nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00901-3

The distribution of common-variant effect sizes

Luke J. O'Connor ✉

The genetic effect-size distribution of a disease describes the number of risk variants, the range of their effect sizes and sample sizes that will be required to discover them. Accurate estimation has been a challenge. Here I propose Fourier Mixture Regression (FMR), validating that it accurately estimates real and simulated effect-size distributions. Applied to summary statistics for ten diseases (average $N_{eff} = 169,000$), FMR estimates that 100,000–1,000,000 cases will be required for genome-wide significant SNPs to explain 50% of SNP heritability. In such large studies, genome-wide significance becomes increasingly conservative, and less stringent thresholds achieve high true positive rates if confounding is controlled. Across traits, polygenicity varies, but the range of their effect sizes is similar. Compared with effect sizes in the top 10% of heritability, including most discovered thus far, those in the bottom 10–50% are orders of magnitude smaller and more numerous, spanning a large fraction of the genome.

Genome-wide association studies (GWAS) have detected thousands of disease-associated loci[1–4], yet they remain far from saturation: significant associations explain a fraction of SNP heritability[5], polygenic risk scores capture a fraction of heritable risk[6], and statistical power remains an obstacle for downstream analyses such as fine mapping[7–9] and heritability partitioning[10–12]. What would it take for every genetic association to be discovered? What fraction of the genome would be associated, and how would effect sizes of newly discovered SNPs compare with those of SNPs discovered thus far?

These are questions about the genetic effect-size distribution, which describes how heritability accumulates across SNPs with different effect sizes[1,13–15]. Various statistical methods estimate this distribution. Park et al.[16] used cross-validation to estimate the number of loci with effect sizes similar to those already discovered. Palla and Dudbridge[17] fit a two-component mixture model, which assumes that all non-null loci have similar effect sizes (following a normal distribution). Similar models have been used in several applications[13,18–21]. Moser et al.[22] and Zhang et al.[14] fit more flexible models, permitting both small- and large-effect SNPs. O'Connor et al. quantified polygenicity nonparametrically[15]. A related but distinct objective is heritability partitioning, either across annotations[10–12], regions[23,24] or fine-mapped SNPs[25].

I introduce FMR, a method-of-moments estimator of the common-variant effect-size distribution. Similar to linkage disequilibrium (LD) score regression (LDSC)[10,26], FMR operates on GWAS summary statistics, and it reduces the estimation problem to linear regression. Whereas LDSC coefficients quantify the heritability explained by different annotations, FMR coefficients quantify the heritability explained by components of a mixture model for the effect-size distribution.

Results

Definition of the effect-size distribution. Several studies[14,18,22] have sought to estimate the distribution of causal effect sizes: the phenotypic effect of changing a single allele, without changing other alleles in the same haplotype (in practice, 'causal' effect sizes can be confounded by population stratification, biased ascertainment and family-related effects[27–29]). Causal effect sizes of individual SNPs are estimated in fine-mapping studies[7–9,25,30,31], and they differ from marginal effect sizes or associations, which are duplicated across tag SNPs in high LD with one that is causal. Because of duplication, the distribution of marginal effect sizes across SNPs does not determine the number of associated loci or the proportion of heritability that will be explained by GWAS, and it seems necessary to estimate the causal effect-size distribution, a challenging problem. However, power calculations are actually determined by marginal effect sizes.

Instead of counting the number of SNPs with various marginal effect sizes and potentially overcounting tag SNPs, a different approach is to measure the heritability that they explain. The heritability distribution of marginal effect sizes (HDM) is the distribution of marginal effect sizes over proportions of heritability: for example, 50% of heritability is explained by SNPs with effect sizes up to the median of the HDM, and 90% of heritability is explained by SNPs up to the 90th percentile. Effect sizes are also measured in units of variance explained (which is larger on average for common SNPs). This definition avoids overcounting tag SNPs, as the number of non-causal tag SNPs has no effect on the amount of heritability explained by a locus. Although the HDM is defined under a random-effect model (Methods), reported estimates are converted into proportions of fixed-effect SNP heritability[24], which are more readily interpreted. Genome-wide significant SNPs explain a larger fraction of fixed-effect heritability (Methods).

Overview of FMR. To estimate the HDM, I developed FMR. Similar to stratified (S)-LDSC[10,26], it involves regressing a function of GWAS summary statistics on SNP-specific scores computed from an LD reference panel. Whereas S-LDSC partitions heritability across annotations via multiple regression on annotation-specific LD scores[10], FMR partitions heritability across mixture components via multiple regression on component-specific 'Fourier LD scores' that depend on the effect-size distribution of each mixture component.

FMR relies on the Fourier transform, which maps between a time-varying signal (for example, sound waves) and a distribution of frequencies (notes). The Fourier transform of a probability density function is called a characteristic function (CF), denoted as $\phi(t)$. Suppose the HDM is a mixture distribution for which mixture components have CFs $\Phi_1,\ldots,\Phi_K$ and weights (proportions of heritability) $w_1, \ldots, w_K$. The FMR regression equation is

$$\frac{d}{dt}E\left(e^{itz+\frac{1}{2}\chi_0^2t^2}\right) \approx itN\overline{\sigma^2}\sum_K w_K\ell_K(t)\,,\ \ell_K(t) = \sum_j r_j^2\phi_K\left(r_jt\right). \quad (1)$$

Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ✉e-mail: loconnor@broadinstitute.org

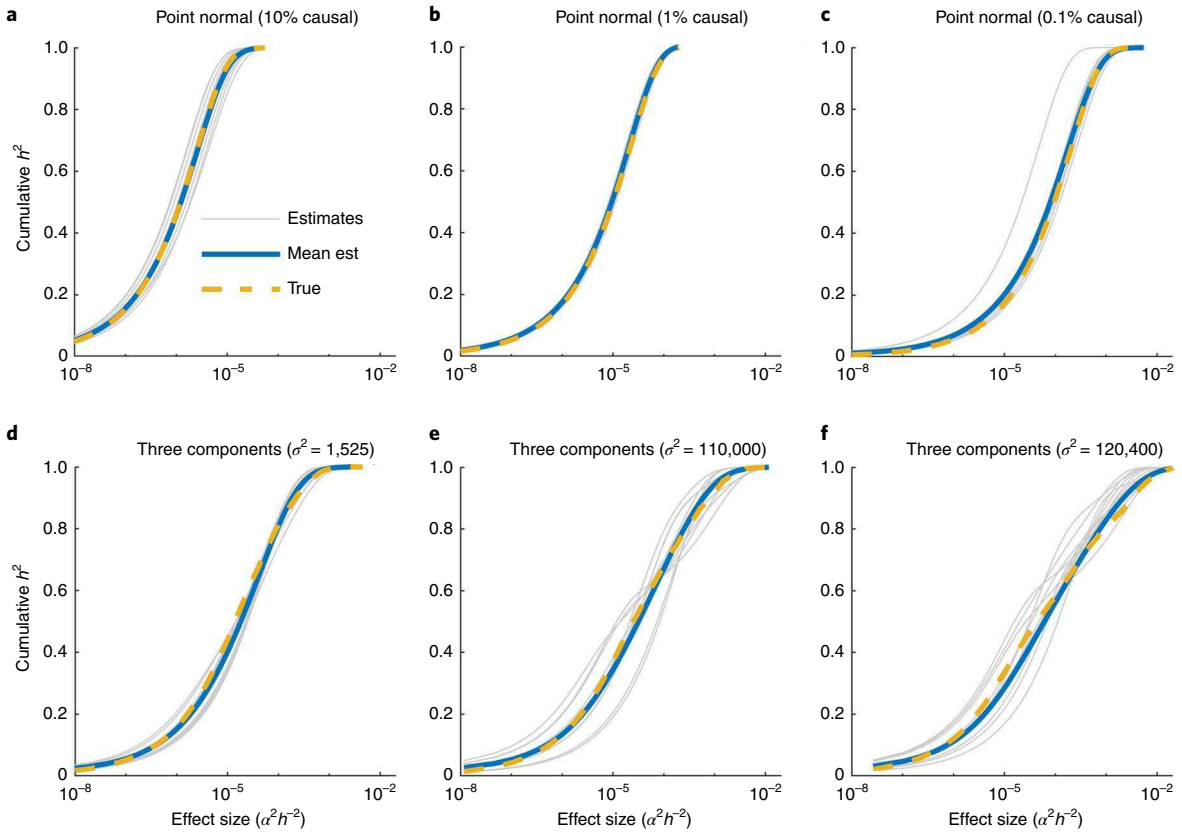


Fig. 1 | Performance of Fourier regression in well-powered simulations ($N = 460{,}000$) with real LD. Yellow, blue and gray curves indicate the true HDM, the mean estimated (est) HDM across 100 replicates and ten individual estimates (chosen at random). **a**–**c**, Point-normal genetic architecture with different proportions of causal SNPs. **d**–**f**, Gaussian mixture models with small-, medium- and large-effect SNPs. Each component explains one-third of heritability, and relative effect-size variances of the large- versus small-effect components are 25×, 100× and 400× in **d**–**f**, respectively.

The left-hand side is the derivative of the CF of the GWAS z score with noise correction. It depends on the sampling time t and the LDSC intercept (χ_0^2)[26]. The right-hand side is a linear combination of Fourier LD scores, $\ell_k(t)$, one for each mixture component, with coefficients proportional to mixture weights. r_j is the LD coefficient between the regression SNP and SNP j. N is the effective GWAS sample size, $\overline{\sigma^2}$ is the average per-SNP heritability, and i is the imaginary unit (both sides of the equation are imaginary). The equation relies on an LD approximation, roughly that LD between causal SNPs is either strong or weak (Methods).

In brief, FMR involves the following steps (see Methods and Supplementary Note for details): first, a K-component mixture model is specified, and Fourier LD scores for mixture components are computed at K convenient sampling times. I used 13 mixture components with variance parameters spanning a wide range of effect sizes (a factor of 4,000 difference between the largest and smallest effect-size variances). The human leukocyte antigen (HLA) region is excluded. Second, the left-hand side of equation (3) is evaluated at K sampling times for all M regression SNPs, and these MK values are regressed on the corresponding Fourier LD scores. Additional moment equations from LDSC and LD fourth-moment regression are also included[15,26]. Regression weights are chosen using a heuristic. The regression is constrained to produce non-negative coefficients. Third, the estimated mixture weights are converted into estimates of various quantities. Open-source software for FMR is available (Code availability).

Performance of FMR in simulations. I evaluated FMR in simulations using real LD from UK Biobank typed SNPs ($M = 455{,}000$ SNPs), generating effect sizes from two types of models. First, I simulated from the commonly used point-normal model[17,20,32], which is parameterized by the fraction of causal SNPs, M_c/M. At larger values of M_c/M, the HDM cumulative distribution function is shifted to the right (Fig. 1a–c). I applied FMR to summary statistics generated under these models at $N = 460{,}000$ and heritability $h^2 = 0.1$ (similar to UK Biobank; see below) and determined that it produces unbiased and robust estimates of the HDM across all values of M_c/M (Fig. 1a–c).

Second, I simulated from a mixture model including small-, medium- and large-effect SNPs. Each non-null mixture component explained one-third of heritability. Unlike the point-normal model, the width of this distribution can vary, depending on the relative variance of the three non-null components. I simulated a 25×, 100× or 400× difference between the small- and large-effect components (Fig. 1d–f). Under the 400× model (Fig. 1f), the cumulative distribution function (CDF) has a shallow slope, and heritability is dispersed across a wide range of effect sizes, in contrast to Fig. 1a–c, where most heritability is explained by SNPs with similar effect sizes. FMR produced approximately unbiased estimates under these models, although its estimates were noisier than those produced under the point-normal model.

