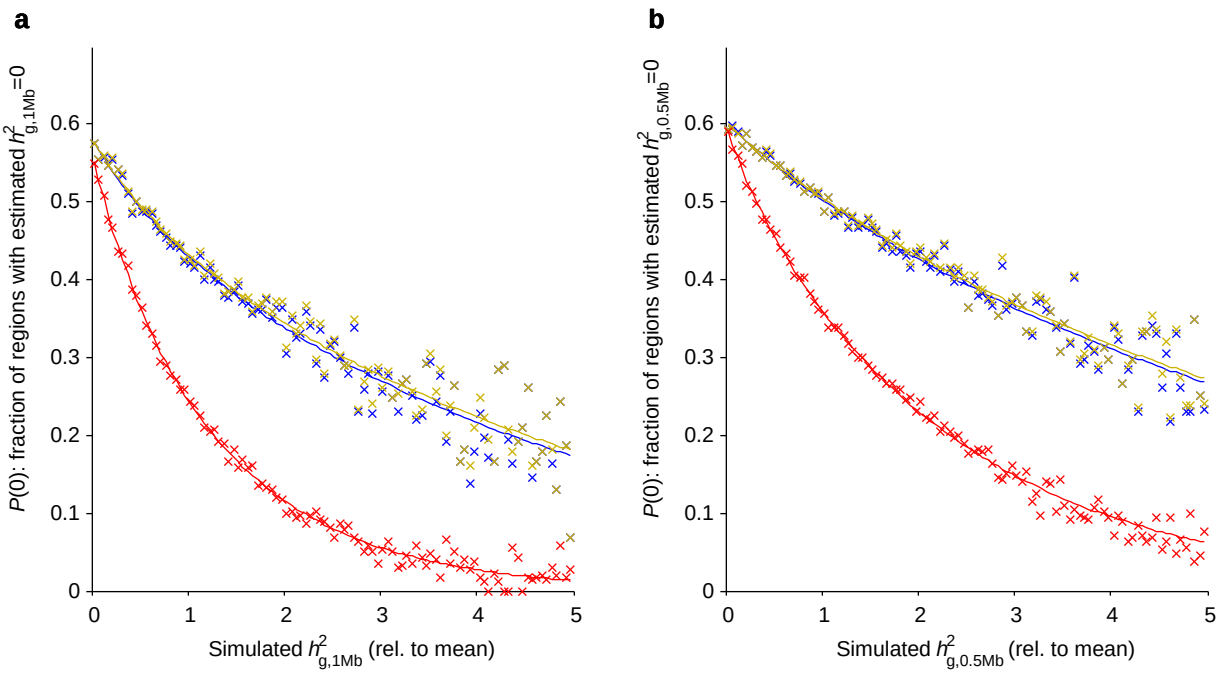
Roffman, Panos Roussos, Douglas M. Ruderfer, Veikko Salomaa, Alan R. Sanders, Ulrich Schall, Christian R. Schubert, Thomas G. Schulze, Sibylle G. Schwab, Edward M. Scolnick, Rodney J. Scott, Larry J. Seidman, Jianxin Shi, Engilbert Sigurdsson, Teimuraz Silagadze, Jeremy M. Silverman, Kang Sim, Petr Slominsky, Jordan W. Smoller, Hon-Cheong So, ChrisC. A. Spencer, Eli A. Stahl, Hreinn Stefansson, Stacy Steinberg, Elisabeth Stogmann, Richard E. Straub, Eric Strengman, Jana Strohmaier, T. Scott Stroup, Mythily Subramaniam, Jaana Suvisaari, Dragan M. Svrakic, Jin P. Szatkiewicz, Erik Söderman, Srinivas Thirumalai, Draga Toncheva, Paul A. Tooney, Sarah Tosato, Juha Veijola, John Waddington, Dermot Walsh, Dai Wang, Qiang Wang, Bradley T. Webb, Mark Weiser, Dieter B. Wildenauer, Nigel M. Williams, Stephanie Williams, Stephanie H. Witt, Aaron R. Wolen, Emily H. M. Wong, Brandon K. Wormley, Jing Qin Wu, Hualin Simon Xi, Clement C. Zai, Xuebin Zheng, Fritz Zimprich, Naomi R. Wray, Kari Stefansson, Peter M. Visscher, Wellcome Trust Case-Control Consortium, Rolf Adolfsson, Ole A. Andreassen, Douglas H. R. Blackwood, Elvira Bramon, Joseph D. Buxbaum, Anders D. Børglum, Sven Cichon, Ariel Darvasi, Enrico Domenici, Hannelore Ehrenreich, Tõnu Esko, Pablo V. Gejman, Michael Gill, Hugh Gurling, Christina M. Hultman, Nakao Iwata, Assen V. Jablensky, Erik G. Jönsson, Kenneth S. Kendler, George Kirov, Jo Knight, Todd Lencz, Douglas F. Levinson, Qingqin S. Li, Jianjun Liu, Anil K. Malhotra, Steven A. McCarroll, Andrew McQuillin, Jennifer L. Moran, Preben B. Mortensen, Bryan J. Mowry, Markus M. Nöthen, Roel A. Ophoff, Michael J. Owen, Aarno Palotie, Carlos N. Pato, Tracey L. Petryshen, Danielle Posthuma, Marcella Rietschel, Brien P. Riley, Dan Rujescu, Pak C. Sham, Pamela Sklar, David St Clair, Daniel R. Weinberger, Jens R. Wendland, Thomas Werge, Mark J. Daly, Patrick F. Sullivan, and Michael C. O'Donovan.



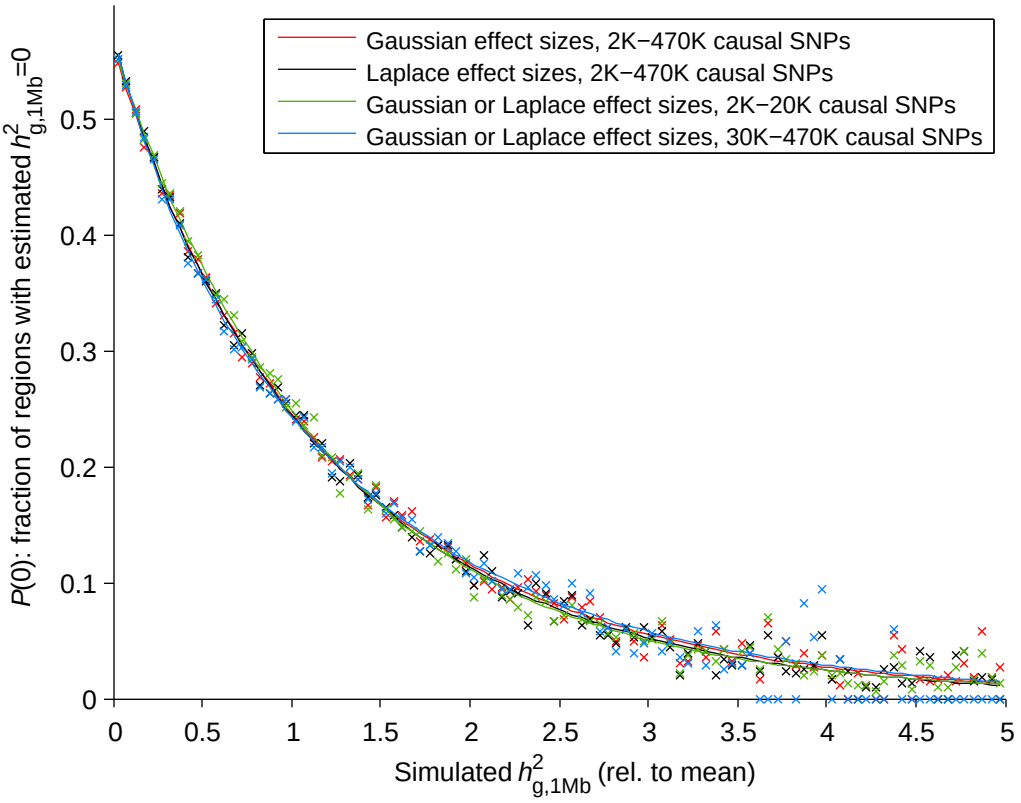
Supplementary Figure 1. Bounds on polygenicity for type 2 diabetes. See caption of Figure 2 for detailed explanations of panels. The smaller h^2_{g-cc} of type 2 diabetes (Supplementary Table 3) resulted in low inferential power: in panel (b), all simulated architectures are consistent with the observed data, and in panel (c), the concentration of SNP-heritability into top regions is virtually unknown.



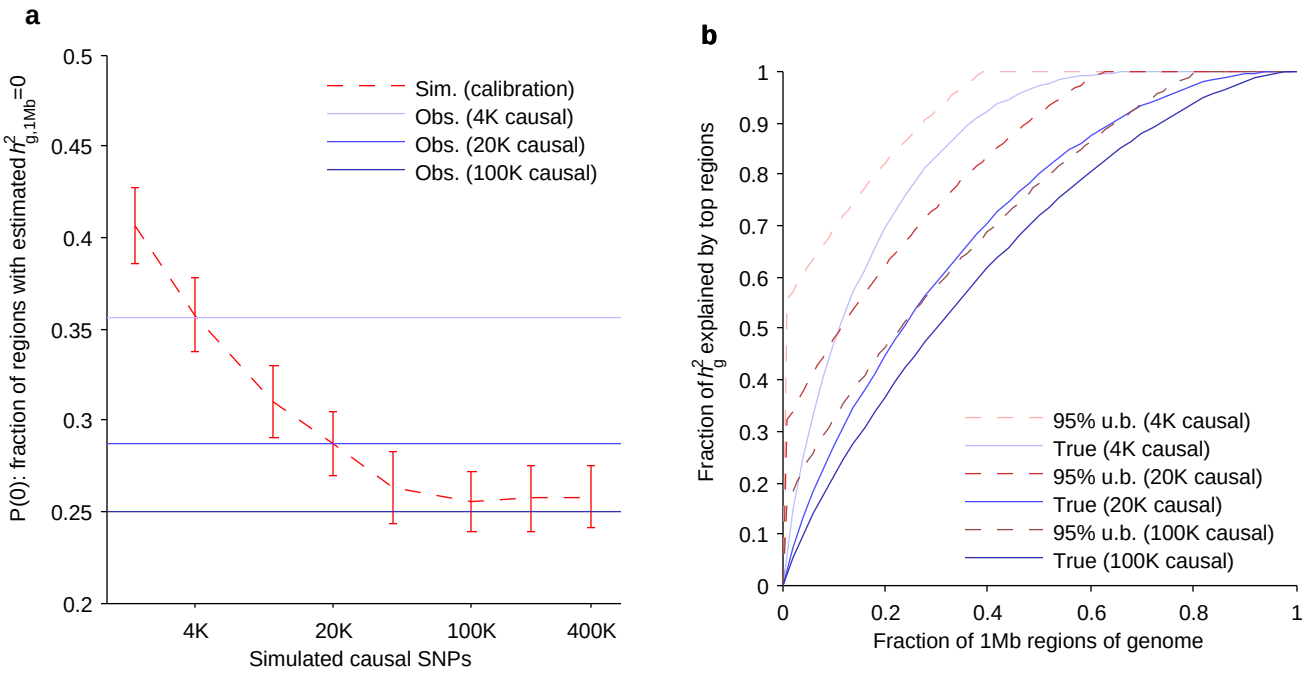
Supplementary Figure 2. Histograms of raw REML heritability parameter estimates per region. Panels (**a–d**), 1Mb regions; panels (**e–h**), 0.5Mb regions. The first bin in each histogram corresponds to zero estimated heritability explained. These plots show the bulk distributions of the parameter estimates; outliers (which would extend far to the right, especially for dyslipidemia) are not plotted, as they are visible in the Manhattan-style plots (Fig. 2a and Supplementary Figures 1a and 7a).



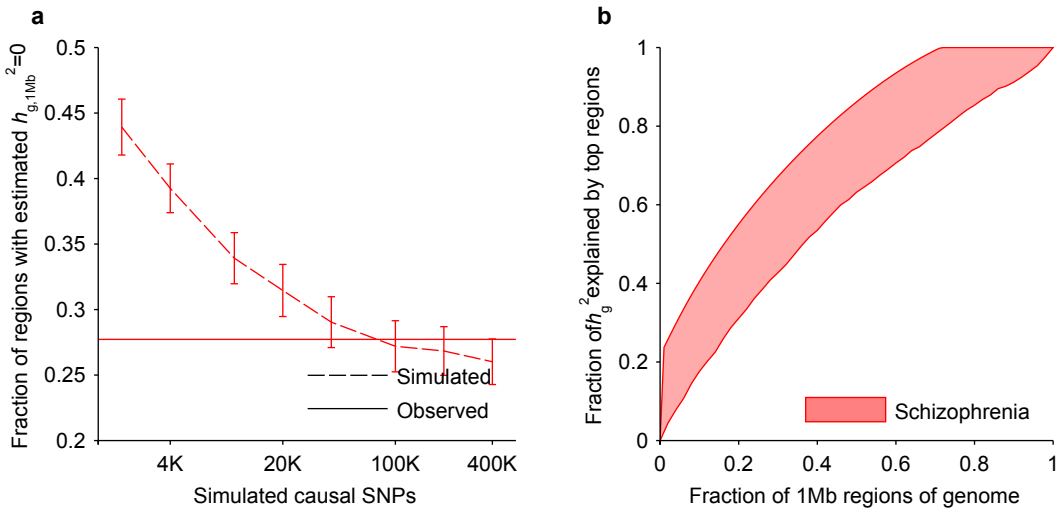
Supplementary Figure 3. Probability of a zero heritability estimate as a function of actual heritability explained per region. Panel (**a**), 1Mb regions; panel (**b**), 0.5Mb regions. Plotted data points (×) correspond to empirical data from simulations; solid curves are best-fit sums of two exponentials. Colors correspond to simulations with h_g^2 calibrated to the h_{g-cc}^2 of dyslipidemia (blue), hypertension (gold), and schizophrenia (red).



