

Supplemental information

**Multivariate GWAS of psychiatric disorders
and their cardinal symptoms reveal two
dimensions of cross-cutting genetic liabilities**

Travis T. Mallard, Richard Karlsson Linnér, Andrew D. Grotzinger, Sandra Sanchez-Roige, Jakob Seidlitz, Aysu Okbay, Ronald de Vlaming, S. Fleur W. Meddens, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Abraham A. Palmer, Lea K. Davis, Elliot M. Tucker-Drob, Kenneth S. Kendler, Matthew C. Keller, Philipp D. Koellinger, and K. Paige Harden

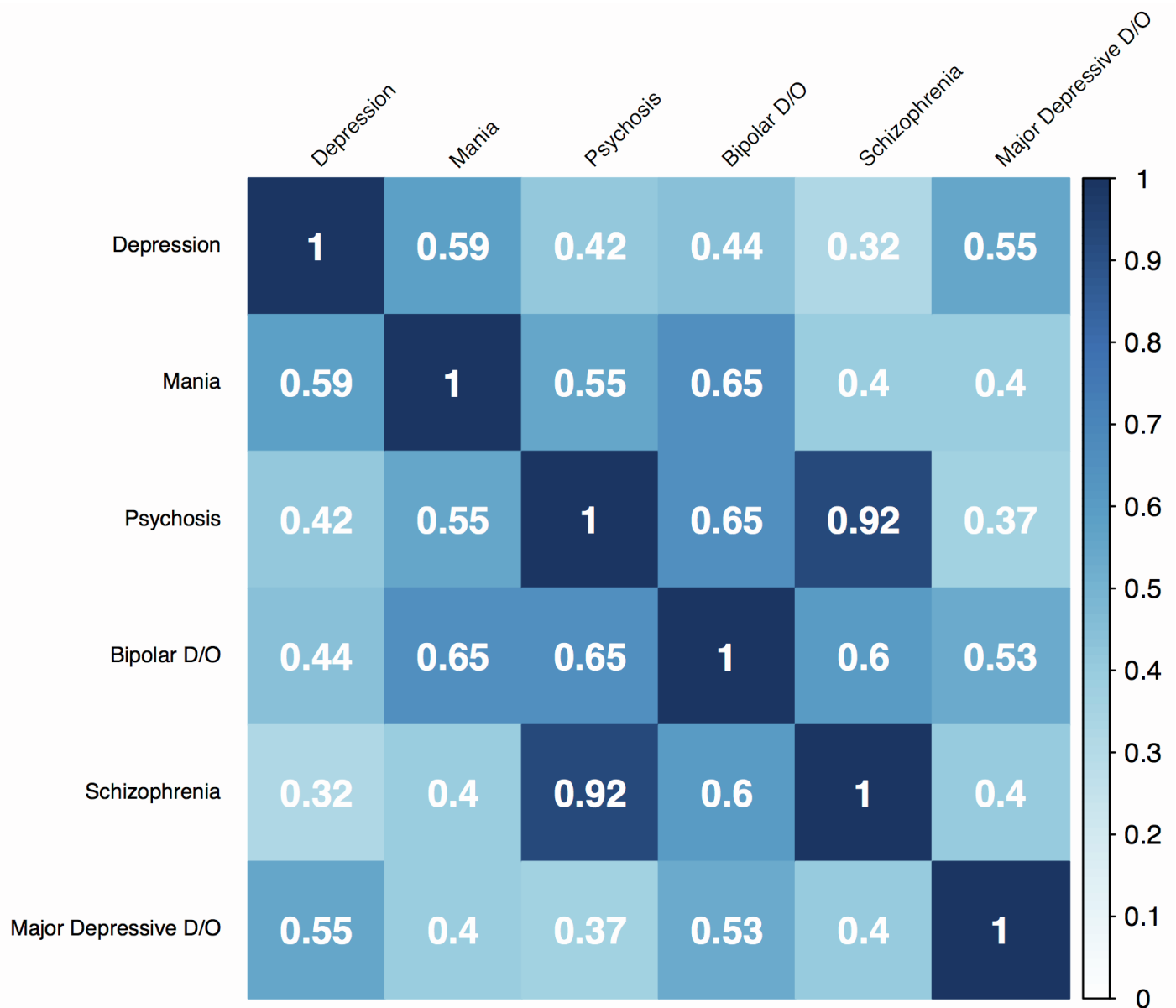


Figure S1. Phenotypic correlations between psychiatric symptoms and disorders, related to Figure 2. Matrix of phenotypic correlations for available psychiatric symptoms and disorders in the UK Biobank discovery sample.

