ARTICLE

Improving genetic prediction by leveraging genetic correlations among human diseases and traits

Robert M. Maier [1,2,3], Zhihong Zhu[4], Sang Hong Lee[1,5], Maciej Trzaskowski[4], Douglas M. Ruderfer[6], Eli A. Stahl[7], Stephan Ripke[2,3,8], Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Naomi R. Wray [1,4], Jian Yang [1,4], Peter M. Visscher [1,4] & Matthew R. Robinson[4,9,10]

Genomic prediction has the potential to contribute to precision medicine. However, to date, the utility of such predictors is limited due to low accuracy for most traits. Here theory and simulation study are used to demonstrate that widespread pleiotropy among phenotypes can be utilised to improve genomic risk prediction. We show how a genetic predictor can be created as a weighted index that combines published genome-wide association study (GWAS) summary statistics across many different traits. We apply this framework to predict risk of schizophrenia and bipolar disorder in the Psychiatric Genomics consortium data, finding substantial heterogeneity in prediction accuracy increases across cohorts. For six additional phenotypes in the UK Biobank data, we find increases in prediction accuracy ranging from 0.7% for height to 47% for type 2 diabetes, when using a multi-trait predictor that combines published summary statistics from multiple traits, as compared to a predictor based only on one trait.

[1] Queensland Brain Institute, University of Queensland, Queensland QLD 4072, Australia. [2] Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA 02142, USA. [3] Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA. [4] Institute for Molecular Bioscience, University of Queensland, Queensland QLD 4072, Australia. [5] Centre for Population Health Research, School of Health Sciences and Sansom Institute of Health Research, University of South Australia, Adelaide, SA 5000, Australia. [6] Division of Genetic Medicine, Department of Medicine, Psychiatry and Biomedical Informatics, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37235, USA. [7] Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [8] Department of Psychiatry and Psychotherapy, Charité, Campus Mitte, 10117 Berlin, Germany. [9] Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland. [10] Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland. Correspondence and requests for materials should be addressed to R.M.M. (email: rmaier@broadinstitute.org) P.M.V.(email: peter.visscher@uq.edu.au) M.R.R.(email: matthew.robinson@unil.ch). A full list of consortium members appears at the end of the paper.

Personalised medicine, in which genetic testing is the basis for informing future health status and determining intervention, is effectively applied for a number of monogenic disorders[1]. For common complex disorders, which are those that are underlain by multiple genetic and environmental factors[2], predictive genetic testing that can discriminate individuals who are most at risk is currently limited, mainly because much of the genetic variation remains poorly understood[3,4]. The potential of genetic risk prediction to (i) inform early interventions and (ii) aid diagnosis by identifying individuals with an increased genetic risk of disease could be improved substantially by increasing the accuracy of genetic risk predictors[5]. While genome-wide association studies (GWASs) of increased sample size will continue to unravel the role of genetic factors for complex diseases[6], improved prediction models are also required to maximise the accuracy of a risk predictor.

GWASs use linear regression to independently estimate the effects of single-nucleotide polymorphisms (SNPs) across the genome, and commonly, these estimated SNP effects are then used to create a genetic risk predictor in independent samples[7–9]. However, this approach is not optimal because it either ignores linkage disequilibrium (LD) between markers or accounts for LD by discarding potentially informative SNPs[10]. Prediction accuracy of complex phenotypes can be improved by methods that jointly estimate the SNP associations to obtain SNP effect estimates with best linear unbiased predictor (BLUP) properties within a linear mixed model (LMM) approach, a model termed genomic BLUP (GBLUP)[7,11,12]. A multi-trait extension of the LMM approach, yielding multivariate BLUP (MT-BLUP) predictors of the SNP effects, can further improve prediction accuracy when phenotypes are genetically correlated, because measurements on each trait provide information on the genetic values of the other correlated traits[13–16]. MT-BLUP has been shown to improve prediction accuracy for genetically correlated common psychiatric disorders when combining individual-level data across independent data sets[16,17]. However, the application of MT-BLUP to complex common disorders is limited as combining individual-level genotype-phenotype data across case–control studies of all complex diseases is generally not feasible due to data protection concerns and restrictions on data sharing.

Here we overcome this limitation by developing a framework that combines publically available GWAS summary statistics across multiple studies of different traits together in a weighted index to generate approximate multi-trait summary statistic BLUP (wMT-SBLUP) predictors (Supplementary Table 1). We show through theory and simulation study that MT-BLUP predictors, which traditionally require individual-level phenotype–genotype data for all traits, can be approximated accurately by wMT-SBLUP predictors in a computationally efficient manner using only summary statistic data and an independent genomic reference sample. We also show how multi-trait summary statistic predictors can be created directly from GWAS summary statistics (wMT-GWAS) or from predictors obtained using the software LDPred[18] that extends a single-trait summary statistic BLUP model (SBLUP) by assuming that marker effects come from a mixture of distributions. We apply our approach to multiple phenotypes in the Psychiatric Genomics Consortium (PGC) to compare summary statistic approaches to direct estimation on individual-level data. We further apply our approach to summary statistics of several other phenotypes to create predictors that we evaluate using the UK Biobank data. We show that, for most traits, our multi-trait predictors improve prediction accuracy as compared to a single-trait predictors.

Results

Overview of the approach. Standard GWAS summary statistics are ordinary least squares (OLS) estimates of the SNP effects and do not have optimal properties for prediction[11]. Even when LMM association analysis is used, the estimated SNP effects still represent marginal effects and not effects conditional on other SNPs, which is what is desirable for prediction[19]. Previous studies have shown how OLS summary statistics can be reanalysed in a mixed model framework to produce approximate BLUP predictors (summary statistic BLUP: SBLUP, implemented in the most recent release of GCTA)[18,20,21] or approximate mixture model predictors (LDPred). We first extend the SBLUP approach to a multi-trait framework (MT-SBLUP) and find a computational limitation associated with the inversion of a SNP-by-SNP-by-trait matrix. To overcome this, we then derive theory to show how single-trait predictors with BLUP properties can be combined together in a weighted index to generate predictors with equivalent properties to those gained from a MT-BLUP analysis (Fig. 1).

Consider two genetically correlated traits for which we have individual-level genetic predictors with BLUP properties. For each individual, i, and focal trait of interest, f, we have a genetic prediction $\left(\widehat{\mathbf{g}}_{\mathrm{BLUP}_{i,k}}\right)$ for each trait, k, that we can combine together using the index weights, $w_{i,k}$, for each $\widehat{\mathbf{g}}_{\mathrm{BLUP}_{i,k}}$ effect to produce a weighted multi-trait BLUP genetic predictor:

$$\widehat{\mathbf{g}}_{\mathrm{wMT-BLUP}_{i,f}} = \sum_k w_{i,k}\widehat{\mathbf{g}}_{\mathrm{BLUP}_{i,k}} = \mathbf{w}_i\widehat{\mathbf{g}}_{\mathrm{SBLUP}_i} \tag{1}$$

In the Methods section, we show that the optimal index weights can be calculated as:

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} R_1^2 & \frac{r_{\mathrm{G}}R_1^2R_2^2}{\sqrt{h_1^2h_2^2}} \\ \frac{r_{\mathrm{G}}R_1^2R_2^2}{\sqrt{h_1^2h_2^2}} & R_2^2 \end{bmatrix}^{-1} \begin{bmatrix} R_1^2 \\ r_{\mathrm{G}}\sqrt{\frac{h_1^2}{h_2^2}}R_2^2 \end{bmatrix} \tag{2}$$

where h_k^2 is the SNP heritability of trait k (proportion of phenotypic variance explained by genome-wide SNPs), r_{G} is the genetic correlation between trait k and the focal trait and R_k^2 is the expected squared correlation between a phenotype and a BLUP predictor, calculated as:

$$R_k^2 = \frac{h_k^2}{1 + M_{\mathrm{eff}}\frac{1-R_k^2}{N_kh_k^2}} \tag{3}$$

where M_{eff} is the effective number of chromosome segments and N_k is the sample size of trait k. These weights will ensure that the contribution of each added trait is approximately proportional to the square root of its sample size, its SNP heritability and its genetic correlation with the focal trait (trait 1), while accounting for different variances of single-trait BLUP predictors.

Both h_k^2 and r_{G} can be estimated from GWAS summary statistics using LD score regression[22,23]. Following[20], individual-level genetic predictors with BLUP properties can also be obtained from GWAS summary statistics ($\widehat{\mathbf{g}}_{\mathrm{SBLUP}_k}$, where SBLUP represents summary statistic approximate BLUP). Therefore, for any given trait, genetic predictors with BLUP properties $\left(\widehat{\mathbf{g}}_{\mathrm{SBLUP}_k}\right)$ can be created from GWAS summary statistics and these can then be placed in a weighted index to produce approximate multi-trait summary statistic BLUP (wMT-SBLUP) predictors, using only LD score regression and an independent reference sample. This approach, provided in the freely available software SMTPred (see Code availability section), approximates MT-BLUP predictors without the need for individual-level
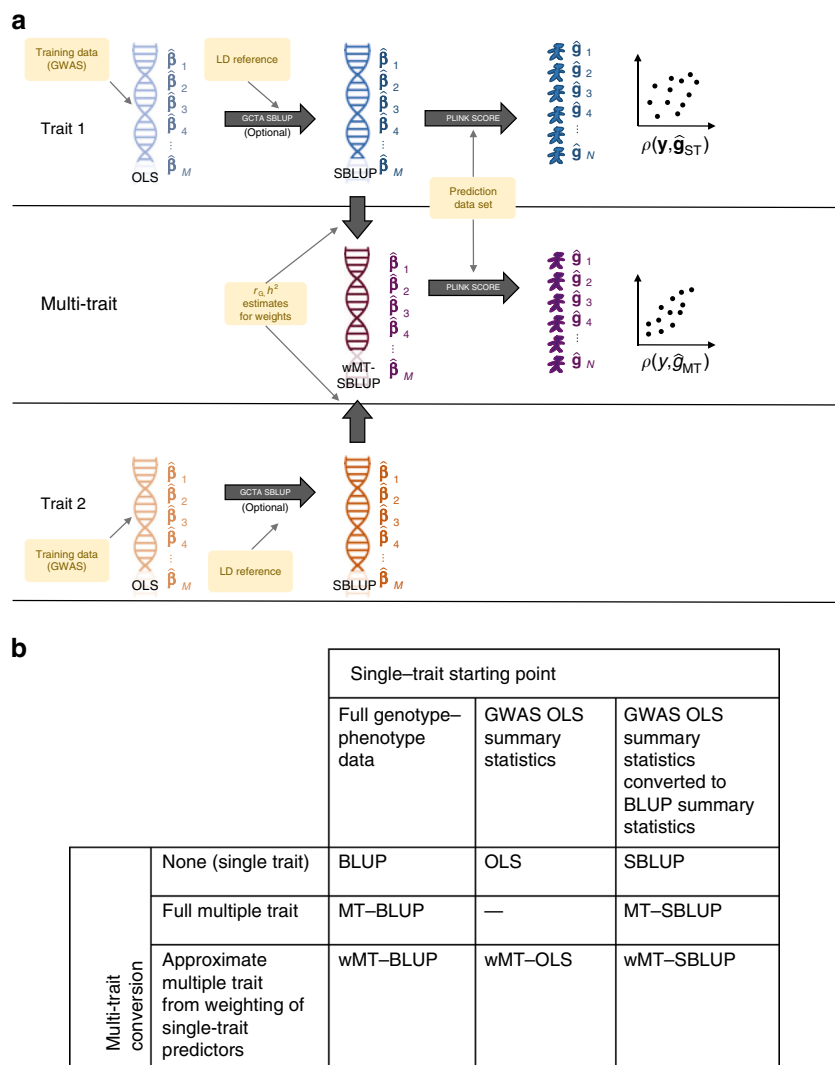
Fig. 1 Schematic of the methods. **a** Data and programs used to create predictors. **b** Terminology to refer to different types of predictors. OLS, ordinary least squares. The most common GWAS methodology to estimate SNP effects is to estimate the effect sizes of one SNP at a time using linear regression. BLUP, best linear unbiased prediction. SNP effects are estimated simultaneously for all SNPs. The estimates depend on the other SNPs included in the analysis, since the contribution from correlated SNPs will be shared between them

phenotype–genotype data for all traits, enabling prediction accuracy to be improved by fully utilising all of the publically available GWAS summary statistic data. We also show how weighted indices can be calculated for GWAS summary statistics (wMT-GWAS) or from predictors obtained using the software LDPred[18] (wMT-LDPred), therefore depending upon the genetic architecture of the trait approximate multi-trait summary statistics can be created to maximise genomic prediction accuracy.

Simulation study. We first conducted a simulation study using observed SNP genotype data to confirm the expectations from our theory. We show through theory (see Methods section) that a wMT-SBLUP genetic predictor has the same expected prediction accuracy as one created from a multivariate mixed-effects model (multi-trait BLUP: MT-BLUP) if the linkage disequilibrium among SNP markers in the individual-level analysis is well approximated by a reference genotype panel (see Methods section). We demonstrate that a wMT-SBLUP predictor increases prediction accuracy over a single-trait predictor, with the

magnitude of increase being proportional to the ratio of the SNP heritability of the added traits relative to that of the predicted trait, the sample size of the added traits relative to that of the predicted trait and the genetic correlation between the added traits and the predicted trait (Fig. 2, Supplementary Figs. 1 and 2). We also demonstrate how genetic predictors generated by LDPred[18] can be combined in an approximate multi-trait weighting (Supplementary Fig. 3).

We also provide a theoretical expectation for the loss in prediction accuracy that occurs when using an independent reference sample to compute SBLUP effects compared to a predictor based on BLUP effects (see Methods section), and we detail the loss of prediction accuracy in our simulation study (Fig. 2b, Supplementary Figs. 1 and 4).

Application to psychiatric disorders. We then applied our approach to the PGC schizophrenia[24,25] and bipolar data, two psychiatric disorders known to have a high genetic correlation[26]. The availability of combined individual-level data for both disorders enabled a direct comparison of the MT-BLUP[16] and
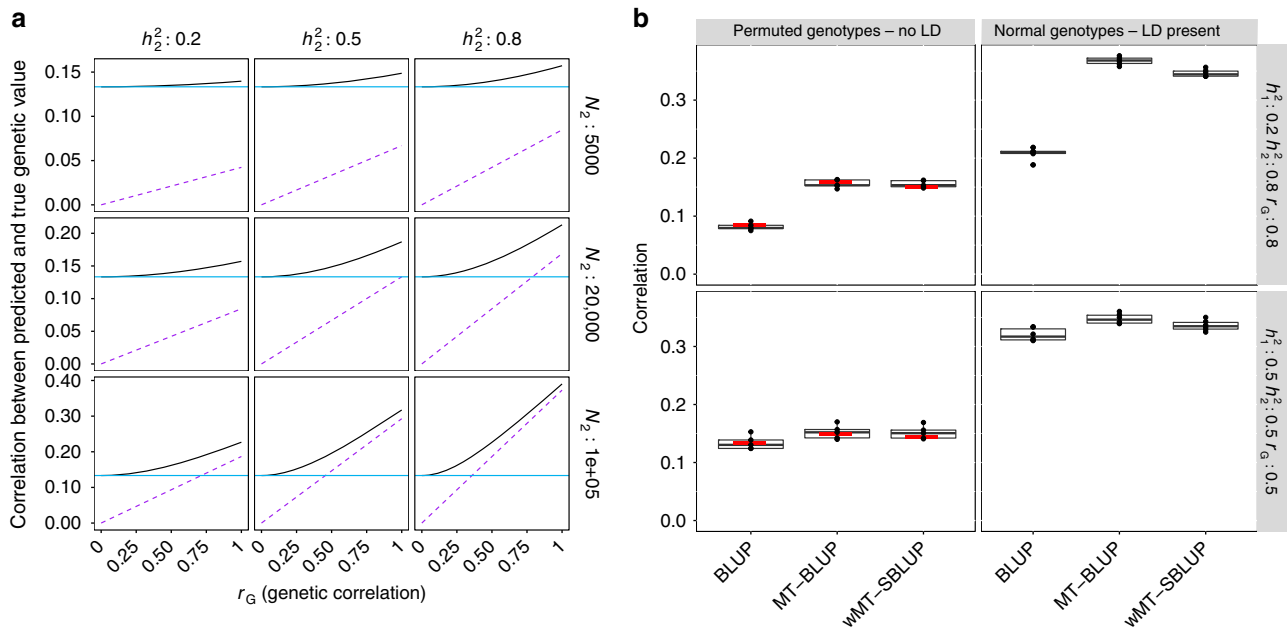
Fig. 2 Improving prediction accuracy using information from multiple traits. **a** Expected gain from multi-trait vs cross-trait predictors as a function of r_G. Two traits are considered. The first trait has a sample size of 20,000 and a SNP heritability of 0.5. The sample size and SNP heritability of the second trait vary between panels. The blue line shows the expected prediction accuracy of a single-trait predictor. The black line shows the expected prediction accuracy of a multi-trait predictor. The purple line shows the expected prediction accuracy of a cross-trait predictor (using only trait 2 to predict trait 1). The advantage of a multi-trait predictor over a cross-trait predictor decreases with increasing r_G, h^2, and sample size of the second trait. **b** Simulation results. Prediction accuracy is shown as correlation between simulated genetic value and predicted phenotype of individuals. Genotypes from European individuals in the GERA cohort were used for simulation. Boxplots show results across six replicates. In the left panels, the LD structure was removed by permuting dosage values for each SNP across all individuals. In the right panels, the original genotypes were used for simulation. Expected prediction accuracies were derived for the case of unlinked genotypes and are shown as red horizontal bars. In each section, the prediction accuracy of three predictors is shown: (1) single trait BLUP, (2) multi-trait BLUP (MT-BLUP), and (3) weighted approximate BLUP (summary statistic-based multi-trait predictor: wMT-SBLUP). Simulation in genotypes without LD results in prediction accuracies, which conform to expectations. In the presence of LD, the expected prediction accuracy depends very much on the choice of M_{eff}

wMT-SBLUP approaches. We calculated all predictors for the previously used[16] PGC wave 1 (PGC1) data sets[24] and compared the prediction accuracy (correlation between predicted values and phenotypes adjusted for sex, cohort and the first 20 principal components) across diseases and approaches. We find comparable but slightly lower accuracies in the wMT-SBLUP predictors as compared to the MT-BLUP predictors (0.151 vs 0.156 in bipolar disorder and 0.217 vs 0.219 in schizophrenia) and an increase in prediction accuracy as compared to the single-trait (BLUP) predictors (0.128 in bipolar disorder, 0.198 in schizophrenia) (Fig. 3). Our results demonstrate that creating SBLUP genetic predictors using an independent LD reference sample and combining these in a weighted sum results in prediction accuracy comparable to a full MT-BLUP prediction for common complex disease traits, at a much lower computational burden.

We then applied our approach to the larger PGC wave 2 (PGC2) data sets for schizophrenia[25] and bipolar disorder (see Methods section), which included the PGC1 data. To test whether the addition of more cohorts improved prediction accuracy, we estimate wMT-SBLUP predictors in the PGC2 data. Having shown the resemblance of wMT-SBLUP and MT-BLUP by theory, simulation and in the PGC1 data, we refrained from running a MT-BLUP model in the PGC2 data to avoid the computational burden of analysing the combined schizophrenia bipolar data set. For schizophrenia, there were 36 cohorts (26,412 cases and 32,440 controls in total) and for bipolar disorder there were 23 cohorts (18,865 cases and 30,460 controls in total). We conducted a cohort-wise leave-one-out cross-validation approach to examine variation in prediction accuracy across cohorts.

For schizophrenia, we find that prediction accuracy increases in 20 of the 36 cohorts of the PGC2 data when using a wMT-SBLUP predictor as compared to a SBLUP predictor (Supplementary Fig. 5). However, the median correlation (0.300 with an SBLUP predictor, and 0.304 with a wMT-SBLUP predictor) and mean correlation (0.295 with a SBLUP predictor and 0.294 with a wMT-SBLUP predictor) across the 36 PGC2 cohorts did not improve with a wMT-BLUP predictor. For bipolar disorder, we find an improvement of the wMT-SBLUP predictor over the SBLUP predictor in 17 out of the 23 cohorts (Supplementary Fig. 6), with a mean correlation increase from 0.212 to 0.229 and a median correlation increase from 0.210 to 0.225. To evaluate whether this is because the weights we used for schizophrenia and bipolar disorder do not represent the mixing proportions that lead to the highest accuracy in this data set or whether other factors explain the variable results across cohorts, we created multi-trait predictors using not only weights calculated from Eq. (17) but also weights corresponding to any other mixing proportion of the two disorders (Supplementary Figs. 5, 6 and 7). This demonstrates (i) that our calculated weights are very close to the empirically optimal weights when averaged across cohorts (Supplementary Fig. 7), (ii) that there is substantial heterogeneity across cohorts as shown by the variable prediction accuracies of single-trait and cross-trait predictors across cohorts, which is supported by previous studies[25], and (iii) that, for some test set cohorts, there is no mixing proportion that will lead to a multi-trait predictor which outperforms a single-trait predictor. The larger gain in accuracy that results from supplementing a bipolar disorder predictor with schizophrenia data compared to supplementing a schizophrenia predictor with bipolar disorder
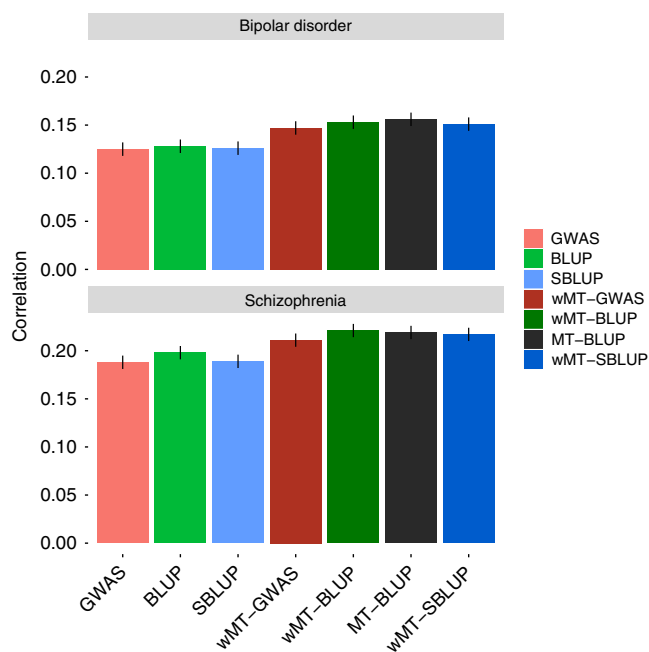
Fig. 3 Prediction accuracy for schizophrenia and bipolar disorder from several single-trait and multi-trait predictors. Prediction accuracy of seven different types of predictors using PGC1 schizophrenia and bipolar disorder data. Single-trait predictor (lighter colours) are on the left, multi-trait predictors (darker colours) are on the right. Black error bars indicate correlation coefficient standard errors, calculated as $se_r = \sqrt{\frac{1-r^2}{n-2}}$

data is consistent with greater power of the schizophrenia discovery sample. We find that for both single-trait and multi-trait predictors the SBLUP predictors outperform the OLS predictors in almost all cohorts (Supplementary Figs. 5 and 6).

Application to traits recorded in a large population study. In principle, any number of traits can be combined into a multi-trait predictor at almost no computational cost. We therefore extended our approach to create wMT-SBLUP predictors from 34 phenotypes for which we could access summary statistics. In order to calculate wMT-SBLUP weights, we used LD score regression to estimate SNP heritability and genetic correlations of the 34 summary statistics traits. The results are mostly in line with previous reports[23] (Supplementary Fig. 8, Supplementary Data 1). As test set, we used 112,338 individuals in the UK Biobank data. We matched 6 of the 34 discovery traits to traits in the UK Biobank (Supplementary Table 1) and created wMT-SBLUP predictors. For the wMT-SBLUP predictor of each focal trait, we included predictor traits with genetic correlation p-value < 0.05. For all traits, wMT-SBLUP genetic predictors were more accurate than any single-trait (SBLUP) predictor (Fig. 4). wMT-SBLUP predictors generally improved prediction accuracy over single-trait GWAS OLS predictors (Supplementary Fig. 9) and were similar to wMT-LDPred predictors (Supplementary Figs. 10 and 11.) We observe the largest increases in accuracy for Type 2 diabetes (47.8%) and depression (34.8%). Accuracy for height (0.7%) and body mass index (BMI) (1.4%) increase only marginally. As shown in our theory and simulation study, the magnitude of increase in prediction accuracy of a wMT-SBLUP predictor over a single-trait SBLUP predictor depends upon the prediction accuracies of all the traits included in the index and the genetic correlation among phenotypes. As GWAS sample sizes increase and genomic predictors increase in accuracy, a wMT-SBLUP approach will likely become increasingly beneficial.

Discussion

In summary, we demonstrate that multivariate predictors derived from GWAS summary statistics can increase prediction accuracy in a wide range of traits. This approach has particular utility in risk prediction of traits for which it is hard to generate large sample sizes for GWAS, as SNP heritability and sample size are the two factors that determine prediction accuracy of a polygenic trait, when using a single-trait predictor. The increase in prediction accuracy of a multi-trait over a standard single-trait genetic predictor is therefore greatest when the additional traits included in the predictor have higher SNP heritability and sample size than the trait to be predicted, as well as a high genetic correlation with the trait to be predicted. We show how genetic predictors from GWAS OLS effects, LDPred effects or SBLUP effects can be combined, yielding an approach that is general across different phenotypes.

Special consideration should be given to the risk of sample overlap between the summary statistics data used to create the predictor and the prediction target. Sample overlap will lead to inflated estimates of accuracy, and while here we were able to take steps to avoid individuals being recorded across multiple data sets, further work is required to negate these effects within this framework. In principle, assuming perfect homogeneity between training and test set and perfect estimates of SNP heritability and genetic correlation, there is no limit to the number of traits that can be combined using our approach. In practice, however, there will be little benefit of combining traits with low genetic correlation, as they will not influence the predictor much. Some added traits might even reduce accuracy, if the genetic correlation is not estimated accurately. The focus of our analysis was the prediction of genetic risk and we aimed to provide a fast, computationally efficient, general framework for genomic prediction. This sets it apart from other multi-trait approaches like phenome-wide association studies, which focus on the effects of individual SNPs on multiple phenotypes. We note, however, that a multi-trait testing approach can in principle also be used to increase the power to identify loci associated with specific traits as demonstrated in the recently developed MTAG method[27]. Another potential caveat of our analysis is that prediction accuracy increases for a focal trait may come from the addition of traits that are standardly measured on patients, and improved frameworks are required to identify marker effects conditionally on known health risk factors. Despite these limitations, current evidence suggests that genetic correlations among phenotypes are pervasive[23], sample sizes of GWAS are increasing[6] and public availability of genome-wide summary statistics is becoming the norm[28], meaning that genomic prediction of complex common disease will continually improve especially when predictors of multiple phenotypes are integrated across studies within this framework.

Methods

General model. We consider a general linear mixed model:

$$\mathbf{y} = \mathbf{Wb} + \quad (4)$$

where $\mathbf{y}$ is the phenotype, $\mathbf{W}$ a matrix of SNP genotypes, where values are standardised to give the ij^{th} element as: $w_{ij} = (x_{ij} - 2p_j)/\sqrt{2p_j(1 - p_j)}$, with x_{ij} the number of minor alleles (0, 1 or 2) for the ith individual at the jth SNP and p_j the minor allele frequency. $\mathbf{b}$ are the genetic effects for each SNP, and the residual error. The dimensions of $\mathbf{y}$, $\mathbf{W}$, $\mathbf{b}$ and are dependent upon the number of phenotypes, k, the number of SNP markers, M, and the number of individuals, N, and are described in the sections below. We denote the distributional properties var($\mathbf{b}$) = $\mathbf{B}$, var() = $\mathbf{R}$ and var($\mathbf{y}$) = $\mathbf{WBW'} + \mathbf{R}$.

For human complex diseases and quantitative phenotypes, GWASs have typically estimated the solutions for $\mathbf{b}$ of Eq. (1) one SNP at a time using OLS regression[29] as:

$$\widehat{\mathbf{b}}_{OLS} = diag[\mathbf{W'W}]^{-1}\mathbf{W'y} \quad (5)$$
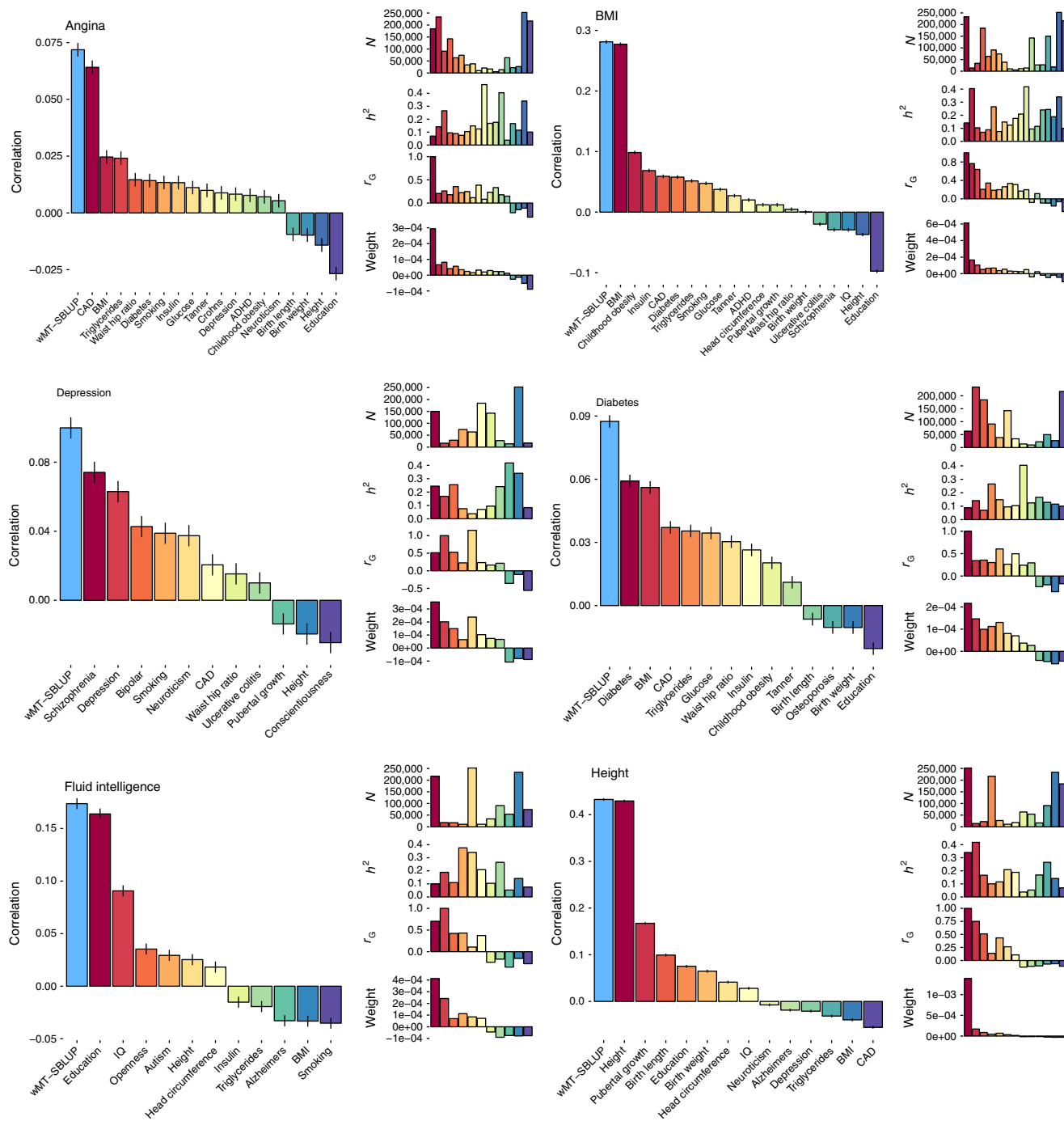
Fig. 4 Prediction accuracy for single-trait and multi-trait predictors in UK Biobank traits. Prediction accuracy for six traits in the UK Biobank for multi-trait predictors (light blue bars, wMT-SBLUP) and single-trait predictors (colourful bars on the right, SBLUP). Black bars show the correlation coefficient standard error. The multi-trait predictors for each trait are composed of all traits for which colourful bars are shown (r_G p-value < 0.05). Smaller bars on the right show, from top to bottom, sample size, SNP heritability, r_G, and weights (given by Eq. (15)) for each trait

where diag[$\mathbf{W'W}$] has diagonal elements $w_j'w_j$ and off-diagonal elements of zero. However, by analysing one SNP at a time, GWAS effect size estimates do not account for the covariance structure among SNPs and they are not unbiased in the sense that $E\left[\mathbf{b}|\hat{\mathbf{b}}\right]=\hat{\mathbf{b}}$[12]. BLUP of the SNP effects have the property $E\left[\mathbf{b}|\hat{\mathbf{b}}\right]=\hat{\mathbf{b}}$, are used in genomic prediction in animal and plant breeding[30] and more recently in human medical genetics, yielding improved prediction accuracy for a number of traits over genetic predictors created from OLS SNP estimates[16,17]. In a general form, BLUP solutions for $\mathbf{b}$ of Eq. (1) can be written using Henderson's mixed model equations[31] as:

$$\hat{\mathbf{b}}_{\mathrm{BLUP}} = \left[\mathbf{W'R}^{-1}\mathbf{W}+\mathbf{B}^{-1}\right]^{-1}\mathbf{W'R}^{-1}\mathbf{y} \qquad (6)$$

and if $\mathbf{R}$ is diagonal, then Eq. (6) can be reduced to:

$$\hat{\mathbf{b}}_{\mathrm{BLUP}} = \left[\mathbf{W'W}+\mathbf{B}^{-1}\mathbf{R}\right]^{-1}\mathbf{W'y} \qquad (7)$$

Below, we describe how Eqs. (6) and (7) can be used to estimate BLUP SNP effects for a single trait and for multiple traits jointly from individual-level phenotype–genotype data. We then show how Eqs. (6) and (7) can be approximated to obtain BLUP SNP effects for single and multiple traits in the absence of individual-level data from publically available GWAS summary statistics and an independent reference sample.

Estimation of BLUP SNP effects for a single trait. For a univariate analysis of trait k, $\mathbf{y}$ of Eq. (4) is a column vector of length $N \times 1$ and $\mathbf{W}$ has dimension $N \times M$. Assuming $\mathbf{b}$ is an $M \times 1$ vector of random SNP effects for trait k, with distribution $\mathbf{b} \sim N\left(0, \mathbf{I}_M \sigma_{b_k}^2\right)$, then $\mathbf{B} = \mathbf{I}_M \sigma_{b_k}^2$ with $\mathbf{I}_M$ is an identity matrix of dimension M. of Eq. (1) is a column vector of independent residual effects, with distribution $\sim N\left(0, \mathbf{I}_N \sigma_{\epsilon_k}^2\right)$, giving $\mathbf{R} = \mathbf{I}_N \sigma_{\epsilon_k}^2$, with $\mathbf{I}_N$ an identity matrix of dimension N. Substituting these expressions into Eq. (6) means that Eq. (7) can then be written as:

$$\widehat{\mathbf{b}}_{\mathrm{BLUP}_k} = \left[\mathbf{W}_k'\mathbf{W}_k + \mathbf{I}_M \lambda_k\right]^{-1} \mathbf{W}_k'\mathbf{y}_k \quad (8)$$

with $\lambda_k = \sigma_{\epsilon_k}^2 / \sigma_{b_k}^2$.

Joint estimation of BLUP SNP effects for multiple traits. Phenotypic measurements of a trait can be informative for the genetic values of other traits, if the traits are genetically correlated with one another[14,15,32]. Recent studies have shown that prediction accuracy of common complex disease can be improved by estimating SNP effects for multiple traits jointly within a multivariate mixed-effects model[16,17].

If k traits are measured on different individuals, with N_k observations for trait k, the elements of Eq. (4) become: $\mathbf{y}' = [\mathbf{y}'_1 \dots \mathbf{y}'_k]$, $\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{W}_k \end{bmatrix}$, and $\mathbf{R} = \mathrm{diag}[\mathbf{R}_k] = \mathrm{diag}\left[\mathbf{I}_{N_k}\sigma_{\epsilon_k}^2\right]$, a diagonal matrix of length $N = \sum_k N_k$. $\mathbf{B} = \mathbf{\Sigma}_b \otimes \mathbf{I}_M$, where $\mathbf{\Sigma}_b$ is a $k \times k$ matrix, with diagonal elements $\sigma_{b_k}^2$ and off-diagonal elements the covariances of SNP effects between traits k and l, $\sigma_{b_{k,l}}$. For Kronecker products, $\mathbf{B}^{-1} = \mathbf{\Sigma}_b^{-1} \otimes \mathbf{I}_M$ and substituting these expressions directly into Eq. (6) means that multi-trait BLUP solutions for $\mathbf{b}$ can be obtained in Eq. (7) as:

$$\widehat{\mathbf{b}}_{\mathrm{MT-BLUP}} = \left[\mathbf{W}'\mathbf{W} + \mathbf{\Sigma}_\epsilon \mathbf{\Sigma}_b^{-1} \otimes \mathbf{I}_M\right]^{-1} \mathbf{W}'\mathbf{y} \quad (9)$$

with $\mathbf{\Sigma}_\epsilon = \mathrm{diag}\left[\sigma_{\epsilon_k}^2\right]$, a diagonal $k \times k$ matrix. For a two-trait example, Eq. (9) expands to:

$$\widehat{\mathbf{b}}_{\mathrm{MT-BLUP}} = \left[\begin{bmatrix} \mathbf{W}_1'\mathbf{W}_1 & 0 \\ 0 & \mathbf{W}_2'\mathbf{W}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{I}_M \sigma_{\epsilon_1}^2 & 0 \\ 0 & \mathbf{I}_M \sigma_{\epsilon_2}^2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_M \sigma_{b_1}^2 & \mathbf{I}_M \sigma_{b_{1,2}} \\ \mathbf{I}_M \sigma_{b_{2,1}} & \mathbf{I}_M \sigma_{b_2}^2 \end{bmatrix}^{-1}\right]^{-1} \begin{bmatrix} \mathbf{W}_1 & 0 \\ 0 & \mathbf{W}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad (10)$$

Multi-trait BLUP SNP effects from summary statistics. Estimating SNP effects for multiple traits jointly in Eq. (9) requires individual-level genotype and phenotype data across a range of common complex diseases and quantitative phenotypes, which are not readily available in human medical genetics due to privacy concerns and data sharing restrictions. Additionally, Eq. (9) requires a series of computationally intensive $M \times k$ equations to be solved. However, these issues can be overcome by approximating Eq. (9) using publically available GWAS summary statistic data and an independent genomic reference sample.

Single-trait approximate BLUP SNP effects can be obtained from GWAS summary statistics (SBLUP: summary statistic approximate BLUP) by replacing $\mathbf{W}_k'\mathbf{W}_k$ and $\mathbf{W}_k'\mathbf{y}_k$ of Eq. (8) by their expectation, which are $\mathbb{E}[\mathbf{W}_k'\mathbf{W}_k] = N_k \mathbf{L}$ and $\mathbb{E}[\mathbf{W}_k'\mathbf{y}_k] = N_k \widehat{\mathbf{b}}_{\mathrm{OLS}_k}$, respectively, where $\mathbf{L}$ is an $M \times M$ scaled SNP LD correlation matrix estimated from a reference SNP data set and $\widehat{\mathbf{b}}_{\mathrm{OLS}_k}$ are obtained from publically available GWAS summary statistics[20]. GWAS summary statistics report effect estimates of SNPs on an unstandardised scale and not $\widehat{\mathbf{b}}_{\mathrm{OLS}}$ as it is defined here. To obtain $\widehat{\mathbf{b}}_{\mathrm{OLS}}$ from GWAS summary statistics, the effect of each SNP must be multiplied by the standard deviation of each SNP: $\widehat{\mathbf{b}}_{\mathrm{OLS}_j} = \widehat{\mathbf{b}}_{\mathrm{OLS-UNSCALED}_j} \times \sqrt{2p_j(1-p_j)}$. Equation (8) can then be written as:

$$\widehat{\mathbf{b}}_{\mathrm{SBLUP}_k} = \left[N_k\mathbf{L} + \mathbf{I}_M\lambda_k\right]^{-1} N_k \widehat{\mathbf{b}}_{\mathrm{OLS}_k}$$
$$= \left[\mathbf{L} + \mathbf{I}_M \lambda_k/N_k\right]^{-1} \widehat{\mathbf{b}}_{\mathrm{OLS}_k} \quad (11)$$

The shrinkage parameter is $\lambda_k = \sigma_{\epsilon_k}^2/\sigma_{b_k}^2 = M\sigma_{\epsilon_k}^2/h_{\mathrm{SNP}_k}^2 = M\left(1 - h_{\mathrm{SNP}_k}^2\right)/h_{\mathrm{SNP}_k}^2$, under the assumption of phenotypic variance of 1 that makes the proportion of phenotypic variance of trait k attributable to the SNPs $h_{\mathrm{SNP}_k}^2 = M\sigma_{b_k}^2$.

This approach was implemented in ref. [21] and is similar to the LDpred model presented by Vilhjálmsson et al.[18] but with a few differences. The first is that it only considers the infinitesimal case, where all SNPs are considered to be causal and

their effect sizes follow a normal distribution. This corresponds to the LDpred-Inf model. The second difference is that the shrinkage parameter of Vilhjálmsson et al.[18] is $\lambda_k = M/h_{\mathrm{SNP}_k}^2$ as they assume that the error variance is 1 rather than $1 - h_{\mathrm{SNP}_k}^2$ in our implementation. The third difference is that in the LDpred-Inf model, Vilhjálmsson et al.[18] calculate BLUP effects for blocks of a certain number of SNPs following a tiling window approach giving a block diagonal structure to $\mathbf{L}$, whereas our implementation within the software GCTA (see URLs) follows a sliding window approach giving a banded diagonal to $\mathbf{L}$. Assuming an error variance of $1 - h_{\mathrm{SNP}_k}^2$ is more appropriate because cumulatively the SNP markers explain $h_{\mathrm{SNP}_k}^2$ of the phenotypic variance. In both implementations, a window is used to capture the LD around SNP markers in order to avoid the large computational costs of inverting a dense M dimensional SNP LD matrix, with only little loss of information (see below).

For multiple phenotypes, the elements of Eq. (11) become: $\widehat{\mathbf{b}}_{\mathrm{OLS}'} = \left[\widehat{\mathbf{b}}_{\mathrm{OLS}_1'} \dots \widehat{\mathbf{b}}_{\mathrm{OLS}_k'}\right]$ and $\mathbf{N} = \begin{bmatrix} \mathbf{N}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbf{N}_k \end{bmatrix}$, meaning that Eq. (11) can be extended as:

$$\widehat{\mathbf{b}}_{\mathrm{MT-SBLUP}} = \left[\mathbf{I}_k \otimes \mathbf{L} + \mathbf{\Sigma}_\epsilon \mathbf{\Sigma}_b^{-1} \mathbf{N}^{-1} \otimes \mathbf{I}_M\right]^{-1} \widehat{\mathbf{b}}_{\mathrm{OLS}} \quad (12)$$

Equation (12) approximates Eq. (9) using only publically available GWAS summary statistic data and an independent genomic reference sample. However, there remains the large computational cost associated with the inversion of the non-diagonal matrix $\left[\mathbf{I}_k \otimes \mathbf{L} + \mathbf{\Sigma}_\epsilon \mathbf{\Sigma}_b^{-1} \mathbf{N}^{-1} \otimes \mathbf{I}_M\right]$.

Index weighted multi-trait BLUP SNP effects. An alternative to Eq. (12), is to obtain k $\widehat{\mathbf{b}}_{\mathrm{MT-SBLUP}}$ effects by combining together k single-trait $\widehat{\mathbf{b}}_{\mathrm{SBLUP}}$ estimates of Eq. (11), using an optimal index weighting for each trait. The index weighting to derive $\widehat{\mathbf{b}}_{\mathrm{MT-SBLUP}}$ from $\widehat{\mathbf{b}}_{\mathrm{SBLUP}}$ estimates is identical to the index weighting to derive $\widehat{\mathbf{b}}_{\mathrm{MT-BLUP}}$ from $\widehat{\mathbf{b}}_{\mathrm{BLUP}}$ estimates.

For SNP j and focal trait f, we have $\widehat{\mathbf{b}}_{\mathrm{SBLUP}}$ values for k traits, and we wish to obtain the index weights, $w_{j,k}$, for each $\widehat{\mathbf{b}}_{\mathrm{SBLUP}_{j,k}}$ effect as:

$$\widehat{\mathbf{b}}_{\mathrm{wMT-SBLUP}_{j,f}} = \sum_k \mathbf{w}_{\mathrm{SBLUP},j,k} \widehat{\mathbf{b}}_{\mathrm{SBLUP}_{j,k}} = \mathbf{w}_{\mathrm{SBLUP},j}' \widehat{\mathbf{b}}_{\mathrm{SBLUP}_j} \quad (13)$$

In animal and plant breeding, selection indices have been developed, which combine many single-trait BLUP predictors of an individual's genetic value together in an index weighting to optimise the selection of individuals with the most favourable multi-trait phenotype for breeding programs[33–36]. Utilising a selection index approach, the solution for $\mathbf{w}_{\mathrm{SBLUP}}$ of Eq. (13) can be obtained as:

$$\mathbf{w}_{\mathrm{SBLUP}} = \mathbf{V}_{\mathrm{SBLUP}}^{-1} \mathbf{C}_{\mathrm{SBLUP}} \quad (14)$$

where $\mathbf{C}_{\mathrm{SBLUP}}$ a $k \times 1$ column vector of the covariance of the $\widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}$ values of the k traits, with the true genetic effects of the SNPs for the focal trait, and $\mathbf{V}_{\mathrm{SBLUP}}$ a $k \times k$ variance–covariance matrix of the $\widehat{\mathbf{b}}_{\mathrm{SBLUP}}$ effects:

$$\mathbf{w}_{\mathrm{SBLUP}} = \mathbf{V}_{\mathrm{SBLUP}}^{-1} \mathbf{C}_{\mathrm{SBLUP}}$$
$$= \begin{bmatrix} \mathrm{var}\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}_1}\right) & \cdots & \mathrm{cov}\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}_1}, \widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}, \widehat{\mathbf{b}}_{\mathrm{SBLUP}_1}\right) & \cdots & \mathrm{var}\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}\right) \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{cov}\left(\mathbf{b}_f, \widehat{\mathbf{b}}_{\mathrm{SBLUP}_1}\right) \\ \vdots \\ cov\left(\mathbf{b}_f, \widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}\right) \end{bmatrix} \quad (15)$$

Therefore, if $\mathbf{V}_{\mathrm{SBLUP}}$ and $\mathbf{C}_{\mathrm{SBLUP}}$ can be approximated then $\widehat{\mathbf{b}}_{\mathrm{MT-SBLUP}}$ of Eq. (12) can be obtained from k single-trait $\widehat{\mathbf{b}}_{\mathrm{SBLUP}}$ estimates from Eq. (11).

To derive the approximations, we first consider the diagonal elements of $\mathbf{V}_{\mathrm{SBLUP}}$, which comprise the variance of the SBLUP SNP solutions, $\mathrm{var}\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}\right)$. These can be approximated from theory under the assumption that $\widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}$ have BLUP properties $E\left[\mathbf{b}|\widehat{\mathbf{b}}\right] = \widehat{\mathbf{b}}$, which in turn implies that $\mathrm{cov}\left(\mathbf{b}_k, \widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}\right) = \mathrm{var}\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}_k}\right)$. Following Daetwyler et al.[37] and Wray et al.[38], the squared correlation between a phenotype, $\mathbf{y}_k$, in an independent sample and a single-trait BLUP predictor of the phenotype, $\widehat{\mathbf{g}}_{\mathrm{BLUP}_k}$, is approximately:

$$R_{\mathbf{y}_k, \widehat{\mathbf{g}}_{\mathrm{BLUP}_k}}^2 = R_k^2 \approx h_k^2 / \left(1 + M_{\mathrm{eff}}\left(1 - R_k^2\right) / \left(N_k h_k^2\right)\right) \quad (16)$$

where $\widehat{\mathbf{g}}_{\mathrm{BLUP}_k} = \mathbf{W}\widehat{\mathbf{b}}_{\mathrm{BLUP}_k}$ and h_k^2 is the proportion of phenotypic variance

attributable to additive genetic effects for trait k. Note that M_{eff} is the effective number of chromosome segments or the number of independent SNPs, which is a function of effective population size (N_e) and can be empirically obtained as an inverse of the variance of genomic relationships[39,40]. Here we use an estimate of M_{eff} of 60,000, which is in line both with our estimates from the genomic relationships in our simulation data and with previously reported estimates[41]. In Eq. (16), R_k^2 occurs on both the left- and righ-hand side. Solving for R_k^2 gives $R_k^2 = \frac{\varphi + h^2 - \sqrt{(\varphi + h^2)^2 - 4\varphi h^4}}{2\varphi}$, where φ is $\frac{M_{eff}}{N}$.

With a phenotypic variance of 1 and individual-level genetic effects $\mathbf{g}_k = \mathbf{W}\mathbf{b}_k$, then $h_k^2 = \sigma_{g_k}^2 = M\sigma_{b_k}^2$ and the squared correlation between the true, $\mathbf{g}_k$, and estimated BLUP effects, $\widehat{\mathbf{g}}_{\text{BLUP}_k}$, is:

$$R^2_{\mathbf{g}_k, \widehat{\mathbf{g}}_{\text{BLUP}_k}} = R_k^2 / h_k^2 \qquad (17)$$

Rearranging Eq. (17) gives $R_k^2 = h_k^2 R^2_{\mathbf{g}_k, \widehat{\mathbf{g}}_{\text{BLUP}_k}} = h_k^2 \frac{\text{cov}(\mathbf{g}_k, \widehat{\mathbf{g}}_{\text{BLUP}_k})^2}{\text{var}(\mathbf{g}_k)\text{var}(\widehat{\mathbf{g}}_{\text{BLUP}_k})}$, which given the BLUP properties $\text{cov}(\mathbf{g}_k, \widehat{\mathbf{g}}_{\text{BLUP}_k}) = \text{var}(\widehat{\mathbf{g}}_{\text{BLUP}_k})$ and $h_k^2 = \sigma_{g_k}^2$ with a phenotypic variance of 1, reduces to $R_k^2 = \text{cov}(\mathbf{g}_k, \widehat{\mathbf{g}}_{\text{BLUP}_k}) = \text{var}(\widehat{\mathbf{g}}_{\text{BLUP}_k}) = M\text{var}(\widehat{\mathbf{b}}_{\text{BLUP}_k})$. Therefore:

$$\text{var}(\widehat{\mathbf{b}}_{\text{BLUP}_k}) = \frac{\text{var}(\widehat{\mathbf{g}}_{\text{BLUP}_k})}{M} = \frac{R_k^2}{M} \qquad (18)$$

Second, we consider the off-diagonal elements of $\mathbf{V}_{\text{SBLUP}}$, which are comprised of the covariance of BLUP SNP solutions among the k traits. These can again be approximated from theory given the covariance of genetic effects among traits k and l is $\text{cov}(\mathbf{b}_k, \mathbf{b}_l) = r_G h_k h_l / M$, with r_G the genetic correlation, and given the squared correlation between the true genetic effects of the SNPs, $\mathbf{b}_k$, and $\widehat{\mathbf{b}}_{\text{BLUP}_k}$ which is given by Eq. (17) as $R^2_{\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{BLUP}_k}} = \frac{R_k^2}{M} / \frac{h_k^2}{M} = R_k^2 / h_k^2$. The covariance of BLUP SNP predictors is then:

$$\text{cov}(\widehat{\mathbf{b}}_{\text{BLUP}_{k'}}, \widehat{\mathbf{b}}_{\text{BLUP}_l}) = \frac{R_k^2}{h_k^2} \cdot \frac{R_l^2}{h_l^2} \text{cov}(\mathbf{b}_k, \mathbf{b}_l) = \frac{r_G R_k^2 R_l^2}{h_k h_l M} \qquad (19)$$

Finally, we can consider the column vector $\mathbf{C}_{\text{SBLUP}}$, which is composed of the covariance between the true genetic effects of the SNPs for the focal trait, $\mathbf{b}_f$, and $\widehat{\mathbf{b}}_{\text{SBLUP}_k}$ for all of the k traits. The first element of $\mathbf{C}_{\text{SBLUP}}$ is covariance between the true genetic effects of the SNPs for the focal trait $\mathbf{b}_f$ and $\widehat{\mathbf{b}}_{\text{SBLUP}_f}$ for the focal trait $\text{cov}(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{BLUP}_f}) = \text{var}(\widehat{\mathbf{b}}_{\text{BLUP}_f}) = \frac{R_f^2}{M}$. The remaining elements of $\mathbf{C}_{\text{SBLUP}}$ are $\text{cov}(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{BLUP}_k})$, which can be approximated from theory by considering a regression of $\mathbf{b}_f$ on $\mathbf{b}_k$ where the regression coefficient $\beta_{f,k} = r_G \sqrt{\text{var}(\mathbf{b}_f)/\text{var}(\mathbf{b}_k)}$. The covariance of $\mathbf{b}_f$ and $\widehat{\mathbf{b}}_{\text{BLUP}_k}$ can then be written as:

$$\text{cov}(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{BLUP}_k}) = \text{cov}(\beta_{f,k}\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{BLUP}_k}) = r_G \frac{R_k^2}{M} \cdot \frac{h_f}{h_k} \qquad (20)$$

If we consider a two-trait example where the focal trait that we want to predict is matched to the first of the two traits, Eqs. (18), (19) and (20) combine as:

$$\mathbf{w}_{\text{SBLUP}} = \mathbf{V}_{\text{SBLUP}}^{-1}\mathbf{C}_{\text{SBLUP}} = \begin{bmatrix} \frac{R_1^2}{M} & \frac{r_G R_1^2 R_2^2}{h_1 h_2 M} \\ \frac{r_G R_1^2 R_2^2}{h_1 h_2 M} & \frac{R_2^2}{M} \end{bmatrix}^{-1} \begin{bmatrix} \frac{R_1^2}{M} \\ r_G \frac{R_2^2}{M} \cdot \frac{h_1}{h_2} \end{bmatrix} \qquad (21)$$

giving the index for the focal trait as: $\widehat{\mathbf{b}}_{\text{wMT-SBLUP}_f} = w_1 \widehat{\mathbf{b}}_{\text{SBLUP}_f} + w_2 \widehat{\mathbf{b}}_{\text{SBLUP}_2}$ with solutions for the index weights of:

$$\begin{aligned} w_f &= \left(1 - \frac{r_G^2 R_2^2}{h_2^2}\right) / \left(1 - \frac{r_G^2 R_1^2 R_2^2}{h_1^2 h_2^2 M}\right) \\ &= \left(1 - r_G^2 R^2_{\mathbf{b}_2, \widehat{\mathbf{b}}_{\text{BLUP}_2}}\right) / \left(1 - r_G^2 R^2_{\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{BLUP}_f}} R^2_{\mathbf{b}_2, \widehat{\mathbf{b}}_{\text{BLUP}_2}}\right), \text{ and} \\ w_2 &= \left(r_G / (h_f h_2)\left(h_f^2 - R_1^2\right)\right) / \left(1 - \frac{r_G^2 R_f^2 R_2^2}{h_f^2 h_2^2 M}\right) \\ &= r_G (h_f / h_2)\left(1 - R^2_{\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{BLUP}_f}}\right) / \left(1 - r_G^2 R^2_{\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{BLUP}_f}} R^2_{\mathbf{b}_2, \widehat{\mathbf{b}}_{\text{BLUP}_2}}\right) \end{aligned} \qquad (22)$$

For traits with low power, R_k^2 is usually very small. In that case, $\mathbf{V}_{\text{SBLUP}}$ can be well approximated by a diagonal matrix with entries $\frac{R_k^2}{M}$. w_f will become 1 and w_k for all other traits will be $r_{G_{f,k}} \frac{h_f}{h_k}$. It may appear surprising that traits with higher SNP

heritability have smaller weights than traits with lower SNP heritability. This can be explained by the fact that the variance of each BLUP predictor (R_k^2) is approximately proportional to $h_k^4 N$ if M_{eff} is large, and thus a trait with higher SNP heritability will still have a larger contribution to the multi-trait predictor than a trait with lower SNP heritability.

Equation (17) implies $R^2_{\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{BLUP}_k}} = R^2_{\mathbf{g}_k, \widehat{\mathbf{g}}_{\text{BLUP}_k}} = R_k^2 / h_k^2$ and thus the index weights of Eq. (15) can be applied equally to BLUP solutions for the SNP effects or BLUP predictors for individuals of each trait as described in the main text in Eq. (1) through (3). Both $r_{G,l}$ and h_k^2 of Eq. (15) can be obtained from summary statistic data using LD score regression[22] and therefore $\widehat{\mathbf{b}}_{\text{MT-BLUP}}$ effects of Eq. (10), which would traditionally require individual-level phenotype–genotype data for all traits, can be approximated accurately in a computationally efficient manner using only publically available GWAS summary statistic data and an independent genomic reference sample.

Index weighted multi-trait OLS SNP effects. In the previous section, we have shown how $\widehat{\mathbf{b}}_{\text{SBLUP}}$ estimates for multiple traits can be combined to yield more accurate $\widehat{\mathbf{b}}_{\text{wMT-SBLUP}}$ SNP effects, which can be turned into $\widehat{\mathbf{g}}_{\text{wMT-SBLUP}}$ individual predictors that approach $\widehat{\mathbf{g}}_{\text{MT-BLUP}}$ accuracy. However, using a similar weighting we can also combine $\widehat{\mathbf{b}}_{\text{OLS}}$ estimates for multiple traits into $\widehat{\mathbf{b}}_{\text{wMT-OLS}}$.

For SNP j and focal trait f, we have $\widehat{\mathbf{b}}_{\text{OLS}}$ values for k traits, and we wish to obtain the index weights, $w_{j,k}$, for each $\widehat{\mathbf{b}}_{\text{OLS},k}$ effect as:

$$\widehat{\mathbf{b}}_{\text{wMT-OLS}_{j,f}} = \sum_k w_{j,k}\widehat{\mathbf{b}}_{\text{OLS}_{j,k}} = \mathbf{w}_j'\widehat{\mathbf{b}}_{\text{OLS}_j} \qquad (23)$$

Just like before, the optimal weights can be derived as: $\mathbf{w}_{\text{OLS}} = \mathbf{V}_{\text{OLS}}^{-1}\mathbf{C}_{\text{OLS}}$, where $\mathbf{C}_{\text{OLS}}$ is now a $k \times 1$ column vector of the covariances of the $\widehat{\mathbf{b}}_{\text{OLS},k}$ values of the k traits with the true genetic effects of the SNPs for the focal trait, and $\mathbf{V}_{\text{OLS}}$ is a $k \times k$ variance–covariance matrix of the $\widehat{\mathbf{b}}_{\text{OLS}}$ effects:

$$\begin{aligned} \mathbf{w}_{\text{OLS}} &= \mathbf{V}_{\text{OLS}}^{-1}\mathbf{C}_{\text{OLS}} \\ &= \begin{bmatrix} \text{var}(\widehat{\mathbf{b}}_{\text{OLS}_1}) & \cdots & \text{cov}(\widehat{\mathbf{b}}_{\text{OLS}_1}, \widehat{\mathbf{b}}_{\text{OLS}_k}) \\ \vdots & \ddots & \vdots \\ \text{cov}(\widehat{\mathbf{b}}_{\text{OLS}_k}, \widehat{\mathbf{b}}_{\text{OLS}_1}) & \cdots & \text{var}(\widehat{\mathbf{b}}_{\text{OLS}_k}) \end{bmatrix}^{-1} \\ &\quad \begin{bmatrix} \text{cov}(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{OLS}_1}) \\ \vdots \\ \text{cov}(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{OLS}_k}) \end{bmatrix} \end{aligned} \qquad (24)$$

The diagonal elements of $\mathbf{V}_{\text{OLS}}$ are:

$$\text{var}(\widehat{\mathbf{b}}_{\text{OLS}_k}) = \frac{h_k^2}{M} + \frac{1}{N_k} \qquad (25)$$

The off-diagonal elements for trait k and l are

$$\text{cov}(\widehat{\mathbf{b}}_{\text{OLS}_k}, \widehat{\mathbf{b}}_{\text{OLS}_l}) = \frac{r_G h_k h_l}{M} \qquad (26)$$

$\mathbf{C}_{\text{OLS}}$ now has elements

$$\text{cov}(\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{OLS}_k}) = \frac{r_G h_k h_l}{M} \qquad (27)$$

If we again consider a two-trait example, Eqs. (25), (26) and (27) combine as:

$$\mathbf{w}_{\text{OLS}} = \mathbf{V}_{\text{OLS}}^{-1}\mathbf{C}_{\text{OLS}} = \begin{bmatrix} \frac{h_1^2}{M} + \frac{1}{N_1} & \frac{r_G h_1 h_2}{M} \\ \frac{r_G h_1 h_2}{M} & \frac{h_2^2}{M} + \frac{1}{N_2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{h_1^2}{M} \\ \frac{r_G h_1 h_2}{M} \end{bmatrix} \qquad (28)$$

These weights are considerably different from the BLUP weights, which reflects the different variances of BLUP effects and OLS effects. Here we include this section for completeness but focus our analyses on multi-trait BLUP effects, because they are more accurate in expectation than multi-trait OLS effects.

Index weighted multi-trait SNP effects using LDPred. For phenotypes with a genetic architecture characterised by a few loci of very large effect sizes, this approach may not be ideal. Models that assume a mixture distribution for SNP effects, such as LDpred or BayesR, can yield higher prediction accuracies in traits of non-infinitesimal genetic architecture[18,42]. As outlined above, Eq. (17) implies $R^2_{\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{BLUP}_k}} = R^2_{\mathbf{g}_k, \widehat{\mathbf{g}}_{\text{BLUP}_k}} = R_k^2 / h_k^2$ and thus the index weights of Eq. (15) can be applied equally to BLUP solutions for the SNP effects or BLUP predictors for individuals of each trait as described in the main text in Eq. (1) through (3). LDpred aims to estimate the posterior mean phenotype that provides best unbiased prediction.

Therefore, single-trait individual-level predictors obtained from LDPred can also be weighted together to create an approximate multi-trait predictor.

Prediction accuracy of weighted multi-trait BLUP predictors. The prediction accuracy of $\widehat{\mathbf{b}}_{\text{wMT-BLUP}}$ effects obtained from Eq. (15) can be derived by considering the correlation of $\mathbf{b}_f$ and $\widehat{\mathbf{b}}_{\text{wMT-BLUP}_k}$ as:

$$r_{\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}} = \frac{\text{cov}\left(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}\right)}{\sqrt{\text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}\right)\text{var}(\mathbf{b}_f)}} \qquad (29)$$

Equation (13) gives $\widehat{\mathbf{b}}_{\text{wMT-BLUP}_f} = \mathbf{w}'\widehat{\mathbf{b}}_{\text{BLUP}}$ and thus the covariance of $\mathbf{b}_f$ and $\widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}$ is:

$$\text{cov}\left(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}\right) = \text{cov}\left(\mathbf{b}_f, \mathbf{w}'\widehat{\mathbf{b}}_{\text{BLUP}}\right) = \mathbf{w}'\text{cov}\left(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{BLUP}}\right) = \mathbf{w}'\mathbf{C} \qquad (30)$$

The variance of the $\widehat{\mathbf{b}}_{\text{wMT-BLUP}}$ effects obtained from Eq. (15) is:

$$\text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}}\right) = \text{var}\left(\mathbf{w}'\widehat{\mathbf{b}}_{\text{BLUP}_k}\right) = \mathbf{w}'\text{var}\left(\widehat{\mathbf{b}}_{\text{BLUP}_k}\right)\mathbf{w} = \mathbf{w}'\mathbf{V}\mathbf{w} \qquad (31)$$

Additionally, $\mathbf{w} = \mathbf{V}^{-1}\mathbf{C}$ and $\mathbf{V}\mathbf{w} = \mathbf{C}$, and thus $\mathbf{w}'\mathbf{C} = \mathbf{w}'\mathbf{V}\mathbf{w}$ or written another way $\text{cov}\left(\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}\right) = \text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}}\right)$ following BLUP properties. Substituting into Eq. (19), the correlation of $\mathbf{b}_f$ and $\widehat{\mathbf{b}}_{\text{wMT-BLUP}_k}$ can then be written as:

$$\begin{aligned} r_{\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}} &= \text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}}\right)/\sqrt{\text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}}\right)\text{var}(\mathbf{b}_f)} \\ &= \sqrt{\text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}}\right)/\text{var}(\mathbf{b}_f)} \end{aligned}, \qquad (32)$$

which gives the squared correlation as $R^2_{\mathbf{b}_f, \widehat{\mathbf{b}}_{\text{wMT-BLUP}_f}} = \text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}}\right)/\text{var}(\mathbf{b}_f) = \frac{R_f^2}{M}/\frac{h_k^2}{M} = R_f^2/h_k^2$. Therefore, the squared correlation between a phenotype and a multiple trait index weighted BLUP predictor of the phenotype is approximately:

$$R^2_{\mathbf{y}_k, \widehat{\mathbf{g}}_{\text{wMT-BLUP}_k}} = M\text{var}\left(\widehat{\mathbf{b}}_{\text{wMT-BLUP}}\right) = M\mathbf{w}'\mathbf{V}\mathbf{w} \qquad (33)$$

If we consider a two-trait example then prediction accuracy for a focal trait $R^2_{\mathbf{y}_f, \widehat{\mathbf{g}}_{\text{wMT-BLUP}_f}}$ can be written as:

$$R^2_{\mathbf{y}_f, \widehat{\mathbf{g}}_{\text{wMT-BLUP}_f}} = \mathbf{w}_f^2 R^2_{\mathbf{y}_f, \widehat{\mathbf{g}}_{\text{BLUP}_f}} + \mathbf{w}_2^2 R^2_{\mathbf{y}_2, \widehat{\mathbf{g}}_{\text{BLUP}_2}} + 2\mathbf{w}_f\mathbf{w}_2\mathbf{V}_{1,2} \qquad (34)$$

where $\mathbf{V}_{1,2}$ is the off-diagonal element of the matrix $\mathbf{V}$ of Eqs. (15) and (21). The value of $R^2_{\mathbf{y}_f, \widehat{\mathbf{g}}_{\text{wMT-BLUP}_f}}$ can then be compared to the prediction accuracy of the single-trait BLUP predictor of Eq. (16) and to the prediction accuracy of a cross-trait predictor[43], where a BLUP predictor of the second trait is used to predict the focal trait phenotype, which is given by: $R^2_{\mathbf{y}_f, \widehat{\mathbf{b}}_{\text{BLUP}_2}} = R^2_{\mathbf{y}_2, \widehat{\mathbf{g}}_{\text{BLUP}_2}} r_G \sqrt{(h_2/h_f)}$. This comparison is of interest, because we expect the multi-trait predictor to be more accurate than any available single-trait predictor, even if the most accurate single-trait predictor is across two different traits. Cross-trait prediction is equivalent to the proxy-phenotype method, which has been used to predict cognitive performance from educational attainment GWAS data[44].

Loss of prediction accuracy from BLUP approximation. Equations (16), (17), (18), (19), (20), (21), (22), (23), (24), (25), (26), (27), (28), (29), (30), (31), (32), (33) and (34) assume that $\text{cov}\left(\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{SBLUP}_k}\right) = \text{var}\left(\widehat{\mathbf{b}}_{\text{SBLUP}_k}\right) = \text{var}\left(\widehat{\mathbf{b}}_{\text{BLUP}_k}\right)$, or in other words that SBLUP SNP solutions have BLUP properties. The use of an independent LD reference sample to create an approximate single-trait BLUP predictor in Eq. (11) does not affect the covariance between the true SNP effect sizes and the approximate BLUP SNP solution, meaning that the approximate single-trait BLUP predictors have BLUP properties. However, the variance of $\widehat{\mathbf{b}}_{\text{SBLUP}}$ is likely affected, which may potentially result in a loss of prediction accuracy of a weighted multi-trait BLUP predictor. The variance of $\widehat{\mathbf{b}}_{\text{SBLUP}}$ is:

$$\begin{aligned} \sigma^2_{\mathbf{b}_{\text{SBLUP}}} &= \left[[N\mathbf{L}+\mathbf{I}_M\lambda]^{-1}\mathbf{W}'\right]\left[\mathbf{W}'\mathbf{W}\sigma_b^2 + \mathbf{I}\sigma_e^2\right] \\ &\quad \left[\mathbf{W}[N\mathbf{L}+\mathbf{I}_M\lambda]^{-1}\right] \\ &= [N\mathbf{L}+\mathbf{I}_M\lambda]^{-1}\left[(\mathbf{W}'\mathbf{W})(\mathbf{W}'\mathbf{W})\sigma_b^2\right. \\ &\quad \left.+\mathbf{W}'\mathbf{W}\sigma_e^2\right][N\mathbf{L}+\mathbf{I}_M\lambda]^{-1} \\ &= \left[([N\mathbf{L}+\mathbf{I}_M\lambda]^{-1}(\mathbf{W}'\mathbf{W})(\mathbf{W}'\mathbf{W})[N\mathbf{L}+\mathbf{I}_M\lambda]^{-1}\right. \\ &\quad \left.+[N\mathbf{L}+\mathbf{I}_M\lambda]^{-1}\mathbf{W}'\mathbf{W}\lambda_k)[N\mathbf{L}+\mathbf{I}_M\lambda]^{-1}\right]\sigma_b^2 \end{aligned} \qquad (35)$$

The loss of information from using an independent data set as an LD reference to

obtain $\mathbf{L}$, rather than directly using the individual-level data to calculate $\mathbf{W}'\mathbf{W}$, can be approximated by considering the scenario where SNP makers are unlinked, resulting in diag[$\mathbf{L}$]. The diagonal elements of $\sigma^2_{\mathbf{b}_{\text{SBLUP}_{jj}}}$ for SNP j are then:

$$\sigma^2_{\mathbf{b}_{\text{SBLUP}_{jj}}} = \left([N+\lambda]^{-2}\text{diag}[(\mathbf{W}'\mathbf{W})(\mathbf{W}'\mathbf{W})] + N\lambda[N+\lambda]^{-2}\right)\sigma_b^2 \qquad (36)$$

The diagonal elements of diag[$(\mathbf{W}'\mathbf{W})(\mathbf{W}'\mathbf{W})$] can be approximated as diag[$(\mathbf{W}'\mathbf{W})(\mathbf{W}'\mathbf{W})$] $\approx N^2(1 + \mathbb{E}[r^2]M) + N^2(1 + M/N)$, where the expectation of the LD correlation of the SNPs, $\mathbb{E}[r^2]$, is $1/N$ as the SNP markers are unlinked. Equation (36) can then be written as:

$$\begin{aligned} \sigma^2_{\mathbf{b}_{\text{SBLUP}_{jj}}} &= \left((N^2 + NM + N\lambda)/(N+\lambda)^2\right)\sigma_b^2 \\ &= \sigma_b^2 N/(N+\lambda) + \sigma_b^2 NM/(N+\lambda)^2 \end{aligned} \qquad (37)$$

From Eq. (37), the squared correlation between true SNP effects and SBLUP SNP effects can be written as:

$$R^2_{\mathbf{b}, \widehat{\mathbf{b}}_{\text{SBLUP}}} = N/(N+M+\lambda) = N/\left(N+M/h^2\right) \qquad (38)$$

This can be contrasted to Eq. (17), which gives the squared correlation between the true genetic effects of the SNPs, $\mathbf{b}_k$, and $\widehat{\mathbf{b}}_{\text{BLUP}_k}$ as:

$$\begin{aligned} R^2_{\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{BLUP}_k}} &= \frac{R_k^2}{M}/\frac{h_k^2}{M} = R_k^2/h_k^2 \\ &= 1/\left(1 + M\left(1 - R_k^2\right)/\left(N_k h_k^2\right)\right) \\ &= N_k/\left(N_k + M\left(1 - R_k^2\right)/h_k^2\right) \end{aligned} \qquad (39)$$

Equation (39) is similar to Eq. (38) apart from the factor $1 - R_k^2$. Therefore, the relative loss of prediction accuracy from using an SBLUP predictor is given as a ratio of Eqs. (39) and (38) as:

$$\frac{R^2_{\mathbf{b}, \widehat{\mathbf{b}}_{\text{SBLUP}}}}{R^2_{\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{BLUP}_k}}} = \frac{Nh^2 + M}{Nh^2 + M\left(1 - R_k^2\right)} \qquad (40)$$

For a phenotype of SNP heritability 0.5, with effective number of independent markers (independent genomic segments), M_{eff}, of $\sim 60,000$ and sample size, N, of 500,000, $R^2_{\mathbf{b}, \widehat{\mathbf{b}}_{\text{SBLUP}}}$ from summary statistics in an independent reference sample will be 91% of the value of $R^2_{\mathbf{b}_k, \widehat{\mathbf{b}}_{\text{BLUP}_k}}$ if individual-level data were available. Likewise, for a two-trait example where both traits have $h^2 = 0.5$ and $N = 500,000$, the accuracy of the multi-trait SBLUP predictor will also be 91% of the accuracy of the multi-trait BLUP predictor.

It should be noted that here we assume $\mathbf{L}$ to be a a diagonal matrix, which will lead to a conservative estimate of the accuracy of SBLUP relative to the accuracy of BLUP, and that this estimate is in fact equivalent to the expected accuracy of a polygenic risk predictor based on marginal OLS effects[28]. In practice, approximating $\mathbf{L}$ through an external reference data set leads to SBLUP predictors, which are more accurate than predictors based on marginal OLS effects but less accurate than predictors based on BLUP effects.

Computation time. Computing weights and combing up to 930,000 SNP effects of 34 traits takes <10 min on an Intel Xeon E7-8837 processor with 2.76 GHz. Memory requirements do not extend much beyond the amount necessary to read in the summary statistics. Calculation of single-trait SBLUP effects is more computationally demanding, so we split the data by chromosome. Runtime for chromosome one was <40 min, with memory usage just under 10 GB.

Simulation study. To compare the accuracy of single-trait and multi-trait genetic predictors created from SNP effects obtained from both individual-level and summary statistic data, we conducted a simulation study based on real genotypes from the Kaiser Permanente study (Genetic Epidemiology Research on Adult Health and Aging: GERA cohort) and simulated phenotypes.

From the GERA cohort, we selected 50,000 individuals of European ancestry (for definitions of European individuals and quality control (QC) of the genotypic data, see ref. [45]). SNP QC procedures on the initial data sets consisted of per-SNP missing data rate of <0.01, minor allele frequency >0.01, per-person missing data rate <0.01 and Hardy–Weinberg disequilibrium p-value < 1×10^{-6}. For the subsequent imputation, the data were first haplotyped using HAPI-UR. After that, Impute2 was used to impute the haplotypes to the 1000 genomes reference panel (release 1, version 3). Best-guess genotypes at common SNPs included in the HapMap 3 European sample were then extracted and filtered for imputation info score >0.5, missing data rate of <0.01, minor allele frequency >0.01, per-person missing data rate <0.01 and Hardy–Weinberg disequilibrium p-value < 1×10^{-6}. Next, we performed principal component analysis and

removed individuals with principal eigenvector values that were >7 SD from the mean of the European cluster. Lastly, we identified pairs of individuals with genetic relatedness matrix >0.05 and removed one individual from each of these pairs.

The Atherosclerosis Risk in Communities study (ARIC data) was used as an independent LD reference when estimating SBLUP SNP effects of Eq. (11). Eight thousand seven hundred and forty-four European individuals were selected and the data were imputed and QC conducted in the same way as described above for the GERA cohort. We then reduced the SNPs used in both the GERA and ARIC cohorts to overlapping HapMap3 SNPs, which gave 557,034 SNPs that were used in the simulation study.

We then randomly assigned 20,000, 20,000 and 10,000 individuals from the GERA cohort to create three data sets: training set one, training set two, and a testing set. We simulated two genetically correlated traits by randomly selecting 2000 causal SNPs. Effect sizes for the causal markers were simulated from a bivariate normal distribution with mean 0, variances of $\frac{h_1^2}{M}$ and $\frac{h_2^2}{M}$ and covariance of $r_G\sqrt{h_1^2 h_2^2}$. These effect sizes were then multiplied with the standardised genotype dosages (mean 0 and variance 1) to create a genetic value for each individual. Normally distributed environmental effects $e \sim N(0, 1-h^2)$ were added to this genetic value for each individual to create phenotypes with mean 0 and variance of 1. To remove any effects of population stratification, the simulated phenotypes were then regressed against the first 20 genetic principal components, and the residuals from this regression were used in all subsequent analyses.

In training set 1, we simulated trait 1 and we then estimated: (i) OLS SNP effects using Eq. (5) $\left(\widehat{\mathbf{b}}_{\mathrm{OLS}}\right)$, (ii) BLUP SNP effects from the individual-level data using Eq. (8) $\left(\widehat{\mathbf{b}}_{\mathrm{BLUP}}\right)$, and (iii) approximate SBLUP effects using the OLS SNP effects from Eq. (5) and the ARIC data as a reference $\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}}\right)$. In training set 2, we simulated trait 2 and estimated $\widehat{\mathbf{b}}_{\mathrm{OLS}}$, $\widehat{\mathbf{b}}_{\mathrm{BLUP}}$ and $\widehat{\mathbf{b}}_{\mathrm{SBLUP}}$ in the same manner. We then estimated multi-trait BLUP SNP effects using Eq. (9) $\left(\widehat{\mathbf{b}}_{\mathrm{MT-BLUP}}\right)$ from individual-level data by combining trait 1 from training set 1 and trait 2 from training set 2.

In the testing set, we then used the estimated SNP effects from the training sets to produce genetic predictors for both traits. Single-trait genetic predictors were created for both simulated traits from (i) the OLS SNP effects $\left(\widehat{\mathbf{g}}_{\mathrm{OLS}}\right)$, (ii) the BLUP SNP effects $\left(\widehat{\mathbf{g}}_{\mathrm{BLUP}}\right)$ and (iii) the SBLUP SNP effects $\left(\widehat{\mathbf{g}}_{\mathrm{SBLUP}}\right)$. We then created multi-trait predictors where trait 1 was the focal trait from: (i) individual-level multi-trait BLUP predictor $\left(\widehat{\mathbf{g}}_{\mathrm{MT-BLUP}}\right)$, (ii) weighted multi-trait SBLUP predictor $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-SBLUP}}\right)$, (iii) a weighted multi-trait BLUP predictor-based individual-level single-trait BLUP estimates $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-BLUP}}\right)$, and (iv) a weighted multi-trait GWAS predictor based on GWAS OLS estimates $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-OLS}}\right)$. We simulated phenotypic values for both traits using the same effect sizes as those used to generate the phenotypes in the training sets and normally distributed environmental effects sampled independently for each trait as $e \sim N(0, 1-h^2)$. We also compared estimates obtained using LDPred with the proportion of SNPs in the model of 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1 or using the LDPred-Inf option.

We created two simulation scenarios. Heritability of the first and second trait and genetic correlations were $h_1^2 = 0.2$, $h_2^2 = 0.8$ and $r_G = 0.8$, respectively, in the first scenario and were $h_1^2 = 0.5$, $h_2^2 = 0.5$, and $r_G = 0.5$, respectively, in the second scenario. In each set-up, six replicates were conducted, each with a different set of randomly selected causal markers. We then repeated all analyses on a permuted data set, where the values of the genotype matrix were permuted across all individuals, for each SNP. This creates a genotype matrix where the allele frequency distribution remains the same, but all LD structure is removed, allowing us to determine the degree to which differences between the simulations results are driven by the LD structure in the real genotype data. Finally, because prediction accuracy is expected to be reduced by the error introduced by using an external LD reference data set and a restricted LD window when implementing Eq. (8) (see above), we examined how changing the LD reference and restricting the LD window size influences to optimal value of shrinkage parameter λ when implementing Eq. (8) (see Supplementary Fig. 3).

Application to PGC schizophrenia and bipolar disorder. We then applied our approach to the schizophrenia (SCZ) and bipolar disorder (BIP) samples from both wave 1 and wave 2 data of the PGC (PGC1 and PGC2). A description of the data collection and imputation of the SNP genotype data can be found elsewhere[25,26,46].

We selected these two disorders because there is a high genetic correlation between them (estimate for r_G between schizophrenia and bipolar disorder using ldsc: 0.72, SE: 0.03; estimated using meta-analysis of all PGC2 schizophrenia and bipolar cohorts, excluding cohorts which were used as test set in the initial PGC1 analysis), and it enabled us to draw a direct comparison between the approach described here and a previous study, which estimated multi-trait BLUP SNP effects $\left(\widehat{\mathbf{b}}_{\mathrm{MT-BLUP}}\right)$ from individual-level data in an approach equivalent to Eq. (9). The previous study used PGC1 data in the training set and selected four cohorts for schizophrenia and three cohorts for bipolar disorder as test sets. For schizophrenia, the training set comprised 17 cohorts (8826 cases, 6106 controls) and for bipolar disorder the

training set comprised 11 cohorts (5867 cases, 3328 controls). The test set of 4 cohorts for schizophrenia contained 4068 cases and 5471 controls, and the test set of 3 cohorts for bipolar disorder contained 2029 cases and 5338 controls. The analyses on the PGC1 data were performed on 745,705 HapMap3 SNPs in common across all data sets. To have a direct comparison with our previous study, we began by re-analysing the same PGC1 training set data to estimate: (i) OLS SNP effects using Eq. (5) $\left(\widehat{\mathbf{b}}_{\mathrm{OLS}}\right)$, (ii) BLUP SNP effects from the individual-level data using Eq. (8) $\left(\widehat{\mathbf{b}}_{\mathrm{BLUP}}\right)$, and (iii) approximate SBLUP effects using the OLS SNP effects from Eq. (5) and the ARIC data as a reference $\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}}\right)$ using Eq. (11). For the estimation of schizophrenia SBLUP effects, λ was set to 1,100,000, corresponding roughly to 1,000,000 markers and an observed scale SNP heritability estimate of 0.47, and for the estimation of bipolar disorder SBLUP effects, lambda was set to 1,200,000, corresponding roughly to 1,000,000 markers and an observed scale SNP heritability estimate of 0.45. For the four SCZ testing cohorts and the three BIP testing cohorts used in the previous study, we created: (i) weighted multi-trait SBLUP predictors $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-SBLUP}}\right)$, (ii) weighted multi-trait BLUP predictor-based individual-level single-trait BLUP estimates $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-BLUP}}\right)$, and (iii) weighted multi-trait GWAS predictor based on GWAS OLS estimates $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-OLS}}\right)$. We then compared the prediction accuracy we obtained using the weighted multi-trait SBLUP predictors to the individual-level multi-trait BLUP predictor $\left(\widehat{\mathbf{g}}_{\mathrm{MT-BLUP}}\right)$ used in the previous study[16].

We then extended our analysis to the PGC2 data set. There were 36 cohorts for schizophrenia (26,412 cases and 32,440 controls in total) and 23 cohorts for bipolar disorder (18,865 cases and 30,460 controls in total) available to us. The number of SNPs used in the PGC2 analyses varied between cohorts. Summary statistics for each of the PGC2 cohorts was available to an imputed SNP set of >10,000,000 SNPs. After intersecting this set of SNPs with the HapMap3 SNPs and the ARIC SNPs, 932,344 SNPs remained that were used to create predictors.

We applied a cross-validation approach as we observed that prediction accuracy as well as accuracy differences between predictors can be highly dependent on the choice of the test set in the extended PGC2 data set (Supplementary Figs. 5 and 6), which is supported by previous results showing highly variable prediction accuracy across cohorts in the PGC2 data set[25]. A cross-validation approach allowed us to get a more robust estimate of the increase of prediction accuracy achieved by our multi-trait prediction method compared to a single-trait predictor. We employed a leave-one-out cross-validation approach, where, for each test set cohort, all cohorts of the same disease without any highly related individuals were chosen to be in the training set for the single-trait predictor and all cohorts of both diseases without any highly related individuals were chosen to be in the training set for the multi-trait predictor. To identify pairs of cohorts with highly related individuals, genetic relatedness for all pairs of individuals (across all pairs of cohorts) was calculated based on chromosome 22, and whenever at least one pair of individuals had relatedness >0.8, that pair of cohorts was not simultaneously used in the training set and the test set.

The full genotypes from the PGC2 cohorts that were used as test sets underwent stringent QC and only comprised 458,744–860,576 SNPs for schizophrenia and 556,278–859,034 SNPs for bipolar disorder. We refrained from using the intersection between all these cohorts to not reduce the number of SNPs used in prediction by too much. This meant that different iterations in the cross-validations were based on predictions using a different number of SNPs. However, each comparison between a single-trait predictor and a multi-trait predictor is based on the same number of SNPs.

In each iteration of the cross-validation, a different cohort acts as the test set and a different set of cohorts comprises the training set. To create a predictor from a particular set of cohorts, we first had to obtain effect size estimates from this particular set of cohorts. This is achieved by performing a meta-analysis of the summary statistics of the cohorts that comprise the training set. The meta-analysed beta values b_{META} are calculated as:

$$b_{\mathrm{META}} = \frac{\sum_s \frac{b_s}{\mathrm{SE}_s^2}}{\sum_s \frac{1}{\mathrm{SE}_s^2}} \qquad (41)$$

where b_s is the effect size in cohort s and SE_s is the standard error in cohort s. Conversion between beta values and odds ratios (OR) simply follows the equality $b = \log(\mathrm{OR})$. The weights derived for each trait make assumptions about the variance of SNP effects. We found that, in the summary statistics we used, the observed variance across SNP effects often departed from the expected value. To correct for that, we scaled the SNP effect estimates for each trait to have a variance of one and multiplied the weights for the unscaled SNP effects by the expected standard deviation across all SNPs.

We created approximate SBLUP effects $\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}}\right)$ using the OLS SNP effects from Eq. (5) and the ARIC data as an LD reference using Eq. (11) and set the shrinkage parameter, λ, to 1,300,000 for schizophrenia and to 2,000,000 for bipolar disorder, corresponding to observed scale SNP heritability estimates of 0.43 and 0.33 for schizophrenia and bipolar disorder, respectively. We then used the PLINK "--score" function to turn SNP effects $\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}}, \widehat{\mathbf{b}}_{\mathrm{GWAS}}\right)$ into individual predictors

$\left(\widehat{\mathbf{g}}_{\mathrm{SBLUP}}, \widehat{\mathbf{g}}_{\mathrm{GWAS}}\right)$ for each meta-analysed schizophrenia or bipolar disorder cross-validation set. For the multi-trait weighting, we estimated the heritability of schizophrenia and bipolar disorder and their genetic correlation using LD score regression from publicly available PGC2 schizophrenia summary statistics and the PGC1 bipolar disorder summary statistics. These estimates were then used to calculate the index weights of Eq. (15) for the weighted multi-trait SBLUP predictors $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-SBLUP}}, \widehat{\mathbf{g}}_{\mathrm{wMT-GWAS}}\right)$ of SCZ and BIP, and these were not altered between different cross-validation sets.

To test the degree to which the choice of weights affects the accuracy of the multi-trait predictor, we compared the accuracy of multi-trait predictors based on a spectrum of other weights (Supplementary Figs. 5 and 6). For this, we took advantage of two things: First, when individual predictors $\left(\widehat{\mathbf{g}}_{\mathrm{SBLUP}}, \widehat{\mathbf{g}}_{\mathrm{GWAS}}\right)$ are weighted rather than SNP effects $\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}}, \widehat{\mathbf{b}}_{\mathrm{GWAS}}\right)$, the conversion from SNP effects to individual effects does not have to be repeated for different weights. Second, the scaling of a predictor does not influence its accuracy in terms of correlation between prediction and outcome. Therefore, rather than testing each combination of weights of schizophrenia and bipolar disorder, it is sufficient to vary the relative weight of schizophrenia to bipolar disorder to explore the whole range of possible multi-trait predictors for these two traits. For each test cohort, this enabled us to test whether the weights of our multi-trait predictor derived from theory deviate from the weights that would result in the highest prediction accuracy for that data set.

Application to phenotypes in the UK Biobank study. We applied our approach to a large range of phenotypes for which GWAS summary statistics are publicly available. We started with GWAS summary statistics for 46 phenotypes. However, in some circumstance the same studies (i.e., based on the same individuals) had generated summary statistics for multiple similar phenotypes, so we chose only one phenotype per study, which left us with 34 phenotypes. For example, out of "Cigarettes per day" and "Smoking Ever" we only selected the latter to have only one trait for smoking. We used 112,338 unrelated individuals of European descent in the UK biobank data as the testing set. We paired 6 phenotypes out of the 34 summary statistic phenotypes to phenotypes in the UK Biobank: Height, BMI, fluid intelligence score, depression, angina, and diabetes. The first three are quantitative traits and the latter three are disease traits for which we could identify at least 1000 cases in the UK Biobank data. For details, see Supplementary Table 1.

For the disease traits, we used the self-reported diagnoses rather than ICD10 diagnoses, as they tend to have larger sample sizes. For depression, we used a more refined definition of cases and controls, where individuals were not counted as cases if they had any history of psychiatric symptoms or diagnoses other than depression or if they were prescribed drugs that are indicative of such diagnoses. Individuals were selected as controls only when there was an absence of any psychiatric symptoms or diagnoses and only when they were not prescribed any drugs that could be indicative of such diagnoses. All 6 traits in the UK Biobank were corrected for age, sex and the first 10 principal components by regressing the phenotype on these covariates and using the residuals from that regression for further analysis. For each trait, the SNPs that went into the analysis were based on the overlap between the GWAS summary statistics, the HapMap3 SNPs, the GERA data set, which was used as an LD reference in the SBLUP analysis, and the imputed SNPs from the UK Biobank. (For details on the QC process and imputation, see URLs.) Depending on the trait, the total number of SNPs ranged from around 660,000 to around 930,000.

We created single-trait $\left(\widehat{\mathbf{g}}_{\mathrm{SBLUP}}\right)$ as well as multi-trait $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-SBLUP}}\right)$ predictors for the six paired phenotypes. To create SBLUP SNP effects $\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}}\right)$ from summary statistic trait, we used a λ value of $M\left(1 - h_{\mathrm{SNP}_k}^2\right)/h_{\mathrm{SNP}_k}^2$ for each trait k, where M is assumed to be 1,000,000. As LD reference set, we used a random subset of 10,000 people of European descent from the GERA data set, and we set the LD window size to 2000 kb. We then used the PLINK "--score" function to turn SNP effects $\left(\widehat{\mathbf{b}}_{\mathrm{SBLUP}}\right)$ into individual predictors $\left(\widehat{\mathbf{g}}_{\mathrm{SBLUP}}\right)$ for each trait. For the multi-trait weighting, we used LD score regression to calculate SNP heritability and genetic correlation between all pairs of cohorts. For dichotomous disease traits, SNP heritability was calculated on the observed scale. For each phenotype for which a multi-trait predictor was created, we selected all phenotypes that had a genetic correlation estimate significantly different from 0 at $p = 0.05$ with the focal trait, as well as the focal trait itself. The summary statistics based single-trait SBLUP predictors of the selected phenotypes were then combined into multi-trait SBLUP $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-SBLUP}}\right)$ predictors. The weights for each phenotype were calculated according to Eq. (15). These weights require the single-trait predictors to have exactly the right variance. Since the summary statistics data slightly diverged from this expectation, we scaled each single-trait SBLUP predictor to have mean 0 and variance 1 and then multiplied it with its expected standard deviation, to ensure everything is on exactly the correct scale. We followed the same approach when using single-trait LDpred predictors.

We compared the performance of the multi-trait predictors $\left(\widehat{\mathbf{g}}_{\mathrm{wMT-SBLUP}}\right)$ not only to the performance of the single-trait predictor $\left(\widehat{\mathbf{g}}_{\mathrm{SBLUP}}\right)$ for the same trait but also to the performance of all other (cross-trait) single-trait predictors for the traits that exhibited significant r_{G} with the focal trait (Fig. 4). This is appropriate because

in some traits the single-trait predictor from the same trait is not the most accurate single-trait predictor.

Code availability. Code reported in this manuscript is available from https://github.com/uqrmaie1/smtpred.
For GCTA, see http://cnsgenomics.com/software/gcta/
For LDSC, see https://github.com/bulik/ldsc
For MTG2, see https://sites.google.com/site/honglee0707/mtg2
For LDpred, see https://github.com/bvilhjal/ldpred/
For UK Biobank, see http://www.ukbiobank.ac.uk/
For PLINK2, see http://www.cog-genomics.org/plink2

References

1. Katsanis, S. H. & Katsanis, N. Molecular genetic testing and the future of clinical genomics. *Nat. Rev. Genet.* **14**, 415–426 (2013).
2. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
3. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* **6**, e1000864 (2010).
4. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **14210**, 14205–14210 (2016).
5. Abraham, G. & Inouye, M. Genomic risk prediction of complex human disease and its clinical application. *Curr. Opin. Genet. Dev.* **33**, 10–16 (2015).
6. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
7. Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
8. Wray, N., Goddard, M. & Visscher, P. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* **17**, 1520–1528 (2007).
9. Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
10. Wray, N. R. et al. Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* **55**, 1068–1087 (2014).
11. de los Campos, G., Gianola, D. & Allison, D. B. Predicting genetic predisposition in humans: the promise of whole-genome markers. *Nat. Rev. Genet.* **11**, 880–886 (2010).
12. Goddard, M. E., Wray, N. R., Verbyla, K. & Visscher, P. M. Estimating effects and making predictions from genome-wide marker data. *Stat. Sci.* **24**, 517–529 (2010).
13. Henderson, C. R. & Quaas, R. L. Multiple trait evaluation using relatives records. *J. Anim. Sci.* **43**, 1188–1197 (1976).
14. Schaeffer, L. R. et al. Sire and cow evaluation under multiple trait models. *J. Dairy Sci.* **67**, 1567–1580 (1984).
15. Thompson, R. & Meyer, K. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest. Prod. Sci.* **15**, 299–313 (1986).
16. Maier, R. et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am. J. Hum. Genet.* **96**, 283–294 (2015).
17. Li, C., Yang, C., Gelernter, J. & Zhao, H. Improving genetic risk prediction by leveraging pleiotropy. *Hum. Genet.* **133**, 639–650 (2014).
18. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
19. Yang, J., Zaitlen, Na, 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).
20. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
21. Robinson, M. R. et al. Genetic evidence of assortative mating in humans. *Nat. Hum. Behav.* **1**, 16 (2017).
22. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
23. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
24. Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).

25. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
26. Lee, S. H. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
27. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
28. Pasaniuc, B. & Price, A. L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18**, 117–127 (2016).
29. Cantor, R. M., Lange, K. & Sinsheimer, J. S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* **86**, 6–22 (2010).
30. Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423–447 (1975).
31. Henderson, C. R. Sire evaluation and genetic trends. *J. Anim. Sci.* **1973** 10–41 (1973).
32. Henderson, C. R. & Quaas, R. L. Multiple trait evaluation using relatives' records. *J. Anim. Sci.* **43**, 11–88 (1976).
33. Smith, H. F. A discriminant function for plant selection. *Ann. Eugen.* **7**, 240–250 (1936).
34. Hazel, L. N. & Lush, J. L. The efficiency of three methods of selection. *J. Hered.* **33**, 393–399 (1942).
35. Hazel, L. N. The genetic basis for constructing selection indexes. *Genetics* **28**, 476–490 (1943).
36. Wientjes, Y. C. J., Bijma, P., Veerkamp, R. F. & Calus, M. P. L. An equation to predict the accuracy of genomic values by combining data from multiple traits, populations, or environments. *Genetics* **202**, 799–823 (2016).
37. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* **3**, e3395 (2008).
38. Wray, N. R. et al. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
39. Lee, S. H., Weerasinghe, W. M., Wray, N. R., Goddard, M. E. & van der Werf, J. H. Using information of relatives in genomic prediction to apply effective stratified medicine. *Sci. Rep.* **7**, 42091 (2017).
40. Goddard, M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257 (2009).
41. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
42. Moser, G. et al. Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *PLoS Genet.* **11**, 1–22 (2015).
43. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
44. Rietveld, C. A. et al. Correction for Rietveld et al., Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci.* **112**, E380–E380 (2015).
45. Banda, Y. et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* **200**, 1285–1295 (2015).
46. Sklar, P. et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011).

Acknowledgements

Author contributions

Additional information

Bipolar Disorder Working Group of the Psychiatric Genomics Consortium

Andreas J. Forstner[11,12,13,14], Andrew McQuillin[15], Vassily Trubetskoy[8], Weiqing Wang[16,17], Yunpeng Wang[18,19], Jonathan R.I. Coleman[20,21], Héléna A. Gaspar[20,21], Christiaan A. de Leeuw[22], Jennifer M. Whitehead Pavlides[23], Loes M. Olde Loohuis[24], Tune H. Pers[25,26], Phil H. Lee[2,3,27], Alexander W. Charney[17], Amanda L. Dobbyn[16,28], Laura Huckins[16,28], James Boocock[29], Claudia Giambartolomei[29], Panos Roussos[16,17,30], Niamh Mullins[20], Swapnil Awasthi[8], Esben Agerbo[31], Thomas D. Als[32], Carsten Bøcker Pedersen[33], Jakob Grove[34,35,36], Ralph Kupka[37], Eline J. Regeer[38], Adebayo Anjorin[39], Miquel Casas[40], Cristina Sánchez-Mora[40], Pamela B. Mahon[41], Judith Allardyce[42], Valentina Escott-Price[42], Liz Forty[42], Christine Fraser[42],

Manolis Kogevinas[43], Josef Frank[44], Fabian Streit[44], Jana Strohmaier[44], Jens Treutlein[44], Stephanie H. Witt[44], James L. Kennedy[45,46], John S. Strauss[47], Julie Garnham[48], Claire O'Donovan[48], Claire Slaney[48], Stacy Steinberg[49], Thorgeir E. Thorgeirsson[49], Martin Hautzinger[50], Michael Steffens[51], Roy H. Perlis[52], Cristina Sánchez-Mora[53], Maria Hipolito[54], William B. Lawson[54], Evaristus A. Nwulia[54], Shawn E. Levy[55], Tatiana M. Foroud[56], Stéphane Jamain[57], Allan H. Young[58], James D. McKay[59], Jakob Grove[32], Diego Albani[60], Peter Zandi[61], James B. Potash[41], Peng Zhang[62], J. Raymond DePaulo[41], Sarah E. Bergen[63], Anders Juréus[63], Robert Karlsson[63], Radhika Kandaswamy[20], Peter McGuffin[20], Margarita Rivera[64], Jolanta Lissowska[65], Cristiana Cruceanu[66], Susanne Lucae[66], Pablo Cervantes[67], Monika Budde[68], Katrin Gade[69], Urs Heilbronner[68], Marianne Giørtz Pedersen[70], Derek W. Morris[71], Cynthia Shannon Weickert[72], Thomas W. Weickert[73], Donald J. MacIntyre[74], Jacob Lawrence[75], Torbjørn Elvsåshagen[76,77], Olav B. Smeland[78], Srdjan Djurovic[79], Simon Xi[80], Elaine K. Green[81], Piotr M. Czerski[82], Joanna Hauser[82], Wei Xu[83], Helmut Vedder[84], Lilijana Oruc[85], Anne T. Spijker[86], Scott D. Gordon[87], Sarah E. Medland[87], David Curtis[88], Thomas W. Mühleisen[11], Judith Badner[89], William A. Scheftner[89], Engilbert Sigurdsson[90], Nicholas J Schork[91], Alan F. Schatzberg[92], Marie Bækvad-Hansen[93], Jonas Bybjerg-Grauholm[94], Christine Søholm Hansen[93], James A. Knowles[95,96], Szabolcs Szelinger[97], Grant W. Montgomery[4], Marco Boks[98], Annelie Nordin Adolfsson[99], Per Hoffmann[13,14], Michael Bauer[100], Andrea Pfennig[100], Markus Leber[101], Sarah Kittel-Schneider[102], Andreas Reif[102], Jurgen Del-Favero[103], Sascha B. Fischer[11], Stefan Herms[13,14], Céline S. Reinbold[11], Franziska Degenhardt[13,14], Anna C. Koller[13,14], Anna Maaser[13,14], Anil Ori[24], Anders M. Dale[104], Chun Chieh Fan[105], Tiffany A. Greenwood[106], Caroline M. Nievergelt[107], Tatyana Shehktman[108], Paul D. Shilling[106], William Byerley[109], William Bunney[110], Ney Alliey-Rodriguez[111], Toni-Kim Clarke[74], Chunyu Liu[112], William Coryell[113], Huda Akil[114], Margit Burmeister[115], Matthew Flickinger[116], Jun Z. Li[117], Melvin G. McInnis[118], Fan Meng[114,118], Robert C. Thompson[118], Stanley J. Watson[118], Sebastian Zollner[118], Weihua Guan[119], Melissa J. Green[73], David Craig[120], Janet L. Sobell[121], Lili Milani[122], Katherine Gordon-Smith[123], Sarah V. Knott[123], Amy Perry[123], José Guzman Parra[124], Fermin Mayoral[124], Fabio Rivas[124], John P. Rice[125], Jack D. Barchas[126], Anders D. Børglum[34,35], Preben Bo Mortensen[33], Ole Mors[127], Maria Grigoroiu-Serbanescu[128], Frank Bellivier[129], Bruno Etain[129], Marion Leboyer[129] & Josep Antoni Ramos-Quiroga[130]

[11]Department of Biomedicine, University of Basel, Basel, CH, Switzerland. [12]Department of Psychiatry (UPK), University of Basel, Basel, CH, Switzerland. [13]Institute of Human Genetics, University of Bonn, Bonn, Germany. [14]Life&Brain Center, Department of Genomics, University of Bonn, Bonn, Germany. [15]Division of Psychiatry, University College London, London, UK. [16]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [17]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [18]Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Copenhagen, Denmark. [19]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [20]MRC Social, Genetic and Developmental Psychiatry Centre, Kings College London, London, UK. [21]National Institute of Health Research Maudsley Biomedical Research Centre, Kings College London, London, UK. [22]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [23]Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia. [24]Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, CA, USA. [25]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. [26]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [27]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [28]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [29]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [30]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [31]BSS, NCRR, CIRRAU, Aarhus University, Aarhus, Denmark. [32]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. [33]National Centre for Register-based Research, Aarhus University, Aarhus, Denmark. [34]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [35]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark. [36]Bioinformatics Research Centre (BiRC), Aarhus University, Aarhus, Denmark. [37]Department of Psychiatry, VU Medisch Centrum, Amsterdam, The Netherlands. [38]Outpatient Clinic for Bipolar Disorder, Altrecht, Utrecht, The Netherlands. [39]Psychiatry, Berkshire Healthcare NHS Foundation Trust, Bracknell, UK. [40]Psychiatric Genetics Unit, Group of Psychiatry Mental Health and Addictions, Vall d'Hebron Research Institut (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [41]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [42]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK. [43]Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. [44]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [45]Department of Psychiatry, University of Toronto, Toronto, ON, Canada. [46]Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada. [47]Centre for Addiction and Mental Health, Toronto, ON, Canada. [48]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada. [49]deCODE Genetics/Amgen, Reykjavik, Iceland. [50]Department of Psychology, Eberhard Karls Universität

Tübingen, Tubingen, Germany. [51]Research Division, Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany. [52]Division of Clinical Research, Massachusetts General Hospital, Boston, MA, USA. [53]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain, Barcelona, Spain. [54]Department of Psychiatry and Behavioral Sciences, Howard University Hospital, Washington, DC, USA. [55]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. [56]Department of Medical & Molecular Genetics, Indiana University, Indianapolis, IN, USA. [57]Faculté de Médecine, Université Paris Est, Créteil, France. [58]Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, Kings College London, London, GB, UK. [59]Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, France. [60]NEUROSCIENCE, Istituto Di Ricerche Farmacologiche Mario Negri, Milano, Italy. [61]Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA. [62]Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [63]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [64]Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, Spain. [65]Cancer Epidemiology and Prevention, M. Sklodowska-Curie Cancer Center and Institute of Oncology, Warsaw, Poland. [66]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [67]Department of Psychiatry, Mood Disorders Program, McGill University Health Center, Montreal, QC, Canada. [68]Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Munich, Germany. [69]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [70]Department of Economics and Business Economics, National Centre for Register-based Research, Aarhus University, Aarhus, Denmark. [71]Neuropsychiatric Genetics Research Group, Department of Psychiatry and Trinity Translational Medicine Institute, Trinity College Dublin, Dublin, Ireland. [72]Neuroscience Research Australia, Sydney, NSW, Australia. [73]School of Psychiatry, University of New South Wales, Sydney, NSW, Australia. [74]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [75]Department of Psychiatry, North East London NHS Foundation Trust, Ilford, UK. [76]Department of Neurology, Oslo University Hospital, Oslo, Norway. [77]NORMENT, KG Jebsen Centre for Psychosis Research, Oslo University Hospital, Oslo, Norway. [78]NORMENT, University of Oslo, Oslo, Norway. [79]NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway. [80]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, USA. [81]School of Biomedical and Healthcare Sciences, Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK. [82]Department of Psychiatry, Laboratory of Psychiatric Genetics, Poznan University of Medical Sciences, Poznan, Poland. [83]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [84]Psychiatry, Psychiatrisches Zentrum Nordbaden, Wiesloch, Germany. [85]Department of Clinical Psychiatry, Psychiatry Clinic, Clinical Center University of Sarajevo, Sarajevo, BA, Bosnia and Herzegovina. [86]Department of Mood Disorders, PsyQ, Rotterdam, The Netherlands. [87]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. [88]UCL Genetics Institute, University College London, London, UK. [89]Department of Psychiatry, Rush University Medical Center, Chicago, IL, USA. [90]Faculty of Medicine, Department of Psychiatry, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [91]Scripps Translational Science Institute, La Jolla, CA, USA. [92]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [93]Statens Serum Institut, Copenhagen, Denmark. [94]Neonatal Genetik, Statens Serum Institut, Copenhagen, Denmark. [95]Department of Cell Biology, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [96]Institute for Genomic Health, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [97]Division of Neurogenomics, TGen, Los Angeles, AZ, USA. [98]Department of Psychiatry, UMC Utrecht Hersencentrum Rudolf Magnus, Utrecht, The Netherlands. [99]Department of Clinical Sciences, Psychiatry, Umeå University Medical Faculty, Umeå, Sweden. [100]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [101]Clinic for Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany. [102]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany. [103]Applied Molecular Genomics Unit, VIB Department of Molecular Genetics, University of Antwerp, Antwerp, Belgium. [104]Departments of Neurosciences, Radiology, Psychiatry, Cognitive Science, University of California San Diego, La Jolla, CA, USA. [105]Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA. [106]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [107]Research/Psychiatry, Veterans Affairs San Diego Healthcare System, San Diego, CA, USA. [108]Institute of Genomic Medicine, University of California San Diego, La Jolla, CA, USA. [109]Department of Psychiatry, University of California San Francisco, San Francisco, CA, USA. [110]Department of Psychiatry and Human Behavior, University of California, Irvine, Irvine, CA, USA. [111]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA. [112]Department of Psychiatry, University of Illinois at Chicago College of Medicine, Chicago, IL, USA. [113]University of Iowa Hospitals and Clinics, Iowa City, IA, USA. [114]Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA. [115]Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [116]Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [117]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [118]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA. [119]Biostatistics, University of Minnesota System, Minneapolis, MN, USA. [120]Translational Genomics, University of Southern California, Los Angeles, CA, USA. [121]Department of Psychiatry and the Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. [122]Estonian Genome Center, University of Tartu, Tartu, Estonia. [123]Department of Psychological Medicine, University of Worcester, Worcester, UK. [124]Mental Health Department, University Regional Hospital. Biomedicine Institute (IBIMA), Málaga, Spain. [125]Department of Psychiatry, Washington University in Saint Louis, Saint Louis, MO, USA. [126]Department of Psychiatry, Weill Cornell Medical College, New York, NY, USA. [127]Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, Denmark. [128]Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania. [129]Department of Psychiatry and Addiction Medicine, Assistance Publique - Hôpitaux de Paris, Paris, France. [130]Instituto de Salud Carlos III, Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain. [11]Department of Biomedicine, University of Basel, Basel, CH, Switzerland. [12]Department of Psychiatry (UPK), University of Basel, Basel, CH, Switzerland. [13]Institute of Human Genetics, University of Bonn, Bonn, Germany. [14]Life&Brain Center, Department of Genomics, University of Bonn, Bonn, Germany. [15]Division of Psychiatry, University College London, London, UK. [16]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [17]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [18]Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Copenhagen, Denmark. [19]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [20]MRC Social, Genetic and Developmental Psychiatry Centre, Kings College London, London, UK. [21]National Institute of Health Research Maudsley Biomedical Research Centre, Kings College London, London, UK. [22]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [23]Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia. [24]Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles, CA, USA. [25]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. [26]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [27]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [28]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [29]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [30]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [31]BSS, NCRR, CIRRAU, Aarhus University, Aarhus, Denmark. [32]iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. [33]National Centre for Register-based Research, Aarhus University, Aarhus,

Denmark. [34]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [35]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark. [36]Bioinformatics Research Centre (BiRC), Aarhus University, Aarhus, Denmark. [37]Department of Psychiatry, VU Medisch Centrum, Amsterdam, The Netherlands. [38]Outpatient Clinic for Bipolar Disorder, Altrecht, Utrecht, The Netherlands. [39]Psychiatry, Berkshire Healthcare NHS Foundation Trust, Bracknell, UK. [40]Psychiatric Genetics Unit, Group of Psychiatry Mental Health and Addictions, Vall d'Hebron Research Institut (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [41]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [42]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK. [43]Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. [44]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [45]Department of Psychiatry, University of Toronto, Toronto, ON, Canada. [46]Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada. [47]Centre for Addiction and Mental Health, Toronto, ON, Canada. [48]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada. [49]deCODE Genetics/Amgen, Reykjavik, Iceland. [50]Department of Psychology, Eberhard Karls Universität Tübingen, Tubingen, Germany. [51]Research Division, Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany. [52]Division of Clinical Research, Massachusetts General Hospital, Boston, MA, USA. [53]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain, Barcelona, Spain. [54]Department of Psychiatry and Behavioral Sciences, Howard University Hospital, Washington, DC, USA. [55]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. [56]Department of Medical & Molecular Genetics, Indiana University, Indianapolis, IN, USA. [57]Faculté de Médecine, Université Paris Est, Créteil, France. [58]Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, Kings College London, London, GB, UK. [59]Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, France. [60]NEUROSCIENCE, Istituto Di Ricerche Farmacologiche Mario Negri, Milano, Italy. [61]Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA. [62]Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [63]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [64]Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, Spain. [65]Cancer Epidemiology and Prevention, M. Sklodowska-Curie Cancer Center and Institute of Oncology, Warsaw, Poland. [66]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [67]Department of Psychiatry, Mood Disorders Program, McGill University Health Center, Montreal, QC, Canada. [68]Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Munich, Germany. [69]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [70]Department of Economics and Business Economics, National Centre for Register-based Research, Aarhus University, Aarhus, Denmark. [71]Neuropsychiatric Genetics Research Group, Department of Psychiatry and Trinity Translational Medicine Institute, Trinity College Dublin, Dublin, Ireland. [72]Neuroscience Research Australia, Sydney, NSW, Australia. [73]School of Psychiatry, University of New South Wales, Sydney, NSW, Australia. [74]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [75]Department of Psychiatry, North East London NHS Foundation Trust, Ilford, UK. [76]Department of Neurology, Oslo University Hospital, Oslo, Norway. [77]NORMENT, KG Jebsen Centre for Psychosis Research, Oslo University Hospital, Oslo, Norway. [78]NORMENT, University of Oslo, Oslo, Norway. [79]NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway. [80]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, USA. [81]School of Biomedical and Healthcare Sciences, Plymouth University Peninsula Schools of Medicine and Dentistry, Plymouth, UK. [82]Department of Psychiatry, Laboratory of Psychiatric Genetics, Poznan University of Medical Sciences, Poznan, Poland. [83]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [84]Psychiatry, Psychiatrisches Zentrum Nordbaden, Wiesloch, Germany. [85]Department of Clinical Psychiatry, Psychiatry Clinic, Clinical Center University of Sarajevo, Sarajevo, BA, Bosnia and Herzegovina. [86]Department of Mood Disorders, PsyQ, Rotterdam, The Netherlands. [87]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. [88]UCL Genetics Institute, University College London, London, UK. [89]Department of Psychiatry, Rush University Medical Center, Chicago, IL, USA. [90]Faculty of Medicine, Department of Psychiatry, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [91]Scripps Translational Science Institute, La Jolla, CA, USA. [92]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [93]Statens Serum Institut, Copenhagen, Denmark. [94]Neonatal Genetik, Statens Serum Institut, Copenhagen, Denmark. [95]Department of Cell Biology, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [96]Institute for Genomic Health, SUNY Downstate Medical Center College of Medicine, Brooklyn, NY, USA. [97]Division of Neurogenomics, TGen, Los Angeles, AZ, USA. [98]Department of Psychiatry, UMC Utrecht Hersencentrum Rudolf Magnus, Utrecht, The Netherlands. [99]Department of Clinical Sciences, Psychiatry, Umeå University Medical Faculty, Umeå, Sweden. [100]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [101]Clinic for Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany. [102]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany. [103]Applied Molecular Genomics Unit, VIB Department of Molecular Genetics, University of Antwerp, Antwerp, Belgium. [104]Departments of Neurosciences, Radiology, Psychiatry, Cognitive Science, University of California San Diego, La Jolla, CA, USA. [105]Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA. [106]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [107]Research/Psychiatry, Veterans Affairs San Diego Healthcare System, San Diego, CA, USA. [108]Institute of Genomic Medicine, University of California San Diego, La Jolla, CA, USA. [109]Department of Psychiatry, University of California San Francisco, San Francisco, CA, USA. [110]Department of Psychiatry and Human Behavior, University of California, Irvine, Irvine, CA, USA. [111]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA. [112]Department of Psychiatry, University of Illinois at Chicago College of Medicine, Chicago, IL, USA. [113]University of Iowa Hospitals and Clinics, Iowa City, IA, USA. [114]Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, USA. [115]Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [116]Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [117]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [118]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA. [119]Biostatistics, University of Minnesota System, Minneapolis, MN, USA. [120]Translational Genomics, University of Southern California, Los Angeles, CA, USA. [121]Department of Psychiatry and the Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. [122]Estonian Genome Center, University of Tartu, Tartu, Estonia. [123]Department of Psychological Medicine, University of Worcester, Worcester, UK. [124]Mental Health Department, University Regional Hospital. Biomedicine Institute (IBIMA), Málaga, Spain. [125]Department of Psychiatry, Washington University in Saint Louis, Saint Louis, MO, USA. [126]Department of Psychiatry, Weill Cornell Medical College, New York, NY, USA. [127]Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, Denmark. [128]Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania. [129]Department of Psychiatry and Addiction Medicine, Assistance Publique - Hôpitaux de Paris, Paris, France. [130]Instituto de Salud Carlos III, Biomedical Network Research Centre on Mental Health (CIBERSAM), Madrid, Spain

Schizophrenia Working Group of the Psychiatric Genomics Consortium

Ingrid Agartz[131,132], Farooq Amin[133], Maria H. Azevedo[134], Nicholas Bass[135], Donald W. Black[136], Douglas H. R. Blackwood[137], Richard Bruggeman[138], Nancy G. Buccola[139], Khalid Choudhury[135], C. Robert Cloninger[140], Aiden Corvin[141], Nicholas Craddock[142,143], Mark J. Daly[2,3], Susmita Datta[144], Gary J. Donohoe[145], Jubao Duan[146], Frank Dudbridge[147], Ayman Fanous[148,149], Robert Freedman[150], Nelson B. Freimer[24], Marion Friedl[151], Michael Gill[141], Hugh Gurling[135], Lieuwe De Haan[152], Marian L. Hamshere[142,153], Annette M. Hartmann[151], Peter A. Holmans[142,153], René S. Kahn[154], Matthew C. Keller[155], Elaine Kenny[141], George K. Kirov[142,143], Lydia Krabbendam[156], Robert Krasucki[135], Jacob Lawrence[135], Todd Lencz[157,158,159], Douglas F. Levinson[92], Jeffrey A. Lieberman[160], Dan-Yu Lin[161], Don H. Linszen[152], Patrik K.E. Magnusson[63], Wolfgang Maier[162], Anil K. Malhotra[157,158,159], Manuel Mattheisen[13,163,34,164], Morten Mattingsdal[131,165], Steven A. McCarroll[2], Helena Medeiros[166], Ingrid Melle[131,167], Vihra Milanova[168], Inez Myin-Germeys[156], Benjamin M. Neale[2,3], Roel A. Ophoff[24,29,169], Michael J. Owen[142,143], Jonathan Pimm[135], Shaun M. Purcell[3,28], Vinay Puri[135], Digby J. Quested[170], Lizzy Rossin[2], Alan R. Sanders[146], Jianxin Shi[171], Pamela Sklar[28], David St. Clair[172], T. Scott Stroup[173], Jim Van Os[156], Durk Wiersma[138] & Stanley Zammit[142,143]

[131]KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [132]Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway. [133]Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Emory University, Atlanta, GA, USA. [134]Faculty of Medicine, University of Coimbra, Coimbra, Portugal. [135]Mental Health Sciences Unit, University College London, London, UK. [136]Department of Psychiatry, University of Iowa, Iowa City, IA, USA. [137]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. [138]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [139]School of Nursing, Louisiana State University Health Sciences Center, New Orleans, LA, USA. [140]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [141]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [142]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK. [143]Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK. [144]Genetics Institute, University College London, London, UK. [145]Cognitive Genetics and Therapy Group, Discipline of Biochemistry and School of Psychology, National University of Ireland, Galway, Ireland. [146]Department of Psychiatry and Behavioral Sciences, NorthShore University Health System and University of Chicago, Evanston, IL, USA. [147]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. [148]Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA. [149]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [150]Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA. [151]Department of Psychiatry, University of Halle, Halle, Germany. [152]Department of Psychiatry, Academic Medical Centre University of Amsterdam, Amsterdam, The Netherlands. [153]Biostatistics and Bioinformatics Unit, Cardiff University, Cardiff, UK. [154]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands. [155]Department of Psychology, University of Colorado, Boulder, CO, USA. [156]Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, Maastricht, The Netherlands. [157]Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore, Long Island Jewish Health System, Glen Oaks, NY, USA. [158]Center for Psychiatric Neuroscience, The Feinstein Institute of Medical Research, Manhasset, NY, USA. [159]Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine of Yeshiva University, Bronx, NY, USA. [160]New York State Psychiatric Institute, Columbia University, New York, NY, USA. [161]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [162]Department of Psychiatry, University of Bonn, Bonn, Germany. [163]The Lundbeck Initiative for Integrative Psychiatric Research, iPSYCH, Roskilde, Denmark. [164]Department of Genomic Mathematics, University of Bonn, Bonn, Germany. [165]Sørlandet Hospital, Kristiansand, Norway. [166]Department of Psychiatry, Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [167]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [168]Department of Psychiatry, First Psychiatric Clinic, Alexander University Hospital, Sofia, Bulgaria. [169]Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands. [170]Academic Department of Psychiatry, University of Oxford, Oxford, UK. [171]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [172]Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK. [173]Department of Psychiatry, Columbia University, New York, NY, USA. [131]KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [132]Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway. [133]Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Emory University, Atlanta, GA, USA. [134]Faculty of Medicine, University of Coimbra, Coimbra, Portugal. [135]Mental Health Sciences Unit, University College London, London, UK. [136]Department of Psychiatry, University of Iowa, Iowa City, IA, USA. [137]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. [138]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [139]School of Nursing, Louisiana State University Health Sciences Center, New Orleans, LA, USA. [140]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [141]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [142]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK. [143]Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK. [144]Genetics Institute, University College London, London, UK. [145]Cognitive Genetics and Therapy Group, Discipline of Biochemistry and School of Psychology, National University of Ireland, Galway, Ireland. [146]Department of Psychiatry and Behavioral Sciences, NorthShore University Health System and University of Chicago, Evanston, IL, USA. [147]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. [148]Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA. [149]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [150]Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA. [151]Department of Psychiatry, University of Halle, Halle, Germany. [152]Department of Psychiatry, Academic Medical Centre University of Amsterdam, Amsterdam, The Netherlands. [153]Biostatistics and Bioinformatics Unit, Cardiff University,

Cardiff, UK. [154]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands. [155]Department of Psychology, University of Colorado, Boulder, CO, USA. [156]Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, Maastricht, The Netherlands. [157]Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore, Long Island Jewish Health System, Glen Oaks, NY, USA. [158]Center for Psychiatric Neuroscience, The Feinstein Institute of Medical Research, Manhasset, NY, USA. [159]Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine of Yeshiva University, Bronx, NY, USA. [160]New York State Psychiatric Institute, Columbia University, New York, NY, USA. [161]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [162]Department of Psychiatry, University of Bonn, Bonn, Germany. [163]The Lundbeck Initiative for Integrative Psychiatric Research, iPSYCH, Roskilde, Denmark. [164]Department of Genomic Mathematics, University of Bonn, Bonn, Germany. [165]Sørlandet Hospital, Kristiansand, Norway. [166]Department of Psychiatry, Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [167]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [168]Department of Psychiatry, First Psychiatric Clinic, Alexander University Hospital, Sofia, Bulgaria. [169]Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands. [170]Academic Department of Psychiatry, University of Oxford, Oxford, UK. [171]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [172]Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK. [173]Department of Psychiatry, Columbia University, New York, NY, USA