Accumulating evidence from genome wide association studies (GWAS) suggests an abundance of shared genetic influences among complex human traits and disorders, such as mental disorders. Here we introduce a statistical tool, MiXeR, which quantifies polygenic overlap irrespective of genetic correlation, using GWAS summary statistics. MiXeR results are presented as a Venn diagram of unique and shared polygenic components across traits. At 90% of SNP-heritability explained for each phenotype, MiXeR estimates that 8.3 K variants causally influence schizophrenia and 6.4 K influence bipolar disorder. Among these variants, 6.2 K are shared between the disorders, which have a high genetic correlation. Further, MiXeR uncovers polygenic overlap between schizophrenia and educational attainment. Despite a genetic correlation close to zero, the phenotypes share 8.3 K causal variants, while 2.5 K additional variants influence only educational attainment. By considering the polygenicity, discoverability and heritability of complex phenotypes, MiXeR analysis may improve our understanding of cross-trait genetic architectures.
In recent years, genome-wide association studies (GWASs) have successfully detected genetic variants associated with multiple complex human traits or disorders, providing important insights into human biology. Understanding the degree to which complex human phenotypes share genetic influences is critical for identifying the etiology of phenotypic relationships, which can inform disease nosology, diagnostic practice, and improve drug development. Most human phenotypes are known to be influenced by multiple genetic variants, many of which are expected to influence more than one phenotype (i.e., exhibit allelic pleiotropy). This has led to a widespread endeavor in genetic research, made possible by the public availability of most GWAS summary statistics (p-values and z-scores).

Currently, the prevailing measure to quantify polygenic overlap is genetic correlation. For a pair of traits, polygenic overlap refers to the fraction of genetic variants causally associated with both traits over the total number of causal variants across the two traits considered, while genetic correlation quantifies the correlation coefficient of additive genetic effects for the two traits. The sign of the correlation indicates whether the shared genetic effects predominantly have the same or the opposite effect directions. Available methods can quantify genetic correlation using raw genotypes or GWAS summary statistics. However, these methods report overall positive, negative, or no genetic correlation, but fail to capture mixtures of effect directions across shared genetic variants. This scenario is exemplified by the genetic relationship between schizophrenia and educational attainment. Despite consistent estimates of a non-significant genetic correlation, many genome-wide significant loci are found to be jointly associated with both phenotypes. Among 25 shared loci, 16 had effects in the opposite direction, while 9 had effects in the same direction. Thus, new statistical tools are needed to improve our understanding of the polygenic architecture of complex traits and their intricate relationships.

Here, we introduce a statistical tool (MiXeR), which quantifies polygenic overlap irrespective of genetic correlation between traits, using summary statistics from GWAS. To evaluate polygenic overlap between two traits, MiXeR estimates the total number of shared and trait-specific causal variants (i.e., variants with nonzero additive genetic effect on a trait). MiXeR bypasses the intrinsically difficult problem of detecting the exact location of causal variants, but rather aims at estimating their overall amount. MiXeR builds upon the univariate causal mixture model, which we extend to four bivariate normal distributions as illustrated in Fig. 1, with two causal components for variants specific to each trait, one causal component for variants affecting both traits, and a null component for variants with no effect on either trait. From the prior distribution of genetic effects, we derive the likelihood function of the observed signed test statistics (GWAS z-scores), incorporating effects of linkage disequilibrium (LD) structure, minor allele frequency (MAF), sample size, cryptic relationships, and sample overlap. The parameters of the mixture model are estimated from the summary statistics by direct optimization of the likelihood function.

We show in simulations that MiXeR provides accurate estimates of model parameters in the presence of realistic LD.

![Fig. 1](image-url) Components of the bivariate mixture in three scenarios of polygenic overlap. All figures are generated from synthetic data, where causal variants were drawn from the MiXeR model, the total polygenicity in each trait is set to 0.01%, SNP heritability is set to 0.4, GWAS N = 100,000. First column shows two traits where causal variants do not overlap. Second column adds a component of causal variants affecting both traits in the same (concordant) direction. Third column shows a scenario of polygenic overlap without genetic correlation. Top row shows simulated bivariate density of additive effects of allele substitution (β1, β2), the bottom row shows bivariate density of GWAS signed test statistics (z1, z2) for GWAS SNPs (genotyped or imputed). Due to linkage disequilibrium, GWAS-signed test statistic has substantially larger volume of SNPs associated with the phenotype. The aim of the MiXeR model is to infer distribution of causal effects (top row), using GWAS data (bottom row) as an input. Figures are generated on a regular grid of 100 × 100 bins, color histogram indicates log10(N) where N is the number of SNPs projected into a bin.
structure. Using GWAS summary data, we quantify polygenic overlap of several psychiatric disorders, including schizophrenia and bipolar disorder, with educational attainment and human height, with large implications for understanding how genetic factors overlap between complex human phenotypes.