I performed three secondary analyses. First, I performed simulations under the point-normal model at smaller GWAS sample sizes (Extended Data Fig. 1). The z score of the mean of the HDM (which is estimated as a preprocessing step using LD fourth-moment regression[15]) is an indicator for whether FMR is well powered; analyses of real traits below are restricted to traits with z scores greater than 2. Second, I investigated the calibration of FMR standard errors

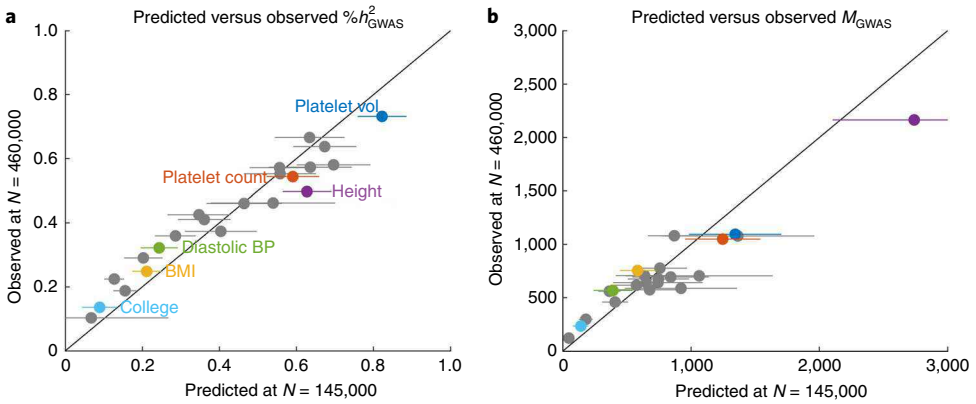


Fig. 2 | Predicted versus observed GWAS results in UK Biobank. FMR was applied to summary statistics from the interim release (N=145,000) to predict the results of the full release (N=460,000). **a**, Proportion of heritability explained by genome-wide significant SNPs ($\%h^2_{GWAS}$). BP, blood pressure; vol, volume. **b**, Number of genome-wide significant loci (M_{GWAS}). Error bars indicate 95% CIs. For numerical values, see Supplementary Table 1.

(estimated using a block jackknife). Estimated standard errors were accurate or conservative at N=50,000–145,000, but they were underestimated at N=460,000, in particular for the proportion of heritability explained by small-effect SNPs (Extended Data Fig. 2). Third, I performed simulations with different numbers of mixture components and sampling times (Extended Data Fig. 3). FMR obtained similar estimates under most parameter settings, except when there were too few sampling times relative to the number of mixture components (Extended Data Fig. 3d).

Performance of FMR predictions in UK Biobank. I applied FMR to summary statistics for 24 complex traits from the UK Biobank interim release (maximum N=145,000) to predict the results of the full UK Biobank release (maximum N=460,000). Predictions were generated for the number of genome-wide significant loci (M_{GWAS}) and the proportion of SNP heritability that they explain ($\%h^2_{GWAS}$) (Supplementary Note). I compared these predictions with observed values (computed using pruning and thresholding; Methods), which were much larger at N=460,000 than at N=145,000 (Extended Data Fig. 4 and Supplementary Table 2). FMR predictions were approximately unbiased (mean 0.42 versus 0.43 for $\%h^2_{GWAS}$, 786 versus 731 for M_{GWAS}), and they were highly correlated with observed values (Fig. 2; r^2=0.94 for $\%h^2_{GWAS}$, r^2=0.92 for M_{GWAS}), indicating that FMR can be used to project the results of future GWAS.

Predictions remained accurate at increasingly stringent significance thresholds (χ^2>100, 300, 1,000) (Extended Data Figs. 5 and 6), indicating that FMR accurately estimates the right tail of the effect-size distribution. The same analysis cannot be used for less stringent significance thresholds because the observed values would be biased due to false positives and winner's curse[25,33]. Conceivably, FMR estimates of the left tail of the effect-size distribution could be biased in a sample size-dependent manner, due to limited power to resolve small effects from zero (although this was not observed in simulations, for example, Fig. 1f). However, FMR estimates of the left tail were nearly identical at N=145,000 versus at N=460,000 (Extended Data Fig. 7). Moreover, a modified version of FMR, FMR-noLD, produced accurate predictions of N=460,000 quantile–quantile plots (Supplementary Fig. 1 and Supplementary Note).

I compared FMR predictions with the state-of-the-art method GENESIS[14], which models the distribution of causal effect sizes across SNPs using a three-component Gaussian mixture model, which is more realistic than the commonly used point-normal (two-component) model[13,17,18,20]. Applied to interim-release UK Biobank summary statistics, GENESIS produced predictions that were highly correlated with the true $\%h^2_{GWAS}$ at N=460,000, similar to FMR; unlike FMR, however, these predictions were upwardly biased by approximately a third (Extended Data Fig. 8 and Supplementary Table 1). This bias could result from model misspecification (Extended Data Fig. 9 and Supplementary Table 3). An additional advantage of FMR is its computational efficiency; it runs in less than a minute, while GENESIS requires more than a day.

FMR models population stratification and cryptic relatedness as excess noise in effect-size estimates. Similar to LDSC, it assumes that this noise is uncorrelated with LD scores (unlike table signal)[26], and, similar to LD fourth-moment regression, it assumes that stratification effects follow a normal distribution[15]. This assumption is expected to hold when stratification arises due to genetic drift[34], and the observed distribution of PC loadings is consistent with normally distributed effects at neutral loci[34]. As a stress test, I applied FMR to summary statistics for height from the Genetic Investigation of Anthropocentric Traits (GIANT) consortium[2] (2010 release, N=131,000). Unlike UK Biobank summary statistics (computed using BOLT-LMM[35,36]), these data contain strong, uncorrected population stratification that has led to confounding results in studies of polygenic selection[37,38]. However, FMR produced nearly identical estimates on the two datasets (Extended Data Fig. 10), indicating that it is not confounded by population stratification within European populations. An additional source of potential confounding in GWAS is assortative and non-random mating, and simulations indicate that non-random mating could distort marginal effect-size distributions and affect FMR (Supplementary Fig. 2).

Sample size targets across 32 diseases and complex traits. I applied FMR to publicly available summary statistics for 32 diseases and complex traits, including 22 UK Biobank traits and ten European-ancestry meta-analyses (Supplementary Table 4). FMR estimates of $\%h^2_{GWAS}$ and M_{GWAS} were concordant with observed values at current sample sizes (Supplementary Figs. 3 and 4 and Supplementary Table 5). FMR predictions varied widely across traits (Fig. 3a and Supplementary Table 6). Assuming a large number of controls, the number of cases required for genome-wide significant SNPs to explain 50% of disease heritability ranged from approximately 100,000 for inflammatory bowel disease (IBD) and hypothyroidism to several hundred thousand for schizophrenia and coronary artery disease (CAD) and millions for Alzheimer's disease (which has low SNP heritability). The number of loci that will be discovered at these sample sizes ranges from a few hundred to several thousand for most traits (Fig. 3b).

To increase $\%h^2_{GWAS}$ from 50% to 90%, a massive increase in sample size is needed: approximately 10× for all traits studied

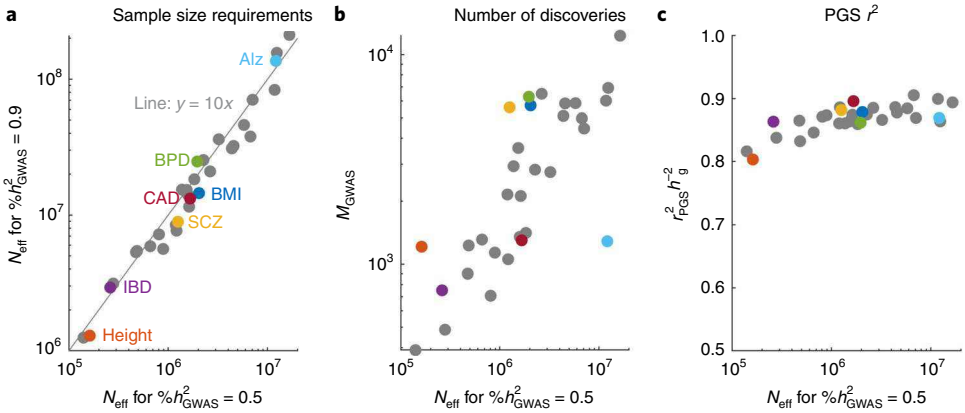


Fig. 3 | Sample size requirements and predictions for future GWAS. a, Sample size required for genome-wide significant SNPs to explain 50% or 90% of SNP heritability. Alz, Alzheimer's disease; BPD, borderline personality disorder; SCZ, schizophrenia. **b**, Predicted number of discoveries at the sample size for which $\%h^2_{GWAS} = 0.5$. **c**, Theoretical maximum PGS accuracy, defined as $r^2_{PGS}(h^2_g)^{-1}$, at the sample size for which $\%h^2_{GWAS} = 0.5$. Existing PGS methods may perform less well than the maximum. For diseases, N_{eff} is twice the harmonic mean of N_{case} and $N_{control}$, and r^2_{PGS} is defined on a liability scale. Numerical results are presented in Supplementary Table 7.

(Fig. 3a). This difference implies that GWAS will have diminishing returns (in $\Delta\%h^2_{GWAS}(\Delta\log(N))^{-1}$) well before $\%h^2_{GWAS}$ nears 100% (Discussion). The number of loci discovered does not diminish, as the number of loci required to explain 90% of heritability is 10× the number required to explain 50% of heritability (Supplementary Table 6). However, distinct loci will become difficult to count, as a substantial fraction of all common SNPs will be genome-wide significant. When $\%h^2_{GWAS}$ is 50%, the fraction of significant common SNPs will be 0.01–0.15 for most traits (Supplementary Fig. 5a), and this fraction increases to 0.1–0.5 when 90% of heritability is explained (Supplementary Fig. 5b).

I derived an upper bound for the prediction r^2 of any polygenic score (PGS) as a function of the HDM (Methods). This bound is expected to accurately predict the within-population performance of an optimal PGS method in large samples, although existing PGS methods may not actually approach it. According to this bound, when $\%h^2_{GWAS}$ reaches 50%, an optimal PGS would explain 80–90% of SNP heritability (Fig. 3c). Four of the analyzed UK Biobank traits already have $\%h^2_{GWAS} \approx 50\%$, and a previous study[35] reported the PGS accuracy of BOLT-LMM[36] trained on these traits (BOLT-LMM improves association power by conditioning on a PGS that assumes a two-component Gaussian mixture model). Reported r^2 values were smaller than the predicted optimum (Supplementary Table 8). This difference indicates that existing PGS methods have substantial room for improvement, possibly because a two-component model does not approximate real effect-size distributions (see below).

Predicted sample size requirements are subject to two notable sources of uncertainty. First, heritability may differ across different studies of the same disease, due to different age distributions, different environments and different patterns of assortative mating. Mixed models also lead to increased effective heritability[35,36,39]. Increasing the study heritability has the same effect on power as increasing the sample size; requirements for the product of effective sample size and heritability are reported in Supplementary Table 7. Second, uncorrected population stratification and cryptic relatedness increase $\%h^2_{GWAS}$, and recalibration using genomic control[40] or the LDSC intercept[26] decreases it. FMR projections assume that the amount of uncorrected population stratification, in particular, the LDSC intercept minus one, will remain constant at increasing sample size and that there will be no recalibration.

True positive rates of non-significant loci. Genome-wide significance is a conservative threshold, and sub-GWAS loci may represent true associations[1]. To quantify true positive rates at different significance thresholds, I define the not-by-chance true positive rate (NTPR) for putatively associated loci: out of all lead SNPs exceeding a threshold, the proportion that are true positives with a correctly estimated direction of effect, not counting 'true-by-chance positives' for which the estimated direction could just as well have been the opposite (Methods). True positive lead SNPs may not be causal; the NTPR quantifies the fraction of loci that are genuinely associated. The NTPR is related to the false sign rate[41], and it can be estimated using FMR (Supplementary Note).