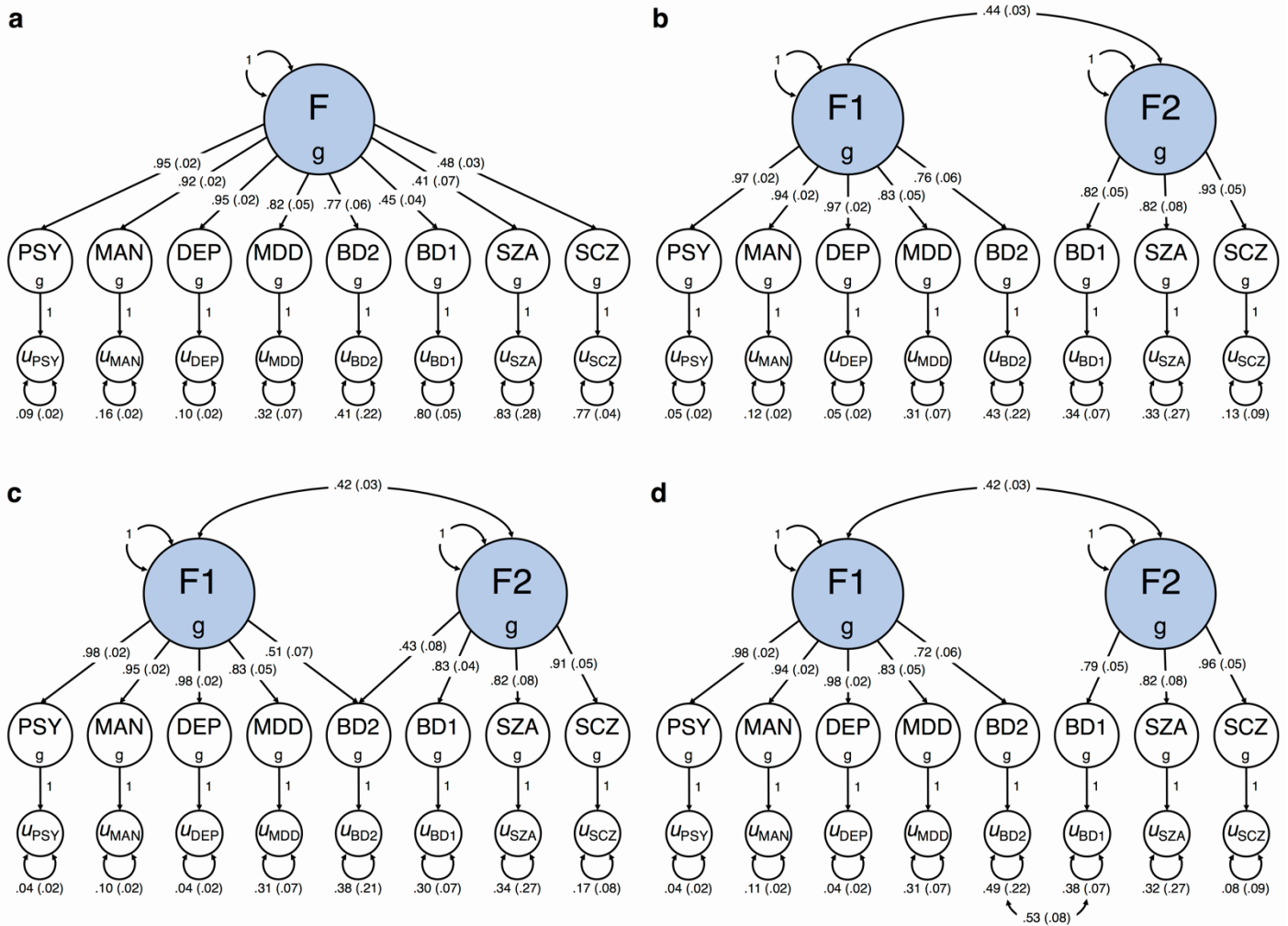


Figure S2. Path diagrams for the four confirmatory factor models, related to Figure 2. a,b,c,d, Path diagrams with standardized parameter estimates for (a) a common factor model, (b) a correlated factors model, (c) a correlated factors model with bipolar II disorder cross-loading on both factors, and (d) a correlated factors model with correlated residuals between bipolar I disorder and bipolar II disorder. Standard errors are reported in parentheses next to each parameter estimate. Latent genetic factors are highlighted in blue.

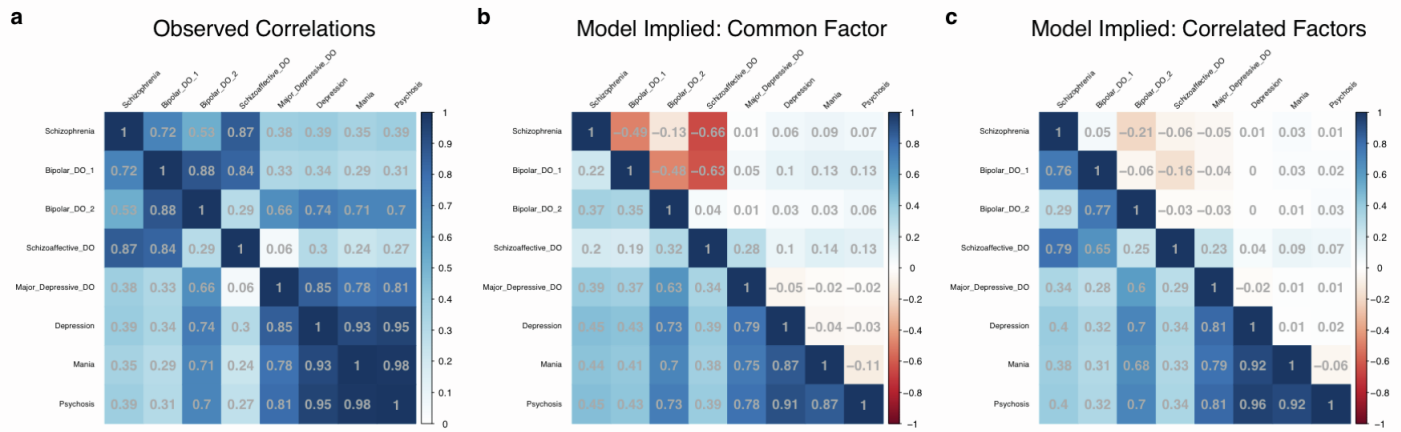


Figure S3. Observed and implied genetic correlation matrices, related to Figure 2. **a**, Matrix of observed genetic correlations for the eight psychiatric symptoms and disorders. **b,c**, Matrices of model implied genetic correlations for **(a)** the common factor (*i.e.*, the p factor) and **(b)** the final correlated factors model (Figure S2d). Model implied correlations are presented below the diagonal while the difference between the model implied and observed correlations (model implied r_g – observed r_g) are presented above the diagonal. Positive and negative values above the diagonal, therefore, reflect upwardly and downwardly biased estimates, respectively.