Results
Simulations studies. In our first set of simulations, we generated synthetic GWAS data that follow model assumptions and assessed the validity of MiXeR estimates (polygenic overlap, \( \pi_{12} \); correlation of effect sizes within the shared polygenic component, \( \rho_{12} \), and genetic correlation, \( r_g \)) in the presence of a realistic LD structure (Fig. 2). We observed no bias in the estimates across a wide range of simulation scenarios (Supplementary Figs. 1–3), except for the scenario with low heritability \( (h^2 = 0.1) \) and high polygenicity \( (\pi^u = 3 \times 10^{-3}) \) in both traits, which represents an insufficiently powered GWAS study. In this scenario, MiXeR shows large variation among the estimates, and reports a non-zero polygenic overlap, when no such overlap exists. Standard errors estimated by the model are shown in Supplementary Table 1. Similarly, we show that univariate estimates of polygenicity and heritability are correct in all scenarios (Supplementary Fig. 4, Supplementary Table 2), except for the same scenario with low heritability and high polygenicity. In this scenario, polygenicity estimates suggest that the biases can be explained by truncated distribution of the errors, as the polygenicity parameters \( \pi_1, \pi_2, \) and \( \pi_{12} \) are bound to be non-negative. Traits with low heritability should obtain correct MiXeR estimates when the GWAS are sufficiently powered.

In addition, we validated that the model accurately predicts GWAS quantile–quantile (Q–Q) plots (Supplementary Fig. 5) and detailed Q–Q plots with SNPs partitioned into disjoint groups according to MAF and LD score (Supplementary Fig. 6a, b). Detailed Q–Q plots show a stronger GWAS signal for SNPs with higher MAF and higher LD score. The model’s prediction follows the same pattern, indicating that it correctly captures dependency of GWAS association statistics on MAF and LD score. To validate the accuracy of the predicted bivariate density, we generated conditional Q–Q plots (Supplementary Fig. 7), showing observed versus expected \(-\log_{10} p\)-values in the primary trait as a function of significance of association with a secondary trait at the level of \( p \leq 0.1, p \leq 0.01, \) and \( p \leq 0.001 \). We note that data Q–Q plots are closely reproduced by the model predictions across all \( p\)-value strata. Interestingly, scenarios without polygenic overlap show an enrichment, arising because GWAS \( p\)-values depend on allele frequency and LD structure, though this effect is generally smaller than enrichment arising due to shared causal variants.

Sensitivity analysis. For sensitivity analysis, we conducted simulations with traits that have a shared pattern of differential enrichment of heritability across genomic categories, which are not accounted for by the MiXeR model. Simulations were informed by the enrichment pattern of schizophrenia, as estimated by stratified LD score regression (Supplementary Fig. 8a). In the univariate analysis, polygenicity was underestimated by about 20% (Supplementary Table 3), suggesting that the model likely groups adjacent causal variants together and interpret their combined effect as arising from a single variant. In the bivariate analysis, we observed a small upwards bias in the estimate of polygenic overlap (Supplementary Table 3), but it did not exceed 10% of the polygenicity across all sufficiently powered scenarios.

Another assumption in the MiXeR model is that effect sizes are independent of allele frequencies. To assess this assumption, we ran simulations with effect sizes drawn from the BayesS model (see Online Methods). It characterizes MAF-dependent architecture in terms of a single parameter \( S \), ranging from \( S = 0 \) (equivalent to the MiXeR assumptions) to \( S = -1 \) (equivalent to the LDSR assumption). We simulated three intermediate values, \( S = -0.25, S = -0.5, \) and \( S = -0.75, \) which cover the range of BayesS estimates observed for real GWAS data. The results (Supplementary Fig. 10, Supplementary Tables 2, 4) highlight certain biases in MiXeR parameter estimates: for the extreme case of \( S = -0.75 \), heritability \( (h^2) \), univariate polygenicity \( (\pi^u) \), and polygenicity of the shared genetic component \( (\pi_{12}) \) are underestimated by up to 25%; correlation of effect sizes \( (\rho_{12}) \) is overestimated by 25%; however, the genome-wide genetic correlation \( (r_g) \) appears to have no bias.

Finally, we ran simulations with an incomplete reference, and simulated phenotypes where causal variants were spread across our entire reference panel of \( N = 11,015,833 \) variants, but only a fraction \( (50, 25, \) or 12.5%) of the variants enter LD structure estimation and fit procedure. The results (Supplementary Table 5) show that the total number of causal SNPs, as well as the heritability, are estimated correctly, while the polygenicity parameter is different from the simulated value, because it reflects the fraction of all tagged causal variants with respect to the reference that went into LD structure estimation.
GWAS summary statistics. We applied MiXeR to summary statistics from GWAS of 14 phenotypes, including five psychiatric,21,23–26 and four autoimmune27,28 diseases, four anthropomorphic traits29–32, and educational attainment12 (see Supplementary Table 6 for metadata about the studies).