The NTPR of nominally significant loci ($\chi^2 > 4$, roughly $P < 0.05$) would be nearly zero in an underpowered GWAS, as thousands of false positives are expected. However, current GWAS are powered to detect a vast number of true positives at this threshold; for height and body mass index (BMI), they even outnumber false positives, and the NTPR is >50% (Fig. 4a). For most other traits, the NTPR of nominally significant loci is between 0.1 and 0.5. The proportion of heritability explained by these SNPs is only 57% on average; even in well-powered studies, close to half of SNP heritability can be explained by SNPs for which effect sizes are not even nominally significant. These estimates underscore the tension between large sample size (which allows GWAS to detect numerous small-effect SNPs) and extreme polygenicity (which causes missing heritability to persist despite ever-larger studies).

The NTPR increases quickly with the significance threshold. For most traits, it was greater than 50% at a χ^2 threshold of ten (Fig. 4b), greater than 95% at a threshold of 20 (Fig. 4c) and nearly 100% for genome-wide significant SNPs ($\chi^2 > 30$) (Fig. 4d). Well-powered studies have larger NPTR values at the same significance threshold, and NTPR increases as a function of sample size. However, just as sample size magnifies true signal, it also magnifies spurious signals due to uncorrected population stratification and cryptic relatedness; these estimates assume that confounders are controlled.

I calculated the χ^2 threshold that corresponds to an NTPR of 99%. Across most traits, this number varied from 14 to 28, and SNPs exceeding this threshold explained 10–50% more heritability than those with $\chi^2 > 30$, with a 30–200% increase in the number of loci (Supplementary Fig. 6a,b). The difference was most pronounced for well-powered, extremely polygenic traits, such as BMI. When genome-wide significant SNPs explain 50% of heritability, the 99% NTPR χ^2 threshold will fall to between ten and 20, and SNPs exceeding this threshold will explain 60–75% of heritability (Supplementary Fig. 6c).

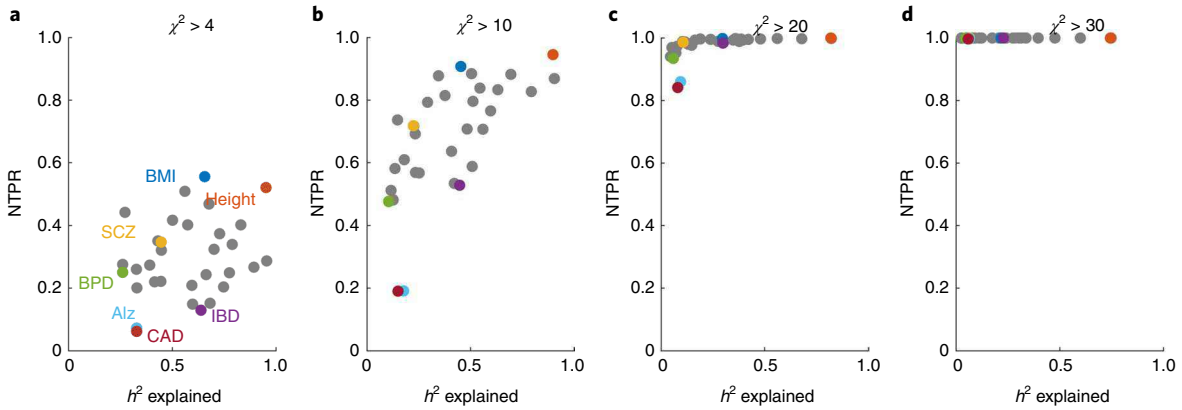


Fig. 4 | The NTPR. At the same significance threshold, the NTPR is greater for well-powered traits, as measured by the proportion of SNP heritability explained by SNPs exceeding that threshold. A χ^2 threshold of four corresponds to nominal significance (**a**), and a threshold of 30 corresponds to genome-wide significance (**d**), while thresholds of ten (**b**) and 20 (**c**) are intermediate. Numerical results are presented in Supplementary Table 9.

The common-variant effect-size distribution. I estimated the distribution of heritability across common SNPs with different effect sizes. The median of this distribution is a measure of polygenicity: it describes the minimum number of loci required to explain 50% of heritability, as well as the per-SNP heritability of a SNP at the 50th percentile. The requisite number of loci ranges from a few hundred to several thousand, and the median effect size is inversely proportional (across 25 well-powered traits; Supplementary Note) (Fig. 5a). Psychiatric and brain-related traits had the greatest polygenicity, consistent with previous estimates[15]. These estimates exclude the major histocompatibility complex region.

At any level of polygenicity, heritability could be concentrated within a narrow range of effect sizes, or it could be dispersed across a wide range of SNPs with effect sizes both large and small. I meta-analyzed quantiles of heritability across all 32 traits (Supplementary Note) (Fig. 5b). In the meta-analysis, there was a factor of ten difference in per-SNP variance between SNPs in the tenth versus 50th and 50th versus 90th percentiles, totaling a factor of 100 difference between the tenth versus 90th percentiles (95% confidence interval (CI), 80–130). These estimates are consistent with the factor of ten difference in sample size required to explain 90% versus 50% of heritability (Fig. 3a). The fifth versus 95th percentiles differed by a factor of ~500 (95% CI, 320–800).

The wide range of effect sizes was similar across traits. The difference in per-SNP variance between the tenth versus 90th percentiles of heritability was ~50–200× for all well-powered traits (Fig. 5c and Supplementary Table 12). This similarity contrasts with order-of-magnitude differences in polygenicity that were observed (Fig. 5a) (Discussion).

Commonly, effect sizes are modeled using a point-normal distribution[13,18,20,21,42], even though it was shown to fit poorly for many traits[14]. Ignoring possible LD between causal SNPs, this model implies that per-SNP heritability in the tenth versus 90th percentiles of heritability differs by only 11× (Supplementary Note); the point-normal misspecifies the width of the effect-size distribution by an entire order of magnitude. An alternative distribution is the log-normal model, where $\log(\alpha^2)$ follows a normal distribution over fixed-effects proportions of heritability. Its mean is inversely related to polygenicity, and its variance quantifies the width of the effect-size distribution. It has convenient mathematical properties (Supplementary Note). Fitting the log-normal model to observed effect-size distributions, with a trait-specific mean and a constant variance parameter, it provided an excellent fit in the meta-analysis (Fig. 5b) and an adequate fit for individual traits (Supplementary Fig. 8).

Discussion

These results suggest plausible sample size targets for future GWAS. For genome-wide significant SNPs to explain 50% of disease heritability in populations of European ancestry, hundreds of thousands of cases will be required for most diseases (Fig. 3a). When these sample sizes are reached, thousands of additional loci below the genome-wide significance threshold will be probable true positives (Supplementary Fig. 6 and Supplementary Table 10), and polygenic risk scores will explain up to 90% of risk due to common variants (Fig. 3c). On the other hand, millions of samples would be required to increase $\%h^2_{GWAS}$ to 90% (Fig. 3a), and, for a disease at ~1% prevalence, this number could exceed the number of cases in Europe. Significant SNPs would span much of the genome (Fig. 3b and Supplementary Fig. 5), limiting their value. Extremely large sample sizes may be more useful in sequencing studies[43] and in studies of diverse populations[44–46].

Methods for polygenic risk prediction[6,47–50] and fine mapping[7–9,25,30,31,51–53] rely on models for the effect-size distribution. Several of them rely on a point-normal or similar model[9,51–53], which assumes that non-null SNPs have a relatively narrow range of effect sizes. However, non-null SNPs actually have a wide range of effect sizes, so more flexible models may perform better in these applications; one candidate is the log-normal model (Fig. 5b and Supplementary Fig. 8).

The width of the effect-size distribution, from the tenth to 90th percentiles of heritability, is approximately two orders of magnitude for all traits analyzed (Fig. 5c), suggesting a shared underlying phenomenon. If heritability is mediated by cellular networks, as hypothesized by Boyle et al.[54], then the size of the range may be determined by network architecture: suppose that the effect size of a gene decays exponentially with its network distance d from a set of disease pathways (or core genes[54]) and that the number of genes at distance d increases exponentially with d. The heritability explained by genes at distance d would be the product of these exponential growth and decay functions, leading to a distribution in which heritability is distributed across a wide range of network distances and effect sizes. Moreover, if network architecture is similar across trait-relevant cell types, it could explain why this range is similar across traits.

Genetic effect-size distributions are flattened by negative selection, which prevents complex-trait heritability from being

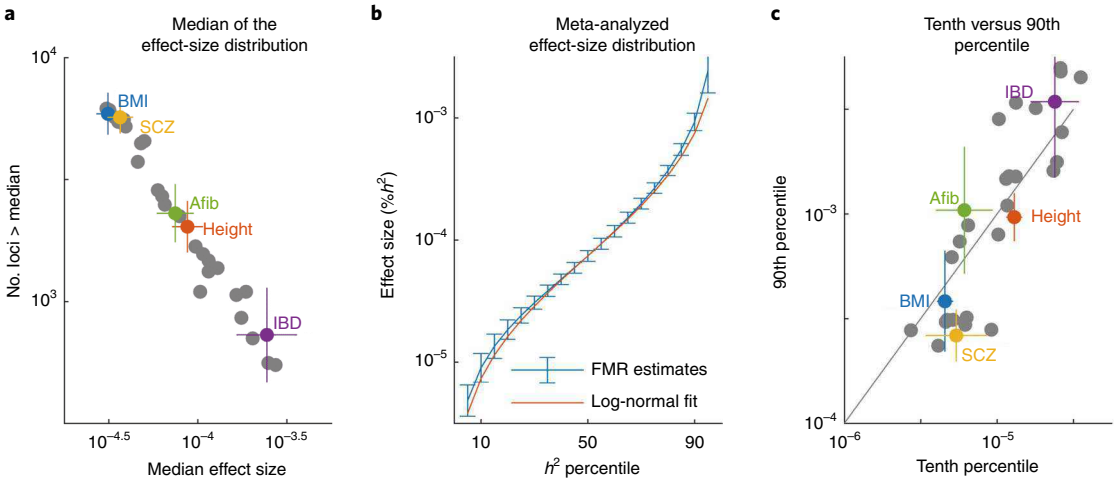


Fig. 5 | The genetic effect-size distribution. a, Median of the effect-size distribution: the proportion of heritability explained by a locus at the 50th percentile of heritability and the number of loci required to explain 50% of heritability. Point estimates are shown for 25 well-powered traits (Methods), with 95% CIs for selected traits (see Supplementary Table 11 for numerical results on all traits). Afib, atrial fibrillation. **b**, Quantiles of heritability, meta-analyzed across all 32 traits, with 95% CIs and a log-normal model fit. **c**, Effect sizes at tenth versus 90th percentiles of heritability for well-powered traits, with 95% CIs.

dominated by large-effect loci[15]. Flattening could result either from direct selection acting on a trait (with selection coefficients approximately proportional to effect sizes) or from highly pleiotropic selection (with relative selection coefficients that vary widely). Direct selection is expected to result in a narrower effect-size distribution, with per-SNP variance concentrated around an equilibrium value that is similar across SNPs[15]. The wide range of effect sizes is more consistent with primarily pleiotropic selection, which varies in strength across SNPs with the same effect size. Simons et al.[55] also provided evidence that highly pleiotropic selection shapes complex-trait genetic architecture.