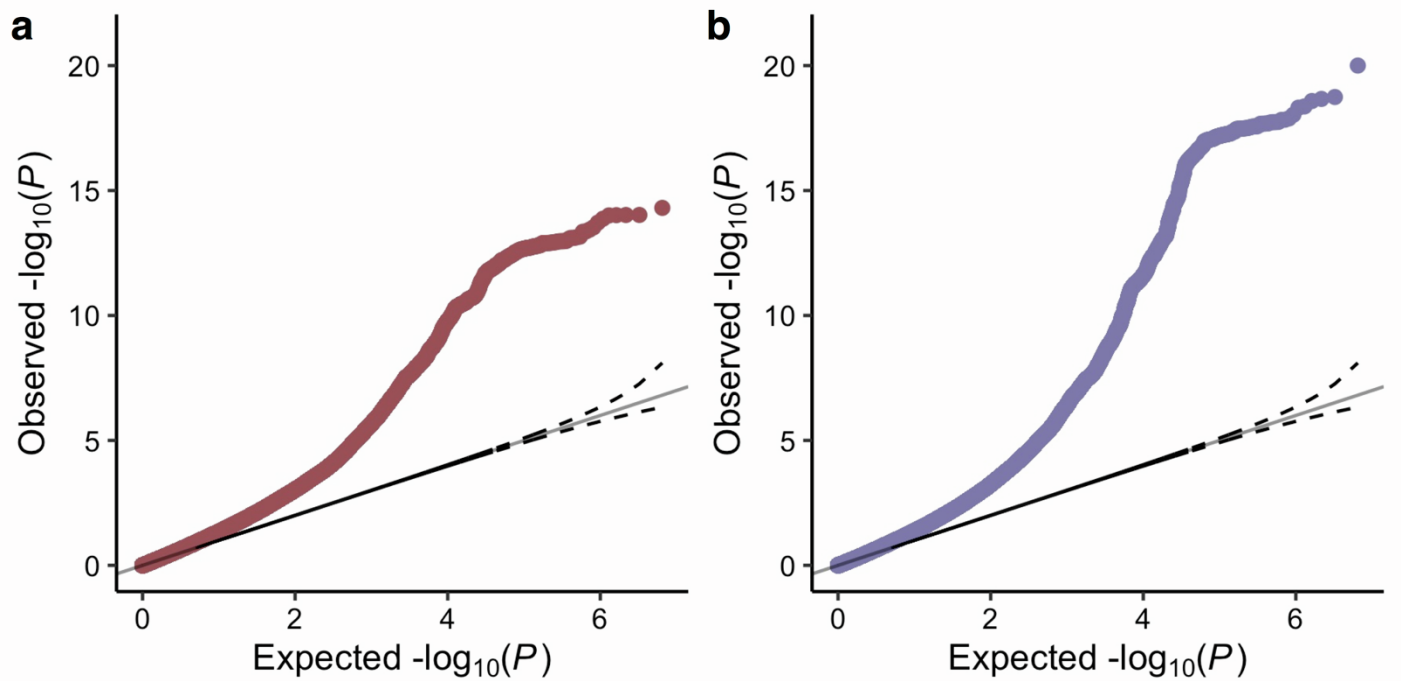


Figure S4. Quantile-quantile plot for the latent genetic factors, related to Figure 3. a,b, Quantile-quantile plot for (a) F1 and (b) F2, illustrating strong polygenic signal for both multivariate GWAS. The y-axis corresponds to the observed distribution of P , while the x-axis corresponds to the expected distribution of P under the null. The null is plotted as a solid gray line, and the accompanying 95% confidence interval is plotted as a dotted black line.

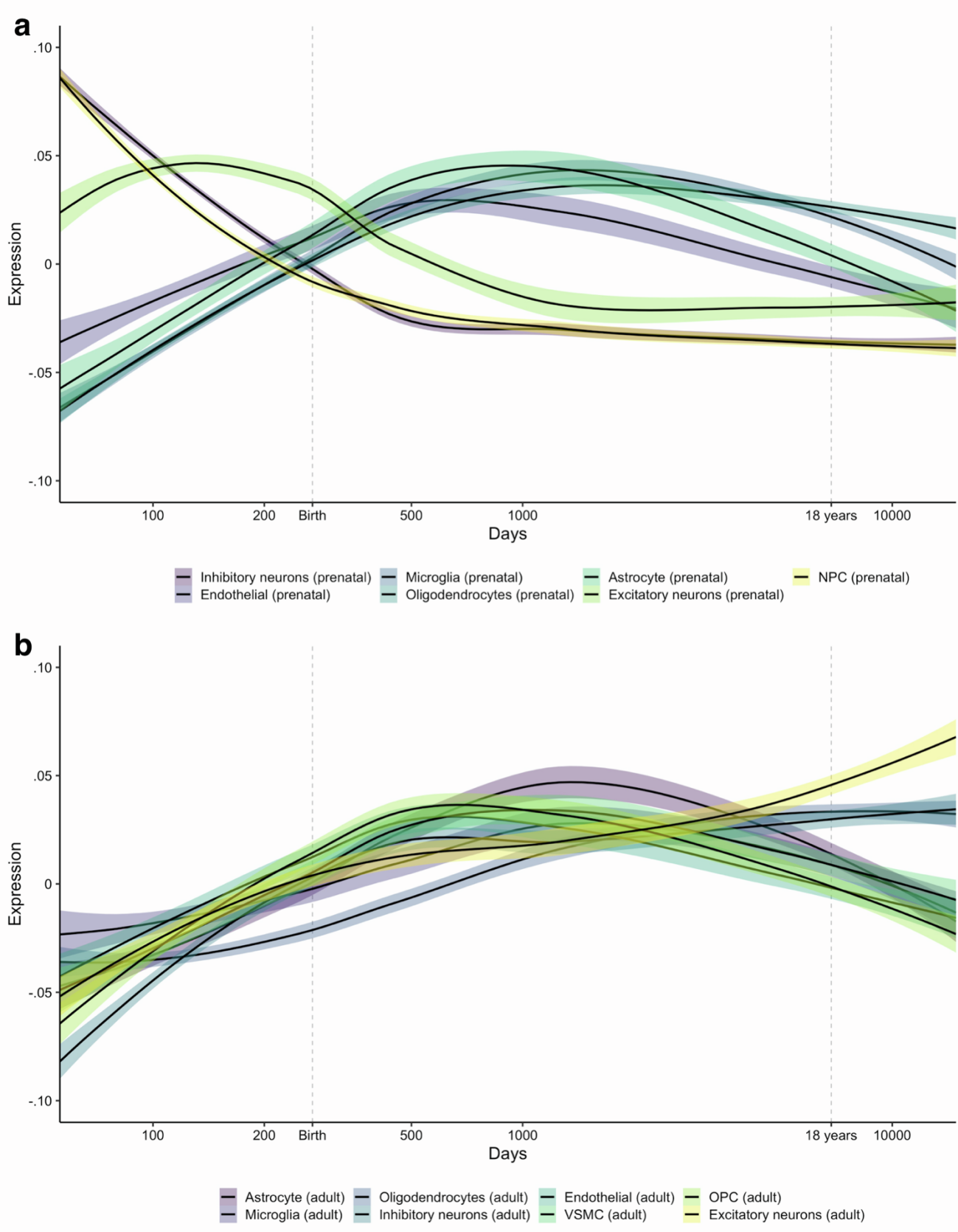


Figure S5. Spatiotemporal gene expression of specific cell types, related to Figure 4. a,b, Developmental trajectories fit via LOESS regression for (a) prenatal and (b) adult cell types in the Brainspan dataset. Cell type data is derived from Li and colleagues¹⁰. Note: inhibitory neurons = mostly GABAergic interneurons, excitatory neurons = mostly glutamatergic excitatory projection neurons, NPC = neural progenitor cells, VSMC = vascular smooth muscle cells, OPC = oligodendrocyte progenitor cells.

Methods S1, Supplemental description of methods employed in this study, related to STAR Methods.

Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities

Travis T. Mallard^{a,1,2}, Richard Karlsson Linnér^{3,4}, Andrew D. Grotzinger¹, Sandra Sanchez-Roige^{5,6}, Jakob Seidlitz^{7,8}, Aysu Okbay³, Ronald de Vlaming³, S. Fleur W. Meddens⁹, Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Abraham A. Palmer^{5,10}, Lea K. Davis^{6,11,12}, Elliot M. Tucker-Drob^{1,13}, Kenneth S. Kendler^{14,15}, Matthew C. Keller^{16,17}, Philipp D. Koellinger^{b,3}, & K. Paige Harden^{b,1,13}

^aLead Contact

^bSenior Author

Affiliations:

1. Department of Psychology, University of Texas at Austin, Austin, TX, USA
2. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
3. Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
4. Autism and Developmental Medicine Institute, Geisinger, Lewisburg, PA, USA
5. Department of Psychiatry, University of California San Diego, La Jolla, CA, USA
6. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
7. Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, Philadelphia, PA, USA
8. Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA
9. Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands
10. Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA
11. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
12. Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
13. Population Research Center, University of Texas at Austin, Austin, TX USA
14. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA
15. Department of Psychiatry, Medical College of Virginia/Virginia Commonwealth University, Richmond, VA, USA
16. Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA
17. Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA

Correspondence to:

Travis T. Mallard, Ph.D., Center for Genomic Medicine, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA. E-mail: tmallard@mgh.harvard.edu

K. Paige Harden, Ph.D., Department of Psychology, The University of Texas at Austin, 108 E. Dean Keeton Street, Stop A8000, Austin, TX 78712, USA. E-mail: harden@utexas.edu

Table of Contents

1	Genome-wide association analyses in UK Biobank	3
1.1	Phenotype construction	3
1.1.1	Depression	3
1.1.2	Mania	3
1.1.3	Psychosis.....	4
1.1.4	Item response theory models	4
1.1.5	Phenotypic factor structure in UK Biobank.....	5
1.2	Principal components of ancestry.....	6
1.3	Quality control.....	6
2	Genomic structural equation modeling	7
2.1	Hierarchical clustering	7
2.2	Factor analysis.....	7
2.3	Structural equation modeling	8
2.3.1	Confirmatory factor analysis	9
2.3.2	Genetic correlation.....	10
2.3.3	Multivariable genetic regression.....	10
2.4	Multivariate genome-wide association analyses.....	11
2.5	Heterogeneity tests	11
2.6	Effective sample size	11
3	References.....	14

1 Genome-wide association analyses in UK Biobank

To investigate the genetic architecture of psychiatric symptoms related to mood disturbance and psychosis, we used a novel combination of Bayesian item response theory and linear mixed models to conduct univariate GWASs in the UK Biobank ($N = 252,252$). Details regarding the phenotypes and extensive quality control procedures are described below.

1.1 Phenotype construction

Lifetime symptoms of depression and mania were assessed with two sets of items administered via in-person (Wave 1) and online (Wave 2) surveys. Lifetime symptoms of psychosis were only assessed during the online follow-up survey. Although items were very similar across assessment occasions, there were slight differences in wording and response options, as described below.

1.1.1 Depression

Wave 1. The in-person surveys indexed depressive symptoms with two screener items that assessed the presence of notable, prolonged feelings of sadness or apathy.

1. "Looking back over your life, have you ever had a time when you were feeling depressed or down for at least a whole week?"
 - Data-Field 4598 (Data-Coding 100349).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".
2. "Have you ever had a time when you were uninterested in things or unable to enjoy the things you used to for at least a whole week?"
 - Data-Field 4631 (Data-Coding 100349).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".

Wave 2. The web-based follow-up survey also indexed depressive symptoms with two screener items that assessed the presence of notable, prolonged feelings of sadness or apathy

1. "Have you ever had a time in your life when you felt sad, blue, or depressed for two weeks or more in a row?"
 - Data-Field 4598 (Data-Coding 503).
 - Possible responses: "Yes", "No", and "Prefer not to answer".
2. "Have you ever had a time in your life lasting two weeks or more when you lost interest in most things like hobbies, work, or activities that usually give you pleasure?"
 - Data-Field 4631 (Data-Coding 503).
 - Possible responses: "Yes", "No", and "Prefer not to answer".

1.1.2 Mania

Wave 1. The in-person surveys indexed manic symptoms with two screener items that assessed the presence of notable, prolonged feelings of (hypo)mania or irritability.

1. "Have you ever had a period of time lasting at least two days when you were feeling so good, "high", excited or "hyper" that other people thought you were not your normal self or you were so "hyper" that you got into trouble?"
 - Data-Field 4642 (Data-Coding 100349).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".

2. "Have you ever had a period of time lasting at least two days when you were so irritable that you found yourself shouting at people or starting fights or arguments?"
 - Data-Field 4653 (Data-Coding 100349).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".

Wave 2. The web-based follow-up survey indexed manic symptoms with two screener items that assessed the presence of notable, prolonged feelings of (hypo)mania or irritability.

1. "Have you ever had a period of time when you were feeling so good, "high", excited or "hyper" that other people thought you were not your normal self or you were so "hyper" that you got into trouble?"
 - Data-Field 20501 (Data-Coding 502).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".
2. "Have you ever had a period of time when you were so irritable that you found yourself shouting at people or starting fights or arguments?"
 - Data-Field 20502 (Data-Coding 502).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".

1.1.3 *Psychosis*

Wave 2. The web-based follow-up survey assessed psychotic symptoms with four screener items. These items indexed various types of unusual beliefs and experiences that are indicative of psychosis or psychotic-like experiences.

1. "Did you ever see something that wasn't really there that other people could not see? Please do not include any times when you were dreaming or half-asleep or under the influence of alcohol or drugs."
 - Data-Field 20471 (Data-Coding 502).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".
2. "Did you ever hear things that other people said did not exist, like strange voices coming from inside your head talking to you or about you, or voices coming out of the air when there was no one around? Please do not include any times when you were dreaming or half-asleep or under the influence of alcohol or drugs."
 - Data-Field 20463 (Data-Coding 502).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".
3. "Did you ever believe that a strange force was trying to communicate directly with you by sending special signs or signals that you could understand but that no one else could understand (for example through the radio or television)? Please do not include any times when you were dreaming or half-asleep or under the influence of alcohol or drugs."
 - Data-Field 20474 (Data-Coding 502).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".
4. "Did you ever believe that there was an unjust plot going on to harm you or to have people follow you, and which your family and friends did not believe existed? Please do not include any times when you were dreaming or half-asleep or under the influence of alcohol or drugs."
 - Data-Field 20468 (Data-Coding 502).
 - Possible responses: "Yes", "No", "Do not know", and "Prefer not to answer".

1.1.4 *Item response theory models*

To construct psychiatric symptom phenotypes in UK Biobank, we used *Mplus*¹ v8 to estimate a two-parameter probit multidimensional item response theory (IRT) model, which simultaneously combined self-report items across waves for each symptom dimension while leveraging all available information and accounting for the

correlations between dimensions. IRT scaling was used as it does not assume that all items are equivalently related to the underlying construct of interest, accommodates differences in the base rate of item endorsement, and does not assume that missing data are missing completely at random. The advantages of IRT scaling using full-information estimation over averaging proportion scores for incomplete longitudinal data is described in more detail in Curran and colleagues². The two parameter probit multidimensional IRT model can be expressed as:

$$P(y_{vij} = 1 | \theta_{vi}, \alpha_{vj}, \gamma_{vj}) = \Phi(\alpha_{vj}\theta_{vi} - \gamma_{vj}) = \int_{-\infty}^{\alpha_{vj}\theta_{vi} - \gamma_{vj}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

where P represents probability; y_{vij} is the observed response for binary item j for individual i in symptom dimension v ; and θ_{vi} , α_{vj} , and γ_{vj} are scalar parameters that represent the latent construct hypothesized to underlie the observed item response patterns, item discrimination (*i.e.*, the degree to which the item discriminates between individuals in different regions on the latent continuum), and item difficulty (*i.e.*, the location where the item provides maximum information) for symptom dimension v . We freely estimated slope parameters for all items, and we modeled items as functions of single time-invariant, symptom-specific θ parameters.

We estimated IRT model parameters (α_{vj} , γ_{vj} , and θ_{vi}) using a Bayesian framework with non-informative priors and multiple imputation, which maximally leveraged all available responses for participants and minimized the impact of missing data. This approach is a full information approach that is asymptotically equivalent to maximum likelihood estimation. We imputed 100 plausible θ values for each participant to create a sampling distribution of θ for each symptom. The median values of these distributions were then used as the phenotype for the GWAS.

1.1.5 Phenotypic factor structure in UK Biobank

To evaluate the phenotypic factor structure of psychiatric symptoms and disorders characterized by mood disturbance and psychosis, we first created diagnostic phenotypes for the 252,252 individuals in the discovery sample, corresponding to diagnoses of major depressive disorder, bipolar disorder, schizoaffective disorder, and schizophrenia. These phenotypes were derived from electronic health records.

1. Diagnoses - main ICD-10
 - Data-Field 41202 (Data-Coding 19)
2. Diagnoses - secondary ICD-10
 - Data-Field 41204 (Data-Coding 19)

Due to low base rates, schizoaffective disorder and schizophrenia cases were combined. These variables were then included as observed variables in the multidimensional IRT model described in Supplementary Section 1.1.4. We first estimated the zero-order correlations between each pairwise combination of the psychiatric symptoms and disorders (Figure S1). To maintain consistency across phenotypic and genetic factor analyses, we then conducted an exploratory factor analysis of the covariance matrix for these phenotypes, where we tested one-, two-, and three-factor solutions using the *factanal* function of R. As described in Supplementary Section 2.2, we retained the highest dimensional solution where each factor explained at least 20% of the variance.

Here, the three-factor solution satisfied this criterion (F1 = 35%, F2 = 29%, F3 = 23%), yielding three factors that were each predominantly characterized by a psychiatric disorder and their cardinal symptom(s). Specifically, we found that the factors corresponded to the following phenotypes (with loadings $\geq .30$):

- F1: schizophrenia (1.08), psychosis (.90), and bipolar disorder (.31)
- F2: bipolar disorder (.54), mania (1.10), and depression (.51)
- F3: major depressive disorder (1.10) and depression (.31)

Collectively, these results suggest that symptoms and disorders tend to correlate most strongly in a canonical fashion at the phenotypic level (*i.e.*, psychosis with schizophrenia, mania with bipolar disorder, depression with major depressive disorder).

1.2 Principal components of ancestry

Although the present analyses were limited to participants who self-reported non-Hispanic European descent, we sought to further account for cryptic relatedness and population stratification specific to this population. To this end, we used flashPCA2³ to extract the first 40 principal components of ancestry in individuals who reported having non-Hispanic European ancestry.

We used an approach that broadly paralleled the original process used to estimate principal components of ancestry in the entire UKB dataset, as well as the default flashPCA2 recommendations as outlined by Abraham and colleagues in their online code repository (<https://github.com/gabraham/flashpca>). Specifically, we first generated a list of UKB participants that (i) were used in original PCA (*i.e.*, passed QC thresholds, pruned for kinship, etc.) and (ii) self-reported 'White British' ethnicity and have very similar genetic ancestry based on original PCA of the genotypes (as determined in the sample QC file provided by UKB investigators). We then extracted the hard genotype calls for those individuals and applied the recommended SNP-level QC thresholds (directly genotyped SNPs outside of long-range LD regions, minor allele frequency $\geq .01$, genotyping call rate $\geq .02$, missingness rate $\leq .05$, and a Hardy-Weinberg equilibrium threshold $\geq 5e-6$). Next, we applied the recommended LD pruning thresholds to produce a sample of 322,886 individuals with 77,355 independent markers before estimating the first 40 principal components with flashPCA2. Principal component loadings for each SNP used in the analysis were computed, exported, and then used to project all remaining participants of non-Hispanic European ancestry (*e.g.*, siblings not used in the original PCA, participants of White Irish ancestry, participants of White Scottish ancestry, etc.) onto the PCs, yielding a final set of PC scores for the entire subsample. Scatterplots of the principal component scores were then manually examined to identify potential ancestral outliers in the sample.

1.3 Quality control

We used an EasyQC⁴ pipeline similar to the one reported by Linnér and colleagues⁵ to check each set of UKB summary statistics for quality control problems. For each results file, we applied the following threshold in order.

1. We removed SNPs if either allele corresponded to a value other than "A", "C", "G", or "T".
2. We excluded SNPs if any of the following values were missing: P value, beta, standard error, effect allele frequency, sample size, and imputation accuracy (for imputed SNPs).
3. We excluded SNPs with values outside of permissible ranges (*e.g.*, negative or infinite standard errors, nonsensical P values, allele frequencies greater than 1 or below 0).
4. We dropped SNPs with minor allele frequencies less than .005.
5. We filtered out SNPs with low imputation accuracy, which was defined as an imputation score $< .90$.
6. We dropped duplicated SNPs based on GRCh37 base pair positions. When duplicated SNPs were identified, we retained the SNP with the largest sample size.

We then inspected several diagnostic plots to further ensure that results were not prone to systematic errors, including:

7. We checked for errors in allele frequencies and strand orientations by inspecting a plot of allele frequencies in our analytic sample against the allele frequency in a non-Hispanic European reference sample.
8. We checked for discrepancies between the reported P values and the reported coefficient estimates and their SEs.
9. We looked for evidence of population stratification that had not been accounted for by checking a Q-Q plot.

2 Genomic structural equation modeling

Genomic structural equation modeling (Genomic SEM)⁶ is a novel statistical method for applying structural equation modeling techniques to GWAS summary statistics to model the joint genetic architecture of complex traits. It is a flexible framework that allows for more accurate modeling of multivariate genetic covariance matrices, such as those derived from LD Score regression. Here, we conducted a series of Genomic SEM analyses to investigate the multivariate genetic architecture of the psychiatric symptoms and disorders characterized by mood disturbance and psychosis. The aim of these analyses is three-fold: (i) to identify the latent genetic factor(s) that best represent the factor structure of these phenotypes, (ii) to estimate the effects of individual SNPs on the latent genetic factor(s), and (iii) to evaluate heterogeneous effects among the discovery phenotypes.

2.1 Hierarchical clustering

Hierarchical clustering is a form of cluster analysis that aims to identify features of a dataset that are similar to one another. It serves as a precursor to factor analysis or structural equation modeling, as its results can be used guide model specification decisions in subsequent analyses. To this end, we applied a hierarchical clustering algorithm to a genetic correlation matrix of our eight psychiatric phenotypes prior to any form of factor analysis. Specifically, we applied the complete-linkage hierarchical clustering algorithm employed by the *hclust* function of R. The algorithm identified two clusters present in the matrix.

The first cluster was comprised of the symptom-level phenotypes (depression, mania, and psychosis), major depressive disorder, and bipolar II disorder. The second cluster was comprised of bipolar I disorder, schizoaffective disorder, and schizophrenia. Beyond these two clusters, there clear evidence of a positive manifold across all items. All point estimates were positive, and all but two of the 28 pairwise genetic correlations were significant following Bonferroni correction.

2.2 Factor analysis

Factor analysis is a multivariate statistical technique used to explain variance and covariance among sets of observed, correlated variables in terms of unobserved latent factors. It is a powerful tool for reducing dimensionality of data and accounting for measurement error in observed variables, often used in structural equation modeling. In factor analysis of genetic covariance matrices, k observed variables are described as linear functions of m latent variables, such that the model can be expressed as

$$y = \Lambda\eta + \varepsilon$$

where y is a $k \times 1$ vector of observed variables, ε is a $k \times 1$ vector of observed variable residuals, η is a $m \times 1$ vector of latent variables, and Λ is a $k \times m$ matrix of factor loadings that relate the observed variables to the latent variables.

In the present study, we used the *factanal* function of R to conduct an exploratory factor analysis with promax rotation. Guided by the results described in Supplementary Section 2.1, we tested factor solutions extracting up to three latent factors, retaining the highest dimensional solution where each factor explained at least 20% of the variance. As the three-factor solution did not meet this criterion (F1 = 43%, F2 = 26%, F3 = 18%), we selected the two-factor solution as the best exploratory factor model (Figure 2b).

In this two-factor solution, we found compelling evidence of approximate simple structure. Phenotypes principally loaded onto one of two latent genetic factors with negligible cross-loadings. The two correlated latent genetic factors explained 81.3% of the total genetic variance across phenotypes.

2.3 Structural equation modeling

Structural equation modeling is a statistical framework comprised of a diverse set of models and methods used to explain variance and covariance among sets of variables. While the background and many applications of structural equation modeling are extensive^{7,8}, the fundamentals as they relate to the Genomic SEM framework are briefly reviewed below.

Structural equation models can be represented in two sets of equations: the *measurement model*, which describes how observed variables relate to latent variables, and the *structural model*, which describes how latent variables relate to one another⁶. As in exploratory factor analysis, k observed variables are again described as linear functions of m continuous latent variables. In confirmatory factor analysis, this is referred to as the measurement model, which is still expressed as

$$y = \Lambda\eta + \varepsilon$$

with the same notation as described in Supplementary Section 2.2.

If theory is used to explain associations between latent variables, a structural model can then be specified to relate latent variables to each other via directed regression coefficients. The structural model can be expressed as

$$\eta = B\eta + \zeta$$

where B is a $m \times m$ matrix of regression coefficients that relate latent variables to one another and ζ is a $m \times 1$ vector of latent variable residuals. In this full structural equation model, the observed sample covariance matrix is represented by a set of parameters that relates observed variables to latent variables, and latent variables to each other in a series of linear equations.

Genomic SEM employs a two-stage SEM approach to model the genetic covariances between a set of phenotypes (*i.e.*, the observed phenotypes). Stage 1 consists of estimating the genetic covariance matrix and the sampling covariance matrix. Stage 2 consists of fitting a structural equation model that minimizes misfit between the model-implied and empirical genetic covariances.

In Stage 1, Genomic SEM uses a multivariable form of LD Score regression to estimate the empirical genetic covariance matrix (S) and its associated sampling covariance matrix (V_S). Here, S is a symmetric matrix of order k with SNP heritabilities on the diagonal and genetic covariances between phenotypes off the diagonal. Comprised of $k^* = \frac{k(k+1)}{2}$ nonredundant elements, S can be written as

$$S = \begin{bmatrix} h_1^2 & & & & \\ \sigma_{g1,g2} & h_2^2 & & & \\ \vdots & & \ddots & & \\ \sigma_{g1,gk} & \sigma_{g2,gk} & \dots & & h_k^2 \end{bmatrix}$$

To obtain unbiased estimates of standard errors and test statistics, Genomic SEM then constructs the asymptotic sampling covariance matrix of the LD Score regression estimates, V_S , by using all nonredundant elements in the S matrix. Here, V_S is a symmetric matrix of order k^* where the diagonal elements are sampling variances and the off-diagonal elements are sampling covariances. Thus, it can be written as

$$V_S = \begin{bmatrix} SE(h_1^2)^2 & & & & & \\ cov(h_1^2, \sigma_{g1,g2}) & SE(\sigma_{g1,g2})^2 & & & & \\ \vdots & \vdots & \ddots & & & \\ cov(h_1^2, \sigma_{g1,gk}) & cov(\sigma_{g1,g2}, \sigma_{g1,gk}) & SE(\sigma_{g1,gk})^2 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ cov(h_1^2, h_j^2) & cov(\sigma_{g1,g2}, h_j^2) & cov(\sigma_{g1,gk}, h_j^2) & & SE(h_j^2)^2 & \\ \vdots & \vdots & \vdots & & \vdots & \\ cov(h_1^2, \sigma_{g_j,gk}) & cov(\sigma_{g1,g2}, \sigma_{g_j,gk}) & cov(\sigma_{g1,gk}, \sigma_{g_j,gk}) & cov(h_j^2, \sigma_{g_j,gk}) & SE(\sigma_{g_j,gk})^2 & \\ cov(h_1^2, h_k^2) & cov(\sigma_{g1,g2}, h_k^2) & cov(\sigma_{g1,gk}, h_k^2) & cov(h_j^2, h_k^2) & cov(\sigma_{g_j,gk}, h_k^2) & SE(h_k^2)^2 \end{bmatrix}$$

The diagonal elements of V_S are then estimated with a jackknife resampling procedure following the original bivariate version of LD Score regression.

In Stage 2, Genomic SEM uses the S and V matrices from Stage 1 to estimate the parameters of the specified structural equation model using a weighted least squares (WLS) fit function, as detailed in Grotzinger et al. (2019). Notably, this method is capable of accounting for differences in GWAS sample size, which is ideal in the present study. Furthermore, the off-diagonal elements of V_S index the extent to which the sampling errors across input GWAS are correlated, which means that Genomic SEM, like LD score regression upon which it is based, is unbiased and robust to varying degrees of, or even complete, sample overlap.

2.3.1 Confirmatory factor analysis

Confirmatory factor analysis is a common application of structural modeling where theoretical models are used to explain the observed covariances among a set of observed variables. Here, we tested a series of competing models to identify the model that best fit the data, where good fit indicated that the specified latent variable structure adequately explained the observed genetic covariances among the set of observed variables.

Guided by the results described in Supplementary Sections 2.1 and 2.2, as well as psychiatric and psychometric theory, we tested a series of confirmatory factor models to identify the solution that best explained the observed genetic covariances among the set of discovery phenotypes. Specifically, we tested four models: (i) a single common factor model (*i.e.*, a p factor), (ii) a correlated factors model (iii) a correlated factors model with bipolar II disorder to cross-loading on both factors, and (iv) a correlated factors model with correlated residuals between bipolar I disorder and bipolar II disorder. Path diagrams for these models are presented in Figure S2. Unit variance identification was used to set the scale of the latent factors.

Model fit was assessed using conventional indices in structural equation modeling: the model χ^2 statistic, the Akaike information criterion (AIC), the comparative fit index (CFI), and the standardized root mean square residual (SRMR). All of these indices retain their standard interpretations within a Genomic SEM framework with the exception of the model χ^2 statistic⁶. In large samples, such as those used here, χ^2 tests are overpowered

and likely to be significant. As such, the model χ^2 statistic was used as a comparative measure of fit to evaluate competing models (akin to AIC), rather than a measure of statistical significance. For CFI and SRMR, values greater than .90 and less than .08, respectively, were considered reflective of good model fit⁹.

Common factor model

As a baseline, we evaluated a common factor model with all eight phenotypes operating as indicators for a single latent factor. While easily interpretable, this particular model exhibited poor fit, as indicated by model fit indices ($\chi^2(20) = 1630.24$, AIC = 1662.24, CFI = .99, SRMR = .20). Inspection of the observed and model implied genetic correlation matrices indicated that the common factor model implied that genetic correlations between schizophrenia, schizoaffective disorder, bipolar I disorder, and bipolar II disorder were severely downwardly biased. Moreover, the genetic correlations between psychiatric symptoms and disorders were modestly upwardly biased in the common factor model.

Correlated factors model

Preliminary results described in Supplementary Sections 2.1 and 2.2 suggested the promise of a two-factor model, where phenotypes loaded onto two distinct-but-correlated latent genetic factors. As an initial test of this model, we estimated a simple correlated factors model with no cross-loadings and no correlated residuals. While this model had better fit than the common factor model, some model fit indices were still suboptimal ($\chi^2(19) = 608.97$, AIC = 642.97, CFI = .99, SRMR = .11). We next fit a model allowing bipolar II disorder to cross-load onto both factors. This model also showed good fit, as indicated by model fit indices ($\chi^2(18) = 390.93$, AIC = 426.93, CFI = .99, SRMR = .08), but the cross-loading resulted in markedly lower loadings for bipolar II disorder (.51 and .43 for F1 and F2, specifically). Finally, we fit a correlated factors model that estimated the correlation between residual variance in bipolar I disorder and bipolar II disorder. This model fit the data best, closely approximating the observed genetic covariance matrix ($\chi^2(18) = 496.16$, AIC = 532.16, CFI = .99, SRMR = .06).

While all variations of the correlated factors model showed improved fit over the common factor model, we identified the model that included correlated residuals as the best fitting model. Notably, this model provided a parsimonious and easily interpretable factor structure that simultaneously minimized the standardized difference between the observed and predicted genetic correlations. Inspection of the observed and model implied genetic correlation matrices indicated that the final correlated factors model fit the data substantially better than the common factor model because it appropriately accounted for different patterns of covariance by segregating the phenotypes into two separate-but-correlated latent factors (Figure S3).

2.3.2 *Genetic correlation*

Structural equation models can also be used to estimate the genetic correlation between an unobserved latent factor and an observed exogenous phenotype not included in the model. Indeed, this method is preferable to using bivariate LD score regression, as it based on the genetic covariances directly rather than the estimated SNP effects, which may not be mediated by the latent factor(s) (see Supplementary Section 2.5 for more on heterogeneous SNP effects). To estimate genetic correlations between a latent factor and an observed exogenous phenotype, we first created a single-item quasi-latent factor for the exogenous phenotype, and fixed the residual variance for the phenotype to zero. We then estimated the correlation between our latent factors and the quasi-latent factor for the phenotype of interest.

2.3.3 *Multivariable genetic regression*

The approach described in Supplementary Section 2.3.2 can be extended to conduct multivariable genetic regression, which yields estimates of genetic associations between two variables after accounting for relationships with additional variables in the model (*i.e.*, partial genetic correlations). Here, this is done by

regressing the exogenous phenotype onto F1 and F2 while simultaneously estimating the genetic correlation between the two latent genetic factors.

2.4 Multivariate genome-wide association analyses

After identifying a confirmatory factor model that best explained the observed genetic covariances among the phenotypes, we conducted a multivariate GWAS by estimating the individual SNP effects on each latent factor in the model. A brief overview of this method is provided below.

To estimate the effect of SNP j on F1 and F2, individual SNP effects are included in both the genetic covariance matrix and the sampling covariance matrix. This is accomplished by expanding the genetic covariance matrix to include covariances between SNP j and the latent genetic components of each phenotype, g_1 through g_k .

$$S = \begin{bmatrix} \sigma_{SNP}^2 & & & \\ \sigma_{SNP,g_1} & h_1^2 & & \\ \sigma_{SNP,g_2} & \sigma_{g_1,g_2} & h_2^2 & \\ \sigma_{SNP,g_k} & \sigma_{g_1,g_k} & \sigma_{g_2,g_k} & h_k^2 \end{bmatrix}$$

The associated sampling covariance matrix, V_S , then includes the following: (i) the sampling variances and sampling covariances of the SNP heritabilities and genetic covariances, (ii) the variance of SNP j as derived from reference panel data, and (iii) the sampling covariances of the SNP-genotype covariances. Finally, Genomic SEM is used to estimate m models to obtain GWAS summary statistic for the latent factors, where m is the number of SNPs present across all included summary statistics.

Note that unit loading identification is used to set the scale of latent factors for models including SNP effects. This is a difference from the structural equation models without SNP effects, where unit variance identification is used to facilitate easy interpretation of factor loadings. This is done for two reasons. First, if the variance of the factor were set to 1, the inclusion of a SNP as a regressor technically changes the variance of the latent factor to be 1 plus the variance explained by the SNP. Second, the use of unit variance identification scales SNP effects as if they were for a phenotype that was entirely heritable (*i.e.*, $h_{SNP}^2 = 1$). This distinction does not change the ratio of effect estimates to standard errors, but it does potentially complicate comparison to other GWAS results. Thus, we find that unit loading identification is the most appropriate method to scale latent factors in models that include SNP effects. Here, we set the scale of F1 and F2 by fixing the factor loadings of psychotic symptoms and schizophrenia, respectively.

2.5 Heterogeneity tests

It is possible that SNP effects might vary across each indicator and not act entirely through the common latent variable. To evaluate this potential heterogeneity in SNP effects, we computed genome-wide Q_{SNP} statistics, which are χ^2 -distributed test statistics estimated for each SNP in the multivariate GWAS. As described by Grotzinger and colleagues⁶, larger values for Q_{SNP} reflect a violation of the null hypothesis that the SNP acts entirely through the latent factor(s).

2.6 Effective sample size

While it can be difficult to estimate effective sample size for a given SNP in a latent factor model, we were able to produce reasonable estimates of effective sample size for the overall multivariate GWAS under a set of reasonable assumptions. First, we assume that the per-allele effect of SNP j on the standardized phenotype is very small, such that it follows

$$\beta_j = \frac{Z_j}{\sqrt{n_j \times 2 \times MAF_j (1 - MAF_j)}}$$

where Z is the Z statistic, n is the unknown effective sample size that we seek to estimate, and MAF is the minor allele frequency of SNP j . Note that the variance of SNP j (σ_j^2) is estimated as $2 \times MAF_j (1 - MAF_j)$. Therefore, if we know the effect and MAF of that SNP, then we can estimate its effective sample size by solving for n_j .

$$\frac{\beta_j}{Z_j} = \frac{1}{\sqrt{n_j \times \sigma_j^2}}$$

$$\frac{Z_j}{\beta_j} = \sqrt{n_j \times \sigma_j^2}$$

$$\left(\frac{Z_j}{\beta_j}\right)^2 = n_j \times \sigma_j^2$$

$$n_j = \frac{(Z_j/\beta_j)^2}{\sigma_j^2}$$

We note that when the phenotype is a latent factor, the choice of scaling the factor will have a nontrivial effect on the estimate of n_j . Here we scale the latent genetic factors with unit loading identification, such that n_j can be intuitively interpreted as the effective sample size in the units of the standardized reference phenotype (As indicated in 2.4, we set the scale of F1 and F2 by fixing the factor loadings of psychotic symptoms and schizophrenia, respectively). If we were to scale the latent genetic factors with unit variance identification, the effective sample size would be interpreted relative to a factor that is 100% heritable, and n_j would be unintuitively very small (because, ceteris paribus, highly heritable phenotypes require smaller sample sizes to detect genetic associations).

This formula will typically produce reasonable estimates of n_j when the factor is scaled using a unit loading identification strategy, but it can be prone to error for SNPs with low MAF. Here, we set a lower and upper MAF limit of approximately 10% and 40%, respectively, when estimating effective N for the overall multivariate GWAS results (N_{eff}). Following this, we estimate that N_{eff} is approximately equal to the mean n_j for m SNPs with a MAF between a and b . This can be expressed as

$$N_{eff} \approx \frac{1}{m} \sum_{MAF=a}^b n_j$$

Here, we apply this to the results for the results for F1 and F2 and estimate that the effective N for each phenotype is 377,518 and 51,276, respectively. We note that this calculation is robust to sample overlap in the multivariate GWAS, as Genomic SEM accounts and corrects for such overlap.

We note that with an effective sample size, it is possible to backout an estimate of genetic variance that is, in one sense, conceptually analogous to SNP heritability in that they reflect the scale of SNP effects for a GWAS

target. However, as there is no information about phenotypic variance of the latent genetic factors, it should not be interpreted as a heritability estimate. Rather, these genetic variance estimates are only useful as an additional metric for comparing GWAS results when paired with effective sample size. Here, the latent factors F1 and F2 have genetic variance estimates of 6% (SE = .28%) and 56% (SE = 2.29%) respectively. The genetic variance estimates for F1 and F2 differ because of the differing SNP heritability estimates of their underlying indicators, with lower heritability estimates observed for self-report symptoms than for clinically-defined disorders. The greater proportion of non-genetic variance in self-report symptoms might be due to greater influence of environmental variation and/or greater unreliability of measurement.

3 References

1. Muthén, L.K., and Muthén, B. (2017). *Mplus User's Guide: Eighth Edition* (Muthén & Muthén).
2. Curran, P.J., Hussong, A.M., Cai, L., Huang, W., Chassin, L., Sher, K.J., and Zucker, R.A. (2008). Pooling data from multiple longitudinal studies: the role of item response theory in integrative data analysis. *Dev. Psychol.* *44*, 365–80.
3. Abraham, G., Qiu, Y., and Inouye, M. (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* *33*, 2776–2778.
4. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Mägi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al. (2014). Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* *9*, 1192–212.
5. Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S.F.W., Wedow, R., Fontana, M.A., Lebreton, M., Tino, S.P., Abdellaoui, A., Hammerschlag, A.R., et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* *51*, 245–257.
6. Grotzinger, A.D., Rhemtulla, M., de Vlaming, R., Ritchie, S.J., Mallard, T.T., Hill, W.D., Ip, H.F., Marioni, R.E., McIntosh, A.M., Deary, I.J., et al. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* *3*, 513–525.
7. Loehlin, J.C., and Beaujean, A.A. (2016). *Latent variable models: an introduction to factor, path, and structural equation analysis* (Taylor & Francis).
8. Kline, R.B. (2015). *Principles and practice of structural equation modeling* (Guilford publications).
9. Hu, L., and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling* *6*, 1–55.