MiXeR estimates of genetic correlation (Table 1; Supplementary Table 7) were generally consistent with those of cross-trait LD Score Regression8, with the highest genetic correlation observed between schizophrenia and bipolar disorder. As expected, these disorders also exhibit substantial polygenic overlap, sharing 6.2 K out of 8.5 K causal variants involved. Here and below, the numbers of causal variants are reported as 22.6% of their total estimate, which jointly accounts for 90% of SNP heritability in each phenotype, to avoid extrapolating model parameters into the area of in infinitesimally small effects (Supplementary Fig. 11).

Furthermore, MiXeR reveals important differences among traits with low genetic correlation, represented as Venn diagrams of shared and unique polygenic components (Fig. 3; Supplementary Figs. 12, 13a–g). For example, schizophrenia and educational attainment exhibit substantial polygenic overlap, sharing 8.3 K out of 10.8 K of causal variants involved. On the contrary, schizophrenia and height share only about 0.8 K out of 10.6 K causal variants. Educational attainment and height also show low polygenic overlap, sharing 1.8 K out of 12.3 K causal variants. Nevertheless, these traits have a high correlation of effect sizes within the shared component, $\rho_{12} = 0.52 (0.04)$, which at genome-wide level is observed as genetic correlation of $rg = 0.16 (0.01)$ according to MiXeR, or $rg = 0.14 (0.01)$ according to LDSR.

MiXeR estimates of the unique polygenic components provide insights into trait-specific genetic architectures. For example, schizophrenia has 2.1 K causal variants not shared with bipolar disorder, less than 0.1 K variants not shared with educational attainment, but as many as 7.5 K variants not shared with height. Also, for the other phenotypes, the number of trait-specific causal variants varies across different pairs of traits (Fig. 3).

Figure 4 and Supplementary Fig. 14a–g visualize the observed bivariate density of the GWAS-signed test statistics ($z_{1j}$, $z_{2j}$), the predicted density ($\tilde{z}_{1j}$, $\tilde{z}_{2j}$) from the MiXeR model, and the estimated bivariate density of the additive causal effects ($\beta_{1j}$, $\beta_{2j}$) that underlie model predictions. Figure 4 gives real examples for the three different scenarios of polygenic overlap (genetically independent traits, polygenic overlap with and without genetic correlation, as shown in Fig. 1). Finally, we use conditional Q–Q plots33,34 to compare the observed and predicted distributions of z-scores, and show that MiXeR-based prediction provides accurate estimates of the data Q–Q plots (Fig. 5), both in univariate and bivariate contexts.

Table 1 The results of cross-trait analysis with the MiXeR model for schizophrenia (SCZ), bipolar disorder (BIP), educational attainment (EDU) and height

| Trait 1 | Trait 2 | $n_2$ (se) | $n_1$ (se) | $n_2$ (se) | $\rho_{12}$ (se) | $rg$ (se) | $rg_{LDSR}$ (se) |
|---------|---------|------------|------------|------------|-----------------|----------|-----------------|
| SCZ     | BIP     | 6.19 (0.99) | 2.10 (1.26) | 0.21 (0.44) | 0.853 (0.019)  | 0.725 (0.071) | 0.725 (0.024)   |
| SCZ     | EDU     | 8.29 (0.84) | 0.00 (0.04) | 2.54 (1.02) | 0.071 (0.015)  | 0.062 (0.014) | 0.079 (0.022)   |
| SCZ     | Height  | 0.83 (0.10) | 7.46 (0.87) | 2.29 (0.12) | $-0.045 (0.060)$ | $-0.007 (0.010)$ | $-0.008 (0.019)$ |
| BIP     | EDU     | 5.72 (1.46) | 0.68 (1.16) | 5.11 (1.58) | 0.278 (0.051)  | 0.191 (0.036) | 0.188 (0.023)   |
| BIP     | Height  | 0.83 (0.11) | 5.57 (1.11) | 2.29 (0.13) | 0.001 (0.067)  | 0.000 (0.013) | $-0.014 (0.021)$ |
| EDU     | Height  | 1.76 (0.11) | 9.07 (0.58) | 1.37 (0.10) | 0.519 (0.040)  | 0.157 (0.010) | 0.141 (0.012)   |