This study has several limitations. First, the HDM is defined as a distribution of marginal, rather than causal, effect sizes. This choice is desirable for projecting the results of future GWAS, and it simplifies the estimation task; however, inferences about causal effect-size distributions are also useful, for example, to make inferences about cross-population genetic architecture[45,56]. Second, FMR currently does not model SNP annotations, including LD-dependent architecture[57], which can potentially bias heritability-style analyses[58,59]. In previous work, estimates of M_e (inversely proportional to the mean of the HDM) were extremely similar when modeling LD-dependent architecture[15]. In principle, FMR could be extended to model any number of annotations, similar to S-LDSC[10]. Third, this study did not explore the joint distribution of effect sizes and allele frequencies, which is an important component of genetic architecture[15,18], and it also did not estimate the effect-size distribution of rare and low frequency variants. Fourth, FMR requires fairly large GWAS sample size, so it is not applicable to small pilot studies of understudied diseases or populations. In general, it is difficult to predict whether FMR will be well powered, but FMR contains an internal check for whether its estimates are well powered (Simulations). Fifth, although FMR does correct for population stratification and cryptic relatedness, it cannot distinguish between true causal effects and those mediated through dynastic effects or assortative mating (Supplementary Fig. 2)[28,29].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00901-3.

Received: 16 September 2020; Accepted: 23 June 2021;
Published online: 29 July 2021

References

1. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
2. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
3. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
4. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
5. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
6. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
7. Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).
8. Ulirsch, J. C. et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019).
9. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. B* **82**, 1273–1300 (2020).
10. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1335 (2015).
11. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
12. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
13. Stahl, E. A. et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44**, 483–489 (2012).
14. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).
15. O'Connor, L. J. et al. Extreme polygenicity of complex traits is explained by negative selection. *Am. J. Hum. Genet.* **105**, 456–476 (2019).
16. Park, J.-H. et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42**, 570–575 (2010).

17. Palla, L. & Dudbridge, F. A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am. J. Hum. Genet.* **97**, 250–259 (2015).
18. Zeng, J. et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50**, 746–753 (2018).
19. Zhu, X. & Stephens, M. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat. Commun.* **9**, 4361 (2018).
20. Holland, D. et al. Beyond SNP heritability: polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genet.* **16**, e1008612 (2020).
21. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
22. Moser, G. et al. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genet.* **11**, e1004969 (2015).
23. Loh, P. R. et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
24. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).
25. Weissbrod, O. et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
26. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
27. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
28. Kong, A. et al. The nature of nurture: effects of parental genotypes. *Science* **359**, 424–428 (2018).
29. Morris, T. T., Davies, N. M., Hemani, G. & Smith, G. D. Population phenomena inflate genetic associations of complex social traits. *Sci. Adv.* **6**, eaay0328 (2020).
30. Hormozdiari, F. et al. Widespread allelic heterogeneity in complex traits. *Am. J. Hum. Genet.* **100**, 789–802 (2017).
31. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
32. Nelson, C. P. et al. Genetically determined height and coronary artery disease. *N. Engl. J. Med.* **372**, 1608–1618 (2015).
33. Palmer, C. & Pe'er, I. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.* **13**, e1006916 (2017).
34. Galinsky, K. J. et al. Fast principal-component analysis reveals convergent evolution of *ADH1B* in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).
35. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
36. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
37. Turchin, M. C. et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* **44**, 1015–1019 (2012).
38. Sohail, M. et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* **8**, e39702 (2019).
39. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).
40. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
41. Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).
42. Zhu, X. & Stephens, M. Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes. *Nat. Commun.* **9**, 4361 (2018).
43. Zuk, O. et al. Searching for missing heritability: designing rare variant association studies. *Proc. Natl Acad. Sci. USA* **111**, E455–E464 (2014).
44. Asgari, S. et al. A positively selected *FBN1* missense variant reduces height in Peruvian individuals. *Nature* **582**, 234–239 (2020).
45. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
46. Kichaev, G. & Pasaniuc, B. Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am. J. Hum. Genet.* **97**, 260–271 (2015).
47. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
48. Chung, W. et al. Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. *Nat. Commun.* **10**, 569 (2019).
49. Chun, S. et al. Non-parametric polygenic risk prediction via partitioned GWAS summary statistics. *Am. J. Hum. Genet.* **107**, 46–59 (2020).
50. Lloyd-Jones, L. R. et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).
51. Chen, W., McDonnell, S. K., Thibodeau, S. N., Tillmans, L. S. & Schaid, D. J. Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. *Genetics* **204**, 933–958 (2016).
52. Kichaev, G. et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* **33**, 248–255 (2017).
53. Benner, C., Havulinna, A., Salomaa, V., Ripatti, S. & Pirinen, M. Refining fine-mapping: effect sizes and regional heritability. Preprint at *bioRxiv* https://doi.org/10.1101/318618 (2018).
54. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
55. Simons, Y. B., Bullaughey, K., Hudson, R. R. & Sella, G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* **16**, e2002985 (2018).
56. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
57. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
58. Speed, D. et al. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).
59. Gazal, S., Marquez-Luna, C., Finucane, H. K. & Price, A. L. Reconciling S-LDSC and LDAK functional enrichment estimates. *Nat. Genet.* **51**, 1202–1204 (2019).

Methods

The non-independent and identically distributed normal model. The HDM is an effect-size distribution over proportions of heritability in a random-effect model. This model, the non-independent and identically distributed (i.i.d.) normal model[15], assigns an effect-size variance parameter to each individual SNP. The vector of normalized (per-s.d.) causal effect sizes, denoted $\boldsymbol{\beta}$, follows a multivariate normal (MVN) distribution with fixed covariance matrix Σ:

$$\boldsymbol{\beta} \sim N(0, \Sigma). \tag{2}$$

I assume for simplicity that Σ is a diagonal matrix with diagonal entries $\sigma_1^2, \ldots, \sigma_M^2$.

The model has a second fixed parameter, the LD matrix:

$$R = E\left(\boldsymbol{X}^T\boldsymbol{X}\right), \tag{3}$$

where $\boldsymbol{X}$ is the random vector of normalized genotypes. The random vector of marginal effect sizes or associations is

$$\boldsymbol{\alpha} = R\boldsymbol{\beta}. \tag{4}$$

$\boldsymbol{\alpha}$ is also MVN (as linear combinations of an MVN random vector are MVN). Its covariance matrix is $E\left(R\boldsymbol{\beta}\left(R\boldsymbol{\beta}\right)^T\right) = R\Sigma R$.

Random-effect heritability. The heritability of the non-i.i.d. normal model is

$$h^2 = Var(\boldsymbol{X}\boldsymbol{\beta}) = E\left(\boldsymbol{\beta}^T R\boldsymbol{\beta}\right) = Tr(R\Sigma). \tag{5}$$

When Σ is a diagonal matrix, this is equal to $M\overline{\sigma^2}$, where M is the number of SNPs and $\overline{\sigma^2}$ is the mean causal effect-size variance. (Tr is the trace.)

The heritability distribution of marginal effect sizes. The random-effect heritability explained by SNP j is σ_j^2, the jth diagonal element of Σ. These values, which sum to h^2, define a probability distribution over proportions of heritability. Let J be the index of a random SNP chosen with probability

$$P(J = j) = \frac{\sigma_j^2}{h^2}. \tag{6}$$

For example, under a point-normal model[13,18,21], a proportion π of SNPs are causal with effect-size variance $\sigma_j^2 = h^2(M\pi)^{-1}$, and a proportion $1-\pi$ are non-causal ($\sigma_j^2 = 0$). Under this model, the distribution of causal effect sizes over proportions of heritability, the distribution of β_J, is normal with variance $h^2(M\pi)^{-1}$.

The HDM is the distribution of α_J: a randomly chosen element of the random vector $\boldsymbol{\alpha}$. The quantiles of the HDM describe what proportion of random-effect heritability is explained by SNPs with less than a given marginal effect size.

Notation. I express probabilities and expectations over proportions of heritability with the notation $P_{h^2}(\cdot)$ and $E_{h^2}(\cdot)$, respectively, dropping the index J. For example, the CDF of the HDM is

$$P_{h^2}\left(\alpha^2 \le x\right) = P\left(\alpha_J^2 \le x\right), \tag{7}$$

and its mean is

$$E_{h^2}\left(\alpha^2\right) = E\left(\alpha_J^2\right). \tag{8}$$

$E_{h^2}\left(\alpha^2\right)$ is equal to $h^2 M_e^{-1}$, where M_e is the effective number of independently associated SNPs[15].

I use the notation $P(\cdot)$ and $E(\cdot)$, with index-free arguments in the conventional fashion, to express a proportion of SNPs or an average across SNPs. For example,

$$h^2 = ME\left(\beta^2\right) = ME\left(\sigma^2\right). \tag{9}$$

Fixed-effect heritability. I distinguish between random-effect heritability, which is a fixed parameter of the model, and fixed-effect heritability, which is a function of $\boldsymbol{\beta}$ and therefore a random variable under the model. The fixed-effect heritability is

$$h^2_{\text{fixed}} = Var(\boldsymbol{X}\boldsymbol{\beta}|\boldsymbol{\beta}) = \boldsymbol{\beta}^T R\boldsymbol{\beta} = \boldsymbol{\beta}^T\boldsymbol{\alpha} = \sum_j \beta_j\alpha_j. \tag{10}$$

The fixed-effect proportion of heritability explained by set of SNPs A is

$$h^2_{\text{fixed}}(A) = \sum_{j\in A} \beta_j\alpha_j. \tag{11}$$

For any fixed set of SNPs A, for example, coding SNPs, the random-effect heritability is the expected value of the fixed-effect heritability:

$$h^2(A) = E\left(h^2_{\text{fixed}}(A)\right). \tag{12}$$

However, A itself can be a random variable under the model, in which case its random-effect heritability may differ from its expected fixed-effect heritability. For example, let $A(\beta)$ be the set of SNPs j with larger-than-expected effect sizes: $\beta_j\alpha_j > \sigma_j^2$. Then $E(h^2_{\text{fixed}}(A(\beta))) > h^2(A(\beta))$.

Similarly, the heritability explained by genome-wide significant SNPs is a random variable under the model, and SNPs that reach genome-wide significance have larger-than-expected effect sizes on average. I define $\%h^2_{\text{GWAS}}$ as the fixed-effect proportion of heritability. Similarly, I describe the shape of the effect-size distribution using quantiles of fixed-effect heritability. Estimates of the HDM can be converted into expected proportions of fixed-effect heritability (Supplementary Note).

Per-standard deviation effect sizes. Causal and marginal effect sizes are measured in per-s.d. units, that is, in standard deviations of the phenotype per standard deviation of the genotype, rather than in per-allele units[5]. In these units, squared effect sizes are equal to the proportion of variance explained. Common SNPs have larger effect sizes on average than rare SNPs (even though rare SNPs have larger per-allele effect sizes)[18]. It would also be of interest to quantify the HDM in per-allele units, but per-s.d. effect sizes determine power in GWAS[60], and they can be interpreted as fractions of heritability.

HDM characteristic function. The CF of the HDM is

$$\phi_{h^2}(t) = E_{h^2}\left(e^{i\alpha t}\right) = E\left(\sigma^2 e^{i\alpha t}\right)\overline{\sigma^{-2}}, \tag{13}$$

where i is the imaginary unit and t is a 'sampling time'. The CF is the Fourier transform of the density function.

If the HDM follows a mixture distribution, the components of which have CFs $\phi_1, \ldots, \phi_K$ and mixture weights (proportions of heritability) $w_1, \ldots, w_K$, then

$$\phi_{h^2} = w_1\phi_1 + \cdots + w_K\phi_K. \tag{14}$$

FMR assumes a finite mixture model for the purpose of estimation (see below).

