

Joint Analysis of Psychiatric Disorders Increases Accuracy of Risk Prediction for Schizophrenia, Bipolar Disorder, and Major Depressive Disorder

Robert Maier,¹ Gerhard Moser,¹ Guo-Bo Chen,¹ Stephan Ripke,² Cross-Disorder Working Group of the Psychiatric Genomics Consortium, William Coryell,³ James B. Potash,³ William A. Scheftner,⁴ Jianxin Shi,⁵ Myrna M. Weissman,⁶ Christina M. Hultman,⁷ Mikael Landén,^{7,8} Douglas F. Levinson,⁹ Kenneth S. Kendler,¹⁰ Jordan W. Smoller,¹¹ Naomi R. Wray,¹ and S. Hong Lee^{1,*}

Genetic risk prediction has several potential applications in medical research and clinical practice and could be used, for example, to stratify a heterogeneous population of patients by their predicted genetic risk. However, for polygenic traits, such as psychiatric disorders, the accuracy of risk prediction is low. Here we use a multivariate linear mixed model and apply multi-trait genomic best linear unbiased prediction for genetic risk prediction. This method exploits correlations between disorders and simultaneously evaluates individual risk for each disorder. We show that the multivariate approach significantly increases the prediction accuracy for schizophrenia, bipolar disorder, and major depressive disorder in the discovery as well as in independent validation datasets. By grouping SNPs based on genome annotation and fitting multiple random effects, we show that the prediction accuracy could be further improved. The gain in prediction accuracy of the multivariate approach is equivalent to an increase in sample size of 34% for schizophrenia, 68% for bipolar disorder, and 76% for major depressive disorders using single trait models. Because our approach can be readily applied to any number of GWAS datasets of correlated traits, it is a flexible and powerful tool to maximize prediction accuracy. With current sample size, risk predictors are not useful in a clinical setting but already are a valuable research tool, for example in experimental designs comparing cases with high and low polygenic risk.

Genome-wide association studies (GWASs) have been highly successful in identifying variants associated with a wide range of complex human diseases.^{1,2} However, most common diseases are highly polygenic and each variant explains only a tiny proportion of the genetic variation. Even when associated SNPs are considered jointly in polygenic approaches such as polygenic risk scores³ or genomic best linear unbiased prediction (GBLUP),^{4,5} the accuracy of risk prediction is low. The use of more advanced methods^{4–8} improved prediction accuracy for traits where a small number of relatively strong associations have been identified, such as type 1 diabetes, ankylosing spondylitis, and rheumatoid arthritis, but not for other traits characterized by small effect size variants, including psychiatric disorders.^{4,5,9}

A major factor determining how well a polygenic model can predict a trait value in an independent sample is the sample size of the discovery data.^{10,11} Using more individuals will provide more information and hence increase the accuracy of the estimated effect size of a specific SNP. Sample size can also be effectively increased through datasets measured for correlated traits. Recently, we estimated

the genetic relationships among five psychiatric disorders from the Psychiatric Genomics Consortium (PGC) by using a bivariate linear mixed model demonstrating that there are significant shared genetic risk factors across the disorders and that measurement of one trait provides information on other genetically correlated traits.¹² Here we extend our bivariate approach to a multivariate linear mixed model and apply multi-trait genomic best linear unbiased prediction (MTGBLUP)^{13,14} for genetic risk prediction of disease. MTGBLUP is expected to be more powerful because it uses correlations between disorders and jointly evaluates individual risk across disorders. To date, the information from other correlated traits has been little exploited in the context of risk prediction although recently Li et al.⁹ applied bivariate ridge regression to two genetically correlated diseases to improve risk prediction.

An important advantage of the MTGBLUP approach is that it does not require multiple phenotypes to be measured on the same individuals and therefore can be readily applied to any number of existing datasets of genetically related traits. This is particularly beneficial for

¹The Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia; ²Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; ³Department of Psychiatry, University of Iowa, Iowa City, IA 52242, USA; ⁴Department of Psychiatry, Rush University Medical Center, Chicago, IL 60612, USA; ⁵Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA; ⁶Department of Psychiatry, Columbia University, and New York State Psychiatric Institute, New York, NY 10032, USA; ⁷Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 171 77, Sweden; ⁸Institute of Neuroscience and Physiology, The Sahlgrenska Academy at the Gothenburg University, 413 45 Gothenburg, Sweden; ⁹Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94305, USA; ¹⁰Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA 23298, USA; ¹¹Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA

*Correspondence: hong.lee@uq.edu.au

<http://dx.doi.org/10.1016/j.ajhg.2014.12.006>. ©2015 The Authors

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

disease studies that are limited to a single phenotype but typically aim for large sample sizes. Moreover, it is not necessary for the datasets to be genotyped with the same SNP array because SNPs can be imputed to a common set of SNPs, such as those available from the HapMap or 1000 Genomes reference panels.^{15,16} Prediction accuracy can be expected to improve as more data from phenotypes with shared etiology are utilized.

In this report, we apply the MTGBLUP approach to the cross-disorder PGC GWAS data and show a significant increase in risk prediction accuracy in independent cohorts of schizophrenia, bipolar disorder, and major depressive disorder. MTGBLUP increased the discriminant power between the top and bottom 10% of individuals ranked on their risk predictor, implying that this approach might be useful for stratified medicine in a research setting, to develop tailored interventions or treatments for individuals having different risks.^{17–19} We further demonstrate a relationship between functionally annotated SNPs and increased prediction accuracy of schizophrenia and bipolar disorder.

As the main method, we use a multivariate linear mixed model for the analyses of GWAS data that estimates the total genetic values of individuals directly by utilizing genomic relationships based on SNP information. In the model, a vector of phenotypic observations for each trait is written as a linear function of fixed effects, random genetic effects, and residuals. For simplicity, we constrain the description to a single component for the random genetic effects, but the model can be readily extended to multiple components of random genetic effects:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1 \mathbf{b}_1 + \mathbf{Z}_1 \mathbf{g}_1 + \mathbf{e}_1 \text{ for trait 1} \\ \mathbf{y}_2 &= \mathbf{X}_2 \mathbf{b}_2 + \mathbf{Z}_2 \mathbf{g}_2 + \mathbf{e}_2 \text{ for trait 2} \\ &\vdots \\ \mathbf{y}_n &= \mathbf{X}_n \mathbf{b}_n + \mathbf{Z}_n \mathbf{g}_n + \mathbf{e}_n \text{ for trait } n \end{aligned}$$

where \mathbf{y} is a vector of trait phenotypes, \mathbf{b} is a vector of fixed effects, \mathbf{g} is a vector of total genetic value for each individual, and \mathbf{e} are residuals. The random effects (\mathbf{g} and \mathbf{e}) are assumed to be normally distributed with mean zero. \mathbf{X} and \mathbf{Z} are incidence matrices for the effects \mathbf{b} and \mathbf{g} , respectively. Subscript 1, ..., n represents trait 1 to trait n . The variance covariance matrix is defined as

$$\mathbf{V} = \begin{bmatrix} \mathbf{Z}_1 \mathbf{A} \sigma_{g_1}^2 \mathbf{Z}' + \mathbf{I} \sigma_{e_1}^2 & \cdots & \mathbf{Z}_1 \mathbf{A} \sigma_{g_{1n}} \mathbf{Z}' + \mathbf{I} \sigma_{e_{1n}}^2 \\ \vdots & \ddots & \vdots \\ \mathbf{Z}_n \mathbf{A} \sigma_{g_n} \mathbf{Z}' + \mathbf{I} \sigma_{e_n}^2 & \cdots & \mathbf{Z}_n \mathbf{A} \sigma_{g_n}^2 \mathbf{Z}' + \mathbf{I} \sigma_{e_n}^2 \end{bmatrix}$$

where \mathbf{A} is the genomic similarity matrix based on SNP information and \mathbf{I} is an identity matrix. The terms $\sigma_{g_i}^2$ and $\sigma_{e_i}^2$ denote the genetic and residual variance of trait i , respectively, and σ_{g_j} and σ_{e_j} the genetic and residual covariance between traits i and j . Multi-trait genomic residual maximum likelihood (MTGREML) estimates (see [Appendix A](#)) are obtained with the average information algorithm.^{20–22}

Next we show that SNP risk predictors can be easily transformed from individual risk predictors with a simpli-

fied BLUP model that uses individual risk predictors as the dependent variable and fits a covariance structure without residual variance (i.e., heritability is 1). Individual risk predictors are the best linear unbiased predictors (BLUPs) of total genetic value of individual subjects contributed by genome-wide SNPs, i.e., \mathbf{g} in the previous section. Analogously, SNP risk predictors are defined as the BLUPs of SNP effects estimated jointly with a linear mixed model that intrinsically accounts for linkage disequilibrium between SNPs. The SNP-BLUP model is computationally more demanding for a large number of SNPs. Therefore, it is desirable to estimate genetic values (GBLUP) for efficiency and to transform them to SNP-BLUP. The SNP-BLUP can be projected to predict genetic risk for independent validation sample without the need to have access to the training individuals. The SNP-BLUP estimates can be applied to independent datasets as the SNP weights used to create a risk profile score, for example using the PLINK-score command. The individual BLUP model is

$$\begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix} = \begin{bmatrix} \sigma_{g_{1n}}^2 & \cdots & \sigma_{g_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma_{g_{n1}} & \cdots & \sigma_{g_n}^2 \end{bmatrix} \otimes \mathbf{A} \cdot \begin{bmatrix} \mathbf{Z}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix}' \cdot \mathbf{V}^{-1} \begin{bmatrix} \mathbf{y}_1 - \mathbf{X}_1 \mathbf{b}_1 \\ \vdots \\ \mathbf{y}_n - \mathbf{X}_n \mathbf{b}_n \end{bmatrix} \quad (\text{Equation 1})$$

SNP-BLUP model is

$$\begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \sigma_{u_{1n}}^2 & \cdots & \sigma_{u_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma_{u_{n1}} & \cdots & \sigma_{u_n}^2 \end{bmatrix} \otimes \mathbf{I} \cdot \begin{bmatrix} \mathbf{W}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{W}_n \end{bmatrix}' \cdot \mathbf{\Omega}^{-1} \begin{bmatrix} \mathbf{y}_1 - \mathbf{X}_1 \mathbf{b}_1 \\ \vdots \\ \mathbf{y}_n - \mathbf{X}_n \mathbf{b}_n \end{bmatrix}$$

where \mathbf{W}_i is a $N \times M$ matrix of standardized SNP coefficients with N being the number of individuals and M the number of SNPs, \otimes is the Kronecker product function, and the variance covariance matrix for SNP-BLUP mode is defined as

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{W}_1 \mathbf{I} \sigma_{u_{1n}}^2 \mathbf{W}' + \mathbf{I} \sigma_{e_1}^2 & \cdots & \mathbf{W}_1 \mathbf{I} \sigma_{u_{1n}} \mathbf{W}' + \mathbf{I} \sigma_{e_{1n}}^2 \\ \vdots & \ddots & \vdots \\ \mathbf{W}_n \mathbf{I} \sigma_{u_{n1}} \mathbf{W}' + \mathbf{I} \sigma_{e_n}^2 & \cdots & \mathbf{W}_n \mathbf{I} \sigma_{g_n}^2 \mathbf{W}' + \mathbf{I} \sigma_{e_n}^2 \end{bmatrix}$$

Replacing \mathbf{y} with \mathbf{g} (individual BLUP) and setting residual (co)variances as zero (because individual BLUP is already adjusted for residuals), the variance covariance matrix can be simplified as

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_{u_{1n}}^2 & \cdots & \sigma_{u_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma_{u_{n1}} & \cdots & \sigma_{u_n}^2 \end{bmatrix} \otimes \mathbf{W} \mathbf{W}' = \begin{bmatrix} \sigma_{u_{1n}}^2 & \cdots & \sigma_{u_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma_{u_{n1}} & \cdots & \sigma_{u_n}^2 \end{bmatrix} \otimes \mathbf{A} \cdot \mathbf{M}$$

Therefore, SNP-BLUP can be written as

$$\begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_n \end{bmatrix}' \otimes \mathbf{A}^{-1} \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix} \cdot \mathbf{M}^{-1}, \quad (\text{Equation 2})$$

and this can be rewritten as

$$\begin{bmatrix} \mathbf{W}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix}.$$

This agrees with Hayes et al.²³ and Yang et al.²² when it reduces to a univariate model. Equation 2, after replacing $[\mathbf{g}_1, \dots, \mathbf{g}_n]'$ with the right-hand side in Equation 1, can be rewritten as

$$\begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_n \end{bmatrix}' \cdot \begin{bmatrix} \sigma_{g_1}^2 & \cdots & \sigma_{g_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma_{g_{n1}} & \cdots & \sigma_{g_n}^2 \end{bmatrix} \otimes \mathbf{I} \\ \cdot \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_n \end{bmatrix}' \cdot \mathbf{V}^{-1} \begin{bmatrix} \mathbf{y}_1 - \mathbf{X}_1 \mathbf{b}_1 \\ \vdots \\ \mathbf{y}_n - \mathbf{X}_n \mathbf{b}_n \end{bmatrix} \mathbf{M}^{-1}. \quad (\text{Equation 3})$$

This agrees with VanRaden²⁴ and Strandén and Garrick²⁵ derived from a matrix inversion theory when it reduces to a univariate model.

We extended our approach to genomic partitions according to gene annotation. An enrichment analysis based on gene annotation categories has shown that SNPs located within genes identified as being differentially expressed in the central nervous system (CNS) explain a significantly larger proportion of phenotypic variance than expected by chance for schizophrenia and bipolar disorder.^{12,26} It is of interest to determine whether the gene/functional annotation information can further increase the prediction accuracy. In the annotation analysis, we grouped SNPs that were located within ± 50 kb from the 5' and 3' UTRs of 2,725 genes differentially expressed in the CNS^{26,27} together, and 21% of the SNPs belonged to this category. We then estimated SNP effects from a two-component model fitting relationship matrices of SNPs in CNS genes and SNPs localized elsewhere. The model is

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1 \mathbf{b}_1 + \mathbf{Z}_1 \mathbf{g}_{1\text{CNS}} + \mathbf{Z}_1 \mathbf{g}_{1\text{non-CNS}} + \mathbf{e}_1 \text{ for trait 1} \\ \vdots \\ \mathbf{y}_n &= \mathbf{X}_n \mathbf{b}_n + \mathbf{Z}_n \mathbf{g}_{n\text{CNS}} + \mathbf{Z}_n \mathbf{g}_{n\text{non-CNS}} + \mathbf{e}_n \text{ for trait n} \end{aligned}$$

where \mathbf{g}_{CNS} is a vector of random genetic effects due to the CNS genes and $\mathbf{g}_{\text{non-CNS}}$ is a vector of random genetic effects resulting from the non-CNS region.

We also tested another gene set that included candidate genes set for schizophrenia, autism, and intellectual disability (SAI).³ We matched these candidate genes with UCSC Genome Browser human genome version 18 (on which the discovery dataset was built) and retained 4,133

autosomal genes. It is noted that we excluded 479 genes flanking GWAS SNPs identified in the Swedish sample²⁸ to avoid artifact inflation in prediction accuracy. We annotated SNPs within the SAI genes (28% of the SNPs) and fitted genomic similarity matrices of the annotated SNPs and the rest of SNPs in the two-component model.