Columns: $n_2$ – estimated number of shared causal variants, reported in thousands; $n_1$ (se) – estimated number of causal variants, unique to trait 1 (trait 2), reported in thousands; $\rho_{12}$ – correlation of effect sizes in shared polygenic component; $rg$ – genetic correlation ($\rho_{12} = \rho_{12}/\sqrt{\rho_{12}^2}$, see Online Methods); $rg_{LDSR}$ – estimate of genetic correlation from LD Score Regression. The number of variants ($n_2$, $n_1$ and $n_2$) are adjusted to explain 90% of heritability in the corresponding component. Parameters are fitted using approximately 1.1 M HapMap3 SNPs.
example, Ruderfer et al.\textsuperscript{40} performed a combined GWAS of schizophrenia and bipolar disorder, and a differential GWAS of schizophrenia cases versus bipolar disorder cases. Our results indicate a higher polygenicity of schizophrenia than bipolar disorder, which is in line with the recent study by Bansal et al.\textsuperscript{14}, who highlighted two schizophrenia subtypes. Our results also indicate that both schizophrenia and bipolar disorder have a small fraction of causal variants conferring disorder-specific risk (Fig. 3). Identifying shared and disorder-specific genetic variants is a subject of our future research, as it could provide critical knowledge about the distinct genetic architectures underlying these psychiatric disorders. Moreover, we find that nearly all causal variants influencing schizophrenia risk also appear to influence educational attainment, despite a genetic correlation close to zero (Table 1). This is in line with recent studies demonstrating shared genetic loci between schizophrenia and educational attainment\textsuperscript{15} and a strong genetic dependence between the phenotypes possibly related to different subtypes of schizophrenia\textsuperscript{14}. In contrast, while 89% of genetic variants influencing bipolar disorder also appear to influence educational attainment, there is in this case a significant positive genome-wide correlation of 0.191 (0.036), in agreement with the cross-trait LD score regression estimate of 0.188 (0.023) (Table 1; Supplementary Table 7).

We show that polygenicity is best expressed as a total number of causal variants (Supplementary Table 5). Previous studies presented it as a fraction, which is highly dependent on the reference panel used (1.1 M hapmap in ref.\textsuperscript{18}, or 484 K Affymetrix SNPs in ref.\textsuperscript{19}). When expressed as a total number, our estimates of polygenicity are consistent with previously reported results\textsuperscript{19} (Supplementary Table 8). In addition, we estimate that under the assumptions of the MiXeR model, only 5% of causal variants are needed to explain 50% of heritability, and 22.6% of causal variants are needed to explain 90% of heritability (Supplementary Fig. 11). These numbers are expected to be less dependent on modeling assumptions, because with finite GWAS samples it is not possible to distinguish small effects from truly null effects. The actual number of causal variants is, potentially, even higher, as our model tends to clump together variants if they are in high LD with each other (Supplementary Tables 3, 4).

Some existing methods can already uncover polygenic overlap in the absence of genetic correlation. For example, conjFDR
Among other notable methods for cross-trait analysis, the GenomicSEM33,34, also aim at detecting genomic loci jointly associated with two traits. MiXeR complements these methods by providing an easily interpretable high-level overview of the shared and unique genetic architectures underlying complex phenotypes. Among other notable methods for cross-trait analysis, the GenomicSEM33,34 and MTAG11 represent a multi-trait extension of the LD score regression. They can handle two or more traits at a time, but are based on the infinitesimal assumption, and quantify polygenic overlap using genetic correlation. For two-trait analysis, these methods are equivalent to LD score regression, thus we did not perform a formal comparison between GenomicSEM, MTAG, and MiXeR.

MiXeR has some notable advantages compared with the existing methods that implement causal mixture. First, our mathematical model for the likelihood term $p(z|\beta)$ is conceptually simpler and more flexible, resulting in unbiased estimates of model parameters across a wide range of simulation scenarios (Supplementary Figs. 1–3) and providing accurate prediction of GWAS z-scores across varying ranges of MAF and LD (Supplementary Fig. 6a, b). Second, MiXeR implementation works well with a large reference of 10 M variants, while other methods have reduced the reference to 1.1 M HapMap SNPs (ref. 18) or 484 K Affymetrix SNPs (ref. 19). Finally, our model individually processes all SNPs, without grouping them into bins (ref. 16).

MiXeR models causal effects as a single Gaussian component, while recent work18,45 suggests that certain phenotypes, including height, require at least two causal components of small and large effects. We note that the MiXeR model still provides a good fit for SNPs not reaching the GWAS significance threshold (Supplementary Fig. 16) and shows deviations only toward the tail of the distribution. To further investigate the effects of model misspecification, we implemented right-censoring of genome-wide significant SNPs (see Online methods). The results (Supplementary Tables 7, 9) are consistent with our main analysis, except for height which received a lower estimate of heritability (65% instead of 70%), a slight increase in polygenicity, and increased polygenic overlap with other traits. We propose that for a better estimate of height’s polygenicity, it would be beneficial to run MiXeR on a
residualized GWAS, after covarying association statistics for genotypes of all genome-wide significant SNPs. Recent work suggests the importance of MAF- and LD-dependent genetic architectures18,46, which are not directly modeled by MiXeR. Our simulations show certain biases in the estimates of polygenicity parameters ($\pi_i^2$ and $\pi_{12}$), which underestimate the true value by up to 25% in the case of extreme MAF-dependent architecture with $S = -0.75$ (Supplementary Fig. 10). However, these biases tend to cancel one another out when considering the relative size of the polygenic overlap ($\pi_{12}/\pi_i^2$ ratio). Also, on real data, most Bayesians18 estimates lay between $S = -0.25$ and $S = -0.5$, where the bias of the MiXeR estimates is in the order of 10% rather than 25%. On real data, we observe effects of MAF-dependent architectures by drawing Q–Q plots for subsets of SNPs (Supplementary Fig. 17a–g) partitioned into nine groups according to MAF and LD score, where the model tends to underestimate z-scores in low MAF bins. This effect, however, is quite subtle, and does not manifest itself on the overall Q–Q plots (Supplementary Fig. 16).