The usefulness of the CF in this application is a relationship between the CF of the HDM and the CF of an effect-size distribution over SNPs. Let $X \approx N(0, \sigma^2)$. Its CF is

$$\phi_X(t) = e^{-\frac{1}{2}\sigma^2 t^2}. \tag{15}$$

Taking the derivative, the variance parameter is pulled out:

$$\frac{d}{dt}\phi_X(t) = -\sigma^2 t\phi_X(t). \tag{16}$$

This equation is useful because the factor σ^2 on the right-hand side is precisely what differs between the CF of a normal mixture model for the effect-size distribution over SNPs and the CF over proportions of heritability. For example, if 1% of SNPs have effect-size variance equal to 0.1% of per-SNP h^2, they explain $0.1\% \div 1\% = 10\%$ of heritability. More generally, suppose the distribution of causal effect sizes β is a mixture of normal distributions with variance parameters $\sigma_1^2, \ldots, \sigma_K^2$ and mixture weights (proportions of SNPs) $a_1, \ldots, a_K$. The distribution of β over proportions of random-effect heritability is also a mixture of normal distributions, with the same variance parameters but with weights proportional to $a_1\sigma_1^2, \ldots, a_K\sigma_K^2$. These two distributions have closely related CFs and, in particular, by equation (16):

$$t\phi_{h^2}(t) \propto \frac{d}{dt}\phi(t). \tag{17}$$

FMR regression equation. FMR relies on a regression equation that relates the distribution of GWAS summary statistics with the HDM. I discuss the choice of estimator and derive the equation in the Supplementary Note. The key step involves differentiating the CF of the marginal effect size of a randomly chosen regression SNP. After using an LD approximation (see below), I obtain

$$\frac{d}{dt}E\left(e^{it\alpha}|\boldsymbol{r}\right) \approx -t\overline{\sigma^2}\sum_j r_j^2\phi_{h^2}\left(r_j t\right). \tag{18}$$

The right-hand side is an LD-weighted sum of LD-scaled CFs: SNPs in weak LD with the regression SNP contribute less amplitude to its CF (proportional to r_j^2), and their contribution has a lower frequency (t is scaled by r_j). The amplitude effect corresponds to the intuition that a SNP in weak LD contributes less to the heritability tagged by the regression SNP (because ϕ_{h^2} is an expectation over proportions of heritability). The frequency effect corresponds to the intuition that weak LD proxies have smaller effect sizes than strong LD proxies (because frequencies correspond to effect sizes).

Modeling the HDM as a mixture distribution with unknown weights, such that $\phi_{h^2} = \sum_k w_k\phi_k$, equation (18) becomes a sum over the mixture components:

$$\sum_j r_j^2\phi_{h^2}\left(r_j t\right) = \sum_k w_k\sum_j r_j^2\phi_k\left(r_j t\right) = \sum_k w_k\ell_k(t), \tag{19}$$

where $\ell_k(t)$ is the 'Fourier LD score' of the regression SNP for mixture component k:

$$\ell_k(t) = \sum_j r_j^2 \phi_{h^2}(r_j t). \quad (20)$$

The left-hand side of equation (18) is a function of the true effect sizes. The sampling distribution of GWAS summary statistics is a normal distribution[61], which allows the regression equation to be restated in terms of the distribution of z scores:

$$\frac{d}{dt} E\left(e^{itz+\frac{1}{2}\chi_0^2 t^2}\right) \approx it N\overline{\sigma^2} \sum_k w_k \ell_k(t), \quad (21)$$

where z is the z score of the regression SNP, χ_0^2 is the LDSC intercept[26] (the sampling variance of z), and N is the GWAS sample size.

LD approximation. In the derivation of FMR, an LD approximation is used to link the observed effect-size distribution of regression SNPs with the unobserved effect-size distribution of the causal SNPs that they tag. Intuitively, it would simplify the estimation task if there were exactly one causal SNP per locus (and no LD between causal SNPs): the marginal effect size of SNP j would be equal to $r_{jc_j}\beta_{c_j}$, where c_j is the causal SNP in LD with SNP j. This assumption may be adequate for traits with less polygenic architectures, but a better assumption is that LD between causal SNPs is either weak or strong ($r^2 \approx 1$), such that the marginal effect size of SNP j is equal to $r_{jc_j}\alpha_{c_j}$, where c_j is any of the causal SNPs in LD with SNP j (all of which have the same causal effect size).

In detail, FMR assumes that, for all SNPs j and k, either $r_{jk} \approx 0$ ('no LD') or $\sigma_k^2 \approx 0$ ('SNP k is not causal') or $E(\alpha^2_j) \approx r^2_{jk} E(\alpha^2_k)$ ('every causal SNP in LD with SNP i is in strong LD with SNP j'). This type of approximation, which holds under two extremes, is expected to work well in practice because the most severe violations of the weak LD approximation satisfy the strong LD approximation and vice versa. A similar assumption is used for LD fourth-moment regression, and it was found to be robust[15].

Choice of mixture model. The HDM is modeled as a mixture of 13 overlapping normal distributions with variance parameters $\sigma^2 = \frac{\sigma^2_{\text{mean}}}{128}, \frac{\sigma^2_{\text{mean}}}{64}, \frac{\sigma^2_{\text{mean}}}{32}, \ldots, 32\sigma^2_{\text{mean}}$, a factor of 4,096 between the smallest and largest variance parameters (this is much larger than the $\sim 100\times$ difference between the tenth and 90th percentiles of heritability observed in Fig. 5). The grid is centered on a trait-specific value: $\sigma^2_{\text{mean}} = E_{h^2}(\alpha^2)$, the mean of the HDM, which is computed as a preprocessing step using LD4M[15]. Instead of recomputing Fourier LD scores for each trait, scores are computed once, and summary statistics are scaled to match them.

Choice of sampling times. I used one sampling time for each mixture component, with $t_k = \sigma_k^{-1}$. This choice is convenient because it reduces the number of scores that must be computed: $\ell_k(t_j)$ only depends on $\sigma_k^2 t_j^2$, so when t and σ are in geometric sequences with the same step sizes, many pairs have identical scores, and the number of Fourier LD scores to be computed is reduced from $K_1 K_2$ to $K_1 + K_2 - 1$.

Additional regression equations. FMR includes second-moment and fourth-moment regression equations, which are equivalent to regression equations for LDSC[10,26] and LD fourth-moment regression[15], respectively (see Supplementary Note for details). For example, the second-moment equation is

$$E(\chi^2) = \chi_0^2 + (w_1 + \cdots + w_{K_1})\ell^{(2)}, \quad (22)$$

where χ_0^2 is the LDSC intercept, and $\ell^{(2)}$ is the LD score.

The effect of including this equation is that the FMR heritability estimate is constrained to be approximately equal to the LDSC estimate.

Regression weights. FMR is a weighted regression with heuristic weights (Supplementary Note). These downweight redundant SNPs in high LD with other regression SNPs. They also reweight the various regression equations (different sampling times and second- and fourth-moment equations) so that they contribute equally to the objective function.

Choice of regression SNPs. FMR is applied to $\sim 10^6$ common SNPs (MAF > 0.05), with the HLA region excluded. SNPs with large effect sizes are not excluded, as doing so would inappropriately truncate the effect-size distribution.

Non-negativity constraint. Because probabilities and proportions of heritability are non-negative, FMR uses non-negativity constraints on regression weights. The computational cost of the constrained regression does not scale with the number of SNPs; therefore it does not contribute to computational complexity (Supplementary Note).

Calculation of Fourier LD scores. I computed Fourier LD scores from European-ancestry samples in 1,000 genomes[62]. Estimated scores were corrected for bias due to finite reference-panel sample size (Supplementary Note).

Implementation of FMR. Open-source software for FMR is available (Code availability). FMR is implemented in MATLAB, and it requires the 'lsqlin' function from the MATLAB optimization toolbox. The software does not require installation, and scripts are provided to replicate the main results.

Simulations. I performed simulations using real LD patterns computed from UK Biobank typed SNPs ($M = 455{,}000$) as previously described[15]. In detail, I simulated per-normalized-genotype causal effect sizes from

1. An i.i.d. point-normal distribution with a specified proportion of causal SNPs (10%, 1% and 0.1% in Fig. 1a–c). The variance parameter was equal to $0.1 M_c^{-1}$, where M_c is the expected number of non-null SNPs.
2. A four-component mixture of normal distributions with a null component and small-, medium- and large-effect non-null components (Fig. 1d–f). In Fig. 1d, proportions of SNPs in each non-null component were 0.01, $0.01(5)^{-1}$ and $0.01(25)^{-1}$, respectively. In Fig. 1e, proportions were 0.01, $0.01(10)^{-1}$ and $0.01(100)^{-1}$. In Fig. 1f, proportions were 0.01, $0.01(50)^{-1}$ and $0.01(2{,}500)^{-1}$. The effect-size variance of each component was equal to $0.1(3M_c)^{-1}$, where M_c is the expected number of SNPs in that component.

Next, I simulated GWAS summary statistics from the asymptotic sampling distribution[61], which depends on the LD matrix, as previously described[15]. I calculated Fourier LD scores directly from the LD matrix that was used to generate summary statistics, with no correction for finite reference-panel size. I applied FMR with the same model and parameter settings as I used for analyses of real traits.

Pruning and thresholding. I calculated the number of genome-wide significant loci and the heritability that they explain by thresholding GWAS summary statistics at a specified significance level and LD pruning to remove SNPs that tag the same signal. Pruning is performed in a greedy iterative manner: at each step, a new lead SNP is added, which is the most significant SNP that is not in LD with any lead SNP selected thus far, until no significant SNPs remain. The LD threshold is $r^2 < 0.01$. The heritability explained is proportional to the sum of their χ^2 statistics. This estimation procedure can be biased due to winner's curse or excessive LD pruning (Supplementary Note).

Risk-prediction accuracy. Under the non-i.i.d. normal model, the optimal PGS r^2 is

$$E(r^2_{\text{PGS}}) = Tr\left(S\left(\tfrac{1}{N}I + S\right)^{-1} S\right), \quad (23)$$

where $S = R\Sigma = E(\boldsymbol{\alpha}\boldsymbol{\beta}^T)$, and N is the GWAS sample size (see Supplementary Note for derivation). This is the accuracy of an optimal predictor given Σ, which is unknown in practice; it gives an upper bound on the accuracy of a real risk-prediction method. (I is the identity.)

Equation (23) can be approximated in three ways. First, when N is small,

$$E(r^2_{\text{PGS}}) \approx N\, Tr(S^2) = \tfrac{N}{M_e} h^4, \quad (24)$$

where M_e is the effective number of independently associated SNPs[15].

Second, when N is large,

$$E(r^2_{\text{PGS}}) \approx h^2 - \tfrac{M_c}{N}, \quad (25)$$

where $M_c = \text{rank}(S)$ is the number of linearly independent SNPs with non-zero effect sizes. This approximation is relevant when most heritability is explained by SNPs with effect-size variance $\gg N^{-1}$ (Supplementary Note). It cannot be computed using FMR, which does not estimate M_c.

Third, it can be approximated using the FMR LD approximation:

$$E(r^2_{\text{PGS}}) \approx E_{h^2}\left(\tfrac{\sigma^2_\alpha}{\sigma^2_\alpha + \frac{1}{N}}\right), \quad (26)$$

where $\sigma^2_\alpha = E(\alpha^2)$. This is the approximation that is used in Fig. 3c, as it can be estimated using FMR and it is appropriate at any sample size (Supplementary Note).

The not-by-chance true positive rate. The NTPR measures the proportion of two-tailed positive tests that correctly reject the null hypothesis in favor of the correct alternative and not by chance. It is designed to satisfy the property that, for any mixture of null and non-null effects, if there is no power to estimate the direction of effect, the NTPR is zero. It thereby avoids an arbitrary distinction between effects that are near zero $\left(0 < \alpha^2 \ll \frac{1}{N}\right)$ and those that are exactly zero. Estimates of the NTPR quantify a rate across loci, as opposed to one across SNPs.