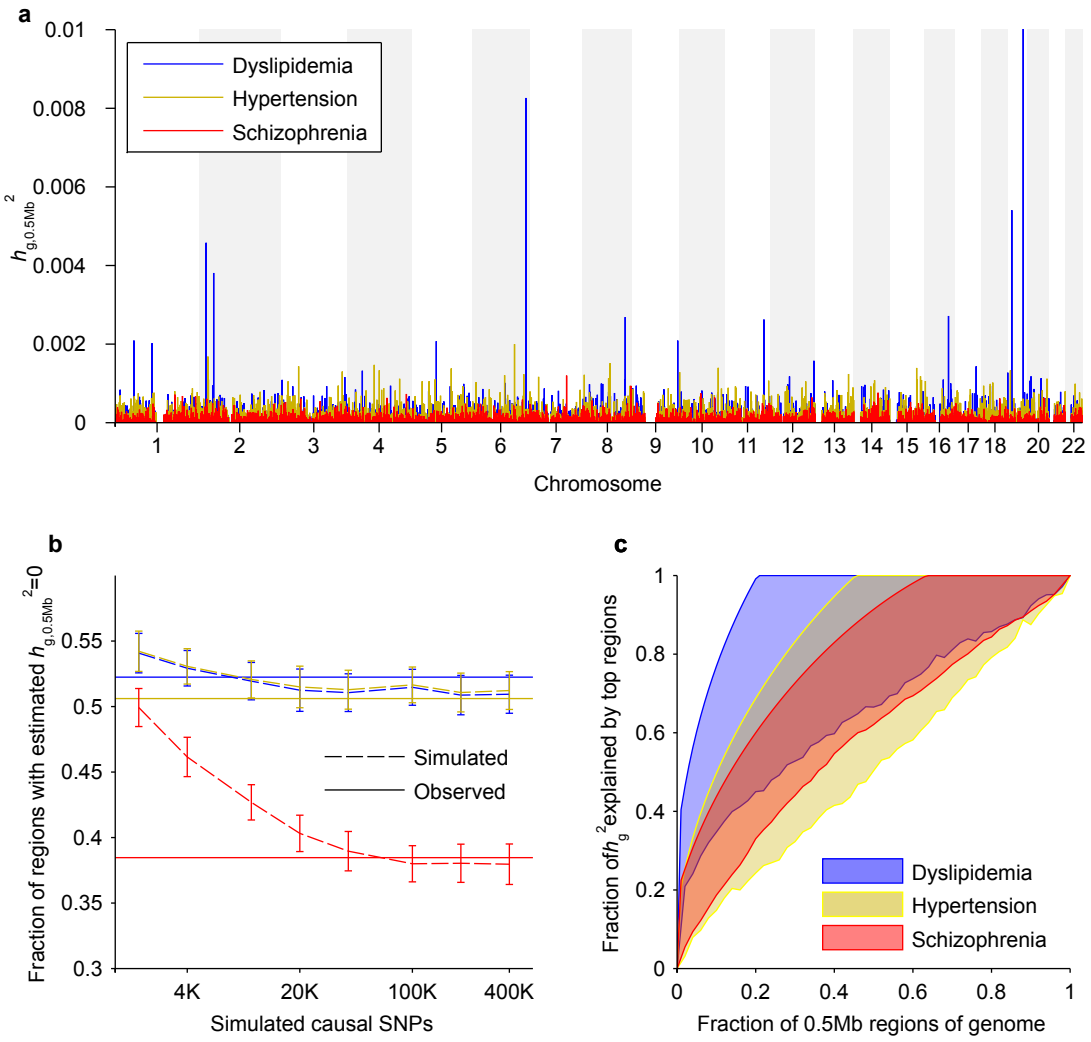
Supplementary Figure 4. Probability of a zero heritability estimate as a function of actual heritability explained per region: Robustness to genetic architecture. Plotted data points ($\times$) correspond to empirical data from simulations; solid curves are best-fit sums of two exponentials. The red curve is the same as the red curve in Supplementary Fig. 3a, which we used in our main per-1Mb heritability analyses of schizophrenia (Figure 2). The simulations used to generate this curve used genetic architectures in which 2K-470K causal SNPs were assigned effect sizes drawn from a normal distribution. The black curve corresponds to an alternative simulated genetic architecture in which effect sizes were drawn from a Laplace distribution (Supplementary Fig. 6). The green and blue curves combine the simulation data from the Gaussian and Laplace effect size distributions but restrict to simulations with $\leq$20K causal SNPs (green) or $\geq$30K causal SNPs (blue). All four curves are nearly identical, indicating that the probability of a zero heritability estimate depends almost solely on the actual heritability explained per region (independent of how the heritability is distributed across SNPs in the region).



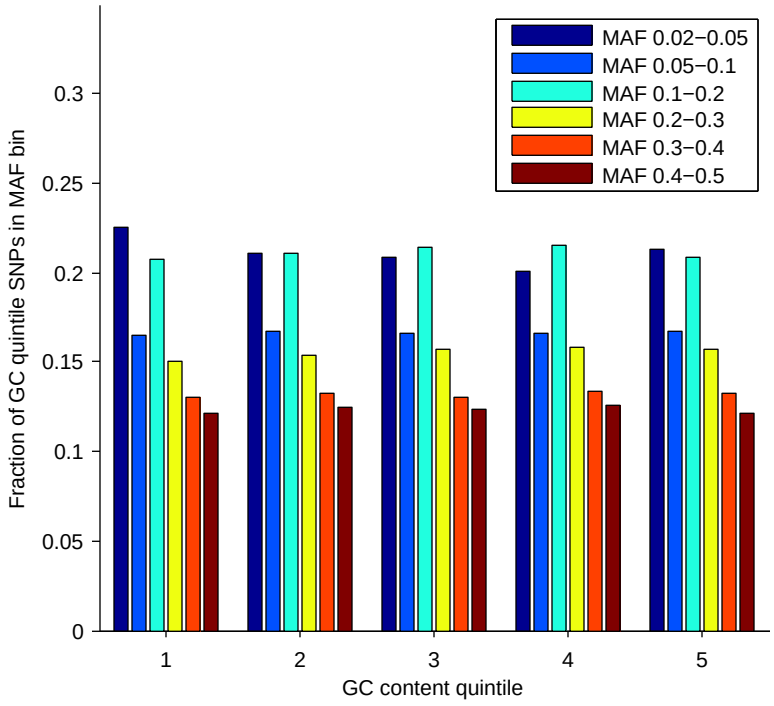
Supplementary Figure 5. Simulations verifying correctness of method for upper-bounding fraction of SNP-heritability explained by top regions. We simulated three case-control data sets with case over-ascertainment according to the following procedure. We simulated genotypes by generating individuals as mosaics of up to 100 random "ancestors" from the PGC2 controls, resampling ancestors every 250 SNPs. We simulated case-control phenotypes using a liability threshold model in which we first generated continuous phenotypes with h_g^2=0.24 explained by either 4,000, 20,000, or 100,000 markers and then defined cases as individuals with phenotypes exceeding a threshold corresponding to 0.4% population risk of disease. In each simulation, we ascertained 22,537 cases and 28,109 controls, matching the case-control ratio of the PGC2 schizophrenia data we analyzed. (The exact numbers of cases and controls differ slightly from our PGC2 data because they match an earlier round of QC.)
(**a**) For each simulated case-control data set, we estimated $h_{g,1Mb}^2$ per 1Mb region and computed the fraction $P(0)$ of regions with zero $h_{g,1Mb}^2$ estimates (solid blue horizontal lines), which we compared to the calibration data we used in our schizophrenia analysis (dashed red curve, as in Fig. 2b) based on quantitative traits simulated for real PGC2 individuals. We observed that the calibration curve closely matches the number of simulated causal SNPs in each simulation. (**b**) We then applied the same procedure we used in our main analyses (Fig. 2c and Online Methods) to bound the fraction of SNP-heritability explained by the 1Mb regions explaining the most heritability (dashed red curves), which we compared to the true fractions of SNP-heritability explained by the top 1Mb regions (solid blue curves), known from the simulation procedure. We observed that the our bounding procedure produced a conservative upper bound in each of the three simulations, as expected for a (conservative) 95% confidence bound.



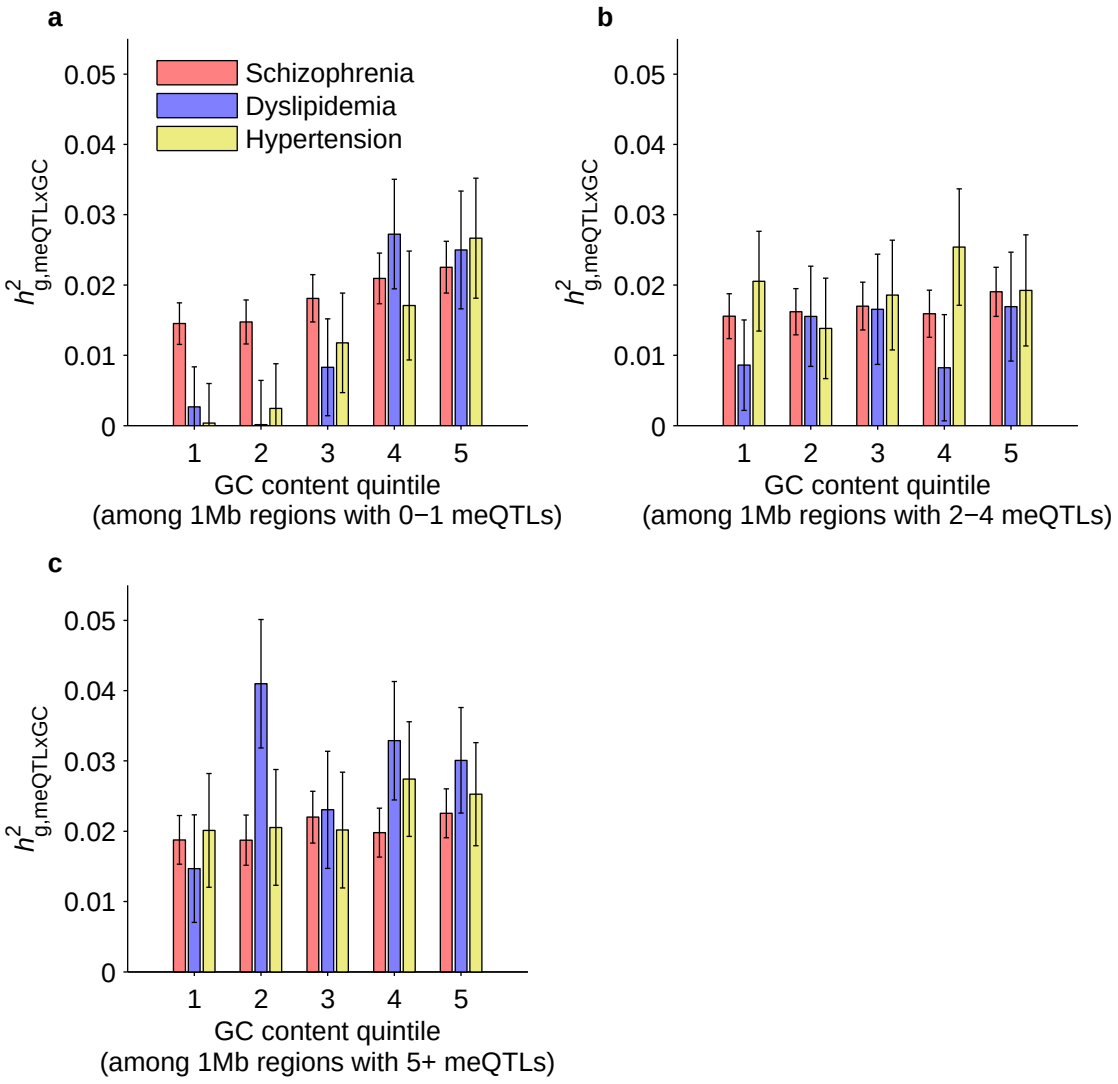
Supplementary Figure 6. Schizophrenia polygenicity analysis using Laplace effect size distribution. (This plot is the analog of (Figure 2b,c) using a Laplace distribution for effect sizes of causal SNPs in simulations used to upper-bound heritability explained by top regions.) (**a**) Fractions of 1Mb regions with estimated $h_{g,1Mb}^2$ equal to its lower bound constraint of zero in schizophrenia (solid) and simulated phenotypes with varying degrees of polygenicity and with h_g^2 matching the h_{g-cc}^2 of schizophrenia (dashed). Simulation data plotted are means over 5 simulations; error bars, 95% prediction intervals assuming Bernoulli sampling variance and taking into account s.e.m. The 95% error bar overlaps the observed fraction of zero $h_{g,1Mb}^2$ estimates for genetic architectures involving >30,000 SNPs. (**b**) Conservative 95% confidence intervals for the cumulative fraction of SNP-heritability explained by the 1Mb regions that contain the most SNP-heritability. Lower bounds are from a cross-validation procedure involving only the schizophrenia phenotype while upper bounds are inferred from the empirical sampling variance of $h_{g,1Mb}^2$ estimates (Online Methods).



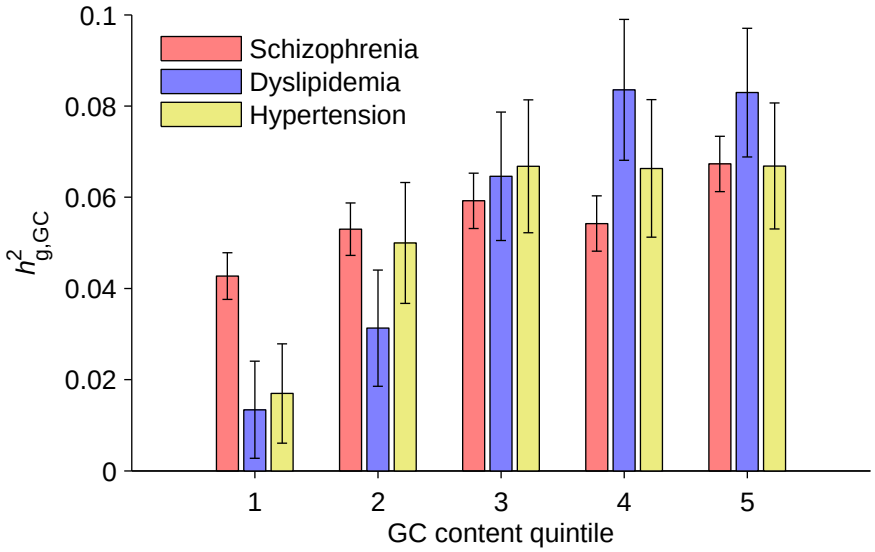
Supplementary Figure 7. Differing levels of polygenicity of complex diseases. (This plot is the analog of Figure 2 using 0.5Mb regions.) (**a**) Manhattan-style plots of estimated SNP-heritability per 0.5Mb region of the genome, $h_{g,0.5Mb}^2$, for dyslipidemia, hypertension, and schizophrenia. The *APOE* region of chromosome 19 is an outlier with an $h_{g,0.5Mb}^2$ estimate of 0.025. (**b**) Fractions of 0.5Mb regions with estimated $h_{g,0.5Mb}^2$ equal to its lower bound constraint of 0 in disease phenotypes (solid) and simulated phenotypes with varying degrees of polygenicity and with h_g^2 matching the h_{g-cc}^2 of each disease (dashed). Simulation data plotted are means over 5 simulations; error bars, 95% prediction intervals assuming Bernoulli sampling variance and taking into account s.e.m. (**c**) Conservative 95% confidence intervals for the cumulative fraction of SNP-heritability explained by the 0.5Mb regions that contain the most SNP-heritability. Lower bounds are from a cross-validation procedure involving only the disease phenotypes while upper bounds are inferred from the empirical sampling variance of $h_{g,0.5Mb}^2$ estimates (Online Methods).



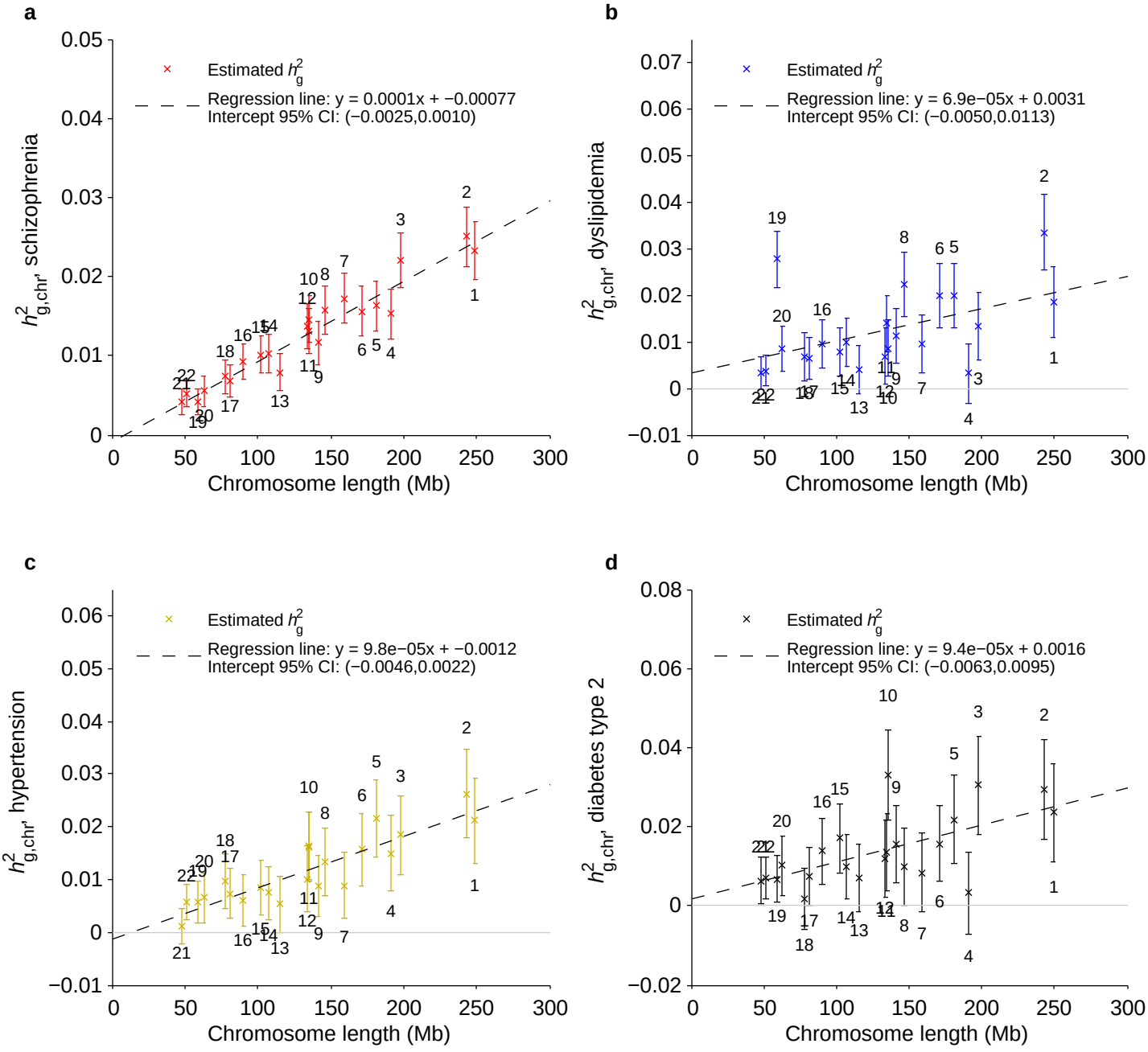
Supplementary Figure 8. MAF spectra of SNPs in each per-1Mb GC content quintile. The allele frequency distribution is nearly independent of GC content.



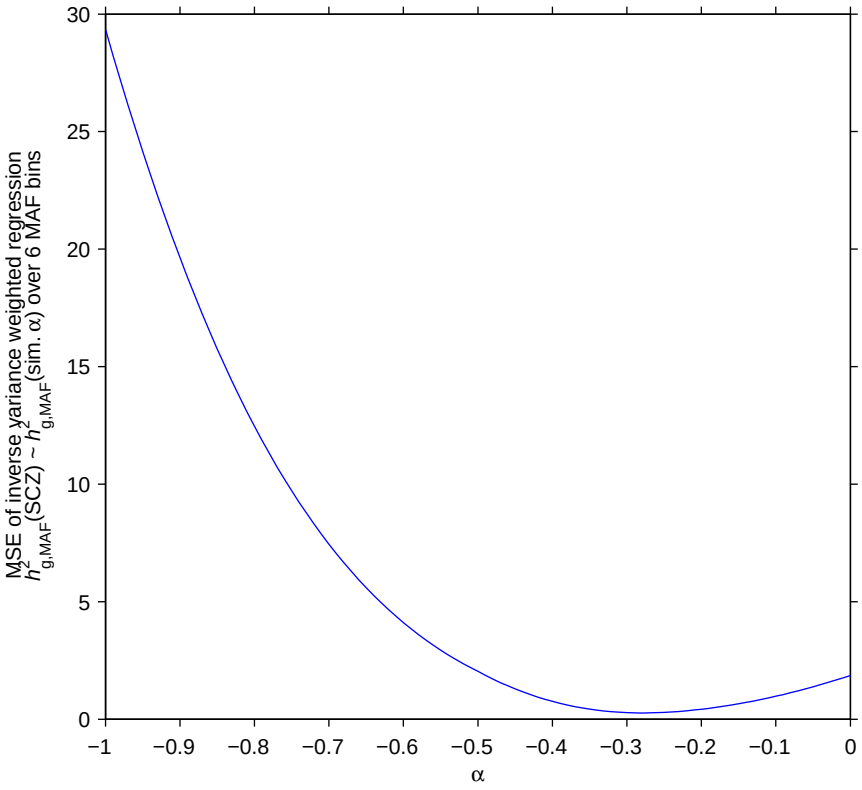
Supplementary Figure 9. SNP-heritability of schizophrenia liability partitioned by per-1Mb GC content after stratifying by meQTLs. We partitioned 1Mb regions into $3 \times 5 = 15$ categories as follows. First, we partitioned the regions into approximate meQTL tertiles (sets containing 1Mb regions with 0–1 meQTLs, 2–4 meQTLs, or 5+ meQTLs). We then partitioned the 1Mb regions in each approximate tertile into GC content quintiles, producing a partition of SNPs into 15 meQTL×GC variance components, on which we performed heritability partitioning using BOLT-REML. Panels **(a-c)** display the results subdivided by approximate meQTL tertile. The enrichment of heritability with higher GC content appears to be driven by regions in the lowest meQTL tertile.



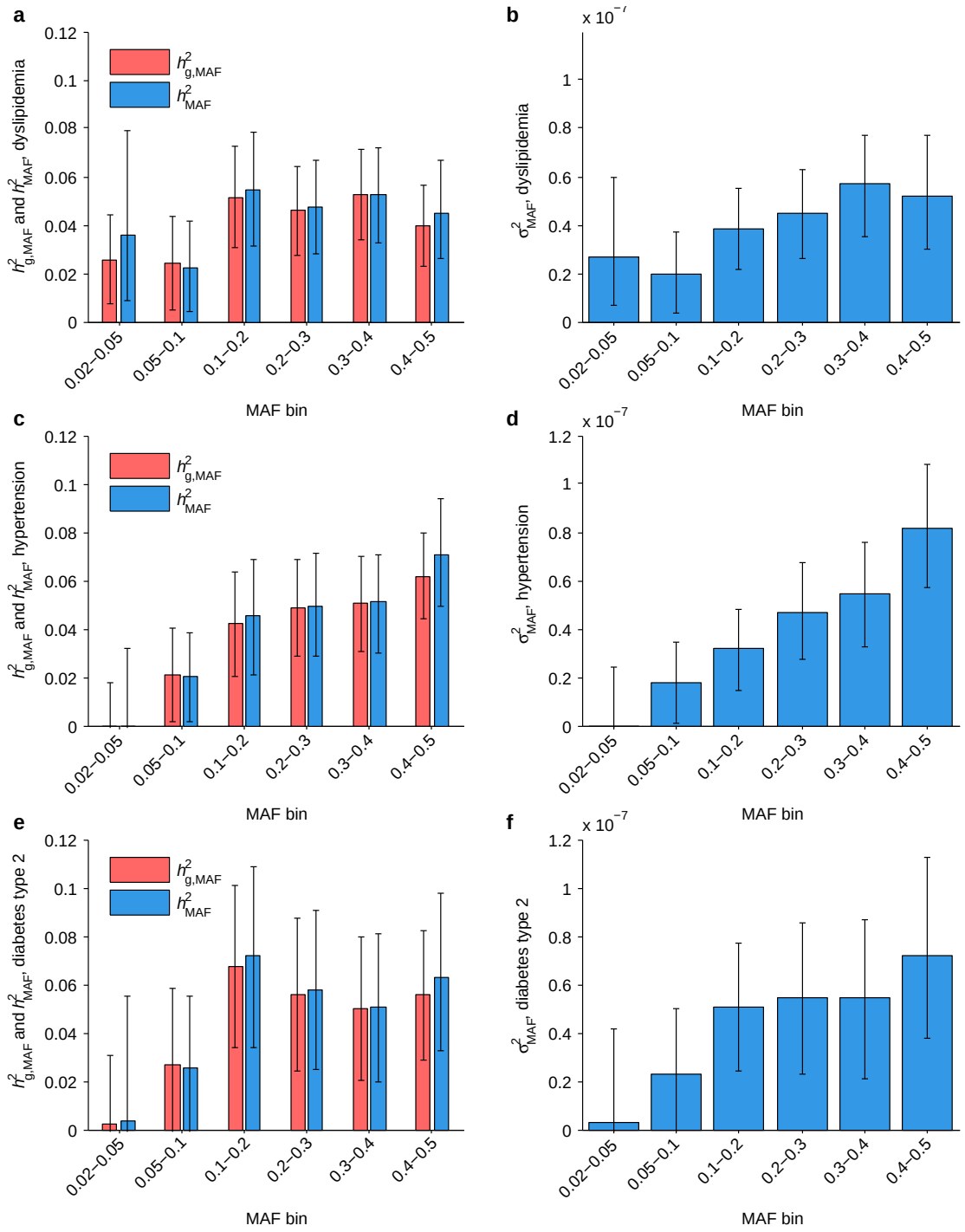
Supplementary Figure 10. SNP-heritability of disease liabilities partitioned by GC content. (This plot is the analog of Figure 3 using 0.5Mb regions.) GC content was computed at 0.5Mb resolution, after which 0.5Mb regions were stratified into GC quintiles for variance components analysis. Quintiles 1 and 5 correspond to the lowest and highest 20%, respectively. Error bars, 95% confidence intervals based on REML analytic standard errors.



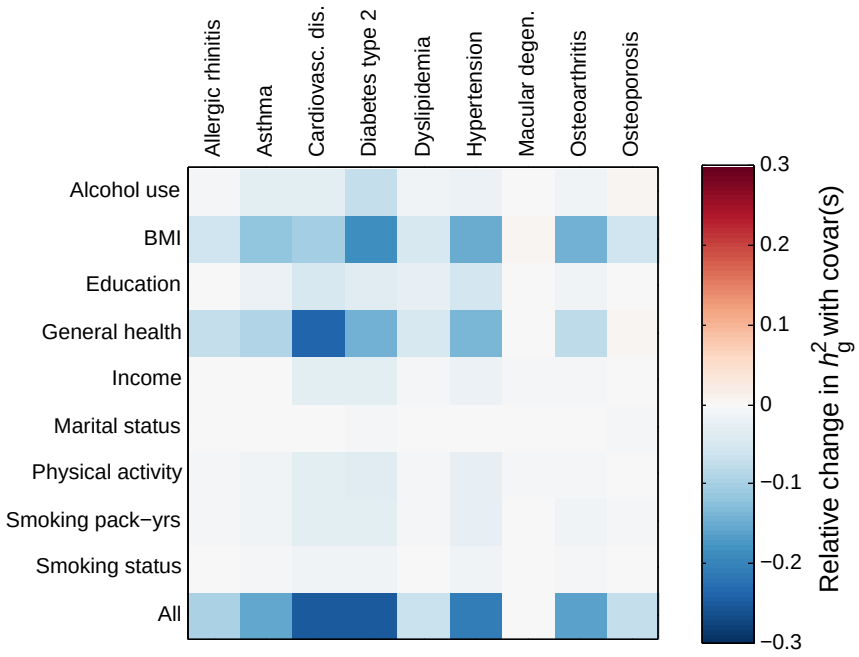
Supplementary Figure 11. SNP-heritability of disease liability partitioned by chromosome. Results are from a single REML analysis for each disease using one variance component per chromosome; $h_{g,chr}^2$, SNP-heritability explained per chromosome, as defined in Online Methods. Error bars, 95% confidence intervals based on REML analytic standard errors.



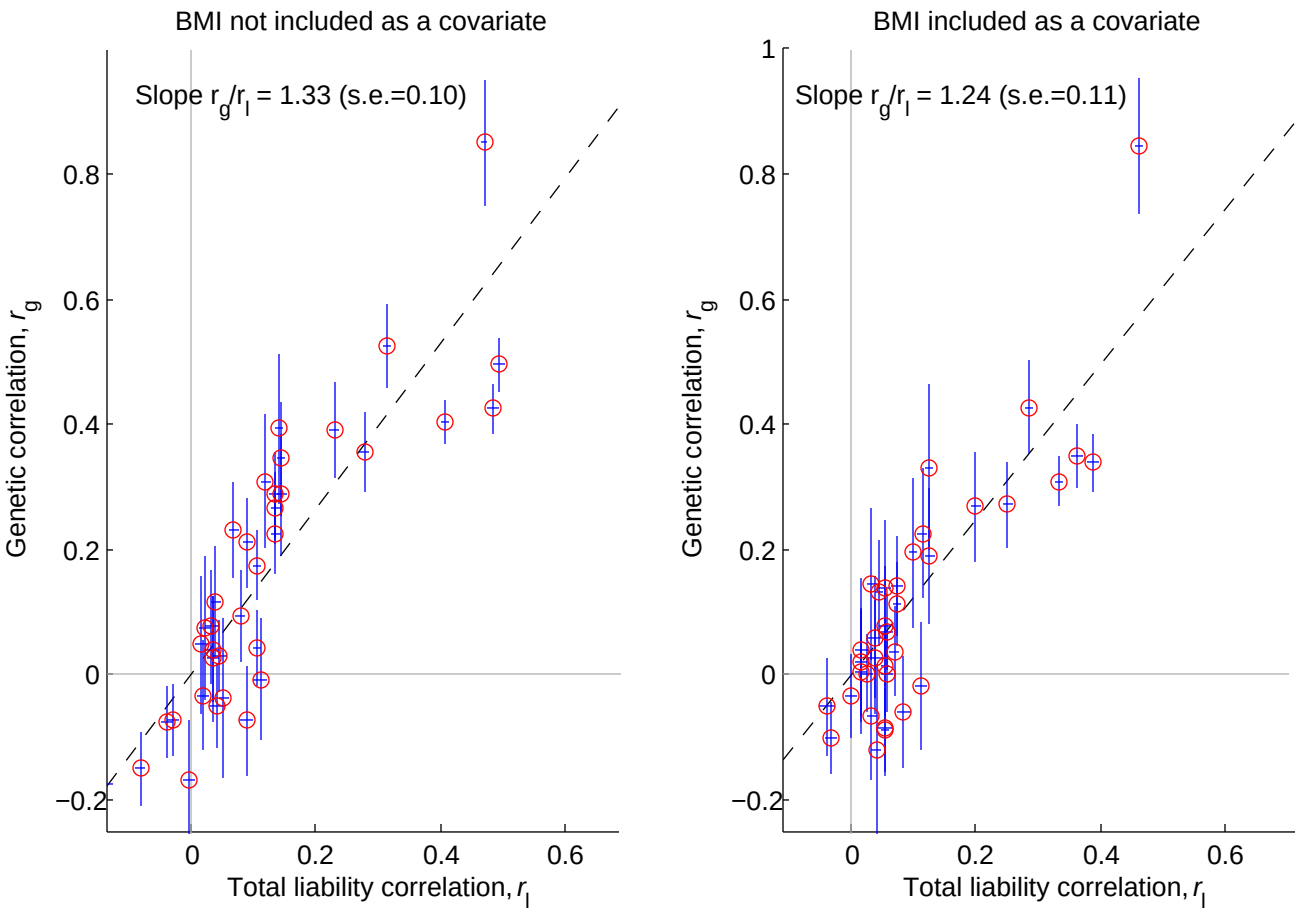
Supplementary Figure 12. Optimizing α (parameterizing MAF-dependence of SNP effects). We performed weighted least-squares regression of estimated $h_{g,MAF}^2$ for schizophrenia in each of six MAF bins (Fig. 4b) using $h_{g,MAF}^2$ from simulations (based on a given value of the parameter α) as a lone regressor (with no intercept term) and using regression weights equal to the inverses of estimated $h_{g,MAF}^2$ variances. For α = 0, –0.25, –0.5, –1, we averaged $h_{g,MAF}^2$ over 4,000 simulation replicates; for other values of α, we interpolated $h_{g,MAF}^2$ from the closest two simulated values of α. We plot the mean-squared residual error as a function of α; the value of α minimizing residual error is –0.28. We applied a leave-one-chromosome-out jackknife to estimate the standard error of this estimate and obtained an s.e. of 0.09.



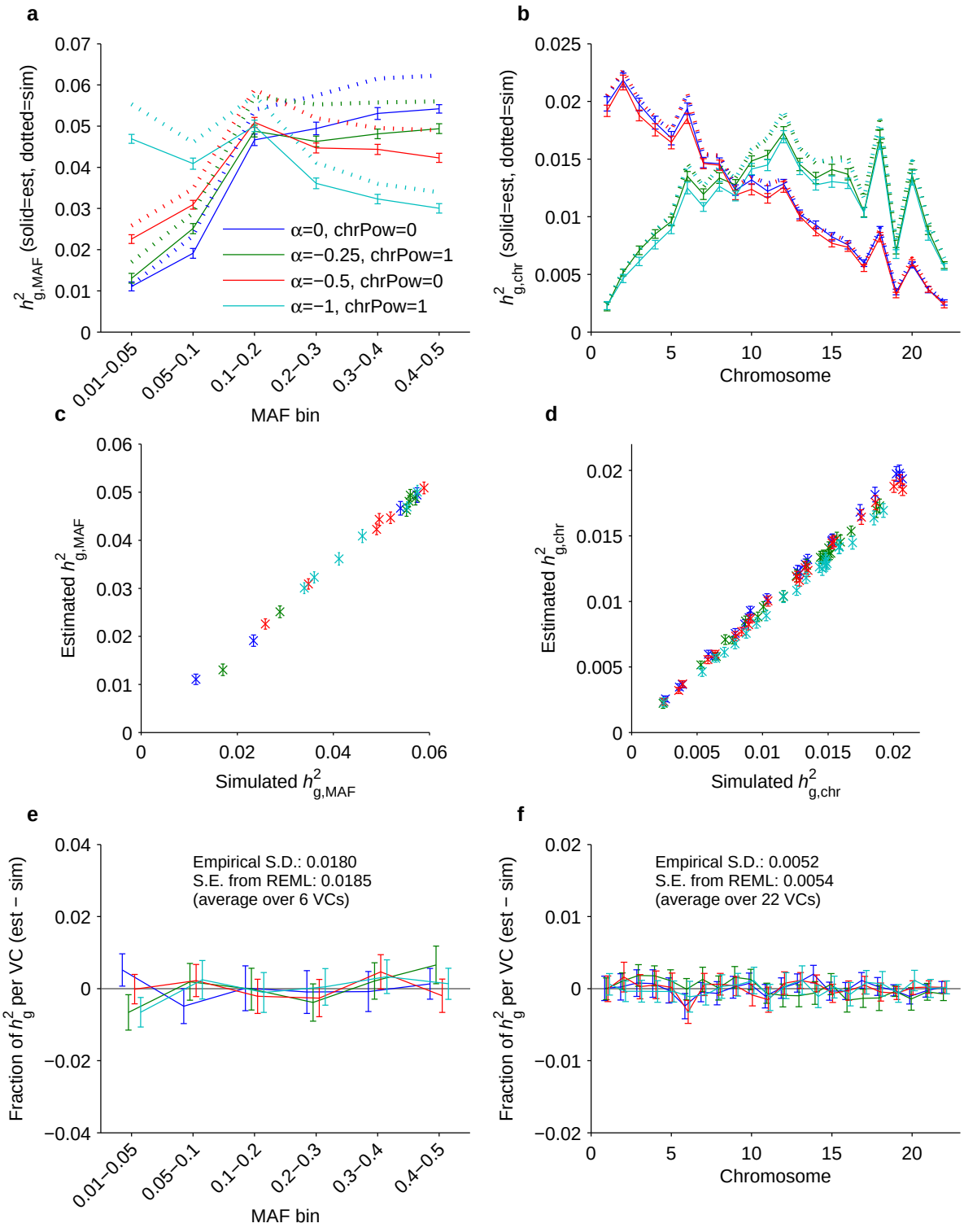
Supplementary Figure 13. Inferred heritability of GERA disease liabilities due to SNPs of various allele frequencies. (This plot is the analog of Fig. 4b,c for the three GERA traits with largest $h_{g-cc}{}^2$; see the caption of Figure 4 for details.) We ran GERA analyses using the 558,778 GERA SNPs with MAF≥2% and ran UK10K simulations (to infer $h_{MAF}{}^2$ from $h_{g,MAF}{}^2$) using the 535,626 SNPs with IDs also present in the UK10K data. We averaged simulation results over the 91% of GCTA runs that converged. Estimates of $h_{MAF}{}^2$ are based on best-fit α = –0.47 (0.25), 0, –0.22 (0.32) (s.e., jackknife) for dyslipidemia, hypertension, and type 2 diabetes. Error bars for $h_{g,MAF}{}^2$, 95% confidence intervals based on REML analytic standard errors; error bars for $h_{MAF}{}^2$ and $\sigma_{MAF}{}^2$, unions of 95% confidence intervals assuming $0 \geq \alpha \geq -1$.



Supplementary Figure 14. Impact of including covariates on SNP-heritability of GERA traits. The heat map depicts the ratio of h_g^2 estimated using additional covariates to h_g^2 estimated using only our standard set of covariates (age, sex, 10 principal components, and Affymetrix kit type). The first nine rows correspond to augmenting the covariate set with each survey-derived covariate in the GERA data individually; the last row corresponds to including all nine additional covariates. The "general health" covariate denotes self-reported health status at date of survey completion.



Supplementary Figure 15. Approximate linear relationship between genetic correlations and total liability correlations. Scatter plots of genetic correlations (r_g) vs. total liability correlations (r_l) for the 36 trait pairs analyzed in Figure 5, (**a**) before adjusting for BMI as a covariate, and (**a**) after adjusting for BMI as a covariate. For the regression fits, we used r_l as the lone regressor (with no intercept term); we note that standard errors on values of r_l are negligible (Supplementary Table 15). Horizontal and vertical bars indicate estimates of r_l and r_g plus or minus 1 s.e.

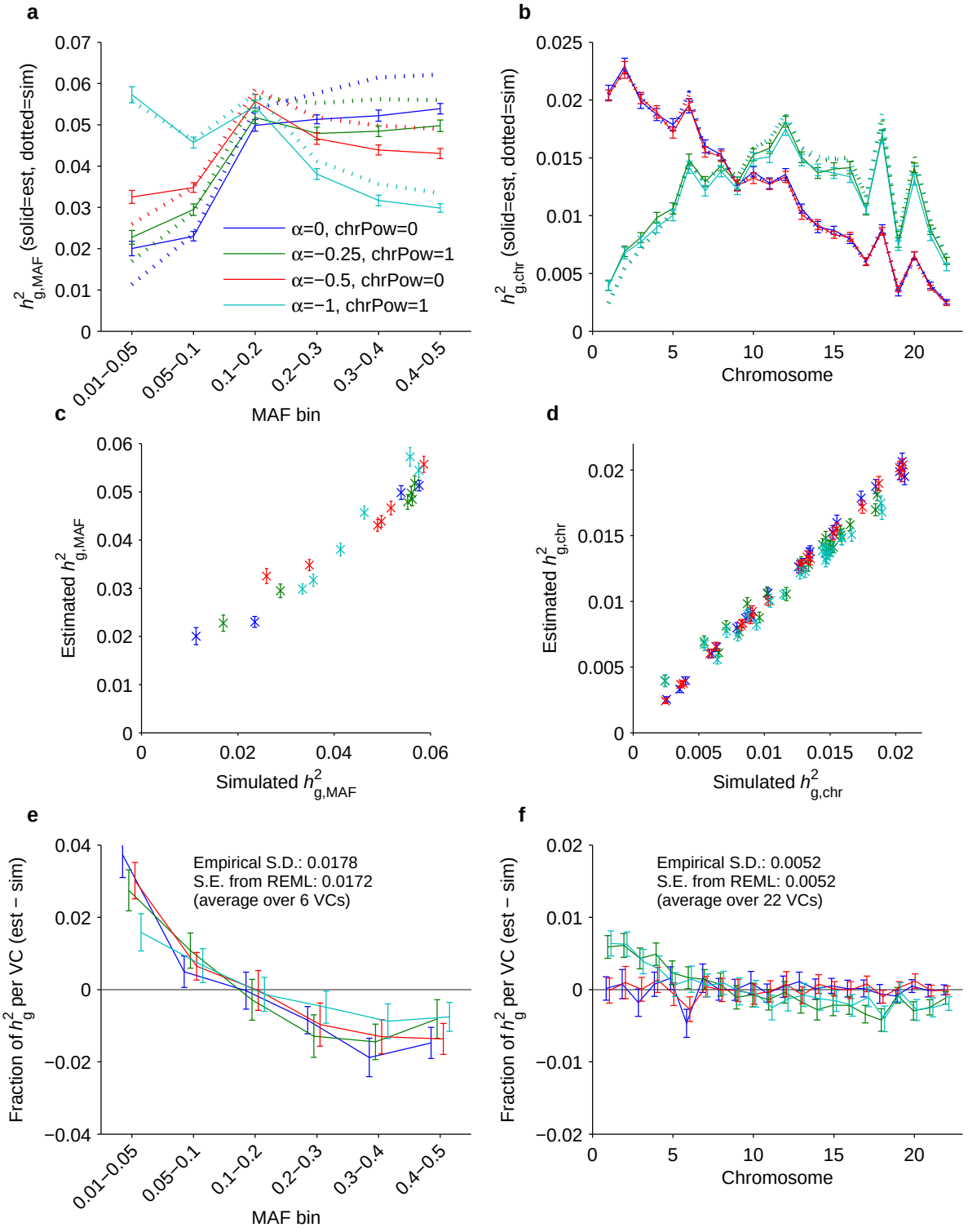


Supplementary Figure 16. REML heritability partitioning under case-control ascertainment: Simulations without stratification. The simulation setup and panels are described on the following page.

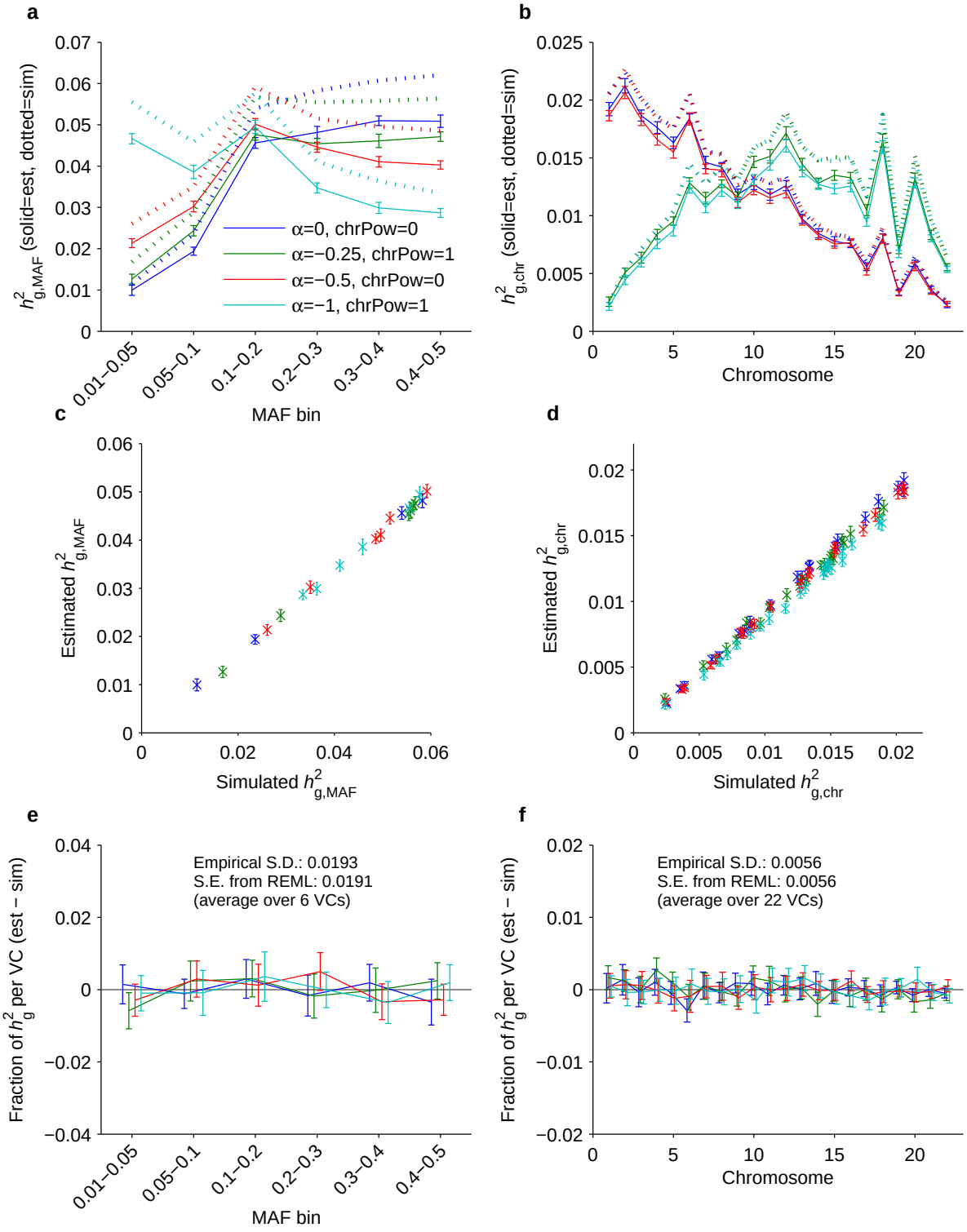
Extended caption for Supplementary Fig. 16. Results are shown for simulations in which genotypes of samples were generated from mosaics of five sub-cohorts of GERA individuals with Northern European, Southern European, Eastern European, Irish, Ashkenazi Jewish ancestry [62]. Explicitly, for each individual, we simulated genotypes as mosaics of chromosomal segments sampled from 100 random "ancestors" chosen from its ancestry group, resampling every 500 SNPs. We then generated genetic components of disease liabilities using a polygenic model (20,000 causal SNPs with total h_g^2=0.27; details described below). The simulations in Supplementary Fig. 16 did not include ancestry effects (i.e., these simulations contain population structure but not stratification), whereas the simulations in Supplementary Figures 17–19 include population effects explaining 5% of variance in liability (thus inducing stratification). We then generated remaining variance in liability from independent normal distributions. We designated individuals with liabilities in the top 1% as cases and sampled 22,177 cases and 27,629 controls in this way, matching our PGC2 data.

To test the robustness of heritability partitioning using REML under case-control ascertainment, we simulated genetic architectures in which SNPs have per-allele average effect size variances proportional to $p(1-p)^{\alpha}(\text{chrNum})^{\text{chrPow}}$, where p denotes minor allele frequency, chrNum denotes the chromosome (1–22) on which the SNP is located, and α and chrPow are parameters. We then partitioned heritability across MAF bins (**a**, **c**, **e**) and across chromosomes (**b**, **d**, **f**) using BOLT-REML with 10 principal component covariates and applying the liability threshold transformation of ref. [3]. In panels (**a**–**d**), we plot absolute estimates based on the Lee et al. [3] transform; in panels (**e**–**f**), we plot relative estimates (minus simulated values). We repeated each simulation 50 times; means and 95% confidence intervals are plotted.

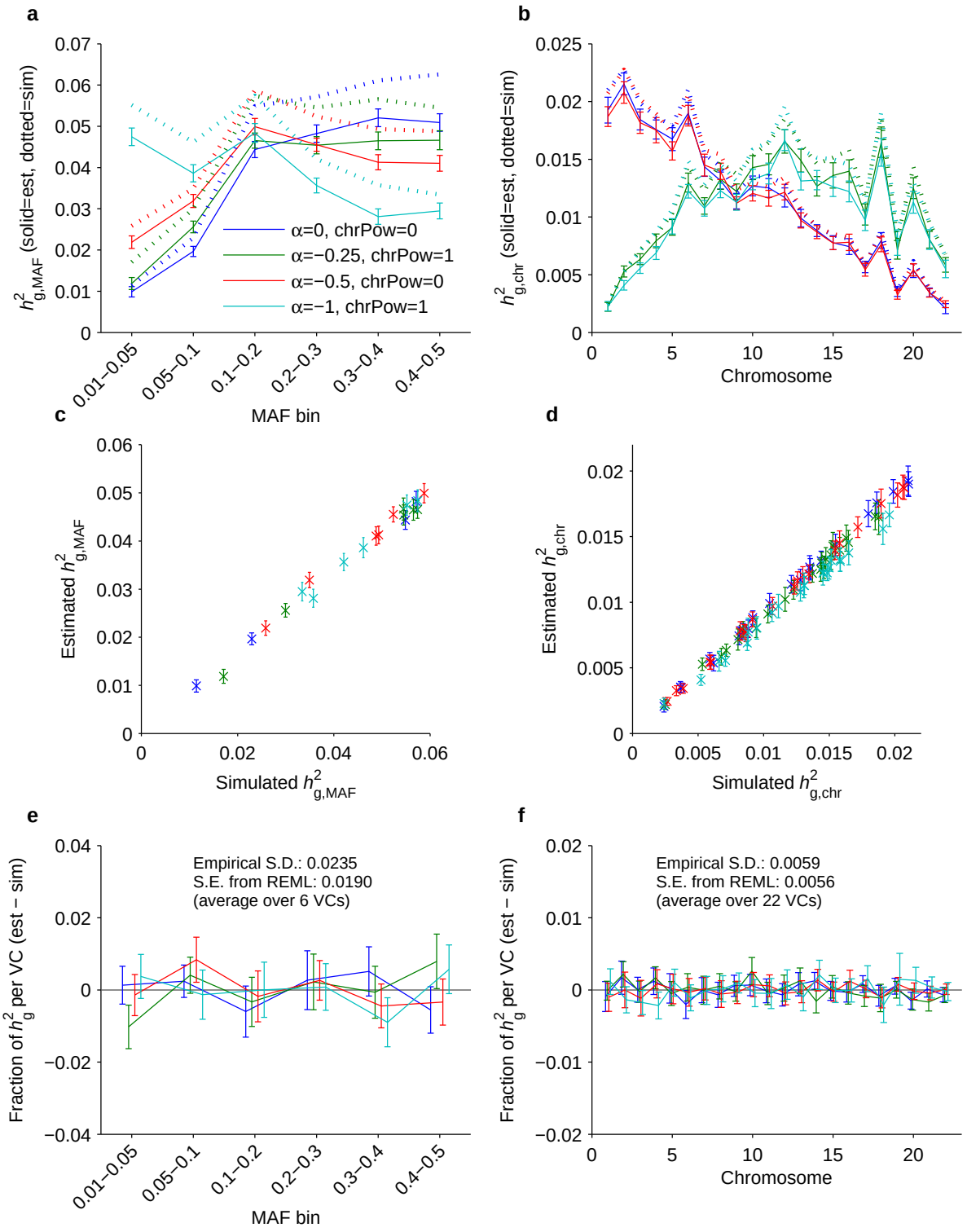
While the absolute h_g^2 estimates exhibit downward bias (**a**–**d**), as expected due to ascertainment [17, 27], the relative h_g^2 contributions of different variance components were correctly estimated, as were their standard errors (**e**–**f**).



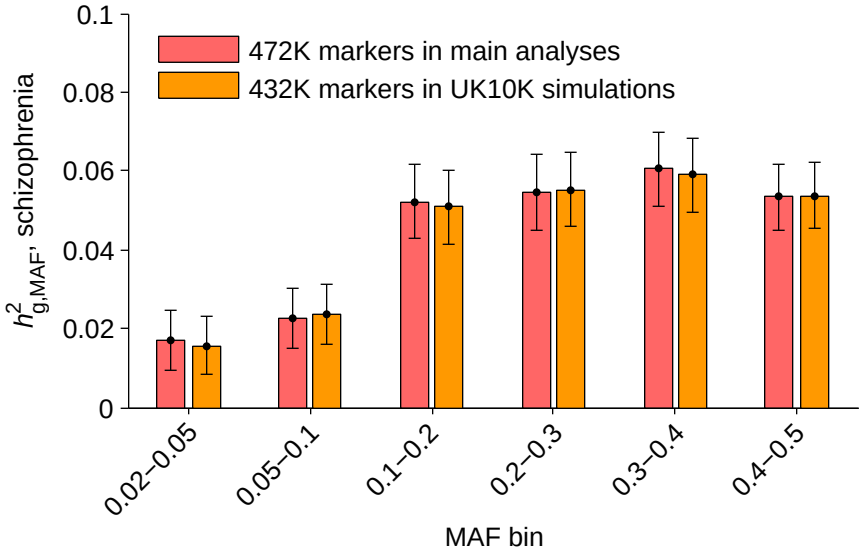
Supplementary Figure 17. REML heritability partitioning under case-control ascertainment: Simulations with uncorrected stratification. The simulation setup and panels are as in Supplementary Fig. 16 (see caption for details) but with 5% of liability explained by ancestry and no principal component covariates. Uncorrected stratification leads to biased results (**e**–**f**).



Supplementary Figure 18. REML heritability partitioning under case-control ascertainment: Simulations with PC-corrected stratification. The simulation setup and panels are as in Supplementary Fig. 16 (see caption for details) but with 5% of liability explained by ancestry and 10 principal component covariates. Principal component covariates successfully correct for stratification (**e**–**f**); cf. Supplementary Fig. 17.



Supplementary Figure 19. REML heritability partitioning under case-control ascertainment: Simulations with fewer causal SNPs. The simulation setup and panels are as in Supplementary Fig. 16 (see caption for details) but with 2,000 instead of 20,000 causal SNPs, 5% of liability explained by ancestry, and 10 principal component covariates.



Supplementary Figure 20. Negligible impact of using 432K vs. 472K SNPs for MAF-partitioning of SNP-heritability. Estimates of ${h_{g,MAF}}^2$ are statistically indistinguishable using the full set of 472,178 well-imputed markers vs. 432,177 SNPs with IDs present in UK10K data. (Of the 40,001 markers lost, we checked that 25,689 are indels.) Error bars, 95% confidence intervals based on REML analytic standard errors.

We note that the sum of ${h_{g,MAF}}^2$ over 6 MAF bins is slightly lower (by a factor of ≈0.95) than ${h_g}^2$ estimated using all SNPs in one variance component. While this slight difference is not statistically significant (given the standard errors of our REML analyses), it is consistent with the observation of Speed et al. [28] that mismatch between the true MAF-dependence of allele effect sizes (parameterized by α; see Online Methods) and the modeled MAF-dependence ($\alpha = -1$) leads to slight bias in estimated SNP-heritability. Indeed, the sign and magnitude of the difference are roughly as predicted by the simulations of ref. [28] (interpolated for our estimated $\alpha = -0.28$ for schizophrenia).

We also note that our MAF-partitioned analyses only approximately allow the variance component model to adjust for MAF-dependent genetic architectures (as all SNPs within a given MAF bin are normalized according to the standard assumption $\alpha = -1$); however, given the small MAF range within each bin, this procedure effectively estimates ${h_{g,MAF}}^2$ [29].

Supplementary Table 1. Detailed computational performance of BOLT-REML and GCTA heritability analysis algorithms.

(a) MAF partitioning (6+1 variance components)

	BOLT-REML	BOLT-REML	GCTA		
N	15 MC trials	100 MC trials	GRM computation	REML analysis	Total CPU
5	0.5 hr / 1.0 GB	1.4 hr / 1.0 GB	0.7 hr / 3.5 GB	0.1 hr / 1.9 GB	0.8 hr
10	1.3 hr / 1.7 GB	3.0 hr / 1.7 GB	2.4 hr / 7.1 GB	0.6 hr / 7.5 GB	3.0 hr
20	4.3 hr / 3.1 GB	8.9 hr / 3.2 GB	8.8 hr / 15.6 GB	6.3 hr / 29.9 GB	15.1 hr
30	6.5 hr / 4.5 GB	16.3 hr / 4.7 GB	19.4 hr / 25.5 GB	20.8 hr / 67.1 GB	40.2 hr
40	11.9 hr / 6.0 GB	25.1 hr / 6.1 GB	NA	NA	NA
50	16.0 hr / 7.4 GB	35.4 hr / 7.6 GB	NA	NA	NA

(b) Chromosome partitioning (22+1 variance components)

	BOLT-REML	BOLT-REML	GCTA		
N	15 MC trials	100 MC trials	GRM computation	REML analysis	Total CPU
5	1.0 hr / 1.4 GB	2.2 hr / 1.4 GB	0.8 hr / 1.5 GB	0.1 hr / 4.9 GB	0.9 hr
10	2.3 hr / 2.1 GB	5.2 hr / 2.1 GB	2.6 hr / 3.3 GB	1.1 hr / 19.4 GB	3.7 hr
20	7.0 hr / 3.5 GB	12.4 hr / 3.5 GB	9.1 hr / 8.0 GB	7.9 hr / 77.5 GB	17.0 hr
30	11.8 hr / 5.0 GB	21.7 hr / 5.0 GB	NA	NA	NA
40	17.6 hr / 6.4 GB	32.9 hr / 6.6 GB	NA	NA	NA
50	24.8 hr / 7.9 GB	44.7 hr / 8.2 GB	NA	NA	NA

(c) Bivariate analysis

	BOLT-REML	BOLT-REML	GCTA		
N	15 MC trials	100 MC trials	GRM computation	REML analysis	Total CPU
5	0.8 hr / 0.9 GB	2.1 hr / 0.9 GB	0.7 hr / 15.1 GB	0.6 hr / 3.4 GB	1.3 hr
10	1.5 hr / 1.6 GB	6.1 hr / 1.7 GB	2.3 hr / 30.3 GB	3.9 hr / 13.5 GB	6.2 hr
20	5.0 hr / 3.0 GB	12.9 hr / 3.2 GB	8.8 hr / 61.8 GB	35.2 hr / 53.7 GB	44.0 hr
30	8.0 hr / 4.4 GB	23.7 hr / 4.7 GB	NA	NA	NA
40	12.8 hr / 5.8 GB	46.6 hr / 6.3 GB	NA	NA	NA
50	16.5 hr / 7.3 GB	47.6 hr / 7.7 GB	NA	NA	NA

(This table provides numeric data plotted in Figure 1 with additional details.) We benchmarked BOLT-REML and GCTA in three heritability estimation scenarios: (a) partitioning across six MAF bins, (b) partitioning across 22 chromosomes, and (c) bivariate analysis. Runs used subsets of the GERA cohort with fixed SNP count M=597,736 and increasing sample size (N) and analyzed dyslipidemia as the phenotype in the univariate analyses and hypertension as the second phenotype in the bivariate analysis. Reported run times are medians of five identical runs using one core of a 2.27 GHz Intel Xeon L5640 processor. Reported run times for GCTA break down total CPU time into time required for computing the GRM and time required for performing REML analysis. "NA" indicates scenarios in which GCTA required more memory than the 96GB available. Software versions: BOLT-REML, v2.1; GCTA, v1.24.
For bivariate analysis (with all SNPs in a single variance component), the default GCTA GRM computation used more memory than GCTA REML analysis at the sample sizes we tested. Given that the memory footprint of GCTA GRM computation can easily be circumvented (e.g., by combining GCTA-computed per-chromosome GRMs or by using PLINK2 [60]), we plotted only the memory use of GCTA REML analysis in Fig. 1b; we report the memory use of GCTA GRM computation in these tables for completeness. We also note that while PLINK2 allows efficient parallelization of the GRM computation, the total CPU time required is similar to that of GCTA GRM computation (data not shown).

Supplementary Table 2. Breakdown of PGC2 schizophrenia cohorts samples analyzed.

Cohort	Cases	Controls	Total
ajsz	885	1570	2455
asrb	452	283	735
buls	195	606	801
butr	599	606	1205
cims	66	64	130
clm2	3418	3873	7291
clo3	2057	1889	3946
cou3	524	673	1197
denm	443	443	886
dubl	263	826	1089
edin	363	280	643
egcu	216	1064	1280
ersw	255	307	562
gras	988	1113	2101
irwt	1278	972	2250
lacw	153	218	371
lie2	129	266	395
lie5	492	380	872
mgs2	2573	2286	4859
msaf	187	92	279
pewb	506	1714	2220
pews	150	233	383
s234	1845	2141	3986
swe5	1726	2508	4234
swe6	924	1098	2022
top8	377	403	780
ucla	656	576	1232
umeb	285	526	811
umes	172	619	791
Total	22177	27629	49806

Four-character abbreviations and full descriptions of cohorts are detailed in ref. [12].

Supplementary Table 3. Estimated proportions of variance in disease liability explained by SNPs for schizophrenia and all GERA diseases.

Disease	Cases	Controls	$h_{g\text{–}cc}^2$ (s.e.)	h_g^2 (s.e.)
Schizophrenia	22,177	27,629	0.415 (0.008)	0.274 (0.007)
Allergic rhinitis	13,437	41,297	0.039 (0.008)	0.074 (0.015)
Asthma	8,929	45,805	0.068 (0.008)	0.152 (0.018)
Cardiovasc. dis.	14,861	39,873	0.051 (0.008)	0.092 (0.015)
Diabetes type 2	6,845	47,889	0.115 (0.009)	0.297 (0.022)
Dyslipidemia	29,511	25,223	0.167 (0.009)	0.263 (0.014)
Hypertension	27,921	26,813	0.162 (0.009)	0.255 (0.014)
Macular degen.	3,700	51,034	0.066 (0.008)	0.242 (0.029)
Osteoarthritis	19,832	34,902	0.059 (0.008)	0.098 (0.014)
Osteoporosis	5,337	49,397	0.066 (0.008)	0.195 (0.024)
Cancer (all)	16,828	37,906	0.072 (0.008)	0.124 (0.014)
Depression	6,957	47,777	0.025 (0.008)	0.063 (0.020)
Dermatophytosis	7,479	47,255	0.020 (0.008)	0.049 (0.019)
Hemorrhoids	8,817	45,917	0.007 (0.008)	0.016 (0.017)
Hernia abdominopelvic	6,132	48,602	0.023 (0.008)	0.064 (0.021)
Insomnia	3,851	50,883	0.013 (0.008)	0.046 (0.028)
Iron deficiency	2,404	52,330	0.016 (0.008)	0.079 (0.038)
Irritable bowel	2,968	51,766	0.015 (0.008)	0.063 (0.032)
Peripheral vasc. dis.	4,278	50,456	0.022 (0.008)	0.076 (0.026)
Peptic ulcers	897	53,837	0.015 (0.008)	0.144 (0.073)
Psychiatric (all)	8,240	46,494	0.052 (0.008)	0.123 (0.019)
Stress (acute reaction)	4,127	50,607	0.020 (0.008)	0.070 (0.027)
Varicose veins	2,435	52,299	0.023 (0.008)	0.113 (0.038)

Schizophrenia cases and controls are from the PGC2 data set [12]; cases and controls for the other 22 diseases are from the GERA data set. For all diseases, the reported $h_{g\text{–}cc}^2$ is the raw REML heritability parameter obtained by running BOLT-REML on the full data set. For schizophrenia, the h_g^2 estimate was computed using PCGC regression (assuming a population risk of 1%) to avoid bias induced by over-ascertainment of cases [17, 27] (Supplementary Table 6). For GERA diseases, h_g^2 estimates and were computed from $h_{g\text{–}cc}^2$ using the liability transform of ref. [3] (assuming random sample ascertainment). We restricted further analyses to the top nine GERA diseases, eliminating 11 diseases with very low $h_{g\text{–}cc}^2 \leq 0.025$ as well as two heterogeneous traits (all types of cancer and all psychiatric diseases).

Supplementary Table 4. Effect of sample size on REML heritability parameter estimates ($h_{\text{g–cc}}^2$) in ascertained and non-ascertained traits.

Subsampling	Schizophrenia	Dyslipidemia	Hypertension	Diabetes type 2
1x	0.415	0.167	0.162	0.116
2x	0.436 (0.009)	0.167 (0.007)	0.161 (0.009)	0.121 (0.007)
3x	0.450 (0.004)	0.184 (0.008)	0.170 (0.009)	0.120 (0.006)
4x	0.462 (0.005)	0.172 (0.006)	0.172 (0.009)	0.118 (0.006)
5x	0.460 (0.007)	0.179 (0.007)	0.169 (0.009)	0.126 (0.005)
6x	0.463 (0.007)	0.182 (0.008)	0.162 (0.009)	0.111 (0.007)
7x	0.468 (0.006)	0.190 (0.009)	0.169 (0.009)	0.126 (0.007)
8x	0.464 (0.007)	0.190 (0.009)	0.165 (0.008)	0.126 (0.008)
9x	0.472 (0.007)	0.194 (0.009)	0.157 (0.008)	0.127 (0.007)
10x	0.467 (0.007)	0.184 (0.008)	0.158 (0.008)	0.122 (0.006)

We report BOLT-REML $h_{\text{g–cc}}^2$ estimates for schizophrenia and the three GERA traits with highest $h_{\text{g–cc}}^2$ on subsamples of the the PGC2 and GERA data sets. The first row contains estimates using all samples (corresponding to $h_{\text{g–cc}}^2$ values reported in Supplementary Table 3). Subsequent rows show the results of randomly splitting the data into $F = 2, \ldots, 10$ folds and computing $h_{\text{g–cc}}^2$ for each fold. For each value of F, we repeated this experiment F times (using different random splits); values reported are means over F^2 estimates (s.e.m. in parentheses). The analytic REML s.e. estimates for the full-sample (1x) estimates are 0.008, 0.009, 0.009, 0.009 for the four traits; we note that there are no significant differences between the full-sample vs. subsampled $h_{\text{g–cc}}^2$ for the GERA traits. The liability-scale estimates of h_{g}^2 for 1x and 10x subsamples of the schizophrenia data are 0.232 and 0.261, respectively (assuming 1% population risk).

Supplementary Table 5. Comparison of published schizophrenia SNP-heritability estimates.

Study	Data set	Cases	Controls	SNPs	h_g^2(s.e.)	Notes
This study	PGC2	22,177	27,629	472K 2x LD-pruned ($r^2 = 0.9$)	0.274 (0.007)	PCGC
					0.261 (0.006)	N=5K at a time
					0.244 (0.006)	N=25K at a time
					0.232 (0.005)	All 50K together
				597K LD-pruned	0.222 (0.005)	All 50K together
				1.6M HapMap3 only	0.195 (0.004)	All 50K together
				3.9M All well-imp. (1000 Genomes imp.)	0.194 (0.004)	All 50K together
Lee et al. 2012 (Table 1 of [5])	ISC	3,220	3,445	915K (HapMap3 imputed)	0.27 (0.02)	N=7K
	MGS	2,571	2,419		0.31 (0.03)	N=5K
	ISC+MGS	5,791	5,864		0.25 (0.01)	N=12K
	PGC1	9,087	12,171		0.23 (0.01)	N=21K
Ripke et al. 2013 [11]	Subset of sw1-6	7,301 (core Swedish)		1.1M HapMap3 only (1000 Genomes imp.)	0.32 (0.03)	N=7K
Golan et al. 2014 (Supp Tab 2, [17])	sw5-6	6,731 (Swedish)		617K MAF$\geq$5% (array)	0.334 (0.029)	N=7K, REML
					0.372 (0.044)	PCGC regression

We survey estimates of schizophrenia h_g^2 from recent studies. All estimates reported above assume population risk of 1% and do not apply an adjustment for incomplete tagging. We observe that our estimate, h_g^2=0.274 (s.e.=0.007) in the top line, is significantly higher than the estimate h_g^2=0.23 (s.e.=0.01) of Lee et al. [5] using the full PGC1 data set. Our analyses suggest that two factors together fully explain this discrepancy: (1) downward bias of large-sample REML h_g^2 estimates due to case over-ascertainment [17, 27] (Supplementary Table 4), which we accounted for in this study; and (2) LD bias of h_g^2 estimates due to LD-dependent genetic architecture [28]. While previous work has shown that REML analysis of LD-pruned data (as in this study) generally produces larger h_g^2 estimates than direct REML analysis of all imputed markers [30] (as in ref. [5, 11]), the question of which procedure is unbiased (or less biased) remains open. We are aware of preliminary studies suggesting REML h_g^2 estimates based on unpruned marker sets incur downward bias due to enrichment of heritability in lower-LD SNPs (B.K.B., H.K.F., V. Anttila; personal communication), but further investigation is needed. We also note that our subsampling experiments (Supplementary Table 4) and simulations (Supplementary Table 6) indicate that at the sample size analyzed by Golan et al. [17], we would expect REML estimates of schizophrenia h_g^2 to be downward biased by $<$0.01, so our interpretation of the large but non-significant difference between REML and PCGC estimates in ref. [17] is that it is mostly due to noise.

Supplementary Table 6. Downward bias of REML h_g^2 estimates for simulated diseases of varying prevalence and schizophrenia.

(a) Simulated case-control data with over-ascertainment of cases matching PGC2 data

Sim. parameters		N=50K				N=5K subsets, 100 rep mean				$h_{g\text{–}cc}^2$(50K / 5K)	
h_g^2	K	$\overline{h_{g\text{–}cc}^2}$	s.d.	$\overline{h_g^2}$	s.d.	$\overline{h_{g\text{–}cc}^2}$	s.d.	$\overline{h_g^2}$	s.d.	$\overline{\text{ratio}}$	s.d.
0.31	0.02	0.406	0.009	0.269	0.006	0.454	0.014	0.301	0.010	0.893	0.012
0.261	0.01	0.408	0.008	0.228	0.005	0.459	0.015	0.256	0.008	0.890	0.024
0.214	0.004	0.404	0.012	0.185	0.005	0.456	0.015	0.209	0.007	0.888	0.016
0.187	0.002	0.403	0.008	0.162	0.003	0.461	0.014	0.185	0.006	0.876	0.021
0.166	0.001	0.400	0.005	0.142	0.002	0.462	0.013	0.165	0.005	0.865	0.016

(b) Analyses of PGC2 schizophrenia data set

Assumed	N=50K		N=5K subsets, mean over 100 reps		Ratio
pop. risk, K	$h_{g\text{–}cc}^2$	h_g^2	$h_{g\text{–}cc}^2$	h_g^2	$h_{g\text{–}cc}^2$(50K) / $h_{g\text{–}cc}^2$(5K)
0.01	0.415	0.232	0.467	0.261	0.889
0.004	0.415	0.190	0.467	0.214	0.889

In table **(a)**, we report data from BOLT-REML runs on simulated case-control data sets with case over-ascertainment. We simulated genotypes by generating individuals as mosaics of up to 100 random "ancestors" from the PGC2 controls, resampling ancestors every 500 SNPs. We simulated case-control phenotypes using a liability threshold model in which we first generated continuous phenotypes with h_g^2=0.166–0.31 explained by 20,000 markers and then defined cases as individuals with phenotypes exceeding thresholds corresponding to the desired prevalences K=0.001–0.02. In each simulation, we ascertained 22,177 cases and 27,629 controls, matching the PGC2 schizophrenia data we analyzed. We ran BOLT-REML on the full simulated data (N=50K), obtaining $h_{g\text{–}cc}^2$, which we transformed to h_g^2 (ref. [3]) using the simulated population prevalence; we also averaged 100 $h_{g\text{–}cc}^2$ estimates on 10x subsamples of the data (Supplementary Table 4) and transformed this estimate to h_g^2, and we compared the estimate based on N=5K subsamples to the N=50K estimate. Reported values are the mean and s.d. over 10 independent simulations. In table **(b)**, we report results on the real PGC2 schizophrenia data assuming population risk of 1% and 0.4%.

From (a), we observe that the N=50K h_g^2 estimates are substantially (11–14%) downward biased, as expected [17, 27], whereas subsampling to N=5K yields h_g^2 estimates with at most faint downward bias. We also see that the downward bias $h_{g\text{–}cc}^2$(50K) / $h_{g\text{–}cc}^2$(5K) worsens with decreasing disease prevalence (i.e., increasing severity of ascertainment), suggesting that prevalence can be inferred from the extent of the bias. However, the observed $h_{g\text{–}cc}^2$(50K) / $h_{g\text{–}cc}^2$(5K) of 0.889 in (b) is within two standard deviations of the mean $h_{g\text{–}cc}^2$(50K) / $h_{g\text{–}cc}^2$(5K) in all our simulations (K=0.001–0.02), indicating that a much larger sample size would be required to make precise inference.

Supplementary Table 7. Minimal cross-tagging of heritability across adjacent 1Mb or 0.5Mb regions.

(a) 1Mb regions

MAF range of causal SNPs	$h_{g,1Mb}^2$, left region	$h_{g,1Mb}^2$, causal region	$h_{g,1Mb}^2$, right region
0.05–0.5	0.004 (0.000)	0.443 (0.001)	0.004 (0.000)
0.01–0.05	0.010 (0.000)	0.440 (0.002)	0.010 (0.001)
0.005–0.01	0.014 (0.001)	0.374 (0.004)	0.014 (0.001)
0.001–0.005	0.008 (0.001)	0.160 (0.004)	0.007 (0.001)

(b) 0.5Mb regions

MAF range of causal SNPs	$h_{g,0.5Mb}^2$, left region	$h_{g,0.5Mb}^2$, causal region	$h_{g,0.5Mb}^2$, right region
0.05–0.5	0.009 (0.000)	0.434 (0.001)	0.009 (0.000)
0.01–0.05	0.027 (0.001)	0.417 (0.002)	0.027 (0.001)
0.005–0.01	0.036 (0.001)	0.357 (0.003)	0.037 (0.001)
0.001–0.005	0.019 (0.001)	0.201 (0.003)	0.020 (0.001)

(a) Using genotypes from 3,413 unrelated UK10K samples, we simulated quantitative phenotypes in which variance-normalized UK10K SNPs in a specific MAF range (5–50%, 1–5%, 0.5–1%, 0.1–0.5%) and in a given 1Mb region were assigned standard normal effect sizes. We then ran BOLT-REML with one variance component containing the PGC2 SNPs in the causal region, one component containing the PGC2 SNPs in the region to the left, and one component containing the PGC2 SNPs in the region to the right. Each reported value is the mean (s.e.m.) over the 2,567 1Mb regions. **(b)** We repeated this experiment using the 5,136 0.5Mb regions containing $\geq$5 SNPs.

The results show that nearly all of the heritability tagged by the PGC2 SNPs is tagged by SNPs in the causal region, with very little "leakage" of heritability into the regions to the left and right (especially for 1Mb regions). As expected, PGC2 SNPs tag most of the heritability when causal SNPs are common (5–50%), whereas PGC2 SNPs do not tag most of the heritability when causal SNPs are rare (0.1–0.5%).

Supplementary Table 8. Minimal cross-tagging of heritability across adjacent regions in samples with heterogeneous ancestry.

(a) 1Mb regions

MAF range of causal SNPs	$h_{g,1Mb}^2$, left region	$h_{g,1Mb}^2$, causal region	$h_{g,1Mb}^2$, right region
0.05–0.5	0.004 (0.000)	0.440 (0.001)	0.004 (0.000)
0.01–0.05	0.009 (0.000)	0.399 (0.002)	0.009 (0.000)
0.005–0.01	0.012 (0.001)	0.264 (0.004)	0.011 (0.001)
0.001–0.005	0.007 (0.000)	0.081 (0.003)	0.007 (0.000)

(b) 0.5Mb regions

MAF range of causal SNPs	$h_{g,0.5Mb}^2$, left region	$h_{g,0.5Mb}^2$, causal region	$h_{g,0.5Mb}^2$, right region
0.05–0.5	0.008 (0.000)	0.433 (0.001)	0.008 (0.000)
0.01–0.05	0.022 (0.001)	0.389 (0.002)	0.022 (0.001)
0.005–0.01	0.026 (0.001)	0.281 (0.003)	0.026 (0.001)
0.001–0.005	0.013 (0.001)	0.114 (0.003)	0.013 (0.001)

These tables are analogs of Supplementary Table 7 in which we restricted the set of UK10K individuals used in simulations to a subset of UK10K individuals with greater heterogeneity in within-Europe genetic ancestry. Specifically, we computed individual-level ancestry fractions (Northwest European, Southeast European, Ashkenazi Jewish) using the SNPweights v2.1 software [56] and restricted our simulation to 1,696 UK10K individuals having the greatest proportions of non-Northwest European ancestry (on average 16%, with a standard deviation of 12%). We observed greater variance in per-chromosome genetic ancestry, indicating that local ancestry differences (i.e., admixture) contribute to the heterogeneity. Again, nearly all of the heritability tagged by the PGC2 SNPs is tagged by SNPs in the causal region, with very little "leakage" of heritability into the regions to the left and right (especially for 1Mb regions). Again, PGC2 SNPs tag most of the heritability when causal SNPs are common (5–50%), whereas PGC2 SNPs do not tag most of the heritability when causal SNPs are rare (0.1–0.5%).

Supplementary Table 9. Autocorrelations of annotations across consecutive 1Mb or 0.5Mb genomic segments.

	PGC2 1Mb	GERA 1Mb	PGC2 0.5Mb	GERA 0.5Mb
GC content	0.80	0.79	0.81	0.81
Genic content	0.30	0.28	0.37	0.39
Replication timing	0.69	0.69	0.85	0.85
Recombination rate	0.57	0.59	0.51	0.54
Background selection	0.61	0.63	0.69	0.70
# of meQTLs	0.61	0.65	0.55	0.58

We filtered out regions with length $<$80% of the nominal region length (e.g., due to region truncation at centromeres and telomeres); these truncated regions accounted for 3-6% of the $\approx$2,600 1Mb regions and $\approx$5,300 0.5Mb regions. The region boundaries and numbers of regions in the PGC2 and GERA analyses were slightly different because PGC2 data was in build 37 coordinates while GERA data was in build 36 coordinates.

Supplementary Table 10. Correlations of per-region SNP-heritability estimates with long-range annotations.

(a) 1Mb regions, regressing out SNP count

	GC content	Genic content	Replic. timing	Recomb. rate	Bkgd. selection	# meQTLs
Schizophrenia	0.093 (p=2E-06)	0.069 (p=0.0005)	0.064 (p=0.001)	0.049 (p=0.01)	-0.042 (p=0.03)	0.047 (p=0.02)
Dyslipidemia	0.089 (p=6E-06)	0.035 (p=0.07)	0.082 (p=2E-05)	0.052 (p=0.008)	-0.013 (p=0.5)	0.039 (p=0.04)
Hypertension	0.116 (p=3E-09)	0.080 (p=4E-05)	0.097 (p=7E-07)	0.088 (p=7E-06)	-0.020 (p=0.3)	0.029 (p=0.1)
Diabetes type 2	0.096 (p=1E-06)	0.069 (p=0.0005)	0.094 (p=2E-06)	0.065 (p=0.0009)	-0.028 (p=0.2)	0.031 (p=0.1)

(b) 1Mb regions, regressing out SNP count and GC content

	Genic content	Replication timing	Recombination rate	Background selection	# meQTLs
Schizophrenia	0.034 (p=0.09)	-0.004 (p=0.8)	0.004 (p=0.9)	-0.026 (p=0.2)	-0.019 (p=0.3)
Dyslipidemia	-0.001 (p=1)	0.027 (p=0.2)	0.005 (p=0.8)	0.003 (p=0.9)	-0.021 (p=0.3)
Hypertension	0.037 (p=0.06)	0.020 (p=0.3)	0.031 (p=0.1)	0.001 (p=0.9)	-0.056 (p=0.005)
Diabetes type 2	0.033 (p=0.09)	0.037 (p=0.06)	0.016 (p=0.4)	-0.011 (p=0.6)	-0.038 (p=0.05)

(c) 0.5Mb regions, regressing out SNP count

	GC content	Genic content	Replic. timing	Recomb. rate	Bkgd. selection	# meQTLs
Schizophrenia	0.077 (p=6E-08)	0.067 (p=2E-06)	0.048 (p=0.0007)	0.053 (p=0.0002)	-0.033 (p=0.02)	0.043 (p=0.002)
Dyslipidemia	0.073 (p=2E-07)	0.024 (p=0.08)	0.060 (p=2E-05)	0.053 (p=0.0001)	-0.007 (p=0.6)	0.029 (p=0.04)
Hypertension	0.085 (p=1E-09)	0.052 (p=0.0002)	0.075 (p=7E-08)	0.071 (p=3E-07)	-0.013 (p=0.3)	0.016 (p=0.2)
Diabetes type 2	0.078 (p=3E-08)	0.033 (p=0.02)	0.061 (p=1E-05)	0.072 (p=3E-07)	-0.003 (p=0.8)	0.028 (p=0.05)

(d) 0.5Mb regions, regressing out SNP count and GC content

	Genic content	Replication timing	Recombination rate	Background selection	# meQTLs
Schizophrenia	0.044 (p=0.002)	-0.008 (p=0.6)	0.021 (p=0.1)	-0.019 (p=0.2)	-0.006 (p=0.7)
Dyslipidemia	0.000 (p=1)	0.013 (p=0.3)	0.019 (p=0.2)	0.005 (p=0.7)	-0.015 (p=0.3)
Hypertension	0.025 (p=0.07)	0.023 (p=0.1)	0.033 (p=0.02)	0.002 (p=0.9)	-0.038 (p=0.006)
Diabetes type 2	0.008 (p=0.6)	0.010 (p=0.5)	0.038 (p=0.006)	0.010 (p=0.5)	-0.019 (p=0.2)

For each region, we computed GC content based on the human reference, genic content as the complement of the "intergenic" annotation of ref. [6], replication timing as the region-wide average of data from ref. [33], recombination rate as the ratio of genetic map distance (according to HapMap phase II [34]) to physical distance spanning each region, and background selection as the region-wide average of data from ref. [35]. We filtered out regions with length $<$80% of the nominal region length (e.g., due to region truncation at centromeres and telomeres); these truncated regions accounted for 3-6% of the $\approx$2,600 1Mb regions **(a,b)** and $\approx$5,300 0.5Mb regions **(c,d)**. (The region boundaries and numbers of regions in the PGC2 and GERA analyses were slightly different because PGC2 data was in build 37 coordinates while GERA data was in build 36 coordinates.) We then computed correlation coefficients between these annotations and per-region SNP-heritability estimates after regressing out SNP count per region **(a,c)** or both SNP count and GC content **(b,d)**. We note that the statistical significance threshold after correcting for multiple hypotheses tested is $p<0.001$.

Supplementary Table 11. Correlations among long-range annotations.

(a) 1Mb regions, regressing out SNP count

	GC content	Genic content	Replic. timing	Recomb. rate	Bkgd. selection	# meQTLs
GC content	1.00	0.40	0.72	0.54	-0.18	0.63
Genic content	0.40	1.00	0.49	0.10	-0.45	0.22
Replication timing	0.72	0.49	1.00	0.25	-0.49	0.40
Recombination rate	0.54	0.10	0.25	1.00	0.35	0.22
Background selection	-0.18	-0.45	-0.49	0.35	1.00	-0.07
# meQTLs	0.63	0.22	0.40	0.22	-0.07	1.00

(b) 0.5Mb regions, regressing out SNP count

	GC content	Genic content	Replic. timing	Recomb. rate	Bkgd. selection	# meQTLs
GC content	1.00	0.33	0.69	0.50	-0.17	0.56
Genic content	0.33	1.00	0.43	0.06	-0.43	0.19
Replication timing	0.69	0.43	1.00	0.21	-0.48	0.36
Recombination rate	0.50	0.06	0.21	1.00	0.34	0.17
Background selection	-0.17	-0.43	-0.48	0.34	1.00	-0.09
# meQTLs	0.56	0.19	0.36	0.17	-0.09	1.00

We report correlation coefficients between pairs of annotations across regions after filtering and regressing out SNP count as described in Supplementary Table 10.

Supplementary Table 12. PGC2 and GERA marker counts per GC quintile.

GC content quintile	# PGC2 markers	# GERA SNPs
1 (lowest GC content)	86671	107591
2	102657	129653
3	105563	138340
4	103925	136804
5 (highest GC content)	73340	85292

We report the total numbers of markers in each GC quintile of PGC2 1Mb regions (based on build 37 coordinates) and GERA 1Mb regions (based on build 36 coordinates): i.e., the 20% of PGC2 1Mb regions with lowest per-region GC content contained 86,671 PGC2 markers. The total marker counts are very slightly less than the sizes of our PGC2 and GERA marker sets because we excluded regions with $<$5 SNPs.

Supplementary Table 13. Comparison of schizophrenia SNP-heritability partitioning using PCGC regression vs. BOLT-REML.

(a) Partitioning across GC content quintiles

	Heritability attributed to GC content quintiles				
Method	1	2	3	4	5
PCGC	0.047 (0.003)	0.052 (0.003)	0.064 (0.004)	0.055 (0.003)	0.066 (0.003)
PCGC	16.6% (1.1%)	18.4% (1.2%)	22.3% (1.3%)	19.4% (1.2%)	23.3% (1.1%)
REML	16.4% (1.0%)	19.0% (1.1%)	21.6% (1.1%)	19.7% (1.1%)	23.1% (1.1%)
REML, 10 PCs per VC	16.4% (1.0%)	19.1% (1.1%)	21.7% (1.1%)	19.7% (1.1%)	23.1% (1.1%)
REML, extra QC	16.6% (1.0%)	20.0% (1.1%)	22.4% (1.2%)	19.7% (1.2%)	21.4% (1.1%)

(b) Partitioning across MAF bins

	Heritability attributed to MAF bins					
Method	0.02–0.05	0.05–0.1	0.1–0.2	0.2–0.3	0.3–0.4	0.4–0.5
PCGC	0.018 (0.005)	0.025 (0.005)	0.055 (0.006)	0.054 (0.006)	0.057 (0.006)	0.047 (0.005)
PCGC	7.2% (2.1%)	9.8% (1.9%)	21.3% (2.4%)	21.0% (2.2%)	22.3% (2.3%)	18.3% (2.0%)
REML	6.6% (1.5%)	8.6% (1.5%)	20.1% (1.9%)	20.9% (1.9%)	23.2% (1.9%)	20.5% (1.7%)
REML, 10 PCs per VC	6.7% (1.5%)	8.6% (1.5%)	20.3% (1.9%)	20.8% (1.9%)	23.1% (1.9%)	20.5% (1.7%)

We compare partitions of schizophrenia h_g^2 across GC quintiles and MAF bins computed using the following techniques:

- PCGC regression. This method is not biased by case-control ascertainment, so we report both absolute and relative (% of h_g^2) estimates. We ran PCGC regression with 10 PCs per variance component (as recommended by ref. [17]) as well as sex and 29 study indicators.
- BOLT-REML. We only report relative (% of h_g^2) estimates, which remain valid despite the downward bias of absolute REML h_g^2 estimates under case-control ascertainment (Supplementary Figures 16–19). REML rows correspond to:
  - BOLT-REML using the set of covariates (sex, 29 study indicators, and 10 genome-wide PCs) and QC procedures used in our main analyses (Online Methods).
  - BOLT-REML using 10 additional PC covariates per variance component.
  - BOLT-REML using our usual set of covariates but run on a subset of 321,567 markers remaining after extremely stringent QC (missing rate <0.01).

All analyses produced consistent results.

Supplementary Table 14. PGC2 marker count and UK10K SNP count per MAF bin.

MAF bin	PGC2 markers	UK10K SNPs	Ratio
0.02-0.05	99607	1315932	13.21
0.05-0.1	78564	1114949	14.19
0.1-0.2	99867	1416648	14.19
0.2-0.3	73460	1060806	14.44
0.3-0.4	62216	930520	14.96
0.4-0.5	58464	867114	14.83

We report the numbers of PGC2 markers and UK10K SNPs in each MAF bin used in our MAF-partitioning analyses (Figure 4 and Supplementary Fig. 13).

Supplementary Table 15. Detailed genetic correlations and total correlations of GERA disease liabilities.

(a) BMI not included as a covariate

	Allergic rhinitis	Asthma	Cardiovasc. dis.	Diabetes type 2	Dyslipidemia	Hypertension	Macular degen.	Osteoarthritis	Osteoporosis
Allergic rhinitis		0.851 (0.101)	-0.037 (0.129)	0.115 (0.089)	0.093 (0.074)	0.231 (0.078)	0.048 (0.111)	0.394 (0.119)	0.075 (0.114)
Asthma	0.470 (0.006)		0.288 (0.098)	0.288 (0.069)	0.175 (0.057)	0.266 (0.059)	0.040 (0.086)	0.345 (0.091)	-0.034 (0.088)
Cardiovasc. dis.	0.054 (0.008)	0.145 (0.008)		0.391 (0.077)	0.356 (0.064)	0.526 (0.067)	-0.007 (0.098)	0.309 (0.106)	0.025 (0.101)
Diabetes type 2	0.040 (0.009)	0.134 (0.008)	0.230 (0.008)		0.425 (0.041)	0.495 (0.042)	-0.051 (0.068)	0.211 (0.073)	-0.176 (0.069)
Dyslipidemia	0.080 (0.007)	0.107 (0.008)	0.278 (0.007)	0.485 (0.010)		0.403 (0.035)	-0.077 (0.057)	0.043 (0.061)	-0.073 (0.058)
Hypertension	0.070 (0.008)	0.136 (0.008)	0.314 (0.006)	0.494 (0.009)	0.406 (0.006)		0.030 (0.058)	0.224 (0.062)	-0.150 (0.060)
Macular degen.	0.018 (0.011)	0.036 (0.011)	0.111 (0.010)	0.043 (0.013)	-0.037 (0.010)	0.047 (0.011)		0.076 (0.091)	-0.073 (0.088)
Osteoarthritis	0.141 (0.007)	0.144 (0.007)	0.120 (0.007)	0.092 (0.009)	0.107 (0.007)	0.137 (0.007)	0.034 (0.010)		-0.167 (0.095)
Osteoporosis	0.022 (0.010)	0.020 (0.011)	0.036 (0.010)	-0.137 (0.014)	-0.029 (0.009)	-0.080 (0.008)	0.089 (0.012)	-0.003 (0.010)	

(b) BMI included as a covariate

	Allergic rhinitis	Asthma	Cardiovasc. dis.	Diabetes type 2	Dyslipidemia	Hypertension	Macular degen.	Osteoarthritis	Osteoporosis
Allergic rhinitis		0.845 (0.109)	-0.122 (0.140)	0.005 (0.100)	0.013 (0.078)	0.131 (0.085)	0.040 (0.114)	0.332 (0.132)	0.145 (0.121)
Asthma	0.463 (0.006)		0.190 (0.109)	0.141 (0.079)	0.068 (0.062)	0.112 (0.067)	0.028 (0.090)	0.226 (0.104)	0.077 (0.096)
Cardiovasc. dis.	0.042 (0.008)	0.127 (0.008)		0.268 (0.088)	0.272 (0.069)	0.427 (0.075)	-0.018 (0.103)	0.195 (0.119)	0.138 (0.109)
Diabetes type 2	0.017 (0.009)	0.074 (0.009)	0.200 (0.008)		0.338 (0.048)	0.350 (0.052)	-0.087 (0.074)	0.020 (0.085)	-0.051 (0.078)
Dyslipidemia	0.056 (0.007)	0.059 (0.008)	0.252 (0.007)	0.387 (0.009)		0.309 (0.040)	-0.101 (0.058)	-0.089 (0.068)	0.002 (0.061)
Hypertension	0.046 (0.008)	0.075 (0.008)	0.286 (0.007)	0.363 (0.009)	0.334 (0.006)		0.002 (0.063)	0.037 (0.072)	-0.034 (0.066)
Macular degen.	0.018 (0.011)	0.039 (0.012)	0.112 (0.010)	0.056 (0.013)	-0.032 (0.010)	0.058 (0.010)		0.059 (0.098)	-0.060 (0.090)
Osteoarthritis	0.127 (0.007)	0.115 (0.007)	0.101 (0.007)	0.017 (0.008)	0.055 (0.007)	0.070 (0.007)	0.039 (0.010)		-0.065 (0.105)
Osteoporosis	0.035 (0.009)	0.055 (0.010)	0.054 (0.010)	-0.037 (0.013)	0.026 (0.009)	0.001 (0.009)	0.085 (0.012)	0.035 (0.009)	

(This table provides numeric data plotted in Figure 5 with additional details.) (**a**) Correlations from bivariate analyses using only age, sex, 10 principal components, and Affymetrix kit type as covariates. (**b**) Correlations from bivariate analyses including BMI as an additional covariate. Genetic correlations are above the diagonals; total liability correlations are below the diagonals.

Supplementary Table 16. Effect of sample size on precision of SNP-heritability partitioning using REML.

(a) Partitioning across GC content quintiles

	Heritability attributed to GC content quintiles				
N	1	2	3	4	5
50K	16.4% (1.0%)	19.0% (1.1%)	21.6% (1.1%)	19.7% (1.1%)	23.1% (1.1%)
25K	17.4% (1.6%)	18.6% (1.7%)	21.3% (1.8%)	19.8% (1.8%)	22.8% (1.8%)
10K	11.6% (3.1%)	22.5% (3.7%)	25.7% (3.8%)	17.7% (3.8%)	22.4% (3.6%)

(b) Partitioning across MAF bins

	Heritability attributed to MAF bins					
N	0.02–0.05	0.05–0.1	0.1–0.2	0.2–0.3	0.3–0.4	0.4–0.5
50K	6.6% (1.5%)	8.6% (1.5%)	20.1% (1.9%)	20.9% (1.9%)	23.2% (1.9%)	20.5% (1.7%)
25K	5.7% (2.7%)	6.2% (2.7%)	21.6% (3.2%)	22.7% (3.2%)	23.9% (3.2%)	19.9% (2.8%)
10K	7.1% (6.1%)	12.0% (6.0%)	23.5% (7.0%)	25.6% (6.7%)	8.7% (6.5%)	23.1% (5.7%)

We compare partitions of schizophrenia ${h_g}^2$ across GC quintiles and MAF bins computed by running BOLT-REML on the full PGC2 data set we analyzed (N=50K) as well as one random N=25K subsample and one random N=10K subsample. Standard errors decrease with increasing sample size, as expected.

Supplementary Table 17. Accuracy of BOLT-REML and GCTA heritability analysis in simulations.

(a) BOLT-REML, 15 Monte Carlo trials

Variance component	True h_g^2	Mean est. h_g^2(s.e.m.)	RMSE (s.e.)	Mean reported SE (s.e.m)
Chromosome 21	0.2	0.1998 (0.0001)	0.0322 (0.0001)	0.0342 (0.0000)
Chromosome 22	0.3	0.2995 (0.0001)	0.0312 (0.0001)	0.0347 (0.0000)

(b) BOLT-REML, 100 Monte Carlo trials

Variance component	True h_g^2	Mean est. h_g^2(s.e.m.)	RMSE (s.e.)	Mean reported SE (s.e.m)
Chromosome 21	0.2	0.1997 (0.0001)	0.0312 (0.0001)	0.0333 (0.0000)
Chromosome 22	0.3	0.2995 (0.0001)	0.0301 (0.0001)	0.0338 (0.0000)

(c) GCTA

Variance component	True h_g^2	Mean est. h_g^2(s.e.m.)	RMSE (s.e.)	Mean reported SE (s.e.m)
Chromosome 21	0.2	0.1995 (0.0001)	0.0310 (0.0001)	0.0331 (0.0000)
Chromosome 22	0.3	0.2997 (0.0001)	0.0299 (0.0001)	0.0336 (0.0000)

Simulations used genotypes from N=2,000 GERA individuals and M=13,971 SNPs on chromosomes 21 and 22. In each of 50,000 simulation replicates, we simulated a quantitative trait with 20% of variance explained by SNPs on chromosome 21 and 30% of variance explained by SNPs on chromosome 22. We then performed REML analysis using one variance component per chromosome. We ran BOLT-REML using 15 Monte Carlo trials **(a)**, BOLT-REML using 100 Monte Carlo trials **(b)**, and GCTA **(c)**.

We observed that all methods computed accurate estimates with negligible bias. We expected from theory (Supplementary Note) that BOLT-REML estimates using S Monte Carlo trials would have variance 1+1/S times larger than exact REML estimates (e.g., computed using GCTA), and our simulations confirmed that indeed, to within the statistical resolution available from 50,000 replicates, the RMSE of BOLT-REML was greater than that of GCTA by a factor of $\sqrt{1+1/S}$. (We also observed that in this particular simulation setup, standard errors reported by REML—computed using analytic approximations based on Fisher information—were slightly conservative compared to observed RMSE.)

Supplementary Table 18. Accuracy of BOLT-REML heritability analysis in bivariate quantitative trait simulations.

Var comp	Parameter	True value	Mean est (s.e.m.)	RMSE (s.e.)	Mean reported SE (s.e.m)
Chr 1	Trait 1 $h_{\text{g,chr}}^2$	0.15	0.144 (0.001)	0.014 (0.001)	0.012 (0.000)
	Gen corr r_{g}	0.4	0.408 (0.003)	0.032 (0.002)	0.036 (0.000)
	Trait 2 $h_{\text{g,chr}}^2$	0.6	0.603 (0.001)	0.015 (0.001)	0.015 (0.000)
Chr 2	Trait 1 $h_{\text{g,chr}}^2$	0.6	0.597 (0.001)	0.014 (0.001)	0.015 (0.000)
	Gen corr r_{g}	-0.2	-0.195 (0.004)	0.037 (0.003)	0.043 (0.000)
	Trait 2 $h_{\text{g,chr}}^2$	0.15	0.150 (0.001)	0.012 (0.001)	0.012 (0.000)
Resid	Trait 1 h_{e}^2	0.25	0.259 (0.002)	0.020 (0.001)	0.016 (0.000)
	Resid corr r_{e}	0.6	0.585 (0.004)	0.040 (0.003)	0.037 (0.000)
	Trait 2 h_{e}^2	0.25	0.246 (0.002)	0.017 (0.001)	0.016 (0.000)

Simulations used genotypes from N=10,000 GERA individuals and M=96,543 SNPs on chromosomes 1 and 2. In each of 100 simulation replicates, we simulated a pair of correlated quantitative traits covariance given by

$$\text{Cov}\begin{pmatrix} y_{\text{trait1}} \\ y_{\text{trait2}} \end{pmatrix} = \begin{pmatrix} 0.15 & 0.12 \\ 0.12 & 0.6 \end{pmatrix} \otimes G_{\text{chr1}} + \begin{pmatrix} 0.6 & -0.06 \\ -0.06 & 0.15 \end{pmatrix} \otimes G_{\text{chr2}} + \begin{pmatrix} 0.25 & 0.15 \\ 0.15 & 0.25 \end{pmatrix} \otimes I_N,$$

where G_{chr1} and G_{chr2} denote genetic relationship matrices computed using SNPs on chromosomes 1 and 2, respectively, and I_N denotes the $N \times N$ identity matrix (corresponding to environmental effects). We then estimated these variance parameters by running bivariate BOLT-REML (with 15 Monte Carlo trials) using one variance component per chromosome.

Supplementary Table 19. Accuracy of BOLT-REML heritability analysis in bivariate case-control trait simulations.

			Case-control	Liability transform		
VC	Parameter	True	Mean raw est	Mean estimate	RMSE	Mean reported SE
Chr 1	Trait 1 $h_{g,chr}^2$	0.15	0.089 (0.001)	0.139 (0.002)	0.022 (0.001)	0.021 (0.000)
	Gen corr r_g	0.4	0.441 (0.006)	0.441 (0.006)	0.075 (0.004)	0.070 (0.001)
	Trait 2 $h_{g,chr}^2$	0.6	0.270 (0.002)	0.614 (0.004)	0.040 (0.003)	0.039 (0.000)
Chr 2	Trait 1 $h_{g,chr}^2$	0.6	0.379 (0.002)	0.596 (0.003)	0.025 (0.002)	0.027 (0.000)
	Gen corr r_g	-0.2	-0.195 (0.006)	-0.195 (0.006)	0.063 (0.004)	0.079 (0.001)
	Trait 2 $h_{g,chr}^2$	0.15	0.068 (0.001)	0.155 (0.003)	0.032 (0.002)	0.031 (0.000)
Resid	Trait 1 h_e^2	0.25	0.532 (0.002)	0.265 (0.003)	0.035 (0.002)	0.034 (0.000)
	Resid corr r_e	0.6	0.123 (0.003)	0.568 (0.014)	0.139 (0.013)	0.130 (0.002)
	Trait 2 h_e^2	0.25	0.662 (0.002)	0.231 (0.005)	0.053 (0.004)	0.049 (0.000)

We simulated 100 pairs of correlated quantitative traits exactly as in Supplementary Table 18 and then converted each trait to a case-control trait according to a liability threshold model. Specifically, for each pair, we used a case cutoff of $z > 0$ for trait 1 (corresponding to 50% population prevalence) and a case cutoff of $z > 1$ for trait 2 (corresponding to 16% population prevalence). We then estimated the variance parameters of the underlying liabilities (i.e., the original quantitative traits) by running bivariate BOLT-REML (with 15 Monte Carlo trials) using one variance component per chromosome and transforming parameter estimates to the liability scale. Explicitly, we used the formula of ref. [3] to transform $h_{g,chr}^2$; we did not transform r_g (ref. [7]); and we transformed total phenotypic correlation to total liability correlation as described in Online Methods, after which we computed r_e by subtracting the two genetic covariances from the total liability covariance.