The MiXeR method requires large GWAS studies. Our recommendation is to apply MiXeR to studies with at least $N = 50,000$ participants, and inspect standard errors reported by MiXeR. In addition, MiXeR applies Bayesian information criterion (BIC) to compare causal mixture model versus the infinite-simial model, as shown in Supplementary Table 9. The cases where BIC selects the infinitesimal model indicate that the GWAS sample size is insufficient to reliably fit the polygenicity parameter. Generally, polygenicity estimation requires more GWAS power than heritability estimation, which can be visually explained by GWAS Q–Q plots (Supplementary Fig. 16): heritability is determined by the overall departure of the GWAS curve from the null line, while polygenicity is determined by its curvature, i.e., the point where the GWAS curve begins to bend upward from the null line, which is harder to estimate when GWAS signal is weak. This is captured by MiXeR standard errors, which show that individual parameters of the mixture model have lower estimation accuracy than their combinations—for example, relative errors for $\pi_i$ and $\pi_{12}^2$ are larger than for the heritability estimate $h^2 \propto \pi_i \sigma^2$, due to inversely-correlated errors (Supplementary Table 2). Despite these limitations, there is still a clear minimum on the energy landscape of cost function (Supplementary Figs. 18, 19, showing log-likelihood as a function of model parameters around the optimum).

In our future work, we plan to incorporate an additional Gaussian component to model small and large effects18, and to explicitly account for MAF-dependent architectures46. Further extensions may account for differential enrichment for true causal effects among studies, and is not currently applicable for analysis across different ethnicities. We also aim to extend the MiXeR modeling framework to be used to improve power for discovery of shared and trait-specific SNPs by estimating the posterior effect size of SNPs associated with one trait given the test statistics in another trait, as well as for improving predictive power of polygenic risk scores.

In conclusion, MiXeR represents a useful addition to the toolbox for cross-trait GWAS analysis. By considering the intricate polygenic architectures of complex phenotypes, MiXeR allows for measures of polygenic overlap beyond genetic correlation. We expect this to lead to new insights into the pleiotropic nature of human genetic etiology.

Methods

Bivariate causal mixture model. Consider a simple additive model of genetic effects, ignoring gene-environment interactions, epistasis and dominance effects. Under these assumptions, the contribution of the genotype to the phenotype is modeled as a sum of individual contributions from genetic variants: $y_i = \sum \beta_j g_{ij}$, where $y_i$ is a quantitative phenotype or disease liability of $i$-th individual, $g_{ij}$ is 0, 1, 2-coded number of reference alleles for $j$-th variant, and $\beta_j$ is the additive genetic effect of allele substitution. We say that a genetic variant is causal for a trait if it has a non-zero effect on that trait ($\beta_j \neq 0$).

MiXeR builds upon the univariate causal mixture model16, $\beta_j \sim N(0,0) + \pi_j N(0, \sigma^2_j)$, which assumes that only a small fraction ($\pi_j$) of variants have an effect on the trait, while the effect of the remaining variants is zero. For the mathematical convenience, we chose a Gaussian distribution for the nonnull arm of the causal mixture. A drawback with the gaussian prior is that a large fraction of causal variants will have effect sizes close to zero. We would prefer to count a variant as causal only if it has a sufficiently large effect size, using for example a bi-modal prior distribution with probability mass separated from zero, but for such prior, it was not feasible to accurately model the effects of the LD structure.

In a joint analysis of two traits, we expect some variants to affect both traits; some variants to affect one trait but not the other; and most variants to have no effect on either trait. We assumed that for a given trait, all causal variants follow the same distribution of effect sizes, regardless of what effect a variant has on the other trait. Within the shared component, we model correlation of effect sizes, to account for genetically correlated traits. Based on these assumptions, MiXeR models additive genetic effects $\beta_i$, $\beta_j$, of variant $j$ on the two traits as a mixture of four bivariate Gaussian components (Fig. 1):

$$
\left( \begin{array}{c}
\beta_i \\
\beta_j
\end{array} \right) \sim N(0,0) + \pi_i N(0, \Sigma_i) + \pi_j N(0, \Sigma_j) + \pi_{12} N(0, \Sigma_{12}),$
$$

(1)

where $\pi_i$ and $\pi_j$ are weights of the unique variants (components with an effect on the first trait only, and on the second trait only); $\pi_{12}$ is a weighting of the component affecting both traits; and $\pi_{00}$ is a fraction of variants that are non-causal for both traits, $\pi_0 + \pi_1 + \pi_2 + \pi_{12} = 1$; $\sigma_i^2$ and $\sigma_j^2$ control expected magnitudes of per-variant effect sizes; and $\rho_{ij}$ is the correlation coefficient of the effect sizes in the shared component. All parameters are assumed to be the same for all genetic variants.

The effects ($\hat{\beta}_i$, $\hat{\beta}_j$) estimated by a GWAS, represent only the true causal effects ($\beta_i$, $\beta_j$), which are distorted by limited sample size (poor statistical power), cryptic relatedness within a GWAS sample, as well as LD between variants. To disentangle these effects, we derive the likelihood term for observed GWAS signed test statistics ($z_i$, $z_j$), incorporating effects of the LD structure (allele correlation $r_{ij}$ between variants $i$ and $j$; heterozygosity $H_i = 2p_i(1-p_i)$, where $p_i$ is the minor allele frequency of the $i$-th variant; the number of subjects genotyped per variant ($N_i$ and $N_j$); and variance distortion parameters $\sigma^2_{1i}$, $\sigma^2_{1j}$, and $\rho_{0i}$). Specifically (see Supplementary Note 1),

$$
z_i = \delta_i = \sqrt{N_i \sum r_{ij} \beta_j}.
$$

(3)

The nine parameters of the model ($\pi_0, \pi_1, \pi_2, \sigma^2_1, \sigma^2_2, \sigma^2_{12}, \sigma^2_{0i}, \rho_{0i}$, $\rho_{0j}$) are fit by direct optimization of the weighted log-likelihood, with standard errors estimated from the Observed Fisher’s Information matrix.

Forcing $\pi_{2} = 1$ (so that $\pi_0 = \pi_1 = \pi_2 = 0$) reduces our model to an infinitesimal assumption that underlies cross-trait LD score regression4. Under this constraint, our model predicts that GWAS-signed test statistics follow a bivariate Gaussian distribution with zero mean and variance-covariance matrix $\Sigma = \left[ \begin{array}{cc}
\sigma^2_1 & \rho_{0i} \sigma^2_{0i} \\
\rho_{0j} \sigma^2_{0j} & \sigma^2_{0j}
\end{array} \right]$, where $\delta_i = \sqrt{N_i \sum r_{ij} \beta_j}$ being proportional to the LD score of $j$-th SNP, and parameters $\rho_{0i}$, $\rho_{0j}$, $\sigma^2_{0i}$ playing the role of LD score regression intercepts45. The only distinction between the two models is that we choose to model effect sizes that are independent of allele frequency, leading to the incorporation of $H_i$ into our model; this factor is absent from the LD score regression model due to the assumption there of effect sizes that are inversely proportional to $H_i$. Thus, MiXeR is a direct extension of cross-trait LD score regression, which relaxes the infinitesimal assumption.

Model for bivariate distribution of GWAS z-scores. We derive two models for GWAS z-scores, which we call “fast model” and “full model”. The “fast model” is quicker to run, and we use it to perform an initial search in the space of the model’s
parameters. The “full model” is slower but more accurate, and we use it for a final tuning of model estimates.

The “full model” for GWAS z-scores approximates \((Z, Z_0)\) distribution of a given GWAS SNP as a mixture of \(k = 20,000\) bivariate normal distributions, all having equal weight in the mixture. For each \(k = 1, \ldots, K\), we randomly draw the location of causal variants \((\pi_1, N)\) causal variants specific to the first trait, \(\pi_2, N\) specific to the second trait, and \(\pi_1, N\) shared causal variants, where \(N\) denotes the total number of variants in the reference panel, and calculate the variance–covariance matrix \(\Sigma_q\) from equation (3), using estimated LD \(R^2\) correlations between the assumed causal variants and the GWAS SNP.

The “fast model” is derived from the method of moments (see Supplementary Note 1):

\[
\delta_{ij}/\sqrt{N_\delta} \sim \left(1 - \pi_0, N(0, 0), \Sigma_q^z \right) \oplus \left(1 - \pi_0, N(0, 0), \Sigma_q^x \right) \oplus \left(1 - \pi_0, N(0, 0), \Sigma_q^x \right) \oplus \left(1 - \pi_0, N(0, 0), \Sigma_q^x \right)
\]

which \(\oplus\) denotes convolution of probabiliy distribution functions (so that right-hand side evaluates to a mixture of eight components), \(\pi_0, N(0, 0)\) is adjusted weight of mixture component \((c = 1, 2, 12)\) \(\Sigma_q^c\) is adjusted variance–covariance matrix, \(\ell_j = H_{ij}^R\) is the LD score, adjusted for heterozygosity, and \(\eta_j = \pi_0, N(0, 0)\) is adjusted weight of mixture component \((c = 1, 2, 12)\) \(\Sigma_q^c\) is adjusted variance–covariance matrix. For each GWAS SNP as a mixture of \(K\) variants in the reference panel, and calculate the variance–covariance matrix \(\Sigma_q\) from equation (3), using estimated LD \(R^2\) correlations between the assumed causal variants and the GWAS SNP.

The log likelihood:

\[
F(\theta) = \sum_{j} w_j \log(pdf(z_j | \theta)) + \sum_{j} w_j \log cdff(\text{quantile}(\theta))
\]

where \(\theta = (\pi_1, \pi_2, \pi_1, \pi_2, \pi_1, \pi_2, \pi_1, \pi_2)\) is a vector of all parameters being optimized, and weights \(w_j\) chosen by random pruning (64 iterations at LD \(r^2 < 0.1\)). Optimization is done by the Nelder–Mead Simplex Method as implemented in MATLAB's \texttt{fminsearch}. First, we fit univariate parameters separately for each trait, i.e., \(\pi_1, \pi_2, \pi_1, \pi_2\) for the first trait, and similarly for the second trait. Univariate fit employs a sequence of optimizations to ensure robust convergence: first, we use the “fast model” under constraint \(\pi_1 = 1\) to find \(\pi_2\), and to initialize \(\pi_2\), second, we use constraint \(\pi_2 = \text{init}\) to find initial values of \(\pi_2\) again with the “fast model”. Finally, we use the “full model” and unconstrained optimization to jointly fit \(\pi_1, \pi_2, \pi_1, \pi_2\) parameters. The same procedure is repeated for the second trait, to fit \(\pi_2, \pi_2, \pi_2, \pi_2\). To improve parameterization, we use logit as a link function to convert it into a cumulated distribution function. Error bars on data Q–Q plots represent the 95% binomial confidence interval \(p = 1.96, \sqrt{1 - q}/n_{\text{total}}\) where \(q\) is the probability of observing a \(p\)-value as extreme as, or more extreme than the chosen \(p\)-value, and \(n_{\text{total}}\) is the effective number of SNPs across all SNPs used to fit the model.
number of SNPs after controlling for LD structure, which in our case was calculated as a sum of random pruning weights across all SNPs.

GWAS power curves. Causal mixture model can project the future of GWAS discoveries, by estimating proportion $S(N)$ of narrow-sense heritability captured by genome-wide significant SNPs at a given sample size $N$. The $S(N)$ is defined as follows:

$$S(N) = \sum_{i=1}^N C(z, N, i) dx \cdot \sum_{x, i} C(z, N, i) dx,$$

where $z_i = 5.45$ gives $z$-score corresponding to the standard genome-wide significance threshold $5 \cdot 10^{-8}$, and $C(z, N, i) \equiv P(z, N, i) \cdot E(B(z, N, i)$ denotes a posterior effect size $E[B(z, N, i)$ of the non-centrality parameter $\theta^i$ for a GWAS SNP, given certain $z$-score, multiplied by a prior probability of observing that $z$-score. Probability density function $P(z, N, i)$ is given by Eq. (4), and $E(B(z, N, i)$ can be calculated from the Bayesian rule. Thus, $C(z, N, i) = \delta^j \mathcal{P}_i | \mathcal{P}(\theta^j), dy$, where $\delta | \sim N(0, i_2^2)$. Analytical expressions for $C(z, N, i)$ and $J_{\sum x, i} C(z, N, i) dx$ are given in the Supplementary Note 1.

SNPs in the analysis. To enable a direct comparison of our model with LD score regression, we use the same set of SNPs in our log-likehood optimisation, which consists of approx. 1.1 million variants, subset of 1000 Genomes and HapMap309, with MAF above 0.05, ambiguous SNPs excluded, imputation data above 0.9, MHC and other long-range LD regions excluded. Calculation of the LD structure, LD scores $\ell_f$ and shapes parameter $\eta_f$ are based on 9997.231 SNPs from 1000 Genomes Phase 3 data, downloaded from LD score regression website. In simulations we generate GWAS and estimate LD structure on a subset of 11,015,833 SNPs from 1000 Genomes Phase 3, with MAF above 0.002, call rate above 90%, excluding duplicated RS numbers; the fit procedure was constrained to ~130 K GWAS SNPs, keeping only HapMap3 SNPs, and pruning SNPs at LD $r^2$ threshold of 0.1.

LD score regression estimates. For dichotomous phenotypes, we used an effective sample size of $N_{eff} = 4/(1/N_{GWAS} + 1/N_{controls})$ to account for imbalanced numbers of cases and controls, both in MixeR and in LD score regression. In addition, we ran LSDR using MixeR MAF model (using --per-allele flags in LD score estimation), and show the results alongside with original LSDR estimates (Supplementary Tables 7, 9). For case/control phenotypes heritability is reported on the observed scale.

Simulations. In our simulations, we use a panel of $N = 100,000$ samples and 11,015,833 SNPs generated by HapGen298 using 1000 Genomes31 data to approximate the LD structure for European ancestry. To avoid relatedness across individuals, we run HapGen2 for small disjoint chunks of about 2900 SNPs at a time, 3920 chunks in total. The chunks were acting as additional recombination hotspots, causing certain changes in the distribution of the LD scores (Supplementary Fig. 20). However, the total amount of allelic correlation in the HapGen2 panel was still substantial, for example the median LD scores in the HapGen2 panel (Supplementary Fig. 19) was still substantial, for example the median LD scores in the HapGen2 panel (Supplementary Fig. 19) was 0.4, and 0.7, and for each combination, generate 20 pairs of genetically independent traits (except for having shared pattern of enrichment). To simulate the enrichment, we keep a constant variance of effect sizes across all SNPs, but modulate the probability of having causal variant proportionally to LDRS regression coefficient. We use --per-allele flags in LD score estimation to run simulations with the MixeR MAF model. For simulations with MAF-dependent architectures, we simulate effect sizes as follows:

$$\beta_i \sim N(\theta_i, H_i^{1/2}) \cdot (1 - \eta)N(0, 0)$$

Parameter $S = 0$ corresponds to the MixeR MAF model, $S = -1$ corresponds to the LDRS MAF model. The same model is implemented in BayesS software28, thus we choose parameter $S$ from $-0.25$, $-0.50$, and $-0.75$, which corresponds to the range of BayesS estimates observed on real GWAS data.

Data availability

The datasets analyzed during the current study are freely available for download from the following URLs: LD scores and reference panel derived from 1000 Genomes phase 3, https://data.broadinstitute.org/alkesgroup/LDSCORE/; Psychiatric Genetics Consortium (psychiatric disease), https://www.med.yale.edu/pgc; results and downloads; SSAGC (educational attainment), https://www.thessagc.org/data; GIANT Consortium, https://portals.broadinstitute.org/collaboration/giant; Early Growth Genetics Consortium, https://egg-consortium.org; birth weight-2016.html; Rheumatoid arthritis, http://plaza.umin.ac.jp/~yokada/datasource/software.htm; International Inflammatory Bowel Disease Genetics Consortium, https://www.ibdgenetics.org/.

Code availability

MixeR software and a tutorial example on how to use it are available online (https://github.com/precimed/mixer, v0.9.1); PLINK software (v1.0b5.2, 64-bit, 9 Jan 2018), https://www.cog-genomics.org/plink2; LD score regression software, v1.0.0, https://github.com/bulik/sdsc; HapGen2 software, v2.2.0, http://methagen.stats.ox.ac.uk/ genetics_software; pipeline to harmonize GWAS summary statistics: https://github.com/precimed/python_convert (v0.9.1); pipeline to simulate synthetic GWAS data from genotypes: https://github.com/precimed/simu (v0.9.3).

Received: 30 November 2018 Accepted: 29 April 2019

Published online: 03 June 2019

References

1. Visscher, P. M. et al. 10 Years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet. 101, 5–22 (2017).
2. Sivakumar, S. et al. Abundant pleiotropy in human complex diseases and traits. Am. J. Hum. Genet. 89, 607–618 (2011).
3. Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. & Smoller, J. W. Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. 14, 483–495 (2013).
4. Schork, A. J., Wang, Y., Thompson, W. K., Dale, A. M. & Andreasen, O. A. New statistical approaches exploit the polygenic architecture of schizophrenia—implications for the underlying neurobiology. Curr. Opin. Neurobiol. 36, 89–98 (2016).
5. Pasanic, B. & Price, A. L. Dissecting the genetics of complex traits using association summary statistics. Nat. Rev. Genet. 18, 117–127 (2017).
6. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. PLoS. Genet. 9, e1003348 (2013).
7. Ang, J., Lee, S. H., Godfrey, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011).
8. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).
9. Verhulst, B., Maes, H. H. & Neale, M. C. GW-SEM: a statistical package to conduct genome-wide structural equation modeling. Behav. Genet. 47, 345–359 (2017).
10. Davey Smith, G. & Hemin, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. 23, R89–R98 (2014).
11. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet. 50, 229–237 (2018).
12. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50, 1112–1121 (2018).
with General Electric Healthcare and Medtronic, Inc. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict of interest policies. Dr. Andreassen is a consultant for HealthLytix. The remaining authors have no competing interest.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Journal peer review information:** *Nature Communications* thanks Paul O’Reilly and other anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019