For a z-score threshold $T \geq 0$, NTPR is defined as the fraction of true positives with the correct sign minus the fraction of sign errors:

$$\text{NTPR}(T) = P(z\alpha > 0 \,|\, z^2 > T^2) - P(z\alpha < 0 \,|\, z^2 > T^2). \quad (27)$$

Under sign symmetry, for every near-zero effect for which the effect direction is estimated incorrectly, there is expected to be a near-zero effect for which the effect direction is estimated correctly due to chance; therefore subtracting the fraction of sign errors (counting them as −1 true positives, rather than as 0 true positives) cancels out 'true-by-chance' positives. False positives (SNPs for which the true effect size is exactly zero) count as 0 true positives; therefore the expected number of true positives is the same for a SNP for which the effect size is exactly zero or nearly zero. The NTPR is related to the false sign rate introduced by Stephens[41] (Supplementary Note). It can be estimated using FMR (Supplementary Note).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

GWAS summary statistics are available at https://alkesgroup.broadinstitute.org/. Numerical results for Figs. 2–5 are reported in the Supplementary Tables.

Code availability

Open-source software is available at https://github.com/lukejoconnor[63]. GENESIS[14] software is available at https://github.com/yandorazhang/GENESIS.

References

60. Belmont, J. W. et al. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
61. Conneely, K. N. & Boehnke, M. So many correlated tests, so little time! Rapid adjustment of *P* values for multiple correlated tests. *Am. J. Hum. Genet.* **81**, 1158–1168 (2007).
62. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
63. O'Connor, L. J. lukejoconnor/FMR: initial release of FMR software. *Zenodo* https://doi.org/10.5281/zenodo.4670516 (2021).

Acknowledgements

I am grateful to A. Price, J. Ballard, A. Nadig, D. Weiner, O. Weissbrod, G. Getz, B. Neale and E. Lander for suggestions and discussions.

Author contributions

L.J.O. wrote the manuscript.

Competing interests

The author declares no competing interests.

Additional information

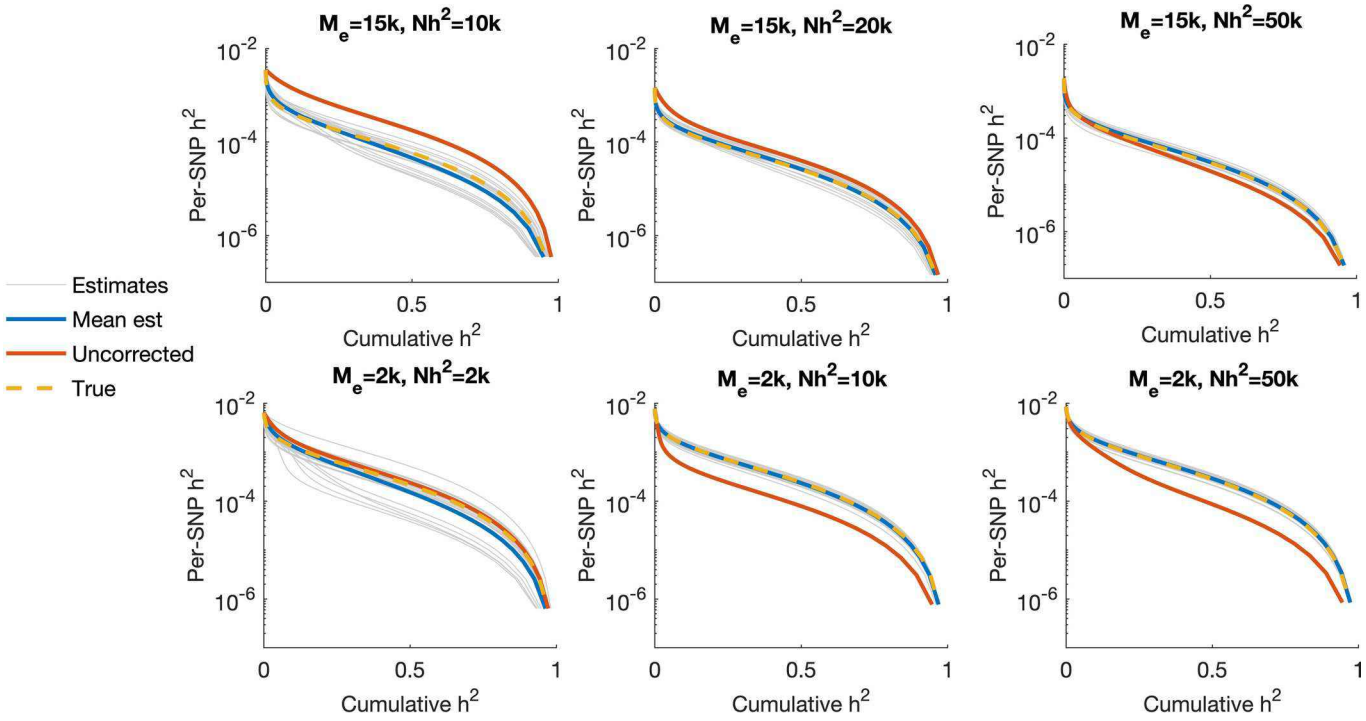
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00901-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00901-3.

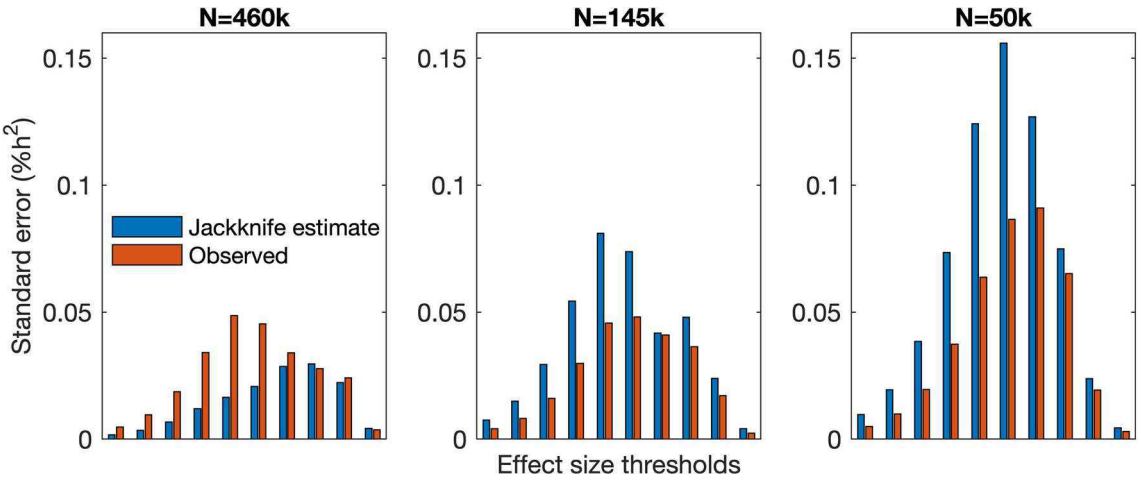
Correspondence and requests for materials should be addressed to L.J.O.

Peer review information *Nature Genetics* thanks Yan Zhang for their contribution to the peer review of this work.

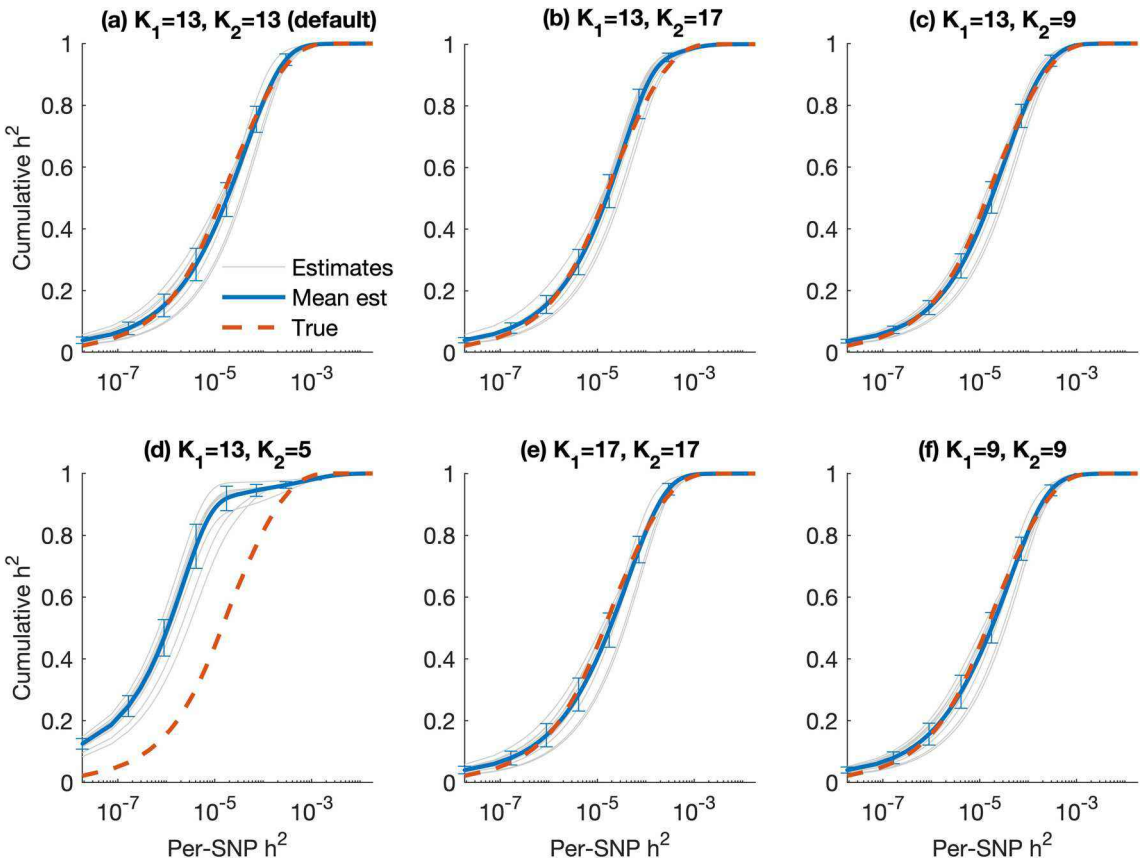
Reprints and permissions information is available at www.nature.com/reprints.



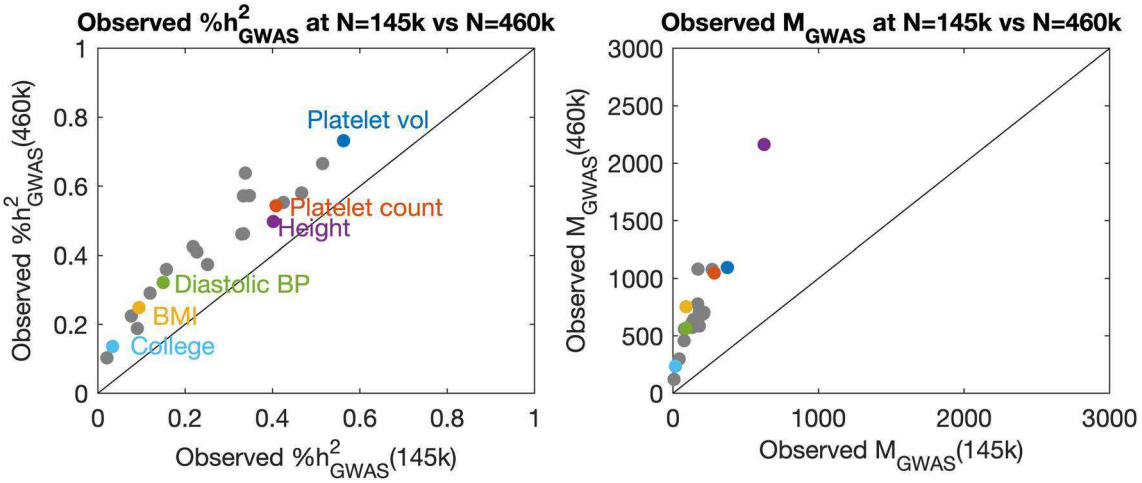
Extended Data Fig. 1 | Performance of FMR in simulations at different sample sizes. I show the true HDM (yellow), estimates for 10 individual simulation replicates (grey), the mean estimate across 20 replicates (blue), and the mean uncorrected estimate. The uncorrected estimate is obtained by running FMR without any correction for sampling variation in the GWAS summary statistics (see Supplementary Note). Data were simulated under a point-normal model with either 1% or 10% of SNPs having nonzero causal effect sizes.