We had access to the PGC Cross-Disorder data and three independent validation datasets. The details of the PGC Cross-Disorder data with additionally available ADHD samples are described elsewhere.¹² The datasets stored in the PGC central server follow strict guidelines with local ethics committee approval. Genotype data from each study cohort were processed through the stringent PGC pipeline and imputation of autosomal SNPs was carried out with the HapMap3 reference sample.²⁹ In each imputation cohort, we retained only SNPs with MAF >0.01 and imputation $R^2 > 0.6$. The number of SNPs used in this study was 745,705. We excluded certain individuals to ensure that all samples from the five disorders were completely unrelated in the conventional sense, so that no pair of individuals had a genome-wide similarity relationship greater than 0.05. The numbers of case and control subjects used in this study are shown in Table 1. All phenotypes were controlled for cohort, sex, and the first 20 principal components estimated from genome-wide SNPs. Adjustments were performed for each trait.

In preliminary analysis, using the multivariate linear mixed model, we estimated genetic variances and genetic correlations between the five psychiatric disorders (Table 1). The estimates agreed with those reported in the previous study¹² (Figure S1) but were slightly less accurate (larger standard errors) because of the smaller sample size due to excluding genetically related samples across all five disorders rather than across only two traits in the bivariate analyses.

To evaluate the risk prediction performance of MTGBLUP, we performed within-study cross-validation of the PCG data, i.e., internal validation. We randomly split the data for each disease into a training sample containing ~80% of individuals and a validation sample containing the remaining ~20%³⁰ and repeated this five times. For assessing predictive performance in the internal validation, we calculated the correlation coefficient between the observed disease status and the predicted genomic risk score of the validation individuals. We also regressed observed disease status on risk scores. If the risk scores are unbiased estimates of genetic risk then the regression coefficient is expected to be 1, i.e., the covariance between true and estimated risks equals the variance of estimated risks. Deviations from 1 reflect the degree of bias of the risk scores. We averaged the correlation and regression coefficients and estimated empirical standard errors over five replicates. Using the empirical standard errors estimates, a t test was performed to assess differences in prediction accuracy between methods. In the within-study cross-validation, MTGBLUP outperformed single-trait genomic best linear unbiased prediction (STGBLUP)

Table 1. Estimates of SNP Heritability and Genetic Correlations from Multivariate Analysis of Five Psychiatric Disorders

Disorders	Cases	Controls	SNP-h ² on the Liability Scale	
			SNP-h ²	SE
SCZ	8,826	6,106	0.235	0.011
BIP	5,867	3,328	0.218	0.017
MDD	8,770	6,506	0.286	0.023
ASD	3,086	3,163	0.130	0.024
ADHD	3,997	8,479	0.281	0.022
			Genetic Correlation	SE
BIP/SCZ	5,867/8,826	3,328/6,106	0.590	0.048
MDD/SCZ	8,770/8,826	6,506/6,106	0.365	0.047
MDD/BIP	8,770/5,867	6,506/3,328	0.371	0.060
ASD/SCZ	3,086/8,826	3,163/6,106	0.194	0.071
ASD/BIP	3,086/5,867	3,163/3,328	0.084	0.089
ASD/MDD	3,086/8,770	3,163/6,506	0.054	0.089
ADHD/SCZ	3,997/8,826	8,479/6,106	0.055	0.046
ADHD/BIP	3,997/5,867	8,479/3,328	0.160	0.059
ADHD/MDD	3,997/8,770	8,479/6,506	0.242	0.059
ADHD/ASD	3,997/3,086	8,479/3,163	-0.044	0.088

Abbreviations are as follows: SE, standard error; SCZ, schizophrenia; BIP, bipolar disorder; MDD, major depressive disorder; ASD, autism spectrum disorder; ADHD, attention deficit disorder.

for all disorders: the gain in prediction accuracy was significant for schizophrenia ($p < 6.0 \times 10^{-8}$) and bipolar disorder ($p < 6.6 \times 10^{-11}$) (Figure S2). The slope from the regression of disease status on predicted risk score ranged from 0.88 to 1.14 (Table S1), indicating that the risk scores are well calibrated.

Results obtained from a within-study validation might not reflect the true performance when SNP effects estimated from the training data are spuriously associated with the diseases. To better assess the true prediction potential of MTGBLUP, risk scores derived from the complete PCG data were validated in independent samples for schizophrenia, bipolar, and major depressive disorder. As independent validation sets, we used Swedish schizophrenia²⁸ and bipolar GWAS data³¹ and the GENRED2

major depressive disorder dataset collected by the same methods as reported for the GENRED1 dataset.³² SNPs in the validation data were processed through the same stringent quality control as the discovery data. The Swedish schizophrenia data were imputed with HapMap3 as reference. The bipolar disorder data and major depressive disorder data were imputed with the 1000 Genomes Project data as reference. Post-imputation quality control was applied to exclude poorly imputed SNPs from the validation sets. Finally, we selected SNPs that matched those in the discovery set. The number of SNPs in each validation set is shown in Table 2. Individuals were removed from the validation datasets if they had relatedness >0.05 to any one of the individuals in the discovery set. Table 2 gives the numbers of case and control subjects in the independent validation datasets before and after excluding related individuals. In the discovery set, we obtained SNP solutions by applying SNP-BLUP (Equation 3) and then projected the SNP solution to the genotypes of the validation individuals (Equation 2). For assessing predictive performance in the independent validation, the correlation and regression coefficients were used as measures of prediction accuracy and biasedness, respectively, similar to the internal validation. A likelihood ratio test (LRT) was used to test for differences in prediction accuracy between methods comparing the likelihood of a logistic regression fitting the STGBLUP to that of a logistic regression fitting the MTGBLUP and STGBLUP jointly. In the logistic regression models, case-control status was used as the dependent variable. In the validation datasets, all phenotypes were controlled for cohort, sex, and the first 20 principal components just as in the discovery dataset. This external validation confirmed the superior performance of MTGBLUP over STGBLUP (Table 3). From the LRT to test differences in prediction accuracy, the model including MTGBLUP fitted the data significantly better ($p = 2.4 \times 10^{-24}$ for schizophrenia, 6.6×10^{-16} for bipolar disorder, and 0.010 for major depressive disorder) (Table 4). We further tested the two-components model fitting similarity matrices based on SNPs annotated in CNS genes and/or SNPs localized elsewhere (MTGBLUP-CNS and STGBLUP-CNS). Including the CNS component resulted in increased prediction accuracy for schizophrenia and bipolar disorder (Tables 3 and 4). We also tested a second annotation model replacing the CNS gene set with a SAI candidate genes set

Table 2. Numbers of Cases and Controls in the Independent Validation Data Sets before and after Removing Related Individuals

	SCZ (Swedish)		BIP (Swedish)		MDD (GENRED2)	
	Cases	Controls	Cases	Controls	Cases	Controls
All	5,193	6,391	2,208	6,056	831	474
After cut-off QC	4,068	5,471	2,029	5,338	822	466
Number of SNPs	745,631		645,237		673,109	

Abbreviations are as follows: SCZ, Swedish schizophrenia GWAS; BIP, Swedish bipolar disorder GWAS; MDD, GENRED2 GWAS.

Table 3. Prediction Accuracy for Schizophrenia, Bipolar Disorder, and Major Depressive Disorder in Independent Validation Data Sets

	Correlation			Regression Slope		
	SCZ	BIP	MDD	SCZ	BIP	MDD
STGBLUP	0.198	0.129	0.045	0.784	0.709	0.304
MTGBLUP	0.222	0.159	0.075	0.815	0.697	0.466
STGBLUP-CNS	0.203	0.132	0.045	0.789	0.719	0.306
MTGBLUP-CNS	0.224	0.162	0.076	0.807	0.690	0.476

Prediction accuracy is given as the correlation coefficient between the observed disease status and the predicted genomic risk score in the validation data. Regression deviated from one reflects the degree of bias of the risk scores.

(4,133 autosomal genes)³ (MTGBLUP-SAI or STGBLUP-SAI), but found little improvement due to SAI genes for three of the disorders (Tables S2 and S3).

When using independent validation samples, the slopes of the regression of the case-control status on the predictor were less than 1 (Table 3). The bias was relatively small for schizophrenia and bipolar disorder but larger for major depressive disorder. A slope less than 1 implies that the difference between the true genetic risks in a pair of individuals is less than that of the predicted genetic risk between them. The bias could be due to low predictive power (e.g., MDD) or to heterogeneity between the discovery and validation sample. In order to assess population differences, we calculated ancestry principal components from the POPRES reference sample^{33,34} and projected them into the discovery and validation samples and found ancestral differences between them for each disorder (Figure S3). We estimated that the SNP correlation²¹ between the discovery and validation datasets was significantly different from 1 for schizophrenia and bipolar disorder (Table S4; the point estimate was lower for major depressive disorder but the small sample size generated a large standard error so it was not significantly different from 1). To explore whether the found heterogeneity reflects real population differences or is caused by other factors that lead to differences between the discovery and validation samples such as batch effects, we looked for evidence of heterogeneity within PGC discovery samples for schizophrenia, bipolar disorder, and major depressive disorder (Appendix B). For each disorder, we divided the discovery sample into four groups based on the 25%, 50%, and 75% quartile of the

first principal component, which reflects ancestral population differences between individuals (Figure S4). Applying a reaction norm model^{35,36} (Appendix B), we found significant heterogeneity attributable to the ancestral population differences for schizophrenia and bipolar disorder (Table S5 and Figure S5). This indicates that for schizophrenia and bipolar disorder, real population heterogeneity rather than batch effects contribute to the reduced SNP correlation between discovery and validation sets. Previously we reported more heterogeneity between major depressive disorder cohorts than between schizophrenia cohorts,¹² where cohorts were defined based on sample collection, genotyping platform, and imputation set. The lack of evidence of population heterogeneity for the depression sample here might reflect that population heterogeneity not detectable given other heterogeneity within these samples.

After a common epidemiological approach to assess a continuous risk factor,³⁷ individuals were stratified into deciles according to the ranked values of the genetic risk predictors. We estimated the odds ratio of case-control status by contrasting each decile to the lowest decile (Figure 1). For all disorders, the odds ratio was highest between individuals in the highest and lowest decile, ranging from 1.3 to 5.5. Generally, odd ratios from MTGBLUP were larger than those from STGBLUP. For example, for bipolar disorder MTGBLUP increased the odds ratio by up to 60% compared to STGBLUP (odds ratio of 4.4 and 2.8, respectively). The discriminant power increased more for the annotation model with the CNS genes, compared to the one-component models without annotation (Figure 1).

Table 4. p Values from the Likelihood Ratio Test Comparing Different Models

x_1	x_2	SCZ	BIP	MDD
		p Values from LRT		
STGBLUP	MTGBLUP	2.4×10^{-24}	6.6×10^{-16}	1.0×10^{-2}
STGBLUP	STGBLUP-CNS	9.1×10^{-6}	4.6×10^{-3}	5.8×10^{-1}
MTGBLUP	MTGBLUP-CNS	2.4×10^{-3}	5.3×10^{-3}	3.3×10^{-1}
STGBLUP	MTGBLUP-CNS	6.7×10^{-26}	1.3×10^{-17}	7.3×10^{-3}

Likelihood ratio $LR = -2 [\log L(x_1) - \log L(x_1 + x_2)]$ where $\log L(x_1)$ ($\log L(x_1 + x_2)$) is the log likelihood from a logistic regression with case-control status as the dependent variable and x_1 (x_1 and x_2) as independent explanatory variable.

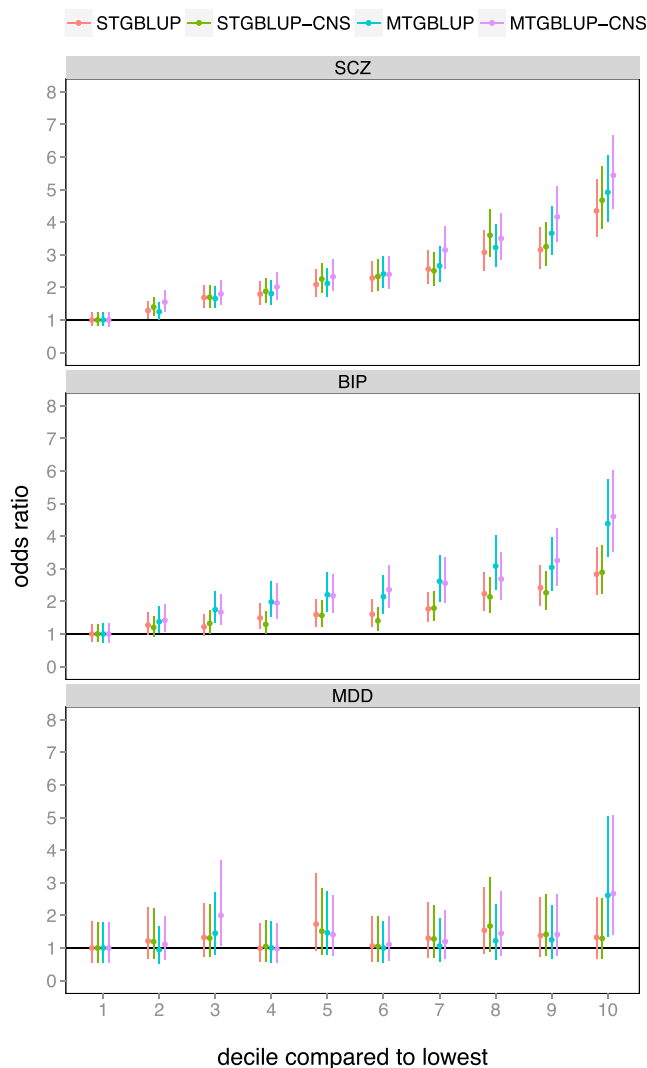


Figure 1. Odds Ratios of Individuals Stratified into Deciles Based on GBLUP Genetic Risk in Independent Samples, using the Decile with the Lowest Risk as the Baseline

The vertical error bars denote 95% CI. We note that the estimates for the different methods are highly correlated, and therefore the vertical error bars cannot be used to infer significance of difference between the methods (see [Appendix C](#)).

With increasing sample sizes, the odds ratio is expected to increase further.³⁷

We also quantified the gain in prediction accuracy from MTGBLUP in terms of sample size. Using recent results on prediction accuracy of polygenic scores derived from quantitative genetic theory,^{11,38} we inferred the sample sizes required to achieve the accuracies observed by the methods ([Figure 2](#)). We assumed prevalence of 1% for schizophrenia, 1% for bipolar disorder, and 15% for major depressive disorder. The proportion of cases in the sample was based on the real structure of the discovery data (59% for schizophrenia, 64% for bipolar disorder, and 57% for major depressive disorder). The effective number of SNPs was assumed to be 69,748 calculated with a weighted SNP method.³⁹ The observed accuracy

was within the theoretical expectation for schizophrenia and bipolar disorder, but not for major depressive disorder where the actual predictive power was lower. Accuracy of risk prediction for individual traits benefited from including the correlated disorders. The gain in accuracy of MTGBLUP compared to STGBLUP was equivalent to increasing the sample size for schizophrenia, bipolar disorder, and major depressive disorder by ~4,660 (95% confidence interval: 3,110–6,270), ~5,560 (2,830–8,640), and ~10,940 (730–24,440) individuals, respectively ([Figure 2](#)). Gains in accuracy were even greater with the CNS annotation model ([Table S6](#)). The 95% confidence interval was obtained according to the sampling error of the difference between the prediction accuracies ([Appendix C](#)).

In order to test how sensitive our results on prediction are against population stratification, we re-estimated the prediction accuracy (correlation), removing potential outliers that were ± 6 SD, 2 SD, 1.75 SD, 1.5 SD, 1.25 SD, or 1 SD away from the mean of the first and second principal component in the validation dataset ([Figure S6](#)). The accuracy of MTGBLUP and STGBLUP remained stable in all three diseases for which independent datasets were available. Restricting the samples to individuals whose values of the first and second principal component lay within one SD of the mean retained between 51% and 70% of the samples ([Figure S6](#)). This shows that the prediction accuracy was not substantially affected by ancestry outliers in the validation dataset.

We compared the performance of MTGBLUP with that of bivariate GBLUP (a special case of MTGBLUP). The accuracy of MTGBLUP was significantly higher than bivariate GBLUP except for a major depressive disorder risk prediction where the accuracy of MTGBLUP and that of the bivariate model involving schizophrenia and major depressive disorder was not significantly different ([Table S7](#) and [S8](#)).

Psychiatry lags behind other fields of medicine in terms of diagnostic tests that could facilitate early diagnosis and accurate classification of disorders. The considerable heritability of psychiatric disorders implies that the genome contains a large amount of information with potential diagnostic utility. However, the highly polygenic nature of psychiatric disorders makes it very hard to exploit this information, mostly because the effect of each individual locus contributing to disease risk can be estimated only with error, and the size of the error depends on factors such as allele frequency, effect size, and (crucially) sample size.

The genetic correlation between several diseases implies that a SNP contributing to risk of one disease will, on average, also be informative of the risk of the correlated diseases. Here, we have developed a multivariate method that can combine data from an arbitrary number of genetically correlated diseases, resulting in better estimates of the disease-specific SNP effects and thus generating more accurate predictors of individual risk. Our results demonstrate a significant advantage of incorporating data from

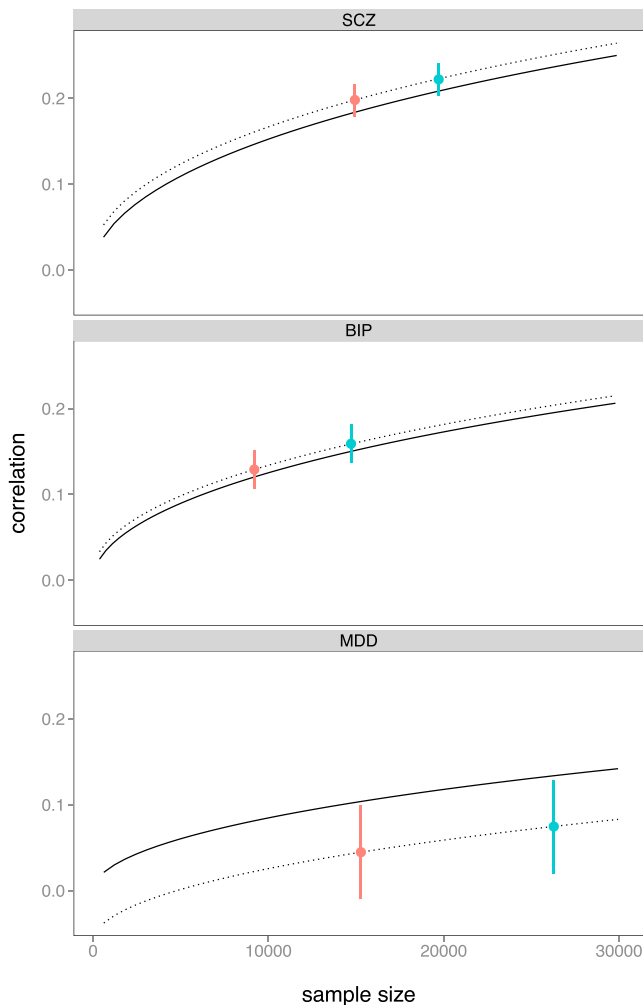


Figure 2. Theoretical and Observed Prediction Accuracy of STGBLUP and MTGBLUP Depending on Sample Size

Theoretical line of prediction accuracy increased with larger sample size (solid line), the observed accuracy achieved by STGBLUP with the actual sample size (red dot), and the observed accuracy achieved by MTGBLUP and inferred sample size (blue dot). The increase from MTGBLUP equates to ~4,660 samples for schizophrenia, ~5,550 samples for bipolar disorder, and ~10,940 for major depressive disorder. The vertical error bars denote 95% CI. We note that the estimates for the different methods are highly correlated, and therefore the vertical error bars cannot be used to infer significance of difference between the methods (see Appendix C).

multiple correlated diseases compared to single-trait analyses. Our estimates of pairwise genetic correlations obtained in independent datasets reconfirm previous results regarding the extent of genetic correlations between the five psychiatric disorders.¹² External validation demonstrated that the predictive models generalize to other populations, confirming that the correlations reflect pleiotropy between the disorders rather than artifacts.

We used a multiple random effects model that fitted two components, one due to annotated SNPs and the other due to the rest of SNPs. The prediction accuracy significantly increased when using an appropriate gene set. For

example, the gain in predictive accuracy in terms of sample size equivalence increased from 4,660 to 5,080 for schizophrenia, from 5,550 to 6,220 for bipolar disorder, and from 10,940 to 11,550 for major depressive disorder when using the CNS genes annotation^{12,26} (Table S6). This demonstrates that the multiple random effects model in MTGBLUP can be useful especially for psychiatric disorders where prediction accuracy is hardly improved by other advanced methods.^{4,5}

Zhou and Stephens⁴⁰ recently introduced a multivariate linear mixed model algorithm that is particularly suited for genome-wide association studies. Their method requires that multiple traits are measured on the same individual or that the level of missingness is sufficiently small so that missing phenotypes can be imputed. However, this algorithm is not useful when phenotypes are collected from independent datasets as in the PGC data where dependent variables are totally missing for the other four traits as is typical of disease-ascertained cohorts. Moreover, the efficiency of Zhou and Stephens' algorithm substantially decreases when fitting multiple random effects (e.g., the annotation model).

Korte et al.⁴¹ proposed a similar model to MTGREML using ASReml⁴² that is as flexible as our method in that it can handle partial overlapping or disjoint sets of phenotypes. However, our algorithm is different from that used in ASReml and is much more efficient when using genomic data²⁰ (see Appendix A). Moreover, Korte et al. did not explore their method with respect to improvements in risk prediction.

Even though sensitivity and specificity of genetic diagnostics to predict an individual's risk of psychiatric disorders are generally low, genetic risk scores can still be a valuable tool for research to stratify a heterogeneous population in groups with shared "genomic" characteristics. It was suggested that psychiatric diagnoses encompass several clinically similar phenotypes with distinct pathophysiology and that stratification according to individual heterogeneity is an important requirement for the development of treatments targeted at specific disease subtypes.¹⁷⁻¹⁹ Our proposed multivariate approach with the annotation model is a flexible and powerful tool for such stratification. The MTGREML and MTGBLUP package and documentation are publicly available online, which we anticipate will be implemented into the GCTA package.²² Using a CPU running at 2.2 GHz, analyzing 58,128 samples with 5 disjoint sets of phenotypes (e.g., the PGC data) takes ~7 hr per each iteration in MTGREML. Convergence is usually achieved within 10 iterations. The virtual memory required for such data is ~45 GB. Good starting values (probably from single-trait GREML⁴¹) can reduce the number of iterations to convergence and our software has the option to provide starting values. The computational time increases cubically with sample size, e.g., analyzing sample size of 10,000 takes a few minutes per each iteration. Our software provides a parallelization option that can reduce computational burden substantially;

for example, speed is increased by a factor of ten when using 20 CPUs. The number of traits hardly affects running time if phenotypes are non-overlapping.

Appendix A. Average of Hessian and Fisher Information Matrix for the Multivariate Model

The log likelihood of the multivariate model is

$$\ln L = -\frac{1}{2} [\ln |\mathbf{V}| + \ln |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}'\mathbf{P}\mathbf{y}]$$

where \ln is the natural log and $||$ the determinant of the associated matrices. The projection matrix is defined as $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$ with

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{X}_n \end{bmatrix}, \text{ and } \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}.$$

The Newton-Raphson algorithm obtains the MTGREML estimates with the following equation⁴³

$$\Theta^{(k+1)} = \Theta^{(k)} - (\mathbf{H}^{(k)})^{-1} \frac{\partial L}{\partial \Theta} \Big|_{\Theta^{(k)}} \quad (\text{Equation A1})$$

where Θ is a column vector of estimated variance components, k is the iteration round, $\partial L / \partial \Theta$ is a column vector of the first derivatives of the log likelihood function with respect to each variance component, and \mathbf{H} is the Hessian matrix, which consists of the second derivatives of the log likelihood function with respect to the variance components. In Fisher's scoring method, the inverse of the Hessian matrix in Equation A1 is replaced by its expected value.⁴³

$$\Theta^{(k+1)} = \Theta^{(k)} + (\mathbf{F}^{(k)})^{-1} \frac{\partial L}{\partial \Theta} \Big|_{\Theta^{(k)}} \quad (\text{Equation A2})$$

The derivation of the Hessian matrix and the Fisher information matrix has been described in several studies.^{43,44} The Hessian matrix for the multivariate model is

$$\mathbf{H} = \frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{1}{2} \left[\text{tr} \left(\frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{P} \right) - \mathbf{y}' \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{P} \mathbf{y} \right] \quad (\text{Equation A3})$$

where \mathbf{y} , \mathbf{P} , and \mathbf{V} are defined in the section "Multivariate Linear Mixed Model" in the main text. The Fisher information (F) matrix is

$$\mathbf{F} = E \left(\frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2} \right) = \frac{1}{2} \left[\text{tr} \left(\frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{P} \right) \right]. \quad (\text{Equation A4})$$

Gilmour et al.⁴² and Johnson and Thompson⁴⁵ used the average of the \mathbf{H} and \mathbf{F} that was estimated based on Henderson's mixed model equation (MME).⁴⁶ The MME-based average information algorithm is efficient particularly when covariance structure fitted in the model is sparse.

Lee and van der Werf²⁰ introduced the direct average information algorithm where average information matrix was derived directly from the \mathbf{V} and \mathbf{P} matrix. When using non-zero elements of covariance structure, this direct average information algorithm is much more efficient than the MME-based average information algorithm. The equation for the iterative AI algorithm is

$$\Theta^{(k+1)} = \Theta^{(k)} + (\mathbf{AI}^{(k)})^{-1} \frac{\partial L}{\partial \Theta} \Big|_{\Theta^{(k)}}$$

where \mathbf{AI} is the average information matrix and that for multivariate model can be written as

$$\mathbf{AI} = \frac{1}{2} \left[\mathbf{y}' \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{P} \mathbf{y} \right].$$

The first derivative for each variance covariance component i can be obtained as^{43,44}

$$\frac{\partial L}{\partial \sigma_i^2} = -\frac{1}{2} \text{tr} \left(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \right) + \frac{1}{2} \mathbf{y}' \mathbf{P} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{P} \mathbf{y}.$$

Appendix B. Reaction Norm Model to Test Heterogeneity across Populations Classified by the Ancestry Principal Component

Reaction norm models have been used in ecology and evolution to study genotype \times environment interaction.^{35,36} Genotype \times environment interaction (G \times E) means that different genotypes respond different to environmental changes, i.e., norms of reaction. In the model, a random intercept and a random slope, as covariance functions, are estimated that can describe genetic and phenotypic variation across different environments. The slope of the reaction norm is often called phenotypic plasticity or environmental sensitivity. The amount of variation in slope in the population indicates the extent of G \times E.^{35,36} Here, we describe a reaction norm model to test heterogeneity across populations. We group each sample set into four populations by splitting them into the four quartiles of the first ancestry principal component. Whereas typically reaction norm models would compare samples with different categories of environmental factors to each other, we use the model to compare the samples in different principal component quartiles to each other. We limit our interpretation to heterogeneity across the groups and do not speculate about potential causes like G \times E or G \times G interaction. We apply the model to each disorder of the PGC data. Incorporating population difference among samples, the linear mixed model can be rewritten as

$$y_{ij} = b_{ij} + g_{ij} + e_{ij},$$

where y_{ij} is the observation for individual i in population class j ($j = 1, \dots, P$ where P is the number of populations classified by the ancestry principal component, in our case four), b_{ij} is fixed effects, g_{ij} is genetic effects, and e_{ij} is

residual effects. We applied a reaction norm model to fit functions of the ancestry principal component as covariables using Legendre polynomials.

$$y_{ij} = b_{ij} + \sum_{m=0}^{k-1} \alpha_{im} f_m(p_{ij}) + e_{ij}$$

p_{ij} is the average of the ancestry principal components in the j^{th} population class containing individual i , $f_m(p_{ij})$ is the m^{th} Legendre polynomials evaluated for p_{ij} , α_{im} is the m^{th} genetic random regression coefficients for the i^{th} individual, and k is the orders of fit. The genetic covariance between individual i in population class j and i' in population class j' is

$$\text{cov}(g_{ij}, g_{i'j'}) = \sum_{m=0}^{k-1} \sum_{l=0}^{k-1} f_m(p_{ij}) f_l(p_{i'j'}) \text{cov}(\alpha_{im}, \alpha_{il}).$$

This can be written in a matrix form as

$$\mathbf{V}_g = \mathbf{F} \mathbf{K} \mathbf{F}'$$

where \mathbf{F} is the matrix of Legendre polynomials evaluated at given ancestry principal components and \mathbf{K} is the covariance coefficient matrix consisting of random regression coefficients, i.e.,

$$\mathbf{K} = \text{cov}(\alpha_{im}, \alpha_{il}) = \begin{bmatrix} \text{var}(\alpha_0) & \cdots & \text{cov}(\alpha_0, \alpha_k) \\ \vdots & \ddots & \vdots \\ \text{cov}(\alpha_k, \alpha_0) & \cdots & \text{var}(\alpha_k) \end{bmatrix}.$$

The optimal order of the polynomial was determined with a likelihood ratio test by comparing the likelihood of models with higher order to the null model with $k = 1$.

Appendix C. Estimating the Sampling Error of the Difference between Prediction Accuracies

It is assumed that there are three normalized variables with the covariance structure as below, mimicking the MTGBLUP, STGBLUP, and outcome variable.

	m	s	y
m	1	0.927	0.222
s	0.927	1	0.189
y	0.222	0.189	1

We are interested in estimating the sampling error of the difference between $\text{cor}(m,y)$ and $\text{cor}(s,y)$. The sampling variance of the difference (σ_d^2) can be expressed as

$$\sigma_d^2 = \sigma_{\text{cor}(m,y)}^2 + \sigma_{\text{cor}(s,y)}^2 - 2 \cdot r \cdot \sigma_{\text{cor}(m,y)} \cdot \sigma_{\text{cor}(s,y)}, \quad (\text{Equation C1})$$

where $\sigma_{\text{cor}(m,y)}^2$ is the sampling variance of $\text{cor}(m,y)$, $\sigma_{\text{cor}(s,y)}^2$ is the sampling variance of $\text{cor}(s,y)$, and r is the correlation

between $\text{cor}(m,y)$ and $\text{cor}(s,y)$. We show here that r is approximately equal to $\text{cor}(m,s)$.

With N records for each variable, correlations among the variables can be written as

$$\text{cor}(m,y) = E(m \cdot y) = \frac{1}{N} \sum_{i=1}^N m_i y_i$$

$$\text{cor}(s,y) = E(s \cdot y) = \frac{1}{N} \sum_{i=1}^N s_i y_i$$

$$\text{cor}(m,s) = E(m \cdot s) = \frac{1}{N} \sum_{i=1}^N m_i s_i.$$

For T replicates, the expected value of the product of $\text{cor}(m,y)$ and $\text{cor}(s,y)$ can be written as

$$\begin{aligned} E[E(m \cdot y) \cdot E(s \cdot y)] &= \frac{1}{T} \sum_j \left[\left(\frac{1}{N} \sum_{i=1}^N m_i y_i \right) \left(\frac{1}{N} \sum_{i=1}^N s_i y_i \right) \right]_j \\ &= \frac{1}{T} \sum_j \left[\left(\frac{m_1 y_1}{N} + \dots + \frac{m_N y_N}{N} \right) \left(\frac{s_1 y_1}{N} + \dots + \frac{s_N y_N}{N} \right) \right]_j. \end{aligned}$$

If m and s are uncorrelated, this reduces to

$$E[E(m \cdot y) \cdot E(s \cdot y)] = E[E(m \cdot y)] \cdot E[E(s \cdot y)]$$

If m and s are correlated, there is an additional term,

$$\begin{aligned} E[E(m \cdot y) E(s \cdot y)] &\cong \\ \frac{1}{T} \sum_j \left[\left(\frac{1}{N^2} \sum_{i=1}^N m_i s_i \right) \right]_{jj} + E[E(m \cdot y)] \cdot E[E(s \cdot y)] &= \\ \frac{1}{T} \sum_j \left[\frac{1}{N} E(m \cdot s) \right]_j + E[E(m \cdot y)] \cdot E[E(s \cdot y)] &= \\ \frac{1}{N} E[E(m \cdot s)] + E[E(m \cdot y)] \cdot E[E(s \cdot y)] \end{aligned}$$

Therefore,

$$\begin{aligned} \text{cov}[E(m \cdot y), E(s \cdot y)] &= E[E(m \cdot y) E(s \cdot y)] \\ &\quad - E[E(m \cdot y)] \cdot E[E(s \cdot y)] \cong \frac{1}{N} E[E(m \cdot s)]. \end{aligned}$$

With $\text{var}[E(m \cdot y)] \cong \text{var}[E(s \cdot y)] \cong 1/N$, the correlation between $\text{cor}(m,y)$ and $\text{cor}(s,y)$ (r) can be approximated as

$$\text{cor}[E(m \cdot y), E(s \cdot y)] \cong E[E(m \cdot s)] \cong \text{cor}(m, s).$$

This expression was checked and validated by simulations (result not shown).

Here we have shown that Equation C1 can be used to estimate the sampling variance (the square of the standard error) of the difference in correlation between the STGBLUP and MTGBLUP predictors (which are themselves correlated with each other) and the outcome variable (the adjusted phenotype). This allows us to estimate the 95% confidence interval of the increase in correlation that MTGBLUP achieves over STGBLUP. Note that because

the two predictors are correlated, this is a smaller confidence interval than that of the correlation between MTGBLUP and the outcome variable (which is shown in [Figures 1 and 2](#)). By using the method described above, we can transform the confidence interval from the correlation scale to the sample size scale, to get estimates of the effective increase in sample size achieved by MTGBLUP ([Table S6](#)).

Supplemental Data

Supplemental Data include six figures, eight tables, and consortium members and affiliations and can be found with this article online at <http://dx.doi.org/10.1016/j.ajhg.2014.12.006>.

Consortia

The members of Cross-Disorder Working Group of the Psychiatric Genomics Consortium are Devin Absher, Ingrid Agartz, Huda Akil, Farooq Amin, Ole A. Andreassen, Adebayo Anjorin, Richard Anney, Dan E. Arking, Philip Asherson, Maria H. Azevedo, Lena Backlund, Judith A. Badner, Anthony J. Bailey, Tobias Banaschewski, Jack D. Barchas, Michael R. Barnes, Thomas B. Barrett, Nicholas Bass, Agatino Battaglia, Michael Bauer, Mónica Bayés, Frank Bellivier, Sarah E. Bergen, Wade Berrettini, Catalina Betancur, Thomas Bettecken, Joseph Biederman, Elisabeth B. Binder, Donald W. Black, Douglas H.R. Blackwood, Cinnamon S. Bloss, Michael Boehnke, Dorret I. Boomsma, Jerome Breen, René Breuer, Richard Bruggeman, Nancy G. Buccola, Jan K. Buitelaar, William E. Bunney, Joseph D. Buxbaum, William F. Byerley, Sian Caesar, Wiepke Cahn, Rita M. Cantor, Miguel Casas, Aravinda Chakravarti, Kimberly Chambert, Khalid Choudhury, Sven Cichon, C. Robert Cloninger, David A. Collier, Edwin H. Cook, Hilary Coon, Bru Cormand, Paul Cormican, Aiden Corvin, William H. Coryell, Nicholas Craddock, David W. Craig, Ian W. Craig, Jennifer Crosbie, Michael L. Cuccaro, David Curtis, Darina Czamara, Mark J. Daly, Susmita Datta, Geraldine Dawson, Richard Day, Eco J. De Geus, Franziska Degenhardt, Bernie Devlin, Srdjan Djurovic, Gary J. Donohoe, Alysia E. Doyle, Jubao Duan, Frank Dudbridge, Eftichia Duketis, Richard P. Ebstein, Howard J. Edenberg, Josephine Elia, Sean Ennis, Bruno Etain, Ayman Fanous, Stephen V. Faraone, Anne E. Farmer, I. Nicol Ferrier, Matthew Flickinger, Eric Fombonne, Tatiana Foroud, Josef Frank, Barbara Franke, Christine Fraser, Robert Freedman, Nelson B. Freimer, Christine M. Freitag, Marion Friedl, Louise Frisén, Louise Gallagher, Pablo V. Gejman, Lyudmila Georgieva, Elliot S. Gershon, Daniel H. Geschwind, Ina Giegling, Michael Gill, Scott D. Gordon, Katherine Gordon-Smith, Elaine K. Green, Tiffany A. Greenwood, Dorothy E. Grice, Magdalena Gross, Detelina Grozeva, Weihua Guan, Hugh Gurling, Lieuwe De Haan, Jonathan L. Haines, Hakon Hakonarson, Joachim Hallmayer, Steven P. Hamilton, Marian L. Hamshere, Thomas F. Hansen, Annette M. Hartmann, Martin Hautzinger, Andrew C. Heath, Anjali K. Henders, Stefan Herms, Ian B. Hickie, Maria Hipolito, Susanne Hoefels, Peter A. Holmans, Florian Holsboer, Witte J. Hoogendijk, Jouke-Jan Hottenga, Christina M. Hultman, Vanessa Hus, Andrés Ingason, Marcus Ising, Stéphane Jamain, Ian Jones, Lisa Jones, Anna K. Kähler, René S. Kahn, Radhika Kandaswamy, Matthew C. Keller, John R. Kelsoe, Kenneth S. Kendler, James L. Kennedy, Elaine Kenny, Lindsey Kent, Yunjung Kim, George K. Kirov, Sabine M. Klauck, Lambertus Klei, James A. Knowles, Martin

A. Kohli, Daniel L. Koller, Bettina Konte, Ania Korszun, Lydia Krabendam, Robert Krasucki, Jonna Kuntsi, Phoenix Kwan, Mikael Landén, Niklas Långström, Mark Lathrop, Jacob Lawrence, William B. Lawson, Marion Leboyer, David H. Ledbetter, Phil H. Lee, Todd Lencz, Klaus-Peter Lesch, Douglas F. Levinson, Cathryn M. Lewis, Jun Li, Paul Lichtenstein, Jeffrey A. Lieberman, Dan-Yu Lin, Don H. Linszen, Chunyu Liu, Falk W. Lohoff, Sandra K. Loo, Catherine Lord, Jennifer K. Lowe, Susanne Lucae, Donald J. MacIntyre, Pamela A.F. Madden, Elena Maestrini, Patrik K.E. Magnusson, Pamela B. Mahon, Wolfgang Maier, Anil K. Malhotra, Shrikant M. Mane, Christa L. Martin, Nicholas G. Martin, Manuel Mattheisen, Keith Matthews, Morten Mattingsdal, Steven A. McCarrroll, Kevin A. McGhee, James J. McGough, Patrick J. McGrath, Peter McGuffin, Melvin G. McInnis, Andrew McIntosh, Rebecca McKinney, Alan W. McLean, Francis J. McMahon, William M. McMahon, Andrew McQuillin, Helena Medeiros, Sarah E. Medland, Sandra Meier, Ingrid Melle, Fan Meng, Jobst Meyer, Christel M. Middeldorp, Lefkos Middleton, Vihra Milanova, Ana Miranda, Anthony P. Monaco, Grant W. Montgomery, Jennifer L. Moran, Daniel Moreno-De-Luca, Gunnar Morken, Derek W. Morris, Eric M. Morrow, Valentina Moskvina, Bryan J. Mowry, Pierandrea Muglia, Thomas W. Mühleisen, Bertram Müller-Myhsok, Michael Murtha, Richard M. Myers, Inez Myin-Germeys, Benjamin M. Neale, Stan F. Nelson, Caroline M. Nievergelt, Ivan Nikolov, Vishwajit Nimgaonkar, Willem A. Nolen, Markus M. Nöthen, John I. Nurnberger, Evaristus A. Nwulia, Dale R. Nyholt, Michael C. O'Donovan, Colm O'Dushlaine, Robert D. Oades, Ann Olincy, Guiomar Oliveira, Line Olsen, Roel A. Ophoff, Urban Osby, Michael J. Owen, Aarno Palotie, Jeremy R. Parr, Andrew D. Paterson, Carlos N. Pato, Michele T. Pato, Brenda W. Penninx, Michele L. Pergadia, Margaret A. Pericak-Vance, Roy H. Perlis, Benjamin S. Pickard, Jonathan Pimm, Joseph Piven, Danielle Posthuma, James B. Potash, Fritz Poustka, Peter Propping, Shaun M. Purcell, Vinay Puri, Digby J. Quedsted, Emma M. Quinn, Josep Antoni Ramos-Quiroga, Henrik B. Rasmussen, Soumya Raychaudhuri, Karola Rehnström, Andreas Reif, Marta Ribasés, John P. Rice, Marcella Rietschel, Stephan Ripke, Kathryn Roeder, Herbert Roeyers, Lizzy Rossin, Aribert Rothenberger, Guy Rouleau, Douglas Ruderfer, Dan Rujescu, Alan R. Sanders, Stephan J. Sanders, Susan L. Santangelo, Russell Schachar, Martin Schalling, Alan F. Schatzberg, William A. Scheftner, Gerard D. Schellenberg, Stephen W. Scherer, Nicholas J. Schork, Thomas G. Schulze, Johannes Schumacher, Markus Schwarz, Edward Scolnick, Laura J. Scott, Joseph A. Sergeant, Jianxin Shi, Paul D. Shilling, Stanley I. Shyn, Jeremy M. Silverman, Pamela Sklar, Susan L. Slager, Susan L. Smalley, Johannes H. Smit, Erin N. Smith, Jordan W. Smoller, Edmund J.S. Sonuga-Barke, David St Clair, Matthew State, Michael Steffens, Hans-Christoph Steinhausen, John S. Strauss, Jana Strohmaier, T. Scott Stroup, Patrick F. Sullivan, James Sutcliffe, Peter Szatmari, Szabolcs Szelinger, Anita Thapar, Srinivasa Thirumalai, Robert C. Thompson, Alexandre A. Todorov, Federica Tozzi, Jens Treutlein, Jung-Ying Tzeng, Manfred Uhr, Edwin J.C.G. van den Oord, Gerard Van Grootheest, Jim Van Os, Astrid M. Vicente, Veronica J. Vieland, John B. Vincent, Peter M. Visscher, Christopher A. Walsh, Thomas H. Wassink, Stanley J. Watson, Lauren A. Weiss, Myrna M. Weissman, Thomas Werge, Thomas F. Wienker, Durk Wiersma, Ellen M. Wijsman, Gonneke Willemsen, Nigel Williams, A. Jeremy Willsey, Stephanie H. Witt, Naomi R. Wray, Wei Xu, Allan H. Young, Timothy W. Yu, Stanley Zammit, Peter P. Zandi, Peng Zhang, Frans G. Zitman, and Sebastian Zöllner. Affiliations of consortium members are available in the [Supplemental Data](#).

Acknowledgments

This study was supported by the Australian Research Council (FT0991360 and DE130100614) and the National Health and Medical Research Council (613608, 1011506, 1047956, and 1080157). The Psychiatric Genomics Consortium is supported by National Institute of Mental Health (NIMH) grant U01 MH085520. We acknowledge the funding that supported the Swedish schizophrenia study (NIMH R01 MH077139), the Stanley Center for Psychiatric Research, the Sylvan Herman Foundation, the Karolinska Institutet, Karolinska University Hospital, the Swedish Research Council, the Stockholm County Council, the Söderström Königsk Foundation, and the Netherlands Scientific Organization (NWO 645-000-003). Statistical analyses were carried out on the Genetic Cluster Computer, which is financially supported by the Netherlands Scientific Organization (NOW; 480-05-003). The GenRED GWAS project was supported by NIMH R01 grants MH061686 (D.F.L.), MH059542 (W.C.), MH075131 (W.B. Lawson), MH059552 (J.B.P.), MH059541 (W.A.S.), and MH060912 (M.M.W.).

Received: October 23, 2014

Accepted: December 8, 2014

Published: January 29, 2015

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, <http://browser.1000genomes.org>
Complex Traits Genomic Group, [http://www.cnsgenomics.com/
software/](http://www.cnsgenomics.com/software/)

GCTA, <http://www.complextaitgenomics.com/software/gcta/>

Genetic Cluster Computer, <http://www.geneticcluster.org/>

International HapMap Project, <http://hapmap.ncbi.nlm.nih.gov/>

Multi-trait GREML and GBLUP Analysis (MTG), <https://github.com/ukrmaie1/mtgblup> or [http://www.cnsgenomics.com/
software/](http://www.cnsgenomics.com/software/)

Psychiatric Genomics Consortium, <http://www.med.unc.edu/pgc/>

UCSC Genome Browser, <http://genome.ucsc.edu>

References

1. Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. *Am. J. Hum. Genet.* *90*, 7–24.
2. Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* *106*, 9362–9367.
3. Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* *506*, 185–190.
4. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* *9*, e1003264.
5. Speed, D., and Balding, D.J. (2014). MultiBLUP: improved SNP-based prediction for complex traits. *Genome Res.* *24*, 1550–1557. <http://dx.doi.org/10.1101/gr.169375.113>.
6. Lee, S.H., van der Werf, J.H., Hayes, B.J., Goddard, M.E., and Visscher, P.M. (2008). Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet.* *4*, e1000231.
7. Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., Reich, C.M., Mason, B.A., and Goddard, M.E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* *95*, 4114–4129.
8. Wei, Z., Wang, K., Qu, H.-Q., Zhang, H., Bradfield, J., Kim, C., Frackleton, E., Hou, C., Glessner, J.T., Chiavacci, R., et al. (2009). From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genet.* *5*, e1000678.
9. Li, C., Yang, C., Gelernter, J., and Zhao, H. (2014). Improving genetic risk prediction by leveraging pleiotropy. *Hum. Genet.* *133*, 639–650.
10. Daetwyler, H.D., Villanueva, B., and Woolliams, J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* *3*, e3395.
11. Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* *9*, e1003348.
12. Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al.; Cross-Disorder Group of the Psychiatric Genomics Consortium; International Inflammatory Bowel Disease Genetics Consortium (IIBDGC) (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* *45*, 984–994.
13. Henderson, C.R., and Quass, R.L. (1976). Multiple trait evaluation using relatives' records. *J. Anim. Sci.* *43*, 1188–1197.
14. Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., and Su, G. (2014). Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* *15*, 30.
15. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* *491*, 56–65.
16. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., et al.; International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* *467*, 52–58.
17. Kapur, S., Phillips, A.G., and Insel, T.R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* *17*, 1174–1179.
18. Trusheim, M.R., Berndt, E.R., and Douglas, F.L. (2007). Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat. Rev. Drug Discov.* *6*, 287–293.
19. Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* *167*, 748–751.
20. Lee, S.H., and van der Werf, J.H.J. (2006). An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genet. Sel. Evol.* *38*, 25–43.
21. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M., and Wray, N.R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* *28*, 2540–2542.

22. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* *88*, 76–82.
23. Hayes, B.J., Visscher, P.M., and Goddard, M.E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* *91*, 47–60.
24. VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* *91*, 4414–4423.
25. Strandén, I., and Garrick, D.J. (2009). Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* *92*, 2971–2975.
26. Raychaudhuri, S., Korn, J.M., McCarroll, S.A., Altshuler, D., Sklar, P., Purcell, S., and Daly, M.J.; International Schizophrenia Consortium (2010). Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet.* *6*, e1001097.
27. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., Wray, N.R., et al.; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS) (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* *44*, 247–250.
28. Ripke, S., O’Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al.; Multicenter Genetic Studies of Schizophrenia Consortium; Psychosis Endophenotypes International Consortium; Wellcome Trust Case Control Consortium 2 (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* *45*, 1150–1159.
29. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* *381*, 1371–1379.
30. Roche, O., Schneider, P., Zuegge, J., Guba, W., Kansy, M., Alanine, A., Bleicher, K., Danel, F., Gutknecht, E.-M., Rogers-Evans, M., et al. (2002). Development of a virtual screening method for identification of “frequent hitters” in compound libraries. *J. Med. Chem.* *45*, 137–142.
31. Bergen, S.E., O’Dushlaine, C.T., Ripke, S., Lee, P.H., Ruderfer, D.M., Akterin, S., Moran, J.L., Chambert, K.D., Handsaker, R.E., Backlund, L., et al. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol. Psychiatry* *17*, 880–886.
32. Shi, J., Potash, J.B., Knowles, J.A., Weissman, M.M., Coryell, W., Scheftner, W.A., Lawson, W.B., DePaulo, J.R., Jr., Gejman, P.V., Sanders, A.R., et al. (2011). Genome-wide association study of recurrent early-onset major depressive disorder. *Mol. Psychiatry* *16*, 193–201.
33. Nelson, M.R., Bryc, K., King, K.S., Indap, A., Boyko, A.R., Novembre, J., Briley, L.P., Maruyama, Y., Waterworth, D.M., Waeber, G., et al. (2008). The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* *83*, 347–358.
34. Chen, C.-Y., Pollack, S., Hunter, D.J., Hirschhorn, J.N., Kraft, P., and Price, A.L. (2013). Improved ancestry inference using weights from external reference panels. *Bioinformatics* *29*, 1399–1406.
35. Kirkpatrick, M., Lofsvold, D., and Bulmer, M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* *124*, 979–993.
36. Meyer, K., and Hill, W. (1997). Estimation of genetic and phenotypic covariance functions for longitudinal or ‘repeated’ records by restricted maximum likelihood. *Livest. Prod. Sci.* *47*, 185–200.
37. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* *511*, 421–427.
38. Lee, S.H., and Wray, N.R. (2013). Novel genetic analysis for case-control genome-wide association studies: quantification of power and genomic prediction accuracy. *PLoS ONE* *8*, e71494.
39. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* *91*, 1011–1021.
40. Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* *11*, 407–409.
41. Korte, A., Vilhjálmsson, B.J., Segura, V., Platt, A., Long, Q., and Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* *44*, 1066–1071.
42. Gilmour, A.R., Thompson, R., and Cullis, B.R. (1995). Average information REML: An efficient algorithm for variance parameters estimation in linear mixed models. *Biometrics* *51*, 1440–1450.
43. Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits* (Sunderland: Sinauer Associates).
44. Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components* (New York: John Wiley & Sons).
45. Johnson, D.L., and Thompson, R. (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J. Dairy Sci.* *78*, 449–456.
46. Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* *31*, 423–447.

The American Journal of Human Genetics

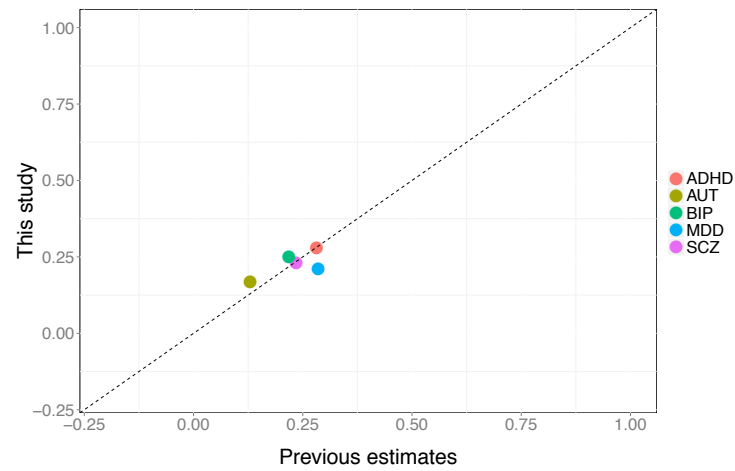
Supplemental Data

**Joint Analysis of Psychiatric Disorders Increases
Accuracy of Risk Prediction for Schizophrenia,
Bipolar Disorder, and Major Depressive Disorder**

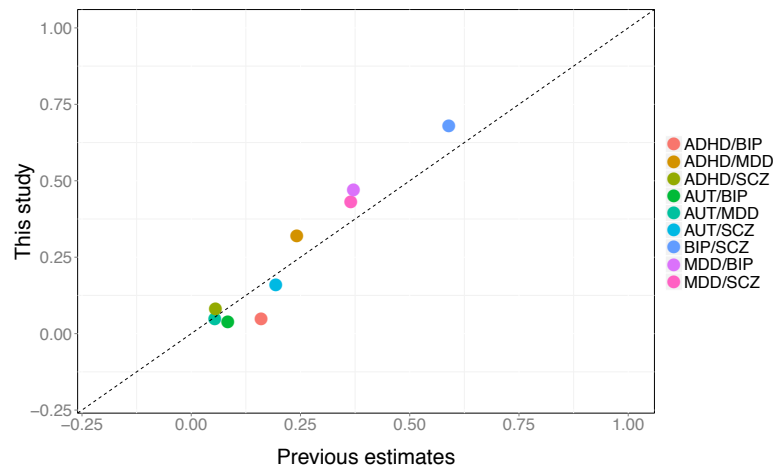
Robert Maier, Gerhard Moser, Guo-Bo Chen, Stephan Ripke, Cross disorder Working group of the Psychiatric Genomics Consortium, William Coryell, James B. Potash, William A. Scheftner, Jianxin Shi, Myrna M. Weissman, Christina M. Hultman, Mikael Landén, Douglas F. Levinson, Kenneth S. Kendler, Jordan W. Smoller, Naomi R. Wray, and S. Hong Lee

Supplementary data

A. Heritability



B. Genetic correlations



C. SNP-coheritability

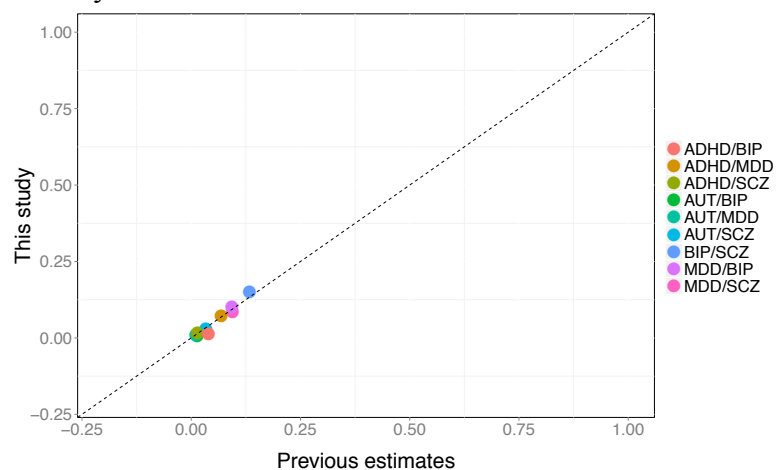


Figure S1. Previous estimates (Lee et al. 2013) plotted against estimates from this study for 5 psychiatric disorders. Both studies and the previous utilised the same data. However, the previous estimates used a bivariate model and so overlapping and closely related samples were excluded on a pairwise basis. In this study overlapping

samples and closely related samples were removed across all 5 disorders generating small samples per disease. A. SNP-heritability (correlation coefficient between previous and current estimates = 0.69); B. Genetic correlations (correlation coefficient between previous and current estimates = 0.98); C. SNP-coheritability (correlation coefficient between previous and current estimates = 0.98)

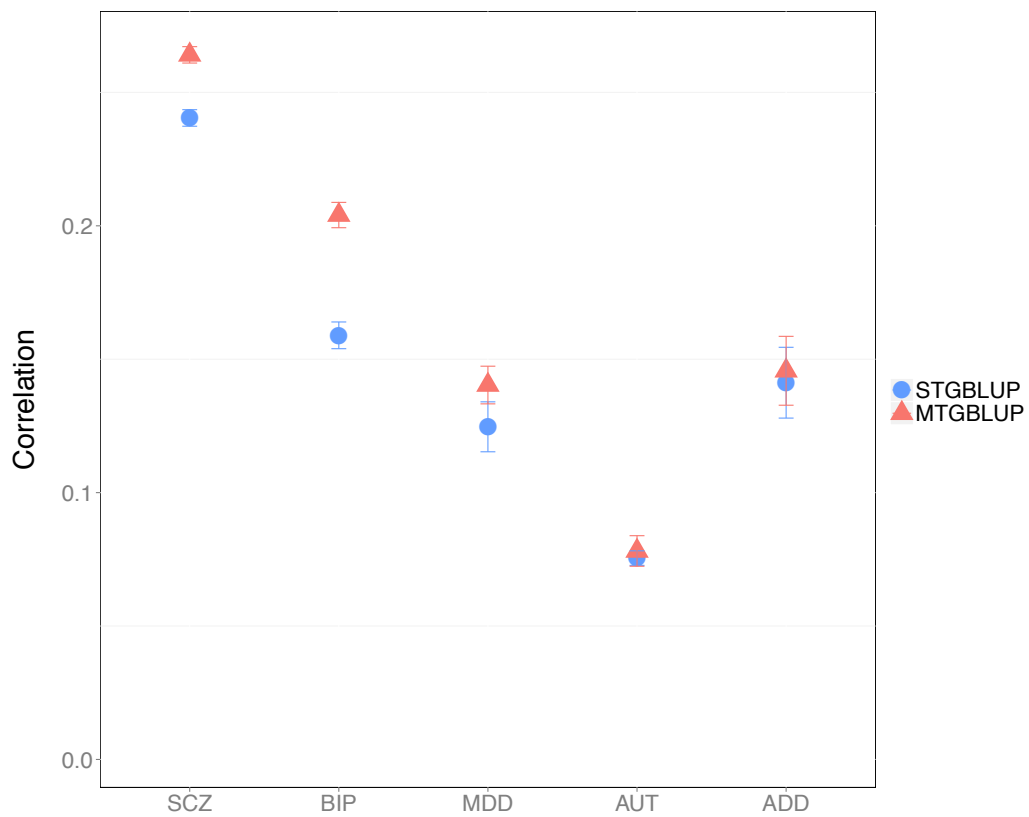
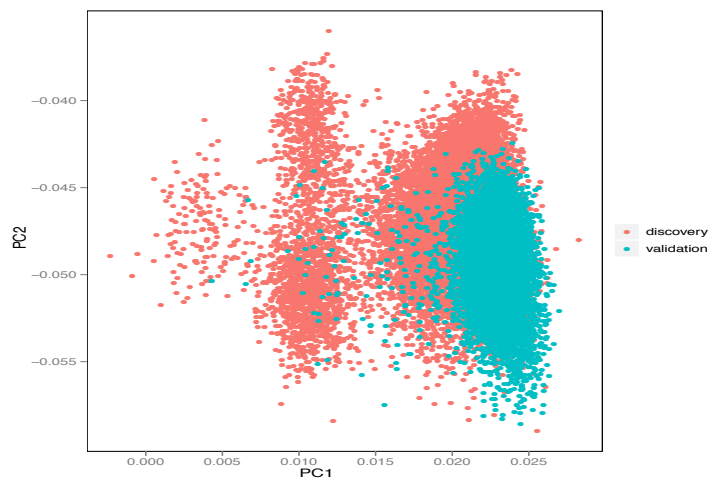


Figure S2. Prediction accuracy of MTGBLUP and STGBLUP for five psychiatric disorders in the within-study validation of PCG. Results are based on 5 replicates. Error bars are \pm empirical standard error. Prediction accuracy is measured as the mean of the correlation coefficient between the true disease status and the predicted genomic risk score in the validation data.

A. Schizophrenia



B. Bipolar disorder



C. Major depressive disorder

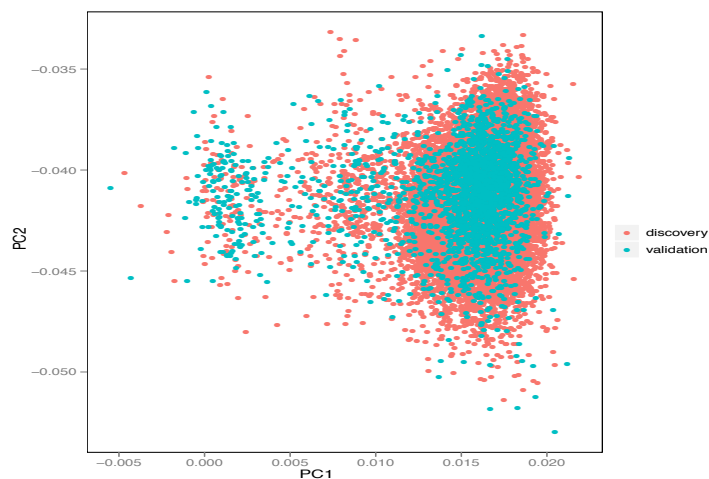
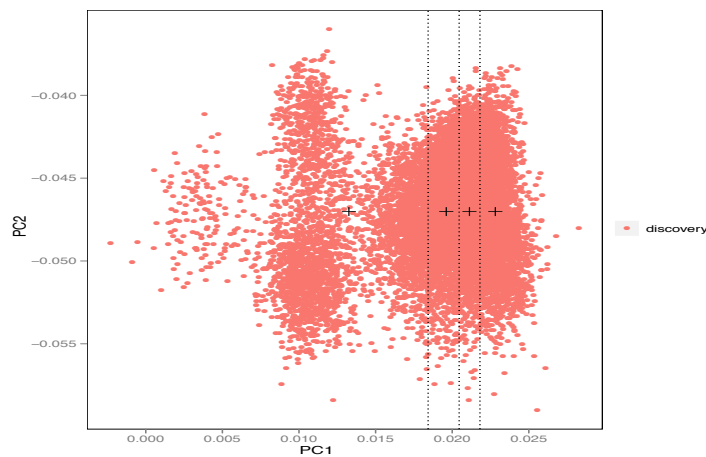
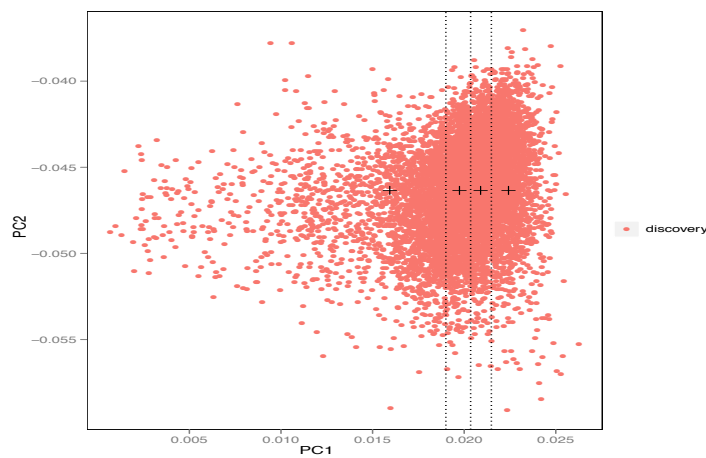


Figure S3. Principal component analysis based on the projected PC from POPRES for SCZ (A), BIP (B) and MDD (C). The same SNPs were selected from the discovery and validation set and used to project PC in each disorder. The number of SNPs used was 745,631 for SCZ, 645,237 for BIP and 673,109 for MDD.

A. Schizophrenia



B. Bipolar disorder



C. Major depressive disorder

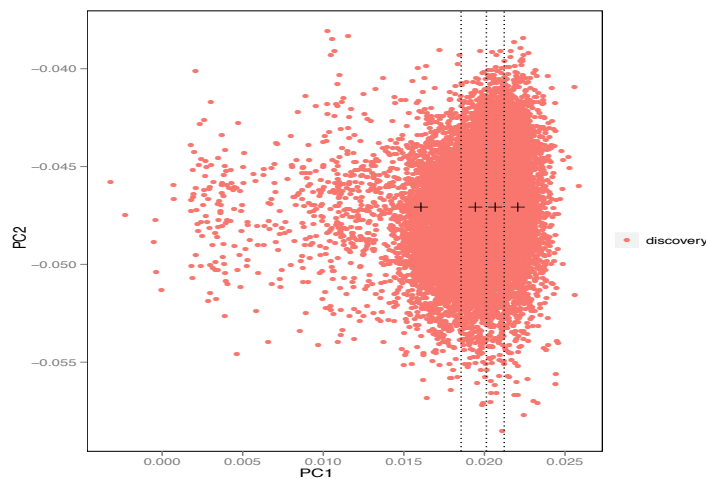
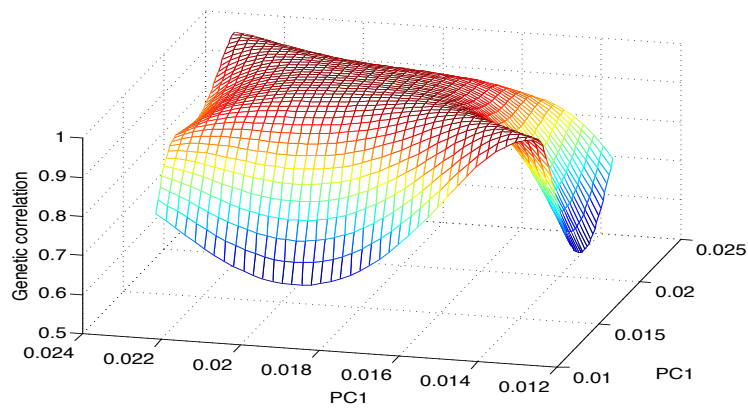
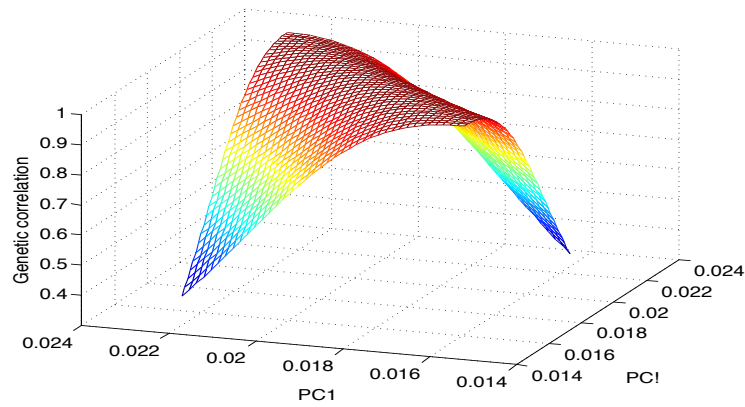


Figure S4. Principal component analysis based on the projected PC from POPRES for the discovery sample of SCZ (A), BIP (B) and MDD (C). The number of SNPs used to project PC was 745,705 for all three disorders. The dashed lines are 25%, 50% and 75% quartiles of the first principal component in the discovery sample (four population classes) and the plus sign is the mean (of PC1) of each population class. The four population classes for each trait were used in the reaction norm model (Appendix B).

A. Schizophrenia



B. Bipolar disorder



C. Major depressive disorder

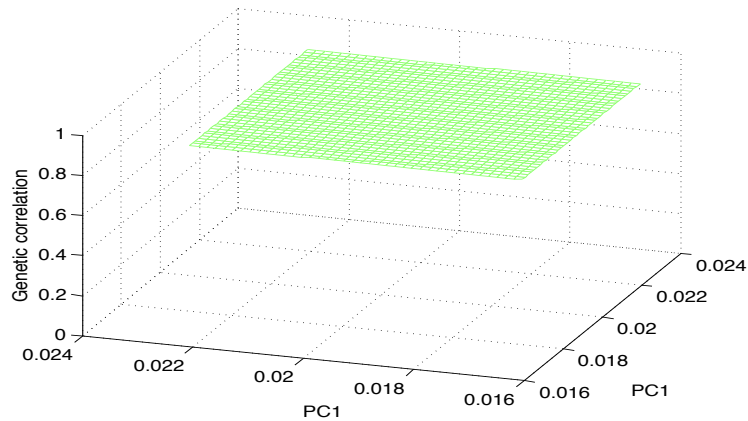
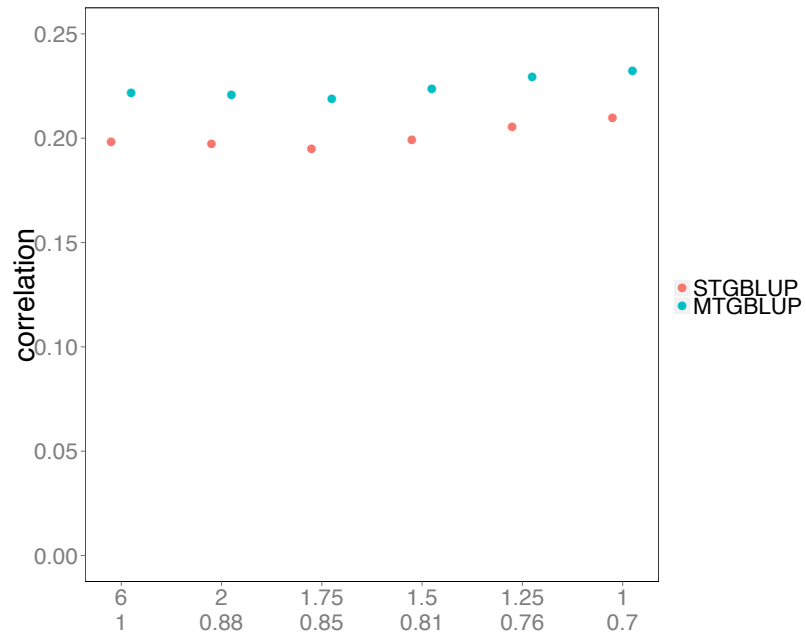


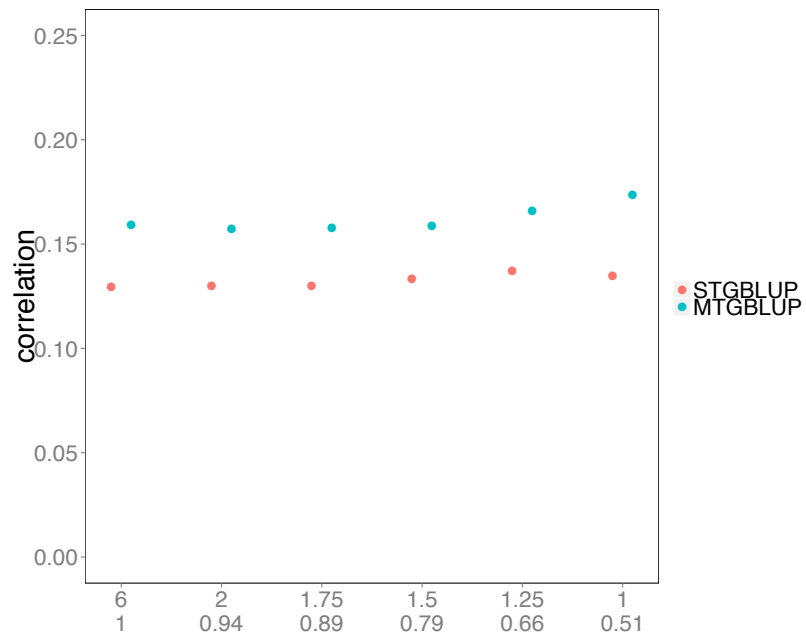
Figure S5. Genetic correlation pattern across different ancestry principal components estimated from the reaction norm model (Appendix B). Order of polynomial (see Table S9): A. $k=3$ for SCZ, B. $k=2$ for BIP, C. $k=1$ for MDD.

A. SCZ



number of standard deviations from mean / fraction included

B. BIP



number of standard deviations from mean / fraction included

C. MDD

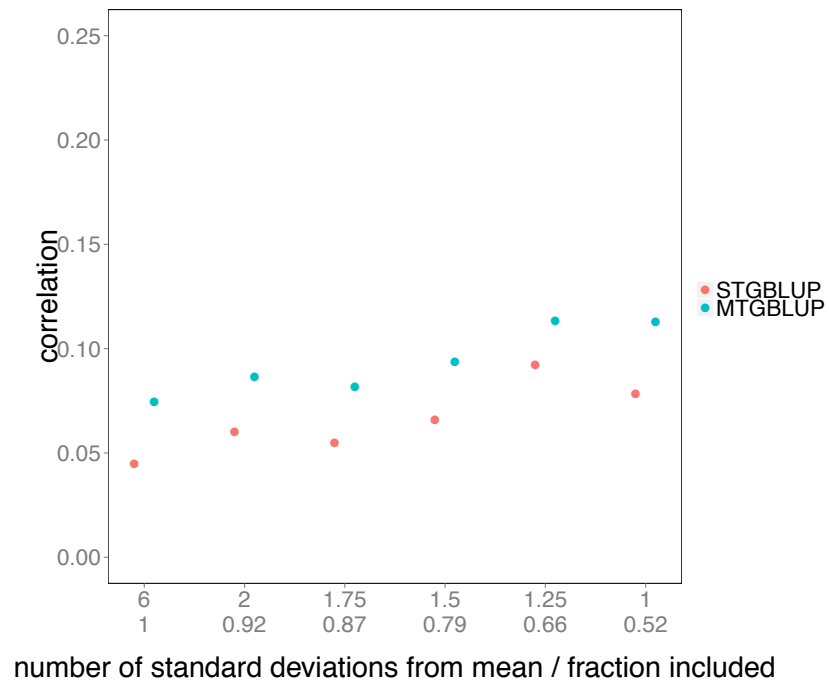


Figure S6. Effect of excluding population outliers on the prediction accuracy from MTGBLUP and STGBLUP. Outliers are defined as points ± 6 , 2, 1.75, 1.5, 1.25 and 1 SD from the mean for both the first and second principal components in the independent (A) SCZ, (B) BIP, (C) MDD samples.

Table S1. Comparison of prediction accuracy (Correlation) and regression coefficient (Regression) from MTGBLUP and STGBLUP for five psychiatric disorders in the within-study validation of PCG

	Correlation	Regression
	Schizophrenia	
STGBLUP	0.240 (0.003)	1.011 (0.022)
MTGBLUP	0.264 (0.003)	1.019 (0.019)
	Bipolar disorder	
STGBLUP	0.159 (0.005)	1.091 (0.054)
MTGBLUP	0.204 (0.005)	0.971 (0.025)
	Major depression	
STGBLUP	0.125 (0.009)	1.078 (0.054)
MTGBLUP	0.140 (0.007)	0.930 (0.038)
	Autism Spectrum Disorders	
STGBLUP	0.075 (0.003)	0.965 (0.080)
MTGBLUP	0.078 (0.006)	0.884 (0.054)
	ADHD	
STGBLUP	0.141 (0.013)	1.144 (0.052)
MTGBLUP	0.146 (0.013)	1.116 (0.050)

Correlation and regression coefficients and empirical standard errors (in brackets) are calculated based on 5 replicates.

Table S2. Prediction accuracy for schizophrenia (SCZ), bipolar disorder (BIP) and major depressive disorder (MDD) in independent validation data sets when using the second annotation model.

	Correlation			Regression		
	SCZ	BIP	MDD	SCZ	BIP	MDD
STGBLUP-SAI	0.199	0.130	0.048	0.787	0.746	0.323
MTGBLUP-SAI	0.222	0.160	0.076	0.817	0.718	0.470

MTGBLUP-SAI or STGBLUP-SAI: in the SAI model SNPs are grouped into schizophrenia / autism / intellectual disability (SAI) candidate genes sets. Prediction accuracy is given as the correlation coefficient between the true disease status and the predicted genomic risk score in the validation data.

Table S3. Comparison of the fit of standard model with the SAI-annotation model for STGBLUP, MTGBLUP and MTGBLUP.

		SCZ	BIP	MDD
x_1	x_2	p-values from LRT		
STGBLUP	STGBLUP-SAI	0.18	0.18	0.070
MTGBLUP	MTGBLUP-SAI	0.22	0.71	0.54
STGBLUP	MTGBLUP-SAI	1.2e-24	9.7E-15	0.0083

Likelihood ratio LR = -2 [$\log L(x_1) - \log L(x_1 + x_2)$]

Table S4. SNP-heritability and genetic correlation between discovery and validation set from bivariate analyses for SCZ, BIP and MDD

Trait 1/ trait 2	Cases T1/T2	Controls T1/T2	Trait 1 h^2 (SE)	Trait 2 h^2 (SE)	r_g (SE)	p-value
SCZ discovery/ SCZ validation	8826/ 4068	6106/ 5471	0.23 (0.01)	0.21 (0.02)	0.80 (0.05)	7.3E-51
BIP discovery/ BIP validation	5867/ 2029	3328/ 5338	0.21 (0.02)	0.22 (0.02)	0.75 (0.08)	7.3E-17
MDD discovery/ MDD validation	8770/ 822	6506/ 467	0.28 (0.02)	0.11 (0.25)	0.51 (0.64)	0.84

h^2 is SNP-heritability on the liability scale. r_g is genetic correlation between discovery/validation set. P-value is for testing if r_g is different from 1, indicating heterogeneity for a lower p-value.

Table S5. Reaction norm model to test heterogeneity across populations classified by the first ancestry principal component

Polynomial order k	log L	Number of parameters	LR	p-value
SCZ				
1	3830.01	5	0.00	1
2	3836.61	7	13.20	0.0014
3	3840.55	10	21.07	0.00078
4	3841.36	14	22.69	0.0069
BIP				
1	2342.89	5	0.00	1
2	2349.27	7	12.76	0.0017
3	2351.64	10	17.49	0.0037
4	2352.77	14	19.75	0.019
MDD				
1	3326.17	5	0.00	1
2	3326.39	7	0.42	0.81
3	3328.24	10	4.14	0.53
4	3330.98	14	9.61	0.38

Populations are classified by the first ancestry principal component. Schizophrenia (p-value=0.00078) and bipolar disorder (p-value=0.0017) show a significant evidence for heterogeneity across different populations.

Table S6. Additional required sample size of STGBLUP to achieve the same prediction accuracy as MTGBLUP and MTGBLUP-CNS.

	SCZ	BIP	MDD
MTGBLUP	4660 (3110 – 6270)	5550 (2830 – 8640)	10940 (730 – 24440)
MTGBLUP-CNS	5080 (3520 – 6690)	6220 (3380 – 9380)	11550 (1220 – 25300)

Values in brackets are Ninety-five percent confidence interval (CI) is in bracket.

Table S7. Comparison of prediction accuracy (correlation) and regression coefficient (regression) from bivariate GBLUP (BVGBLUP)

model	Dependent variable	correlation	regression
BVGBLUP (SCZ, BIP)	SCZ	0.220	0.822
BVGBLUP (SCZ, BIP)	BIP	0.156	0.705
BVGBLUP (SCZ, MDD)	SCZ	0.201	0.785
BVGBLUP (SCZ, MDD)	MDD	0.071	0.461
BVGBLUP (BIP, MDD)	BIP	0.133	0.682
BVGBLUP (BIP, MDD)	MDD	0.040	0.263

Table S8. Results of model comparison between bi-variate (BVGBLUP) and multi-trait GBLUP (MTGBLUP).

x1	x2	Dependent variable	p-value
BVGBLUP (SCZ, BIP)	MTGBLUP	SCZ	1.9E-03
BVGBLUP (SCZ, BIP)	MTGBLUP	BIP	7.4E-03
BVGBLUP (SCZ, MDD)	MTGBLUP	SCZ	1.9E-23
BVGBLUP (SCZ, MDD)	MTGBLUP	MDD	0.50
BVGBLUP (BIP, MDD)	MTGBLUP	BIP	2.5E-14
BVGBLUP (BIP, MDD)	MTGBLUP	MDD	0.00056

Likelihood ratio LR = -2 [$\log L(x_1) - \log L(x_1+x_2)$]

Cross disorder Working group of the Psychiatric Genomics Consortium:

Devin Absher¹, Ingrid Agartz^{2,3}, Huda Akil⁴, Farooq Amin⁵, Ole A Andreassen^{2,6}, Adebayo Anjorin⁷, Richard Anney⁸, Dan E Arking⁹, Philip Asherson¹⁰, Maria H Azevedo¹¹, Lena Backlund¹², Judith A Badner¹³, Anthony J Bailey¹⁴, Tobias Banaschewski¹⁵, Jack D Barchas¹⁶, Michael R Barnes¹⁷, Thomas B Barrett¹⁸, Nicholas Bass⁷, Agatino Battaglia¹⁹, Michael Bauer²⁰, Mònica Bayés²¹, Frank Bellivier^{22,23}, Sarah E Bergen^{24,25,26}, Wade Berrettini²⁷, Catalina Betancur^{28,29}, Thomas Bettecken³⁰, Joseph Biederman³¹, Elisabeth B Binder³⁰, Donald W Black³², Douglas H R Blackwood³³, Cinnamon S Bloss^{34,35}, Michael Boehnke^{36,37}, Dorret I Boomsma^{38,39}, Gerome Breen^{10,40}, René Breuer⁴¹, Richard Bruggeman⁴², Nancy G Buccola⁴³, Jan K Buitelaar⁴⁴, William E Bunney⁴⁵, Joseph D Buxbaum⁴⁶, William F Byerley^{47,48}, Sian Caesar⁴⁹, Wiepke Cahn⁵⁰, Rita M Cantor⁵¹, Miguel Casas^{52,53}, Aravinda Chakravarti⁹, Kimberly Chambert²⁴, Khalid Choudhury⁷, Sven Cichon^{54,55}, C Robert Cloninger⁵⁶, David A Collier¹⁰, Edwin H Cook⁵⁷, Hilary Coon⁵⁸, Bru Cormand^{59,60}, Paul Cormican⁸, Aiden Corvin⁸, William H Coryell³², Nicholas Craddock^{61,62}, David W Craig⁶³, Ian W Craig¹⁰, Jennifer Crosbie⁶⁴, Michael L Cuccaro⁶⁵, David Curtis⁶⁶, Darina Czamara^{30,67}, Mark J Daly^{24,68}, Susmita Datta⁶⁹, Geraldine Dawson^{70,71}, Richard Day⁷², Eco J De Geus^{38,39}, Franziska Degenhardt^{54,73}, Bernie Devlin⁷⁴, Srdjan Djurovic^{2,75}, Gary J Donohoe⁸, Alysa E Doyle⁷⁶, Jubao Duan⁷⁷, Frank Dudbridge⁷⁸, Eftichia Duketis⁷⁹, Richard P Ebstein⁸⁰, Howard J Edenberg^{81,82}, Josephine Elia^{27,83}, Sean Ennis⁸⁴, Bruno Etain^{22,23,85,86}, Ayman Fanous^{87,88}, Stephen V Faraone^{89,90}, Anne E Farmer¹⁰, I Nicol Ferrier⁹¹, Matthew Flickinger^{36,37}, Eric Fombonne^{92,93}, Tatiana Foroud⁸², Josef Frank⁴¹, Barbara Franke⁴⁴, Christine Fraser^{61,62}, Robert Freedman⁹⁴, Nelson B Freimer⁹⁵, Christine M Freitag⁷⁹, Marion Friedl⁹⁶, Louise Frisén¹², Louise Gallagher⁸, Pablo V Gejman⁷⁷, Lyudmila Georgieva^{61,62}, Elliot S Gershon¹³, Daniel H Geschwind^{97,98}, Ina Giegling⁹⁶, Michael Gill⁸, Scott D Gordon⁹⁹, Katherine Gordon-Smith^{49,61}, Elaine K Green¹⁰⁰, Tiffany A Greenwood¹⁰¹, Dorothy E Grice^{102,103}, Magdalena Gross¹⁰⁴, Detelina Grozeva⁶¹, Weihua Guan^{36,37,105}, Hugh Gurling⁷, Lieuwe De Haan¹⁰⁶, Jonathan L Haines¹⁰⁷, Hakon Hakonarson^{108,109}, Joachim Hallmayer¹¹⁰, Steven P Hamilton⁴⁷, Marian L Hamshere^{61,111}, Thomas F Hansen^{112,113}, Annette M Hartmann⁹⁶, Martin Hautzinger¹¹⁴, Andrew C Heath⁵⁶, Anjali K Henders⁹⁹, Stefan Herms^{54,55}, Ian B Hickie¹¹⁵, Maria Hipolito¹¹⁶, Susanne Hoefels¹⁰⁴, Peter A Holmans^{61,111}, Florian Holsboer³⁰, Witte J Hoogendijk¹¹⁷, Jouke-Jan Hottenga^{38,39}, Christina M Hultman²⁶, Vanessa Hus¹¹⁸, Andrés Ingason^{112,113}, Marcus Ising³⁰, Stéphane Jamain^{22,23,85,86}, Ian Jones^{61,62}, Lisa Jones⁴⁹, Anna K Kähler²⁶, René S Kahn⁵⁰, Radhika Kandaswamy⁷, Matthew C Keller¹¹⁹, John R Kelsoe^{101,120}, Kenneth S Kendler^{88,121,122}, James L Kennedy¹²³, Elaine Kenny⁸, Lindsey Kent¹²⁴, Yunjung Kim¹²⁵, George K Kirov^{61,62}, Sabine M Klauck¹²⁶, Lambertus Klei⁷⁴, James A Knowles¹²⁷, Martin A Kohli³⁰, Daniel L Koller⁸², Bettina Konte⁹⁶, Ania Korszun¹²⁸, Lydia Krabbendam¹²⁹, Robert Krasucki⁷, Jonna Kuntsi¹⁰, Phoenix Kwan^{36,37}, Mikael Landén^{26,130}, Niklas Långström²⁶, Mark Lathrop¹³¹, Jacob Lawrence⁷, William B Lawson¹¹⁶, Marion Leboyer^{22,23,85,86}, David H Ledbetter¹³², Phil H Lee²⁵, Todd Lencz^{133,134,135}, Klaus-Peter Lesch^{136,137}, Douglas F Levinson¹³⁸, Cathryn M Lewis¹⁰, Jun Li¹³⁹, Paul Lichtenstein²⁶, Jeffrey A Lieberman¹⁴⁰, Dan-Yu Lin¹⁴¹, Don H Linszen¹⁴², Chunyu Liu¹⁴³, Falk W Lohoff²⁷, Sandra K Loo^{95,144}, Catherine Lord¹⁴⁵, Jennifer K Lowe^{97,98}, Susanne Lucae³⁰, Donald J MacIntyre³³, Pamela AF Madden⁵⁶, Elena Maestrini¹⁴⁶, Patrik KE Magnusson²⁶, Pamela B Mahon¹⁴⁷, Wolfgang Maier¹⁰⁴, Anil K Malhotra^{133,134,135}, Shrikant M Mane¹⁴⁸, Christa

L Martin¹³², Nicholas G Martin⁹⁹, Manuel Mattheisen^{73,113,149,150}, Keith Matthews⁷², Morten Mattingsdal^{2,151}, Steven A McCarroll²⁴, Kevin A McGhee³³, James J McGough¹⁵², Patrick J McGrath¹⁴⁰, Peter McGuffin¹⁰, Melvin G McInnis¹⁵³, Andrew McIntosh^{33,154}, Rebecca McKinney¹⁰¹, Alan W McLean^{33,154}, Francis J McMahon¹⁵⁵, William M McMahon⁵⁸, Andrew McQuillin⁷, Helena Medeiros¹²⁷, Sarah E Medland⁹⁹, Sandra Meier⁴¹, Ingrid Melle^{2,6}, Fan Meng⁴, Jobst Meyer¹⁵⁶, Christel M Middeldorp^{38,39}, Lefkos Middleton¹⁵⁷, Vihra Milanova¹⁵⁸, Ana Miranda¹⁵⁹, Anthony P Monaco^{160,161}, Grant W Montgomery⁹⁹, Jennifer L Moran²⁴, Daniel Moreno-De-Luca¹⁶², Gunnar Morken^{163,164}, Derek W Morris⁸, Eric M Morrow^{165,166}, Valentina Moskvina^{61,111}, Bryan J Mowry^{167,168}, Pierandrea Muglia¹⁶⁹, Thomas W Mühleisen^{54,73,170}, Bertram Müller-Myhsok^{30,67}, Michael Murtha¹⁷¹, Richard M Myers¹, Inez Myin-Germeys¹²⁹, Benjamin M Neale^{24,68}, Stan F Nelson⁹⁵, Caroline M Nievergelt¹⁰¹, Ivan Nikolov^{61,62}, Vishwajit Nimgaonkar^{172,173}, Willem A Nolen¹⁷⁴, Markus M Nöthen^{54,73}, John I Nurnberger^{82,175}, Evaristus A Nwulia¹¹⁶, Dale R Nyholt⁹⁹, Michael C O'Donovan^{61,62}, Colm O'Dushlaine²⁴, Robert D Oades¹⁷⁶, Ann Olincy⁹⁴, Guiomar Oliveira^{11,177}, Line Olsen^{112,113}, Roel A Ophoff^{95,178,179}, Urban Osby¹², Michael J Owen^{61,62}, Aarno Palotie¹⁸⁰, Jeremy R Parr⁹¹, Andrew D Paterson^{181,182}, Carlos N Pato¹²⁷, Michele T Pato¹²⁷, Brenda W Penninx^{39,183,184}, Michele L Pergadia⁵⁶, Margaret A Pericak-Vance⁶⁵, Roy H Perlis^{24,25}, Benjamin S Pickard^{33,154}, Jonathan Pimm⁷, Joseph Piven⁷¹, Danielle Posthuma^{185,186,187}, James B Potash³², Fritz Poustka⁷⁹, Peter Propping⁷³, Shaun M Purcell^{24,68,188}, Vinay Puri⁷, Digby J Quested¹⁸⁹, Emma M Quinn⁸, Josep Antoni Ramos-Quiroga^{52,53}, Henrik B Rasmussen^{112,113}, Soumya Raychaudhuri^{24,68}, Karola Rehnström¹⁸⁰, Andreas Reif¹⁹⁰, Marta Ribasés^{52,191}, John P Rice¹⁹², Marcella Rietschel⁴¹, Stephan Ripke^{24,68}, Kathryn Roeder¹⁹³, Herbert Roeyers¹⁹⁴, Lizzy Rossin²⁴, Aribert Rothenberger¹⁹⁵, Guy Rouleau¹⁹⁶, Douglas Ruderfer¹⁸⁸, Dan Rujescu⁹⁶, Alan R Sanders⁷⁷, Stephan J Sanders^{162,171,197,198}, Susan L Santangelo^{199,200}, Russell Schachar⁶⁴, Martin Schalling¹², Alan F Schatzberg²⁰¹, William A Scheftner²⁰², Gerard D Schellenberg²⁰³, Stephen W Scherer²⁰⁴, Nicholas J Schork^{34,205}, Thomas G Schulze^{147,206}, Johannes Schumacher⁷³, Markus Schwarz²⁰⁷, Edward Scolnick²⁴, Laura J Scott^{36,37}, Joseph A Sergeant²⁰⁸, Jianxin Shi²⁰⁹, Paul D Shilling¹⁰¹, Stanley I Shyn²¹⁰, Jeremy M Silverman¹⁰³, Pamela Sklar¹⁸⁸, Susan L Slager²¹¹, Susan L Smalley^{95,144}, Johannes H Smit^{183,184}, Erin N Smith^{34,205}, Jordan W Smoller^{24,25}, Edmund JS Sonuga-Barke^{194,212}, David St Clair²¹³, Matthew State^{162,171,197}, Michael Steffens²¹⁴, Hans-Christoph Steinhausen^{215,216,217}, John S Strauss²¹⁸, Jana Strohmaier⁴¹, T Scott Stroup²¹⁹, Patrick F Sullivan¹²⁵, James Sutcliffe²²⁰, Peter Szatmari^{221,222,223}, Szabocls Szelinger⁶³, Anita Thapar^{61,62}, Srinivasa Thirumalai²²⁴, Robert C Thompson⁴, Alexandre A Todorov⁵⁶, Federica Tozzi¹⁶⁹, Jens Treutlein⁴¹, Jung-Ying Tzeng²²⁵, Manfred Uhr³⁰, Edwin JCG van den Oord²²⁶, Gerard Van Grootheest^{183,184}, Jim Van Os¹²⁹, Astrid M Vicente^{227,228,229}, Veronica J Vieland²³⁰, John B Vincent²¹⁸, Peter M Visscher^{167,231}, Christopher A Walsh^{232,233,234,235}, Thomas H Wassink³², Stanley J Watson⁴, Lauren A Weiss⁴⁷, Myrna M Weissman²³⁶, Thomas Werge^{112,113,237}, Thomas F Wienker²³⁸, Durk Wiersma⁴², Ellen M Wijsman^{239,240}, Gonneke Willemsen^{38,183}, Nigel Williams^{61,62}, A Jeremy Willsey^{171,197}, Stephanie H Witt⁴¹, Naomi R Wray¹⁶⁷, Wei Xu¹⁸², Allan H Young^{91,241}, Timothy W Yu²⁴², Stanley Zammit^{61,62}, Peter P Zandi²⁴³, Peng Zhang^{36,37,153}, Frans G Zitman²⁴⁴, Sebastian Zöllner^{36,37,153}

¹HudsonAlpha Institute of Biotechnology, Huntsville, Alabama, USA. ²KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. ³Department of Research, Diakonhjemmet Hospital, Oslo, Norway. ⁴Molecular Psychiatry Laboratory, Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, Michigan, USA. ⁵Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Emory University, Atlanta, Georgia, USA. ⁶Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. ⁷Mental Health Sciences Unit, University College London, London, UK. ⁸Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. ⁹McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ¹⁰MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, The Institute of Psychiatry, King's College London, London, UK. ¹¹Faculty of Medicine, University of Coimbra, Coimbra, Portugal. ¹²Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden. ¹³Department of Psychiatry, University of Chicago, Chicago, Illinois, USA. ¹⁴Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada. ¹⁵Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. ¹⁶Department of Psychiatry, Weill Medical College, Cornell University, New York, New York, USA. ¹⁷GlaxoSmithKline, London, UK. ¹⁸Portland Veterans Affairs Medical Center, Portland, Oregon, USA. ¹⁹Stella Maris Institute for Child and Adolescent Neuropsychiatry, Calambrone, Pisa, Italy. ²⁰Department of Psychiatry and Psychotherapy, Carl Gustav Carus University Hospital, Dresden, Germany. ²¹Centro Nacional de Análisis Genómico (CNAG), Parc Científic de Barcelona (PCB), Barcelona, Spain. ²²Institut National de la Santé et de la Recherche Médicale (INSERM) U955, Psychiatrie Génétique, Créteil, France. ²³ENBREC (European Network of Bipolar Research Expert Centres) Group, Fondation FondaMental, Créteil, France. ²⁴Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ²⁵Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. ²⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ²⁷Department of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ²⁸INSERM U952, Paris, France. ²⁹Université Pierre et Marie Curie, Paris, France. ³⁰Max Planck Institute of Psychiatry, Munich, Germany. ³¹Clinical and Research Programs in Pediatric Psychopharmacology and Adult ADHD, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. ³²Department of Psychiatry, University of Iowa, Iowa City, Iowa, USA. ³³Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. ³⁴The Scripps Translational Science Institute, La Jolla, California, USA. ³⁵Scripps Health, La Jolla, California, USA. ³⁶Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. ³⁷Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA. ³⁸Department of Biological Psychology, VU University, Amsterdam, The Netherlands. ³⁹Neuroscience Campus Amsterdam, Amsterdam, The Netherlands. ⁴⁰National Institute of Health Research (NIHR) Biomedical Research Centre for Mental Health, South London, London, UK and Maudsley National Health Service (NHS) Trust and Institute of Psychiatry, London, UK. ⁴¹Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim,

Germany.⁴²Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.⁴³School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA.

⁴⁴Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Centre, Nijmegen, The Netherlands.⁴⁵Department of Psychiatry and Human Behavior, University of California–Irvine, Irvine, California, USA. ⁴⁶Seaver Autism Center for Research and Treatment, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA.⁴⁷Department of Psychiatry, University of California, San Francisco, San Francisco, California, USA. ⁴⁸NCIRE (Northern California Institute of Q Research and Education), San Francisco, California, USA. ⁴⁹Department of Psychiatry, Birmingham University, Birmingham, UK.⁵⁰Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands.⁵¹David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. ⁵²Department of Psychiatry, Hospital Universitari Vall d’Hebron, CIBERSAM (Centro de Investigación Biomédica en el Área de Salud Mental), Barcelona, Spain. ⁵³Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain.⁵⁴Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. ⁵⁵Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland. ⁵⁶Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. ⁵⁷Department of Psychiatry, Institute for Juvenile Research, University of Illinois, Chicago, Illinois, USA.⁵⁸Department of Psychiatry, University of Utah, Salt Lake City, Utah, USA. ⁵⁹Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain. ⁶⁰Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain.⁶¹Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK.⁶²Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK. ⁶³The Translational Genomics Research Institute, Phoenix, Arizona, USA.⁶⁴Neurosciences and Mental Health Program, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.⁶⁵John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, USA. ⁶⁶East London NHS Foundation Trust, Queen Mary, University of London, London, UK. ⁶⁷Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.⁶⁸Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. ⁶⁹Genetics Institute, University College London, London, UK. ⁷⁰Autism Speaks, New York, New York, USA. ⁷¹Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.⁷²Division of Neuroscience, Medical Research Institute, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK. ⁷³Institute of Human Genetics, University of Bonn, Bonn, Germany. ⁷⁴Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA.⁷⁵Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. ⁷⁶Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. ⁷⁷Department of Psychiatry and Behavioral Sciences, NorthShore University Health System and University of Chicago, Evanston, Illinois, USA. ⁷⁸Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. ⁷⁹Department of Child and Adolescent Psychiatry, Psychosomatics and

Psychotherapy, JW Goethe University Frankfurt, Frankfurt, Germany. ⁸⁰Psychology Department, National University of Singapore, Singapore. ⁸¹Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana, USA. ⁸²Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA. ⁸³AI Dupont Hospital for Children, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ⁸⁴School of Medicine, Medical Science University College, Dublin, Ireland. ⁸⁵Université Paris Est, Faculté de Médecine, Créteil, France. ⁸⁶AP-HP, Hôpital H Mondor–A Chenevier, Département de Psychiatrie, Créteil, France. ⁸⁷Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA. ⁸⁸Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, USA. ⁸⁹Department of Psychiatry, State University of New York (SUNY) Upstate Medical University, Syracuse, New York, USA. ⁹⁰Department of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, New York, USA. ⁹¹Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK. ⁹²Department of Psychiatry, Oregon Health & Science University, Portland, Oregon, USA. ⁹³Institute for Development & Disability, Oregon Health & Science University, Portland, Oregon, USA. ⁹⁴Department of Psychiatry, University of Colorado Denver, Aurora, Colorado, USA. ⁹⁵Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, California, USA. ⁹⁶Department of Psychiatry, University of Halle, Halle, Germany. ⁹⁷Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. ⁹⁸Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA. ⁹⁹Queensland Institute of Medical Research, Brisbane, Queensland, Australia. ¹⁰⁰Department of Biomedical and Biological Sciences, Plymouth University, Plymouth, UK. ¹⁰¹Department of Psychiatry, University of California, San Diego, La Jolla, California, USA. ¹⁰²Division of Tics, OCD and Related Disorders, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ¹⁰³Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ¹⁰⁴Department of Psychiatry, University of Bonn, Bonn, Germany. ¹⁰⁵Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota, USA. ¹⁰⁶Department of Psychiatry, Academic Medical Centre, University of Amsterdam The Netherlands. ¹⁰⁷Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA. ¹⁰⁸The Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ¹⁰⁹Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA. ¹¹⁰Department of Psychiatry, School of Medicine, Stanford University, Stanford, California, USA. ¹¹¹Biostatistics and Bioinformatics Unit, Cardiff University, Cardiff, UK. ¹¹²Institute of Biological Psychiatry, Copenhagen University Hospital, Roskilde, Denmark. ¹¹³The Lundbeck Initiative for Integrative Psychiatric Research, iPSYCH, Roskilde, Denmark. ¹¹⁴Department of Clinical and Developmental Psychology, Eberhard Karls University of Tübingen, Tübingen, Germany. ¹¹⁵Brain and Mind Research Institute, University of Sydney, Sydney, New South Wales, Australia. ¹¹⁶Department of Psychiatry and Behavioral Sciences, Howard University College of Medicine, Washington, DC, USA. ¹¹⁷Department of Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands. ¹¹⁸Department of Psychology, University of Michigan, Ann Arbor, Michigan, USA. ¹¹⁹Department of Psychology, University of Colorado, Boulder,

Colorado, USA.¹²⁰Department of Psychiatry, Special Treatment and Evaluation Program (STEP), Veterans Affairs San Diego Healthcare System, San Diego, California, USA.¹²¹Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia, USA.¹²²Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, USA.¹²³Psychiatric Neurogenetics Section, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.¹²⁴School of Medicine, University of St Andrews, St Andrews, UK.¹²⁵Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.¹²⁶Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany.¹²⁷Department of Psychiatry, Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.¹²⁸Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK.¹²⁹Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, Maastricht, The Netherlands.¹³⁰Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden.¹³¹Centre National de Genotypage, Evry, France.¹³²Geisinger Health System, Autism and Developmental Medicine Institute, Danville, Pennsylvania, USA.¹³³Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore, Long Island Jewish Health System, Glen Oaks, New York, USA.¹³⁴Center for Psychiatric Neuroscience, The Feinstein Institute of Medical Research, Manhasset, New York, USA.¹³⁵Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York, USA.¹³⁶Division of Molecular Psychiatry, ADHD Clinical Research Unit, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Würzburg, Germany.¹³⁷Department of Psychiatry and Psychology, School for Mental Health and Neuroscience (MHENS), Maastricht University, Maastricht, The Netherlands.¹³⁸Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA.¹³⁹Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA.¹⁴⁰New York State Psychiatric Institute, Columbia University, New York, New York, USA.¹⁴¹Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.¹⁴²Department of Psychiatry, Academic Medical Centre University of Amsterdam, Amsterdam, The Netherlands.¹⁴³Department of Psychiatry, Institute of Human Genetics, University of Illinois at Chicago, Chicago, Illinois, USA.¹⁴⁴Department of Psychiatry and Biobehavioral Science, University of California, Los Angeles, Los Angeles, California, USA.¹⁴⁵Center for Autism and the Developing Brain, Weill Cornell Medical College, White Plains, New York, USA.¹⁴⁶Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy.¹⁴⁷Department of Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, Maryland, USA.¹⁴⁸Yale Center for Genome Analysis, Orange, Connecticut, USA.¹⁴⁹Department of Biomedicine, Aarhus University, Aarhus, Denmark.¹⁵⁰Department of Genomic Mathematics, University of Bonn, Bonn, Germany.¹⁵¹Sørlandet Hospital, Kristiansand, Norway.¹⁵²Child and Adolescent Psychiatry, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, USA.¹⁵³Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, USA.¹⁵⁴Molecular Medicine Centre, University of Edinburgh, Edinburgh, UK.¹⁵⁵National Institute of Mental Health, US National Institutes of Health, Bethesda, Maryland, USA.¹⁵⁶Department of Neurobehavioral Genetics, Trier University, Trier, Germany.¹⁵⁷Neuroepidemiology and Ageing Research, School of Public Health,

Imperial College London, London, UK. ¹⁵⁸Department of Psychiatry, First Psychiatric Clinic, Alexander University Hospital, Sofia, Bulgaria. ¹⁵⁹Department of Developmental and Educational Psychology, University of Valencia, Valencia, Spain. ¹⁶⁰Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. ¹⁶¹Office of the President, Tufts University, Medford, Massachusetts, USA. ¹⁶²Department of Psychiatry, Yale University, New Haven, Connecticut, USA. ¹⁶³Department of Psychiatry, St. Olavs Hospital, Trondheim, Norway. ¹⁶⁴Department of Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway. ¹⁶⁵Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, Rhode Island, USA. ¹⁶⁶Department of Psychiatry and Human Behavior, Brown University, Providence, Rhode Island, USA. ¹⁶⁷The University of Queensland, Queensland Brain Institute, Brisbane, Queensland, Australia. ¹⁶⁸Queensland Centre for Mental Health Research, Wacol, Queensland, Australia. ¹⁶⁹Neurosciences Centre of Excellence in Drug Discovery, GlaxoSmithKline Research and Development, Verona, Italy. ¹⁷⁰Life & Brain Center, University of Bonn, Bonn, Germany. ¹⁷¹Child Study Center, Yale University, New Haven, Connecticut, USA. ¹⁷²Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ¹⁷³Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. ¹⁷⁴Department of Psychiatry, Groningen University Medical Center, Groningen, The Netherlands. ¹⁷⁵Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana, USA. ¹⁷⁶Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Duisburg-Essen, Essen, Germany. ¹⁷⁷Research and Clinical Training Department, Pediatric Hospital, Centro Hospitalar e Universitário Coimbra, Coimbra, Portugal. ¹⁷⁸Department of Human Genetics, University of California, Los Angeles, Los Angeles, California, USA. ¹⁷⁹Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands. ¹⁸⁰Sanger Institute, Hinxton, Cambridge, UK. ¹⁸¹Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. ¹⁸²Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ¹⁸³EMGO+ (ExtraMuraalGeneeskundig Onderzoek) Institute for Health and Care Research, Amsterdam, The Netherlands. ¹⁸⁴Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands. ¹⁸⁵Department of Functional Genomics, VU University, Amsterdam, The Netherlands. ¹⁸⁶Department of Clinical Genetics, VU Medical Center, Amsterdam, The Netherlands. ¹⁸⁷Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands. ¹⁸⁸Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ¹⁸⁹Academic Department of Psychiatry, University of Oxford, Oxford, UK. ¹⁹⁰Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Goethe-University Frankfurt, Frankfurt am Main, Germany. ¹⁹¹Psychiatric Genetics Unit, Vall d'Hebron Research Institute, Barcelona, Spain. ¹⁹²Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri, USA. ¹⁹³Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. ¹⁹⁴Department of Experimental Clinical & Health Psychology, Ghent University, Ghent, Belgium. ¹⁹⁵Child and Adolescent Psychiatry, University Medicine Göttingen, Göttingen, Germany. ¹⁹⁶Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada. ¹⁹⁷Department of Genetics, Yale University, New Haven, Connecticut, USA. ¹⁹⁸Program on Neurogenetics, Yale University, New Haven, Connecticut, USA. ¹⁹⁹Department of Psychiatry, Maine Medical Center, Portland, Maine,

USA.²⁰⁰ Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA.²⁰¹ Department of Psychiatry and Behavioral Science, Stanford University School of Medicine, Palo Alto, California, USA.²⁰² Rush Ambulatory Behavioral Health, Rush University Medical Center, Chicago, Illinois, USA.²⁰³ Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.²⁰⁴ The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada.²⁰⁵ The Scripps Research Institute, La Jolla, California, USA.²⁰⁶ Department of Psychiatry & Psychotherapy, University of Göttingen, Göttingen, Germany.²⁰⁷ Psychiatric Center Nordbaden, Wiesloch, Germany.²⁰⁸ Department of Clinical Neuropsychology, VU University, Amsterdam, The Netherlands.²⁰⁹ Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA.²¹⁰ Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, USA.²¹¹ Mayo Clinic, Rochester, Minnesota, USA.²¹² Developmental Brain & Behaviour Laboratory, Academic Unit of Psychology, University of Southampton, Southampton, UK.²¹³ Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, UK.²¹⁴ Research Department, Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany.²¹⁵ Research Unit of Child and Adolescent Psychiatry, Aalborg University Hospital, Aalborg, Denmark.²¹⁶ Clinical Psychology and Epidemiology, University of Basel, Basel, Switzerland.²¹⁷ Department of Child and Adolescent Psychiatry, University of Zurich, Zurich, Switzerland.²¹⁸ Molecular Neuropsychiatry and Development Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.²¹⁹ Department of Psychiatry, Columbia University, New York, New York, USA.²²⁰ Vanderbilt Brain Institute, Vanderbilt University, Nashville, Tennessee, USA.²²¹ Department of Psychiatry, University of Toronto, Toronto, Ontario Canada.²²² Neurosciences and Mental Health Program, Hospital for Sick Children, Toronto, Ontario Canada.²²³ Centre for Addiction and Mental Health, Toronto, Ontario, Canada.²²⁴ Oxford Health NHS Foundation Trust, Marlborough House Secure Unit, Milton Keynes, UK.²²⁵ Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA.²²⁶ Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth University, Richmond, Virginia, USA.²²⁷ Instituto Nacional de Saude Dr Ricardo Jorge, Lisbon, Portugal.²²⁸ BioFIG—Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, Campo Grande, Lisbon, Portugal.²²⁹ Instituto Gulbenkian de Ciência, Lisbon, Portugal.²³⁰ Battelle Center for Mathematical Medicine, Nationwide Children's Hospital, Columbus, Ohio, USA.²³¹ The University of Queensland, Diamantina Institute, Brisbane, Queensland, Australia.²³² Howard Hughes Medical Institute, Children's Hospital Boston, Boston, Massachusetts, USA.²³³ Division of Genetics, Children's Hospital Boston, Boston, Massachusetts, USA.²³⁴ Department of Neurology, Harvard Medical School Center for Life Sciences, Boston, Massachusetts, USA.²³⁵ Department of Pediatrics, Harvard Medical School Center for Life Sciences, Boston, Massachusetts, USA.²³⁶ Columbia University College of Physicians and Surgeons, New York, New York, USA.²³⁷ Faculty of Health and Medical Science, University of Copenhagen, Copenhagen, Denmark.²³⁸ Institute of Medical Biometry, University of Bonn, Bonn, Germany.²³⁹ Department of Biostatistics, University of Washington, Seattle, Washington, USA.²⁴⁰ Department of Medicine, University of Washington, Seattle, Washington, USA.²⁴¹ Centre for Affective Disorders, Institute of Psychiatry, King's College London, London, UK.²⁴² Division of Genetics, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts, USA.²⁴³ Department of Mental Health, Johns Hopkins University, Baltimore,

Maryland, USA.²⁴⁴Department of Psychiatry, Leiden University Medical Center,
Leiden, The Netherlands.