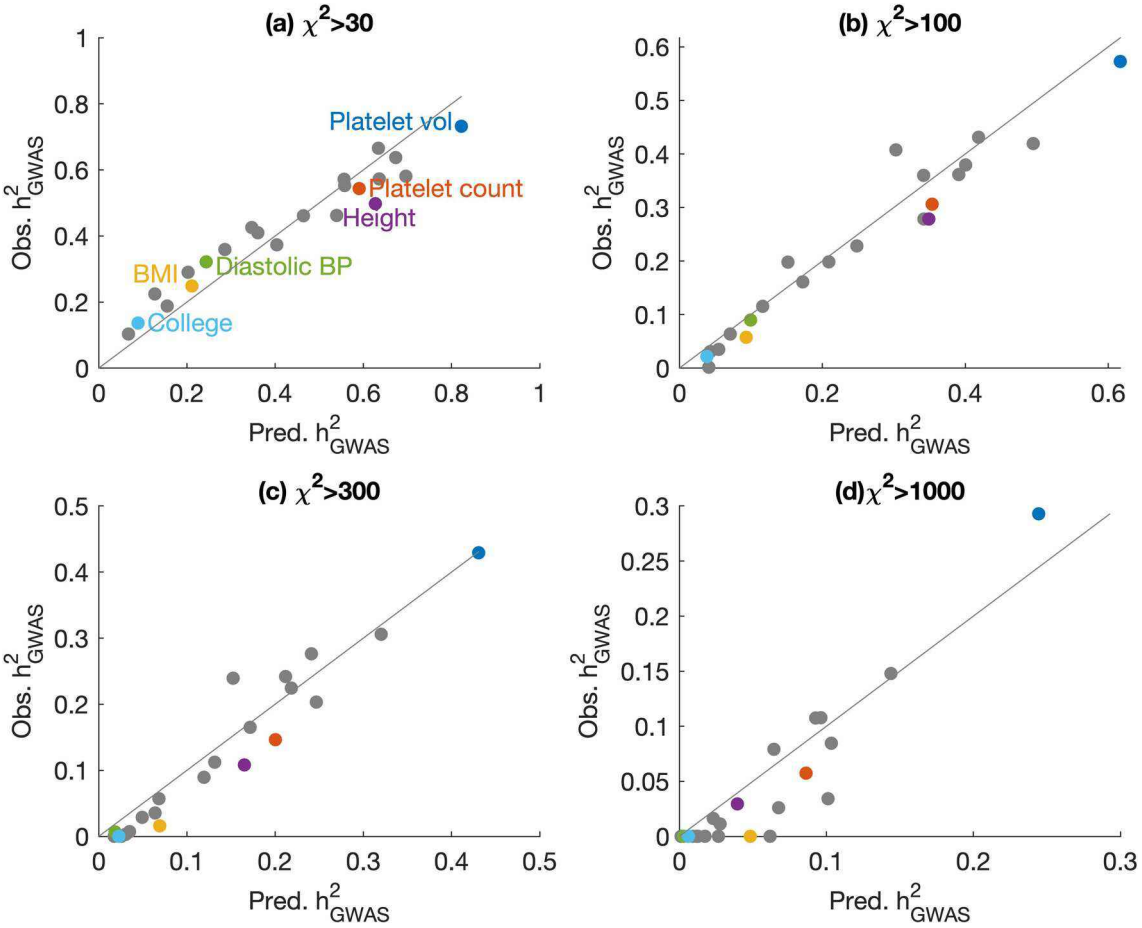
Extended Data Fig. 2 | Calibration of FMR jackknife standard errors. Simulations were performed under a normal mixture model with small-, medium- and large-effect SNPs (similar to Fig. 1d), at sample size N=460k, N=145k or N=50k. For different effect-size thresholds, I calculated the standard error of the proportion of random-effect heritability explained by SNPs with effect sizes less than that threshold. Bar plots show root-mean-squared jackknife standard errors (blue) and empirical standard errors (orange) based on 25 replicates. At large sample size (N=460k), standard errors were sometimes underestimated, probably due to the nonnegativity constraints in the regression. Caution is needed when making comparisons between the genetic architecture of different traits, as underestimated standard errors could lead to false-positive differences.



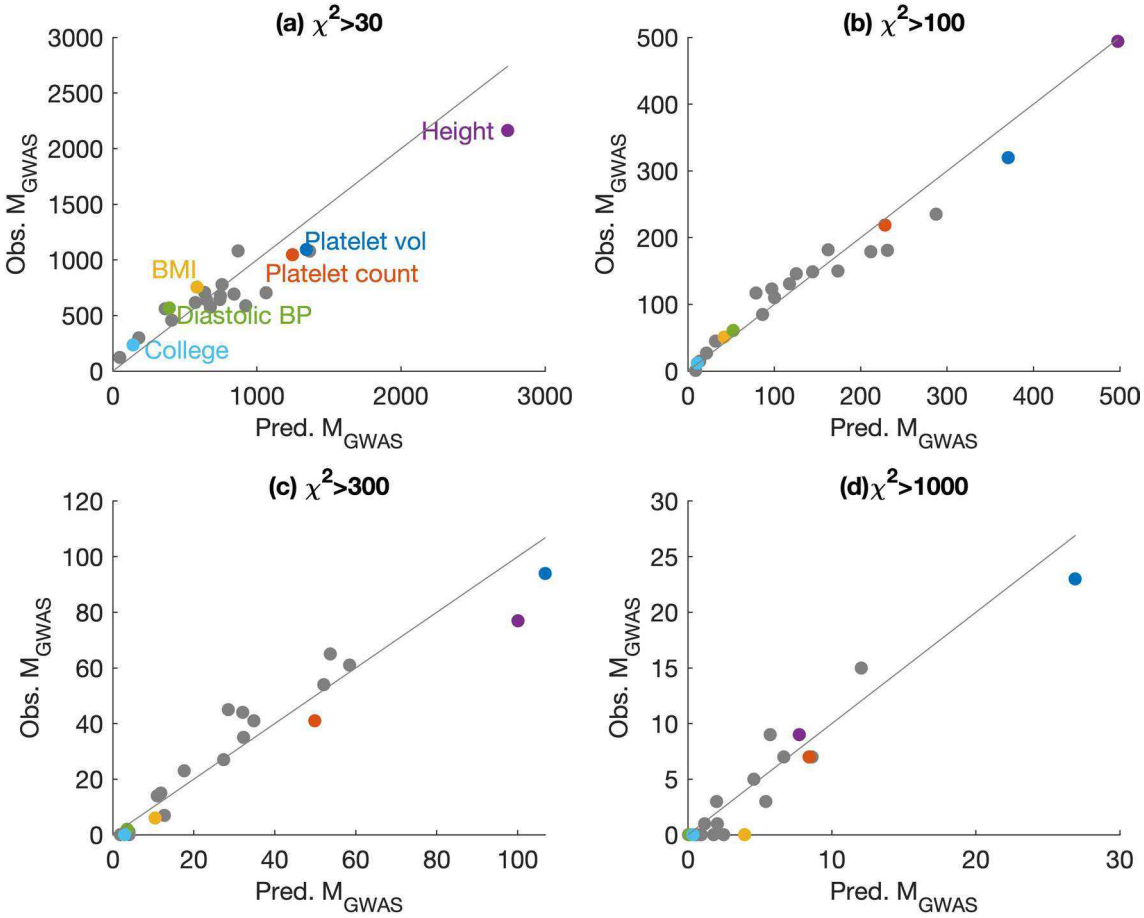
Extended Data Fig. 3 | Effect of changing the FMR sampling times and mixture components in simulations. Simulations were performed under a normal mixture model with small-, medium- and large-effect SNPs (similar to Fig. 1d), at sample size N=460k. I specified a set of 17 mixture components ($\sigma^2 = [2^{-9}, 2^{-8}, ...2^7]$) and 17 sampling times ($t_k = 1/\sigma_k$), and performed simulations with various subsets of the respective values. In panels a-d, I use the same values of σ^2 ($\sigma^2_3, \sigma^2_4, ..., \sigma^2_{15}$, which correspond to the default FMR model) and various values of t. In panels e-f, I vary the values of σ^2. In most cases, very similar results are obtained, except when too few sampling times are used (panel d). 25 replicates are performed (identical between the figure panels), the first 10 of which are plotted in grey. The mean and standard deviation across replicates are shown in blue. (a) $\sigma^2 = [\sigma^2_3, \sigma^2_4, ..., \sigma^2_{15}]$, $\boldsymbol{t} = [t_3, t_4, ..., t_{15}]$; (b) $\sigma^2 = [\sigma^2_3, \sigma^2_4, ..., \sigma^2_{15}]$, $\boldsymbol{t} = [t_1, t_2, ..., t_{17}]$; (c) $\sigma^2 = [\sigma^2_3, \sigma^2_4, ..., \sigma^2_{15}]$, $\boldsymbol{t} = [t_5, t_6, ..., t_{11}]$; (d) $\sigma^2 = [\sigma^2_3, \sigma^2_4, ..., \sigma^2_{15}]$, $\boldsymbol{t} = [t_5, t_7, ..., t_{15}]$; (e) $\sigma^2_1, \sigma^2_2, ..., \sigma^2_{17}, t_1, t_2, ..., t_{17}$; (f) $\sigma^2 = \left[\sigma^2_5, \sigma^2_6, ..., \sigma^2_{13}\right]$, $\boldsymbol{t} = [t_5, t_6, ..., t_{13}]$. I recommend using 13 mixture components and 13 sampling times, even though a smaller number may suffice (Extended Data Fig. 3f).



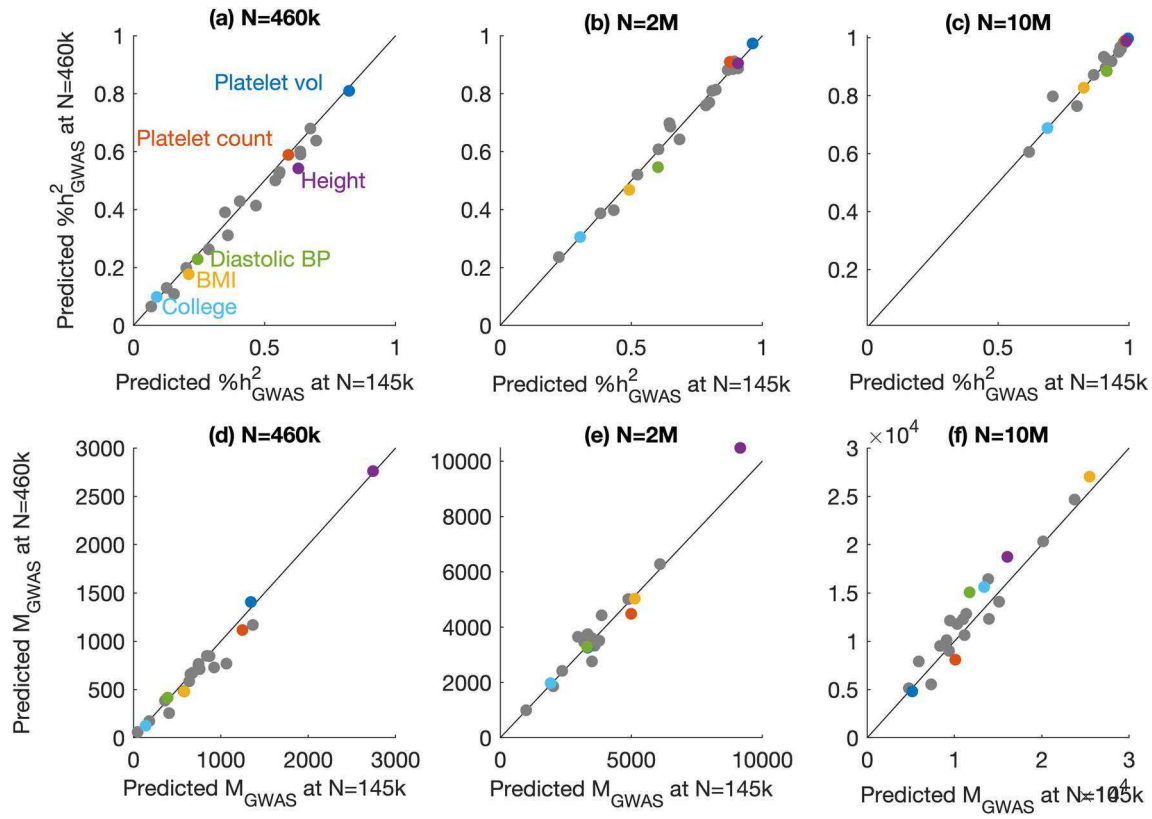
Extended Data Fig. 4 | Observed number of genome-wide significant SNPs and proportion of heritability explained at N=145k vs 460k. For numerical results, see Supplementary Table 2.



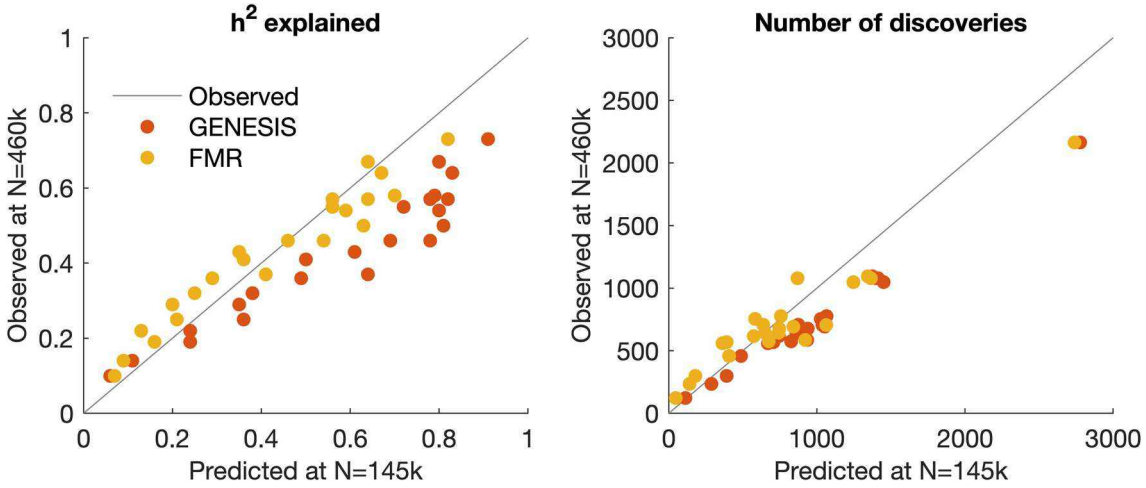
Extended Data Fig. 5 | Predicted vs. observed heritability explained by genome-wide significant SNPs at different significance thresholds. %h2GWAS was predicted using interim-release UK Biobank summary statistics (maximum N=145k) and evaluated in the full release (maximum N=460k). Squared correlations between predicted and observed values were 0.94, 0.95, 0.93, and 0.88 in panels a-d respectively. Lower r^2 at χ^2 > 1000(panel d) could result from the small number of loci with large effect sizes, which may increase the sampling variance of both the FMR predictions and the observed values. In panel d, the data points for several traits are superimposed near the origin. For numerical results, see Supplementary Table 2.



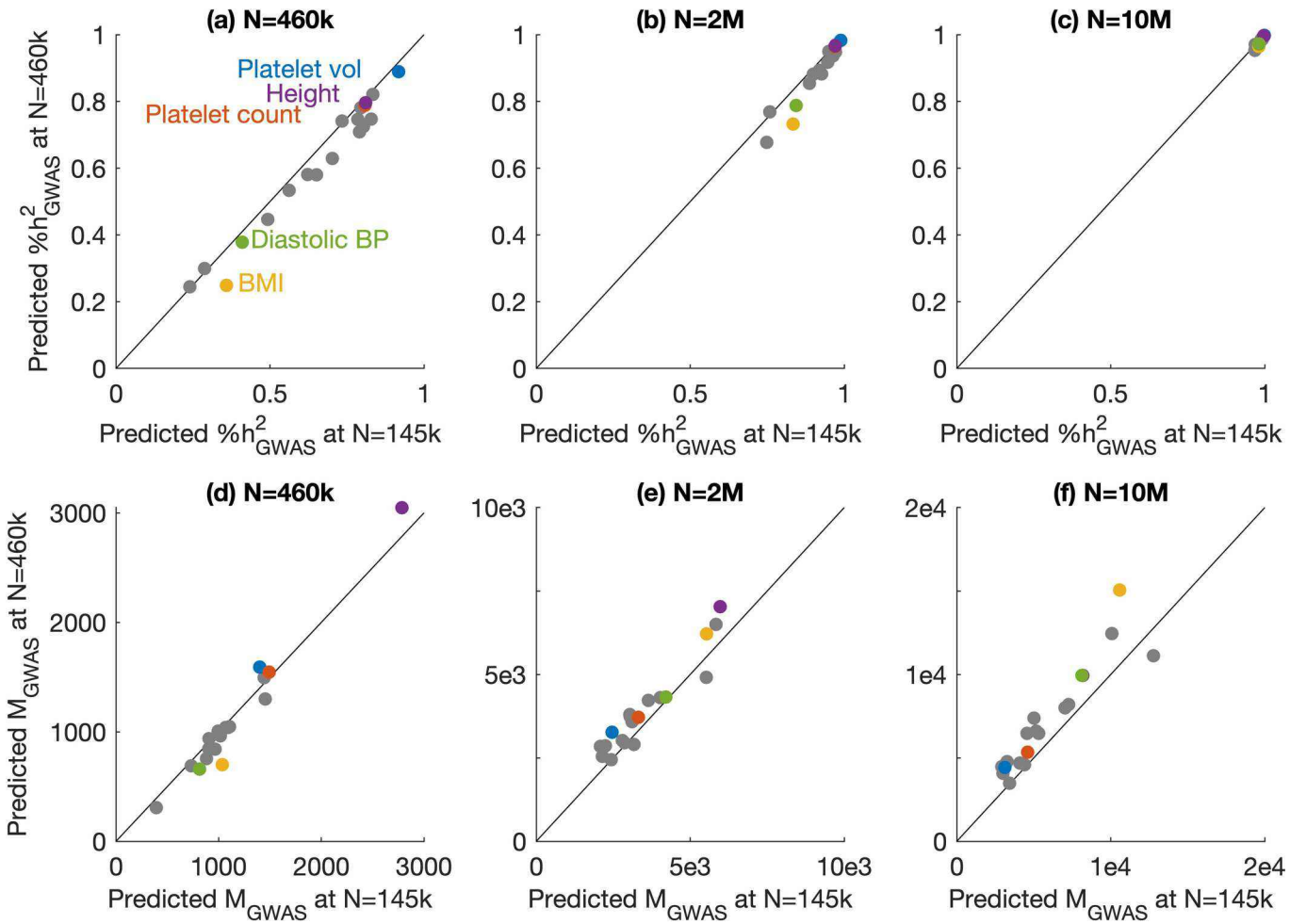
Extended Data Fig. 6 | Predicted vs. observed number of genome-wide significant SNPs at different significance thresholds. MGWAS was predicted using interim-release UK Biobank summary statistics (maximum N=145k) and evaluated in the full release (maximum N=460k). Squared correlations between predicted and observed values were 0.92, 0.97, 0.92 and 0.91 in panels a-d respectively. In panel d, the data points for several points are superimposed near the origin. For numerical results, see Supplementary Table 2.



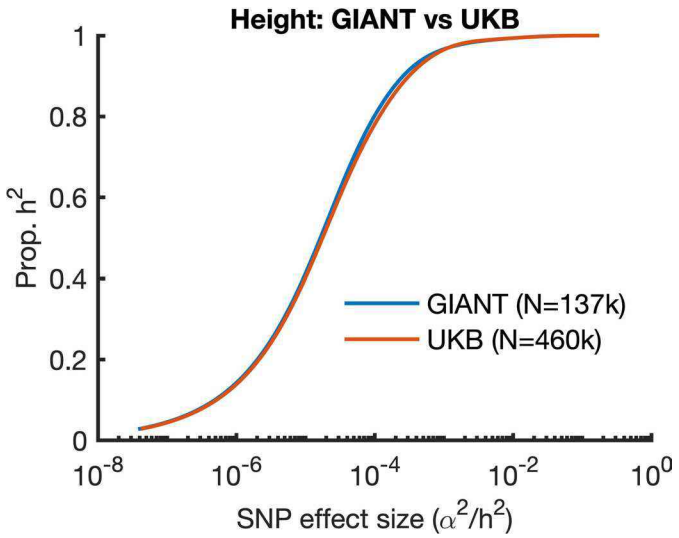
Extended Data Fig. 7 | Consistency of FMR predictions at different sample sizes. $\%h^2_{GWAS}$ and M_{GWAS} were predicted for 22 traits based on N=145k vs. N=460k summary statistics, with target sample size equal to 460k, 2M or 10M. Predictions assume that the LD score regression intercept will be equal to what was observed at N=145k for both sets of estimates. Numerical results are presented in Supplementary Table 2.



Extended Data Fig. 8 | Performance of GENESIS vs. FMR predictions in UK Biobank. FMR and GENESIS were applied to interim-release UK Biobank summary statistics (maximum N=145k) for 22 traits in order to predict the results of the full release (maximum N=460k). Numerical results are presented in Supplementary Table 1.



Extended Data Fig. 9 | Consistency of GENESIS predictions at different sample sizes. $\%h^2_{GWAS}$ and M_{GWAS} were predicted for 19 traits (Supplementary Table 3) based on N=145k vs. N=460k summary statistics, with target sample size equal to 460k, 2M or 10M. At N=460k, predictions of large-N $\%h^2_{GWAS}$ were slightly smaller (panels b-c), while predictions of M_{GWAS} were slightly larger (panel f). This difference could result from a less severe form of the power-dependent bias that is known to affect the point-normal (2-component) model when it is misspecified: as sample size increases, SNPs with smaller effect sizes become detectable, and estimates shift toward a larger number of causal SNPs with smaller effect sizes. (This only occurs when the model is misspecified, with a larger-than-expected number of small-effect SNPs). The 3-component model ameliorates this bias by including a small-effect heritability component even at small sample sizes. However, if this model too is misspecified (for example when there is a mixture of small-, medium- and large-effect SNPs), then it would be affected in the same way as the point-normal model, to a lesser degree. Numerical results are presented in Supplementary Table 3. The same analysis using FMR is presented in Extended Data Fig. 7.



Extended Data Fig. 10 | Estimated HDM of height using summary statistics from GIANT vs. UK Biobank. If results were biased by population stratification, the bottom-left portion of the curve (corresponding to small-effect SNPs) would be inflated for estimates based on GIANT.

nature research

Corresponding author(s): Luke O'Connor

Last updated by author(s): Jun 23, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NA
Data analysis	Software used to perform the analyses is provided, implemented in MATLAB 2020b. Open-source software is publicly available at github.io/lukejoconnor.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Summary statistics are available at https://alkesgroup.broadinstitute.org. Numerical results for Figures 2-5 are reported in supplementary tables.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No data were collected.
Data exclusions	Datasets were excluded from the analysis if they were not sufficiently well powered, as determined by whether the LD4M Z score was greater than 2 (see Supplementary Note).
Replication	The findings are reproducible, and code to reproduce main findings is available (see Code Availability).
Randomization	There was no randomization, as no clinical trial was performed.
Blinding	There was no blinding, as no clinical trial was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging