nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0566-x

Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits

Andrew D. Grotzinger[1*], Mijke Rhemtulla[2], Ronald de Vlaming[3,4], Stuart J. Ritchie[5,6], Travis T. Mallard[1], W. David Hill[5,6], Hill F. Ip[7], Riccardo E. Marioni[5,8], Andrew M. McIntosh[5,9], Ian J. Deary[5,6], Philipp D. Koellinger[3,4], K. Paige Harden[1,10], Michel G. Nivard[7,11] and Elliot M. Tucker-Drob[1,10,11]

Genetic correlations estimated from genome-wide association studies (GWASs) reveal pervasive pleiotropy across a wide variety of phenotypes. We introduce genomic structural equation modelling (genomic SEM): a multivariate method for analysing the joint genetic architecture of complex traits. Genomic SEM synthesizes genetic correlations and single-nucleotide polymorphism heritabilities inferred from GWAS summary statistics of individual traits from samples with varying and unknown degrees of overlap. Genomic SEM can be used to model multivariate genetic associations among phenotypes, identify variants with effects on general dimensions of cross-trait liability, calculate more predictive polygenic scores and identify loci that cause divergence between traits. We demonstrate several applications of genomic SEM, including a joint analysis of summary statistics from five psychiatric traits. We identify 27 independent single-nucleotide polymorphisms not previously identified in the contributing univariate GWASs. Polygenic scores from genomic SEM consistently outperform those from univariate GWASs. Genomic SEM is flexible and open ended, and allows for continuous innovation in multivariate genetic analysis.

Genome-wide association studies (GWASs) are rapidly identifying loci affecting multiple social, behavioural and psychiatric phenotypes[1,2]. Moreover, using cross-trait versions of methods such as genomic-relatedness-based restricted maximum likelihood (GREML)[3] and linkage disequilibrium score regression (LDSC)[4], researchers have identified genetic correlations between diverse traits; for example, age of first birth and risk of smoking[5], insomnia and psychiatric traits (for example, schizophrenia)[6], major depressive disorder (MDD) and number of children[7], and educational attainment and cognitive performance[8]. Widespread statistical pleiotropy appears to be the rule rather than the exception across complex traits. Although these findings are currently suggestive of constellations of phenotypes affected by shared sources of genetic liability, existing methods do not permit the causes of the observed genetic correlations to be investigated systematically. Here, we introduce genomic structural equation modelling (genomic SEM)—a new method for modelling the multivariate genetic architecture of constellations of traits and incorporating genetic covariance structure into multivariate GWAS discovery. Genomic SEM is a flexible framework for formally modelling the genetic covariance structure of complex traits using GWAS summary statistics from samples of varying and potentially unknown degrees of overlap, in contrast with existing methods that model phenotypic covariance structure[9], with specific applications[10], using raw data. Moreover, genomic SEM allows for the specification and comparison of a range of proposed multivariate genetic architectures, which improves on existing approaches for combining information across genetically correlated traits to aid in discovery[11].

One powerful feature of genomic SEM is the capability to model shared genetic architecture across phenotypes with factors representing broad genetic liabilities, and compare the fit of different factor structures to the empirical data. When an appropriate model has been identified at the level of the genome-wide covariance structure, the researcher may incorporate individual single-nucleotide polymorphisms (SNPs) into the model to identify variants with effects on general dimensions of cross-trait liability, boost power for discovery, and calculate more valid and predictive polygenic scores (PGSs). Genomic SEM can also evaluate whether the multivariate genetic architecture implied by a specific model is applicable at the level of individual variants using developed estimates of heterogeneity. When certain SNPs only influence a subset of genetically correlated traits, a key assumption of other multivariate approaches is violated[11]. SNPs with high heterogeneity estimates can be flagged as likely to confer disproportionate liability towards individual traits, be removed when constructing polygenic risk scores, or be studied specifically to understand the nature of heterogeneity.

We validate key properties of genomic SEM with a series of simulations, and illustrate the flexibility and utility of genomic SEM with

[1]Department of Psychology, University of Texas at Austin, Austin, TX, USA. [2]Department of Psychology, University of California, Davis, Davis, CA, USA. [3]Department of Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [4]Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, The Netherlands. [5]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [6]Department of Psychology, University of Edinburgh, Edinburgh, UK. [7]Department of Biological Psychology, Vrije Universiteit University Amsterdam, Amsterdam, The Netherlands. [8]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [9]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [10]Population Research Center, University of Texas at Austin, Austin, TX, USA. [11]These authors jointly supervised this work: Michel G. Nivard, Elliot M. Tucker-Drob. *e-mail: agrotzin@utexas.edu

analyses of real data. These include a joint analysis of GWAS summary statistics from five genetically correlated psychiatric case-control traits: schizophrenia, bipolar disorder, MDD, post-traumatic stress disorder (PTSD) and anxiety. We model their joint genetic architecture using a general factor of psychopathology (*p*), for which we identify 27 independent SNPs not previously identified in the univariate GWASs, 5 of which can be validated based on separate GWASs. PGSs derived using this *p* factor consistently outperform PGSs derived from GWASs of the individual traits in out-of-sample prediction of psychiatric symptoms. Other demonstrations include a multivariate GWAS of neuroticism items, an exploratory factor analysis (EFA) of anthropometric traits and a simultaneous analysis of the unique genetic associations between schizophrenia, bipolar disorder and educational attainment.

Results

Genomic SEM is a two-stage structural equation modelling approach[12–14]. In stage 1, the empirical genetic covariance matrix and its associated sampling covariance matrix are estimated. The diagonal elements of the sampling covariance matrix are squared standard errors. The off-diagonal elements index the extent to which sampling errors of the estimates are associated, as may be the case when there is sample overlap across GWASs. In stage 2, a SEM is specified and parameters are estimated by minimizing the discrepancy between the model-implied genetic covariance matrix and the empirical covariance matrix obtained in the previous stage. We evaluate fit with the standardized root mean square residual (SRMR), model χ^2, Akaike Information Criterion (AIC) and Comparative Fix Index (CFI; see Methods)[13,15]. In a set of simulations, we verify key properties of genomic SEM (Methods). We find that genomic SEM produces unbiased parameter estimates when the correct structural model is specified, and that model fit indices consistently favour the correct model over alternative models. In a second set of simulations, we demonstrate that the inclusion of data from overlapping samples does not bias genomic SEM parameter estimates or their standard errors.

Genomic SEM can be employed as a tool for multivariate GWASs based on univariate summary statistics. First, the genetic covariance matrix and its associated sampling covariance matrix are expanded to include SNP effects. A genomic SEM is then specified in which SNP effects occur at the level of a latent genetic factor defined by several phenotypes, at the level of the genetic components of each of several (potentially genetically correlated) phenotypes, or some combination of the two. The genomic SEM is then run once per SNP (or each set of SNPs, should the user incorporate multiple SNPs into a model) to obtain its effects within the multivariate system.

We provide an index that quantifies the extent to which an observed vector of univariate regression effects of a given SNP on each of the phenotypes can be explained by a common pathway model that assumes that the effects are entirely mediated by the common genetic factor(s). In other words, the index enables the identification of loci that do and do not plausibly operate on the individual phenotypes exclusively by way of their associations with the common factor(s). Because of its intuitive and mathematical similarity to the meta-analytic *Q* statistic used in standard meta-analyses to index heterogeneity of effect sizes[16], we label this heterogeneity statistic Q_{SNP}. Q_{SNP} is a χ^2-distributed test statistic, with larger values indexing a violation of the null hypothesis that the SNP acts entirely through the common factor(s).

Confirmatory factor analysis of genetic covariance matrices. We provide two examples of confirmatory factor analysis (CFA) using genomic SEM. In our first example, we fit a genetic factor model to psychiatric case-control traits. Recent findings indicate that the comorbidity across psychiatric disorders is captured by a general psychopathology factor (that is, the *p* factor), and this is widely supported based on previous results[17–21]. We tested for the presence of a single common genetic *p* factor using genomic SEM with European-only summary statistics for schizophrenia, bipolar disorder, MDD, PTSD and anxiety (Supplementary Table 1 for phenotypes and sample sizes). The model fit was adequate ($\chi^2(5) = 89.55$; AIC = 109.50; CFI = 0.848; SRMR = 0.212). The results indicated that schizophrenia and bipolar disorder loaded the strongest onto the genetic *p* factor (Supplementary Fig. 1)—a pattern of findings that closely replicates previous findings from twin/family studies[19].

In a second example, we tested for the presence of a single common genetic factor of neuroticism using summary statistics from 12 item-level indicators from UK Biobank (UKB; Supplementary Table 1), as estimated using the Hail software[22]. The model fit was good ($\chi^2(54) = 4{,}884.10$; AIC = 4,932.11; CFI = 0.893; SRMR = 0.109). The results indicated strong positive loadings for all indicators (Supplementary Fig. 2). We used this single common factor model for both neuroticism and the *p* factor when estimating SNP effects for discovery under the section 'SNP effects' below.

EFA of a genetic covariance matrix. We provide two examples of how one might use exploratory methods to guide the specification of more nuanced factor models. In the first example, we submitted the LDSC-derived genetic correlation matrix of the 12 neuroticism items in UKB to EFA (see Supplementary Results). Based on these initial EFA results, follow-up CFAs (Supplementary Fig. 3) were specified using genomic SEM (standardized loadings > 0.4 were retained; Supplementary Table 2). The two-factor solution ($\chi^2(53) = 2{,}758.18$; AIC = 2,808.18; CFI = 0.940; SRMR = 0.077) and three-factor solution ($\chi^2(51) = 1{,}879.31$; AIC = 1,933.31; CFI = 0.959; SRMR = 0.057) both provided excellent fit to the data and exceeded the fit of the single, common factor model. Consistent with the superior model fit indices for the two- and three-factor solutions, only 28 and 20 of the 69 Q_{SNP} hits from the single common factor model (described in further detail in the 'SNP effects' section below) continued to surpass genome-wide significance for the two- and three-factor models, respectively (Supplementary Fig. 4 and Supplementary Table 3). In addition, a GWAS of all HapMap3 SNPs for the two- and three-factor models revealed that the average size of Q_{SNP} across all SNPs was largest for the common factor ($\chi^2(1) = 1.68$), followed by the two-factor ($\chi^2(1) = 1.64$) and three-factor models ($\chi^2(1) = 1.51$). Thus, heterogeneity indices of individual SNP effects in the GWAS data agree with model fit indices, with both favouring the three-factor model of neuroticism.

In the second example, EFA was applied to the LDSC-derived genetic correlation matrix for nine anthropometric traits from the Early Growth Genetics and Genetic Investigation of Anthropometric Traits consortia (Supplementary Table 4). EFA results indicated that 2 factors explained 61% of the total genetic variance. Moreover, a heat map of the genetic correlation matrix suggests two primary factors that index overweight and early life-growth phenotypes (Supplementary Fig. 5). A follow-up CFA (Supplementary Fig. 6) within genomic SEM was specified based on the EFA parameter estimates (standardized loadings > 0.25 were retained). The CFA showed good fit to the data ($\chi^2(25) = 12{,}994.71$; AIC = 13,034.71; CFI = 0.962; SRMR = 0.092). The results indicated highly significant loadings and a small genetic correlation (r_g) between the two factors ($r_g = 0.10$; s.e. = 0.03; $P < 0.001$). This indicates that early-life physical growth is modestly associated with later life obesity traits via genetic pathways.

Genetic multivariable regression (replicating genome-wide inferred statistics (GWIS)). Nieuwboer et al.[23] used summary statistics for educational achievement[24] and both schizophrenia and bipolar disorder[25] to determine whether genetic correlations with educational achievement are driven by variation specific to either disorder. Educational achievement is genetically correlated with

schizophrenia ($r_g = 0.148$; s.e. = 0.050; $P = 0.003$) and bipolar disorder ($r_g = 0.273$; s.e. = 0.067; $P < 0.001$). Using the GWIS method, they found that the correlation of educational achievement with schizophrenia unique of bipolar is small ($r_g = 0.040$; s.e. = 0.082; $P = 0.627$), whereas the genetic correlation between bipolar unique of schizophrenia and educational achievement is far less attenuated ($r_g = 0.218$; s.e. = 0.102; $P = 0.032$). We used genomic SEM with the aim of replicating these results using a conceptually similar, but statistically distinct, framework. We present this example to demonstrate that genomic SEM is not limited to factor analytical models, but can be used to construct and test an array of hypotheses using a general SEM approach.

Using the same univariate GWAS summary statistics employed in the original application of GWIS, we used genomic SEM to fit a structural multivariable regression model in which the genetic component of educational achievement was simultaneously regressed onto the genetic components of schizophrenia and bipolar disorder. The results confirmed the findings of Nieuwboer et al.[23]; the conditional standardized association between schizophrenia and educational achievement was small ($b_g = -0.016$; s.e. = 0.096; $P = 0.867$), whereas there was a strong conditional standardized association between bipolar disorder and educational achievement ($b_g = 0.283$; s.e. = 0.113; $P = 0.012$; Supplementary Fig. 7).

SNP effects. *Common factor models.* A powerful application of genomic SEM is to include individual SNP effects in both the genetic covariance matrix and the sampling covariance matrix, to estimate the effect of a given SNP on the latent genetic factor(s). If the summary statistics are composed of M different SNPs, M models are estimated to obtain genome-wide summary statistics for the latent factor. As an example of genomic SEM used for multivariate GWAS, we incorporated SNP effects into the *p*-factor and neuroticism models presented above. Linkage disequilibrium-independent hits are defined below using a clumping threshold of $r^2 < 0.1$ in a 500-kilobase (kb) window, with the exception of a 1-megabase (Mb) window for chromosomes 6 and 8. A total of 128 independent loci were genome-wide significant for the *p* factor ($P < 5 \times 10^{-8}$; Supplementary Figs. 8–10 and Figs. 1a and 2a). Of the 128 loci, 27 independent loci were not previously identified in any of the contributing univariate GWASs (Table 1 and Supplementary Table 5). Of these 27 loci, 5 were identified as either genome-wide significant or suggestive of significance ($P < 1 \times 10^{-5}$) in a separate, previously published GWAS of 1 of the 5 traits. A total of 118 loci were genome-wide significant for neuroticism, with 38 loci not identified in the univariate item-level GWASs (Supplementary Table 6 and Figs. 1b and 2b). Plots of item-level effects for individual SNPs revealed high consistency in magnitude and direction for SNPs identified as genome-wide significant for the common factors (Supplementary Fig. 11). Although there is early lift-off in the *Q*–*Q* plots for both common factors, LDSC analyses of the summary statistics produced by genomic SEM indicated that the results were not due to uncontrolled inflation for either the *p* factor (intercept = 0.987; s.e. = 0.014) or neuroticism (intercept = 0.997; s.e. = 0.001).

General trends. Mean χ^2 statistics were higher for the genomic SEM-derived summary statistics of common factors relative to univariate indicators (Table 1). It is important to note here that, whereas genomic SEM may boost power in many cases, this is not the primary purpose of the method. Rather, it is to identify the relationship between SNPs and observed phenotypes as meditated through a user-specified model and to concurrently evaluate the construct validity of said model. Inspecting the distribution of univariate *P* values for the newly identified SNPs for the general factors indicated that these SNPs were generally characterized by relatively low *P* values, albeit not low enough to cross the genome-wide significance threshold for any individual phenotype (Supplementary Figs. 12 and 13).

Q_{SNP} results. The results revealed 1 and 69 independent Q_{SNP} loci for the *p* factor and neuroticism, respectively (Fig. 2c,d and Supplementary Fig. 14). For neuroticism, significant Q_{SNP} estimates were obtained for SNPs that were highly significant for some traits but not others (Supplementary Table 7 and Supplementary Fig. 15). The association between *P* values for SNP effects and Q_{SNP} estimates were minimal (Supplementary Fig. 16). Comparing the Q_{SNP} estimates for SNPs identified as significant for only the *p* factor or neuroticism relative to SNPs identified as significant for one of the indicators, but not the common factor, indicated that the latter group of SNPs were characterized, as would be expected, by larger Q_{SNP} estimates (that is, greater heterogeneity in individual effects; Supplementary Fig. 17). Intercepts from LDSC analyses of the Q_{SNP} statistics also indicated that the results for the heterogeneity index were not attributable to inflation (*p* factor: intercept = 0.978; s.e. = 0.009; neuroticism: intercept = 0.963, s.e. = 0.009). Slopes from the same LDSC analyses further indicated genetic signal in heterogeneity (*p* factor: $Z = 13.65$; $P = 6.68 \times 10^{-42}$; neuroticism: $Z = 30.23$; $P = 9.98 \times 10^{-201}$).

Comparison with multi-trait analysis of GWAS (MTAG). Existing multivariate methods use summary statistics of genetically correlated phenotypes to boost power for discovery and prediction for a particular trait[11,26,27]. Boosting power is only one application of genomic SEM. That said, a genomic SEM common factor GWAS approach has already been shown by an independent research group to perform comparably to existing multivariate approaches for out-of-sample prediction[28]. Moreover, as a flexible modelling framework, genomic SEM may encompass other multivariate approaches. For example, we show mathematically that genomic SEM can be specified to satisfy the same moment conditions as MTAG[11] (see Supplementary Methods). Simulation results also revealed near-perfect correspondence from a linear regression in which *Z* statistics from MTAG were used to predict those from a genomic SEM specified to satisfy the MTAG moment conditions (Supplementary Fig. 18; unstandardized slope = 0.999; intercept = 2.65×10^{-4}).

Performance in empirical data under controlled missingness. We contrast estimates obtained from the common factor model of neuroticism described above with estimates for a GWAS with an imposed missing structure. We first transformed the binary scale neuroticism items into a smaller number of quantitative scores. To do so, we created three parcels of neuroticism items consisting of four items each with scores ranging from 0–4, at which point it is appropriate to treat the parcel as continuous[29]. Parcels were constructed based on the same EFA results described above and mirrored the composition of the three-factor model, with the exception that the irritability item was included with parcel 2 so as to have an equal distribution of four items per parcel. Of the 300,000 participants, 100,000 non-overlapping participants were removed from 2 of the 3 parcels for missing data models. The best-powered results (indexed by mean χ^2 values) were for genomic SEM of the individual neuroticism items presented above, indicating that construction of composite indices via averaging, although convenient, removes multivariate information that can otherwise be retained with genomic SEM (Supplementary Table 8). Genomic SEM analyses that incorporated supplementary information from parcels containing imposed missing data consistently outperformed GWASs of individual parcels with complete data, and performed nearly as well as analyses of complete data across all three parcels. Thus, the inclusion of summary data from genetically correlated phenotypes in genomic SEM may boost power relative to GWASs of the individual phenotypes,

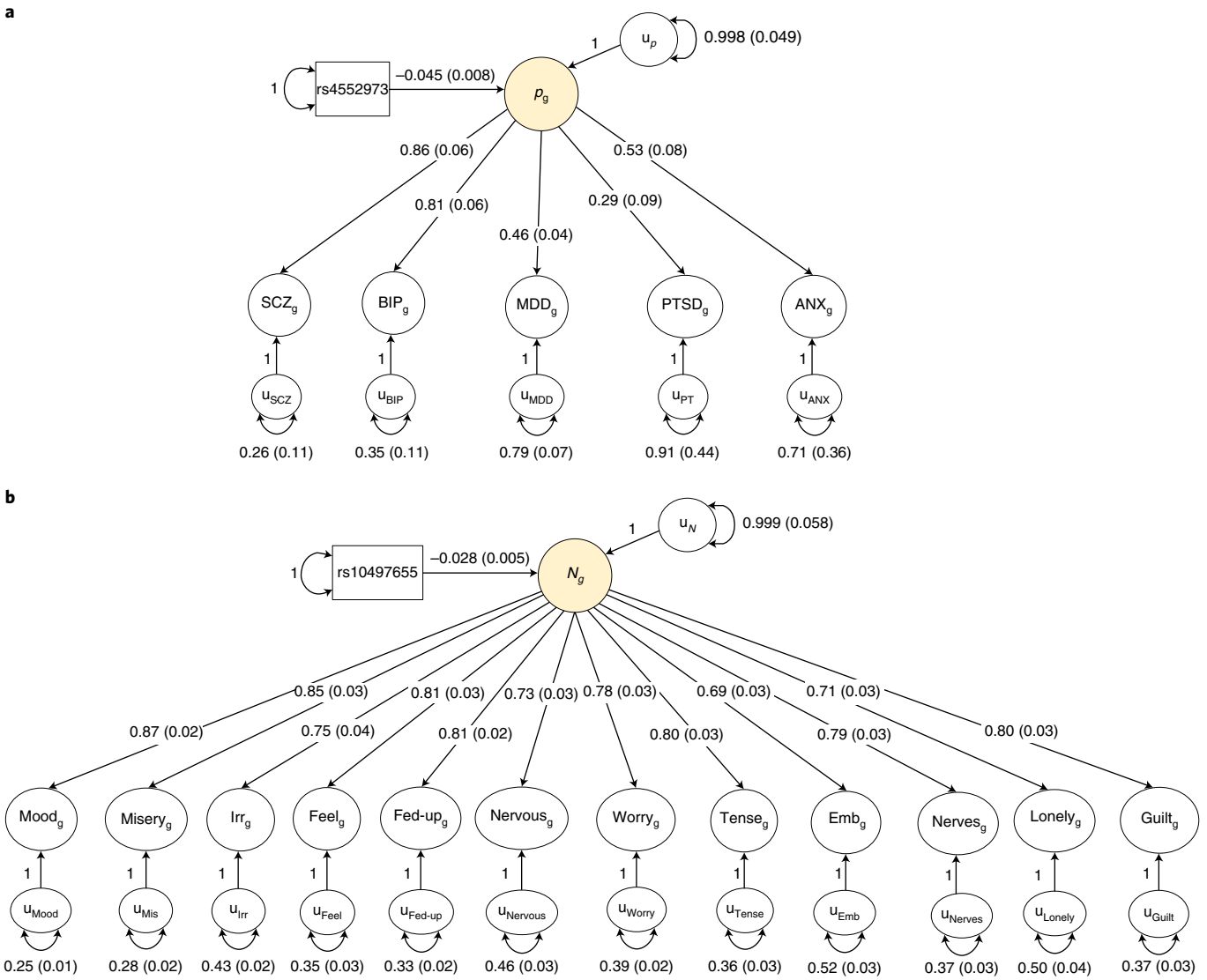

Fig. 1 | Genomic SEM solutions for *p*- and neuroticism-factor models with SNP effect. a,**b**, Standardized results from using genomic SEM (with WLS estimation) to construct a genetically defined *p* factor of psychopathology (**a**) and a genetic neuroticism factor (**b**) with a lead independent SNP predicting the factors (rs4552973 for the *p* factor and rs10497655 for neuroticism). Standard errors are shown in parentheses. For a model that was standardized with respect to the outcomes only, the effect of the SNP was −0.093 (s.e. = 0.017; SNP variance = 0.252) for the *p* factor and −0.042 (s.e. = 0.007; SNP variance = 0.432) for neuroticism. This can be interpreted as the expected standard deviation unit difference in the latent factor per effect allele. The subscript g is used in these path diagrams to denote genetic variables. The *u* variables reflect the residual variance in the genetic indicators not explained by the common factor. ANX, anxiety; BIP, bipolar disorder; Emb, worry too long after embarrassment; Fed-up, fed-up feelings; Feel, sensitivity/hurt feelings; Irr, irritability; SCZ, schizophrenia.

even when there is high sample overlap and sample sizes are uneven across phenotypes.

Parcel comparison of Q_{SNP}. Using the three constructed parcels without any missing data, the distribution of *P* values was compared across SNPs with high ($P < 5 \times 10^{-8}$) and low ($P > 5 \times 10^{-3}$) Q_{SNP} estimates from the item-level genomic SEM analysis of neuroticism for SNPs that were genome-wide significant in at least one of the parcels. These results indicated that, for SNPs with a higher Q_{SNP} for the common factor, there was more discordance of effect sizes among three lower-order factors relative to SNPs that produced lower heterogeneity estimates (Supplementary Fig. 19). The average differences between the highest and lowest $-\log_{10}[P]$ values were 10.56 and 4.96 for high and low Q_{SNP}, respectively. This suggests that Q_{SNP} is appropriately indexing discordance in SNP-level effects across genetically correlated indicators.

Polygenic prediction. We re-estimated the *p* factor model using the summary statistics from the schizophrenia and MDD GWASs that did not overlap with the UKB dataset to predict psychiatric symptoms in UKB (see Supplementary Fig. 20 for the phenotypic model). To produce a reliable set of targets for polygenic prediction, and to focus our analyses on construct validation, latent factors of psychiatric symptoms were specified as the out-of-sample targets. We compared the magnitude of out-of-sample prediction for the *p* factor PGSs predicting the phenotypic *p* factor and factors of individual psychiatric domains relative to the prediction using PGSs derived from univariate summary statistics (Fig. 3 and Supplementary Table 9). The PGS for the genetic *p* factor predicted more variance in symptoms of depression, psychotic experiences, mania, anxiety, PTSD and a phenotypic *p* factor than any univariate PGS.

For neuroticism, univariate PGSs were constructed in data from the Generation Scotland study using summary statistics for

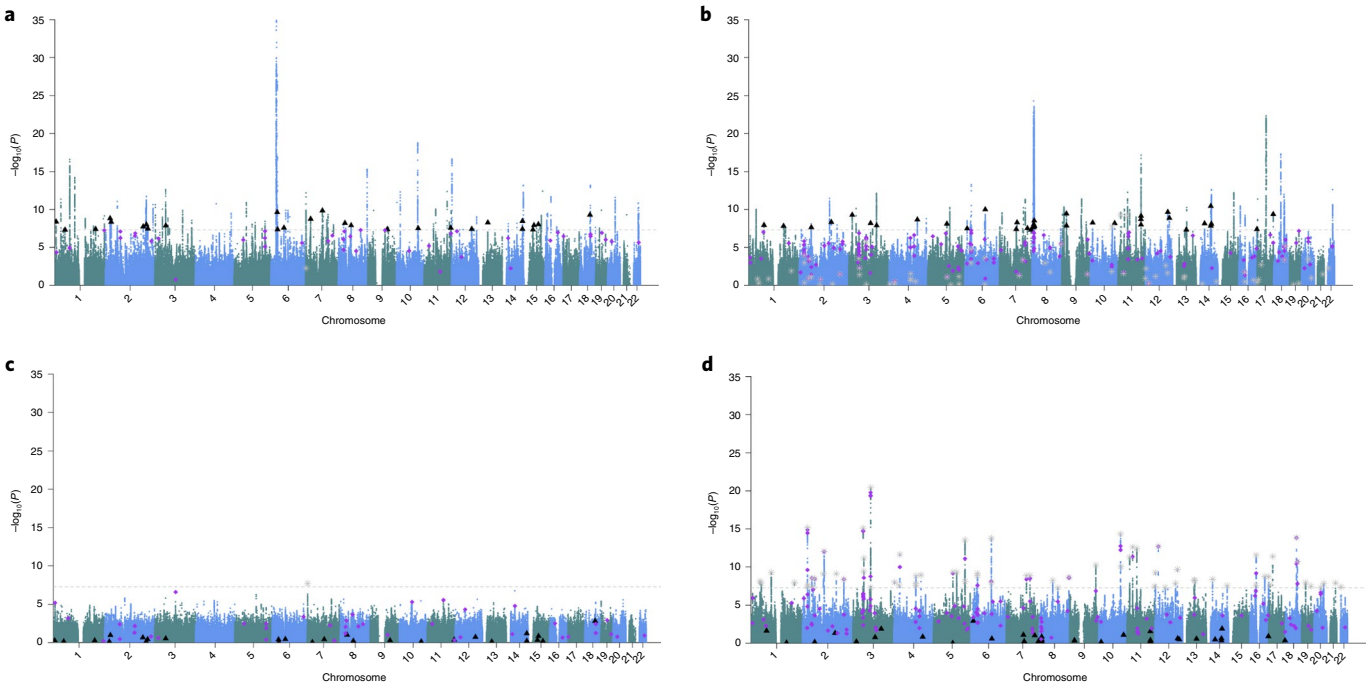

Fig. 2 | Manhattan plots of unique, independent hits from genomic SEM. a–d, Genomic SEM (with WLS estimation) was used to conduct multivariate GWASs of the *p* factor (**a** and **c**) and neuroticism (**b** and **d**). Manhattan plots are shown for SNP effects (**a** and **b**) and Q_{SNP} (**c** and **d**). The grey dashed line marks the threshold for genome-wide significance ($P < 5 \times 10^{-8}$). Black triangles denote independent hits for SNP effects from the GWAS of the general factor that were not in linkage disequilibrium with independent hits for the univariate GWAS or hits for Q_{SNP}. Purple diamonds denote independent hits for the SNP effects from univariate GWASs that were not in linkage disequilibrium with independent hits from the GWAS of the general factor. Grey stars denote independent hits for Q_{SNP}.

the 12 neuroticism items, genomic SEM factor of items, 3 neuroticism parcels, genomic SEM factor of parcels and neuroticism sum score. We used PGSs to predict a sum score composed of the same neuroticism items administered in UKB. We also calculated mean χ^2 values for each of these summary statistics, which we used to infer their relative power. Of all of the summary statistics considered, summary statistics derived from a genomic SEM analysis of a common factor of the neuroticism items produced the largest mean χ^2 in the summary statistics and predicted the greatest variance in the out-of-sample phenotype (Supplementary Fig. 21). In both cases, the superior performance of genomic SEM analysis of the common factor of items relative to the sum score of the items is probably, in part, a reflection of the fact that the sum score in UKB was created using listwise deletion, resulting in a reduced sample size of 274,008. Conversely, genomic SEM uses all available information from neuroticism items, with sample sizes of ~325,000 each. In more severe cases of sample non-overlap, we would expect even larger power benefits of genomic SEM-derived summary statistics relative to individual items or sum scores. Indeed, in instances of minimal sample overlap, it is not possible to compute sum scores, but genomic SEM can still be used to integrate data across phenotypes.

Biological annotation. The biological function of the SNPs related to the *p* factor and neuroticism was examined using DEPICT[30]. Table 1 presents the number of enriched gene sets, prioritized genes, and enriched tissues and cell types across the univariate statistics and common factors (see Supplementary Tables 10–18 for detailed output). Common factors produced more informative results than the individual indicators. As expected, all of the tissue enrichment for the common factors was identified in the nervous system (Supplementary Fig. 22). Neuroticism-prioritized genes indicated a central role of synaptic activity (for example, *STX1B*, *NR4A2* and *PCLO*), including glutamatergic neurotransmission (*GRM3*). The *p* factor gene sets were largely characterized by communication between neurons (for example, 'dendrite development', 'dendritic spine' and 'abnormal excitatory postsynaptic potential'). Biological annotation of Q_{SNP} statistics for neuroticism indicated that genes within the 69 loci related to neuroticism, but not through a single factor, include: *GRIA1*, a glutamate receptor subunit (involved in signalling in excitatory neurons) that has previously been related to schizophrenia[31], chronotype[32] and autism[33]; and *PCDH17*, a gene involved in cellular connections in the brain that has been related to intelligence[34].

General guidelines. When implementing genomic SEM, users should be aware of the limitations and assumptions of the method. First, because genomic SEM is a method for modelling genetic covariance matrices, it relies on the same assumptions as the method used to estimate genetic covariances, and best practices for implementing such a method should be followed. For example, when LDSC is used to construct the genetic covariance matrix, SNPs should not first be pruned for linkage disequilibrium, and summary statistics for different phenotypes should be obtained from ethnically homogeneous samples of similar ancestral backgrounds[4]. With respect to selecting between competing models, users should take into account a variety of both absolute fit indices (for example, SRMR and model χ^2) and relative fit indices (for example, AIC and χ^2 difference). We provide general standards for absolute model fit in the Methods. Finally, a formal power analysis should take into account specific characteristics of the summary data, genetic architecture of the phenotypes and model to be specified. This can typically be achieved with simulation. Generally speaking, we would expect power to detect SNP effects on a common genetic factor

Table 1 | Summary of multivariate (genomic SEM) and univariate GWAS results

	Lead SNPs ($P<5\times10^{-8}$)	Q_{SNP} hits[a]	Unique hits[b]	Number of gene sets	Number of prioritized genes	Number of tissues and cells	Mean χ^2
p factor							
Genomic SEM (WLS)	128	1 (1)	27	71	37	24	1.88
Schizophrenia	127	-	34 (0)	2	25	21	1.82
Bipolar	4	-	4 (0)	0	0	0	1.15
MDD	5	-	5 (0)	0	0	0	1.31
PTSD	0	-	0 (0)	0	0	0	1.01
Anxiety	1	-	1 (0)	0	0	0	1.03
Neuroticism							
Genomic SEM (WLS)	118	69 (5)	38	1	19	20	1.64
Mood	43	-	19 (5)	0	0	15	1.37
Misery	31	-	6 (4)	0	0	0	1.32
Irritability	36	-	17 (4)	0	0	0	1.37
Hurt feelings	24	-	11 (0)	0	0	0	1.33
Fed-up	38	-	21 (6)	0	0	0	1.36
Nervous	41	-	25 (12)	0	0	0	1.36
Worry	56	-	26 (6)	0	13	0	1.46
Tense	19	-	10 (3)	0	0	0	1.32
Embarrass	17	-	6 (2)	0	0	0	1.33
Nerves	12	-	7 (3)	0	0	0	1.26
Lonely	6	-	4 (3)	0	0	0	1.19
Guilt	21	-	8 (1)	0	0	0	1.28

[a]Numbers in parentheses for Q_{SNP} report how many Q_{SNP} hits were in linkage disequilibrium with hits identified as significant for the common factor. [b]Unique hits for the common factor refers to lead SNPs that were not in linkage disequilibrium with hits for the individual indicators. Unique hits for the individual indicators refer to hits for the respective indicator that were not in linkage disequilibrium with hits for the common factor. Unique hits for the common factor excluded hits in linkage disequilibrium with Q_{SNP} hits. For unique hits for indicators, values in parentheses indicate whether any of these hits were identified as significant for Q_{SNP}. For unique hits for the common factor, hits were excluded that were in linkage disequilibrium with previously reported indicator hits that were removed due to missing values across the other phenotypes. The single Q_{SNP} hit for WLS estimation of the *p* factor was significant for both the common factor and schizophrenia. For the common factor and the indicators, independent hits were defined using a pruning window of 500 kb and $r^2>0.1$. For chromosomes 6 and 8, an additional pruning filter of 1 Mb and $r^2>0.1$ was used to account for long-range linkage disequilibrium due to the MHC region and pericentric inversion, respectively. For univariate statistics, we used only the SNPs present across all indicators to facilitate a direct comparison with genomic SEM results.

to increase when the constituent univariate GWASs have larger sample sizes, higher heritabilities, higher genetic correlations with one another and lower sample overlap with one another. That said, we still expect some power benefits relative to univariate GWASs when the constituent phenotypes are only moderately heritable and/or moderately genetically correlated and/or sample overlap is high. The choice of included summary statistics, phenotypes and model(s) will of course depend on the researcher's objectives and the model(s) to be specified.

Discussion

Applications of genome-wide methods to data from large-scale population-based samples have uncovered clear evidence of pervasive statistical pleiotropy. Genomic SEM is a method for modelling the multivariate genetic architecture of constellations of genetically correlated traits and incorporating genetic covariance structure into multivariate GWAS discovery. In contrast with methods[9] that model phenotypic rather than genetic covariance structure, and rely on raw data, genomic SEM employs summary GWAS data to model genetic covariance structure. Genomic SEM is computationally efficient, accounts for potentially unknown degrees of sample overlap and allows for flexible specification of covariance structure, such that several broad classes of structured covariance models can be applied. The genomic SEM approach shares benefits of some existing approaches[11] for boosting power by combining information across genetically correlated phenotypes. However, genomic SEM uniquely allows one to compare different hypothesized genetic covariance architectures and to incorporate such architectures into multivariate discovery. Importantly, shared genetic liabilities across phenotypes can be explicitly modelled as factors that may be treated as broad genetic risk factors with equally broad downstream consequences. Multivariate genetic methods have existed for decades in the twin literature, with Martin and Eaves[35] providing a framework for fitting structural equation models of genetic and environmental variance components to multivariate twin data. Using GWAS summary data from unrelated individuals, genomic SEM can be used to estimate multivariate genetic models similar to those from the existing twin literature. Moreover, genomic SEM offers new promise as a method that allows for modelling genetic covariance even among phenotypes for which phenotypic covariance cannot be estimated.

Genomic SEM is not the first method for multivariate GWAS. Other methods, such as MTAG[11], SHom/SHet[36], metaUSTAT[37], min-P[38] and TATES[27] allow researchers to perform multivariate meta-analyses based solely on summary data. The methods can generally be divided into two distinct classes: methods that aggregate test statistics or effect sizes based on a model (genomic SEM, SHom and MTAG) and those that select from the univariate *P* values while taking care not to inflate type-I error (min-P, TATES and SHet). As we show with respect to MTAG, models on which existing methods are based may can be fit within the genomic SEM framework. We also anticipate that the approaches for selecting the *P* values from a set of analyses while maintaining proper type-I error control could be integrated into the genomic SEM framework. For instance, whereas TATES is currently applied to select *P* values from

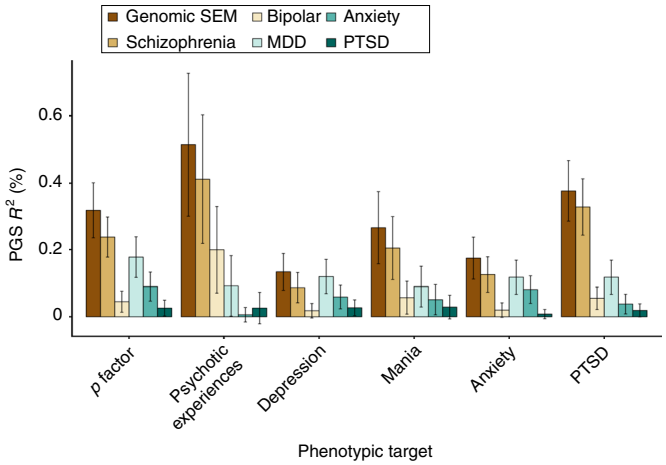

Fig. 3 | Out-of-sample prediction using genomic SEM- and univariate-based PGSs for psychiatric traits. PGSs were constructed using the same set of SNPs for all predictors. R^2 on the *y* axis indicates the percentage of variance (possible range: 0–100) explained in the outcome unique of covariates. The summary statistics for genomic SEM were estimated using WLS. The genomic SEM-based PGS was derived from a model estimating SNP effects on a common *p* factor, constructed from schizophrenia, bipolar, MDD, PTSD and anxiety (as in Fig. 1a). To prevent bias, the genomic SEM summary statistics were produced using schizophrenia and MDD GWAS summary statistics that did not include UKB participants. Error bars indicate 95% confidence intervals estimated using the delta method. Phenotypes were constructed for European participants in the UKB for five symptom domains and for a general *p* factor spanning all five symptom domains.

a series of univariate analyses of correlated traits, the same analysis could be used to select *P* values from a series of genomic SEM models. The multivariate methods available need not be mutually exclusive. With respect to other multivariate analyses of genome-wide data that go beyond multivariate GWAS discovery, the major alternatives to genomic SEM that we are aware of are GWIS[23] and GW-SEM[9]. When considering linear relationships between traits, genomic SEM is more flexible and user friendly than GWIS, and GW-SEM requires access to phenotypic data, which is a substantial limitation for many applications.

Unlike approaches that assume homogeneity of effects across SNPs[11], genomic SEM includes diagnostic indices for its key assumptions, including a test for heterogeneity, Q_{SNP}, that can be applied at the level of the individual SNPs. This offers the unique ability to identify SNPs that confer specific risk to individual phenotypes. This question may be of particular interest as the large degrees of genetic overlap identified across phenotypes (for example, bipolar disorder and schizophrenia) beg the question: what are the genetic causes of phenotypic divergence? Whereas previous GWASs have combined items tapping genetically related phenotypes into a single score, or even combined cases with different diagnoses to obtain a shared genetic effect, genomic SEM allows researchers to interrogate shared genetic effects between diagnoses or indicators, while concurrently testing for causes of divergence (that is, loci that are related only to a specific phenotype, or subset of phenotypes, but not the more general liability). In the context of neuroticism, for example, we identified 69 loci that were significantly involved in one manifestation of neuroticism but whose effects were not shared through a common factor, offering novel evidence of biological heterogeneity in the aetiology of a construct long thought to be unidimensional. Because genomic SEM relies only on GWAS summary data, it can be applied to a broad spectrum of traits, including social, economic, cognitive and psychiatric outcomes.

Methods

Overview of genomic SEM. Genomic SEM is a two-stage structural equation modelling approach[12–14]. In the first stage, the empirical genetic covariance matrix and its sampling covariance matrix are estimated. In principle, these matrices may be obtained using a variety of methods for estimating SNP heritabilities, genetic covariances and their joint estimation errors. Here, we use a novel version of LDSC that accounts for potentially unknown degrees of sample overlap by populating the off-diagonal elements of the sampling covariance matrix. The same strengths, as well as assumptions and limitations, that are known to apply to LDSC[39,40] apply to its extension used here and to genomic SEM. In stage 2, the user specifies a multivariate system of regression and covariance associations involving the genetic components of phenotypes with one another and/or more general latent factors. These associations are represented by parameters that may be fixed or freely estimated, as long as the model is statistically identified (for example, the number of freely estimated parameters does not exceed the number of non-redundant elements in the genetic covariance matrix being modelled). A set of parameters (θ) is estimated such that the fit function indexing the discrepancy between the model-implied covariance matrix, $\sum(\theta)$, and the empirical covariance matrix, *S*, estimated in stage 1 is minimized. Model fit is considered good when $\sum(\theta)$ closely approximates *S*. In the main text, we highlight the results from weighted least squares (WLS) estimation that weights the discrepancy function using the inverse of the diagonal elements of the sampling covariance matrix, and produces model standard errors using the full sampling covariance matrix. In the Supplementary Results, we additionally report the results from an alternative normal theory maximum likelihood estimation method.

Form of structured covariance models. Genomic SEM provides substantial user flexibility with respect to the particular SEM that is specified to produce the model-implied covariance matrix $\sum(\theta)$ that approximates the empirical covariance matrix, *S*. SEMs can be partitioned into two sets of equations—one describing the measurement model and the other describing the structural model. In the measurement model, the genetic components of *k* 'indicator' phenotypes are described as linear functions of a smaller set of *m* (continuous) latent variables, $\mathbf{y} = \Lambda\boldsymbol{\eta} + \boldsymbol{\varepsilon}$. In this equation, $\mathbf{y}$ is a $k \times 1$ vector of indicators, $\boldsymbol{\varepsilon}$ is a $k \times 1$ vector of residuals, $\boldsymbol{\eta}$ is an $m \times 1$ vector of latent variables and Λ is a $k \times m$ matrix of factor loadings (that is, regressions relating the latent variables to the set of indicators). In a typical application of genomic SEM, each indicator is a function of exactly one of the latent variables (although this so-called 'simple structure' restriction may be relaxed). In a CFA model, only the measurement model is specified, and the set of latent variables are allowed to freely co-vary. Thus, the model-implied covariance matrix of a CFA is $\Sigma(\theta) = \Lambda\Psi\Lambda' + \Theta$, where Ψ is an $m \times m$ latent variable covariance matrix and Θ is a $k \times k$ matrix of covariances among the residuals, $\boldsymbol{\varepsilon}$. Typically, Θ is diagonal, which implies that indicators are mutually independent conditional on the set of latent variables. That constraint may be relaxed such that select pairs of indicators are allowed to co-vary over and above their associations via the latent variable structure (that is, residual covariances are allowed). CFA models are typically used to assess the strength of relations between sets of indicators and their respective underlying latent variables, as well as to assess the fit of a measurement model to data. A well-fitting CFA model implies that the latent variable structure is able to account for the observed covariances among a set of indicator variables.

When a theory aims to explain associations among latent variables, a structural model can be added to the measurement model to produce a full SEM. The structural model of a SEM relates latent variables to each other via directed regression coefficients. It can be written in matrix notation as $\boldsymbol{\eta} = B\boldsymbol{\eta} + \boldsymbol{\zeta}$, where *B* is an $m \times m$ matrix of regression coefficients that relate latent variables to each other, and $\boldsymbol{\zeta}$ is an $m \times 1$ vector of latent variable residuals. The model-implied covariance matrix of observed variables is $\Sigma(\theta) = \Lambda(I-B)^{-1}\Psi[(I-B)^{-1}]'\Lambda' + \Theta$, where *I* is an $k \times k$ identity matrix[41]. Thus, in a full SEM, the empirical matrix is represented by a set of parameters that relate observed variables to latent variables, and latent variables to each other in a series of linear equations.

Path diagrams. SEMs can be represented graphically as path diagrams representing regression and covariance relations among variables[42]. In path diagrams, observed variables are represented as squares and unobserved (that is, latent) variables are represented as circles. Regression relationships between variables are represented as one-headed arrows pointing from the independent variable to the dependent variable. Covariance relationships between variables are represented as two-headed arrows linking the two variables. The variance of a variable (that is, the covariance between a variable and itself) is represented as a two-headed arrow connecting the variable to itself. In genomic SEM, we represent the genetic component of each phenotype with a circle, as the genetic component is a latent variable that is not directly measured, but is inferred from LDSC (it is the phenotype itself that is observed in the raw data that is used to produce the summary statistics). SNPs are directly measured, and are therefore represented as squares. When all elements in a SEM are represented in a path diagram, the diagram contains the full system of algebraic equations needed to estimate the full set of SEM parameters, θ, and to produce the model-implied covariance matrix, $\sum(\theta)$.

Stage 1 estimation. In stage 1, the empirical genetic covariance matrix (S_{LDSC}) and its associated sampling covariance matrix ($V_{S_{LDSC}}$) are estimated using our multivariable extension of LDSC. S_{LDSC} is a $k \times k$ symmetric matrix with SNP heritabilities on the diagonal and genetic covariances ($\sigma_{gi,gj}$) between phenotypes i and j off the diagonal. The genetic covariance between phenotypes i and j can be computed as the genetic correlation scaled relative to the total genetic variance of each of the two contributing phenotypes (themselves scaled to unit variances), $\sigma_{gi,gj} = r_{gi,gj} \cdot \sqrt{h_i^2 \cdot h_j^2}$. Thus, the genetic covariance matrix of order k has $k^* = k(k+1)/2$ non-redundant elements. It can be written as:

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

To produce unbiased standard error estimates and test statistics, we require the sampling covariance matrix, $V_{S_{LDSC}}$, of the LDSC estimates that is composed of all non-redundant elements in the S_{LDSC} matrix. Thus, it is a symmetric matrix of order k^*, with $k^*(k^*+1)/2$ non-redundant elements. The diagonal elements of $V_{S_{LDSC}}$ are sampling variances (that is, squared standard errors of the elements in S_{LDSC}). The off-diagonal elements of $V_{S_{LDSC}}$ are sampling covariances that indicate the extent to which the sampling distributions of the variance and covariance estimates in S_{LDSC} co-vary with one another, as would be expected when there is overlap among the samples from which the terms are estimated. This $V_{S_{LDSC}}$ matrix can be written as:

$$V_{S_{LDSC}} = \begin{bmatrix} \text{s.e.}(h_1^2)^2 & & & & & & \\ \text{cov}(h_1^2, \sigma_{g1,g2}) & \text{s.e.}(\sigma_{g1,g2})^2 & & & & & \\ \vdots & \vdots & \ddots & & & & \\ \text{cov}(h_1^2, \sigma_{g1,gk}) & \text{cov}(\sigma_{g1,g2}, \sigma_{g1,gk}) & \text{s.e.}(\sigma_{g1,gk})^2 & & & & \\ \vdots & \vdots & \vdots & \ddots & & & \\ \text{cov}(h_1^2, h_j^2) & \text{cov}(\sigma_{g1,g2}, h_j^2) & \text{cov}(\sigma_{g1,gk}, h_j^2) & \text{s.e.}(h_j^2)^2 & & & \\ \vdots & \vdots & \vdots & & \ddots & & \\ \text{cov}(h_1^2, \sigma_{gj,gk}) & \text{cov}(\sigma_{g1,g2}, \sigma_{gj,gk}) & \text{cov}(\sigma_{g1,gk}, \sigma_{gj,gk}) & \text{cov}(h_j^2, \sigma_{gj,gk}) & \text{s.e.}(\sigma_{gj,gk})^2 & & \\ \text{cov}(h_1^2, h_k^2) & \text{cov}(\sigma_{g1,g2}, h_k^2) & \text{cov}(\sigma_{g1,gk}, h_k^2) & \text{cov}(h_j^2, h_k^2) & \text{cov}(\sigma_{gj,gk}, h_k^2) & \text{s.e.}(h_k^2)^2 \end{bmatrix}$$

The diagonal elements of $V_{S_{LDSC}}$ can be estimated using the jackknife resampling procedure in the bivariate version of LDSC that is currently available from its original developers[4,43]. The LDSC function introduced in the GenomicSEM software package expands the jackknife procedure to the multivariable context to additionally produce sampling covariances (which index dependencies among estimation errors) among the elements of S_{LDSC} needed to populate the off-diagonal elements of $V_{S_{LDSC}}$.

Incorporating individual SNP effects. Several steps are needed to incorporate individual SNP effects into genomic SEM. The first step requires that the inputted genetic covariance matrix be expanded to include covariances between the SNP and each of the phenotypes, g_1 through g_k, by appending a vector of SNP phenotype covariances (S_{SNP}) to S_{LDSC}:

$$S_{Full} = \begin{bmatrix} \sigma_{SNP}^2 & & & & & \\ \sigma_{SNP,g1} & h_1^2 & & & & \\ \sigma_{SNP,g2} & \sigma_{g1,g2} & h_2^2 & & & \\ \sigma_{SNP,g3} & \sigma_{g1,g3} & \sigma_{g2,g3} & h_3^2 & & \\ \vdots & \vdots & & & \ddots & \\ \sigma_{SNP,gk} & \sigma_{g1,gk} & \sigma_{g2,gk} & \sigma_{g3,gk} & \cdots & h_k^2 \end{bmatrix}$$

The sampling covariance matrix, $V_{S_{Full}}$, associated with this expanded S_{Full} covariance matrix, includes a number of components. One block of this $V_{S_{Full}}$ matrix, $V_{S_{LDSC}}$, contains the sampling variances and sampling covariances of the latent genetic variances (SNP heritabilities) and genetic covariances, which are obtained from the multivariable LDSC approach introduced above. A second block of the $V_{S_{Full}}$ matrix, $V_{S_{SNP}}$, is composed of the sampling covariance matrix of the SNP effects on the phenotypes. The SNP variance (derived from reference panel data) is treated as fixed, and its sampling variance and sampling covariance with all other terms are fixed to 0 (or to a very small value to facilitate computational tractability). The sampling covariances of the SNP genotype covariances with one another are obtained using cross-trait LDSC intercepts (which represent sampling correlations weighted by sample overlap) after being rescaled relative to the sampling variances of the respective SNP genotype covariances[11,44]. A final block of the $V_{S_{Full}}$ matrix represents the sampling covariance of the SNP genotype covariances with the genetic variances and genetic covariances. These are fixed to 0, as sampling variation of the SNP genotype covariance is expected to be independent of the test statistics of all linkage disequilibrium blocks except the one it occupies. Because the sampling variance of the heritabilities and genetic correlations derive from sampling variability in the test statistics within all of the linkage disequilibrium blocks, their sampling covariances with a single SNP effect is expected to approach 0. The $V_{S_{Full}}$ matrix can be written in compact form as:

$$V_{S_{Full}} = \begin{bmatrix} V_{S_{SNP}} & \\ 0 & V_{S_{LDSC}} \end{bmatrix}$$

Stage 2 estimation. In stage 2, the genetic covariance matrix obtained in the previous stage, S, is used to estimate the parameters in a SEM. In this stage, we allow for both WLS and normal theory maximum likelihood estimators. WLS does not strictly require positive definite S and V_S matrices, but may still benefit from positive definiteness during optimization. Maximum likelihood estimation requires both S and V_S to be positive definite. The GenomicSEM software package therefore smooths S and V_S to the nearest positive definite matrices before stage 2 estimation using the R function nearPD[45].

The fit function minimized in the diagonally weighted version of WLS estimation that is standard in the GenomicSEM software package is the following:

$$F_{WLS}(\theta) = (s - \sigma(\theta))' D_S^{-1} (s - \sigma(\theta))$$

where S and $\Sigma(\theta)$ have been half-vectorized to produce s and $\sigma(\theta)$, respectively, and D_S is V_S with its off-diagonal elements set to 0. We choose the diagonally weighted version of WLS because it is more tractable to implement for large (highly multivariate) matrices and is more stable than fully weighted WLS in finite samples[46,47].

Maximum likelihood estimation proceeds by minimizing the following fit function:

$$F_{ML}(\theta) = \log |\Sigma(\theta)| - \log |S| + tr\{S\Sigma^{-1}(\theta)\} - k$$

where $\Sigma(\theta)$ is the covariance matrix implied by the set of parameter estimates. Note that, while the formulation of the maximum likelihood fit function does not explicitly include a weight matrix, it is asymptotically equivalent to a more general formulation that is identical to the WLS fit function, with $.5D'_k(\Sigma^{-1}(\theta) \otimes \Sigma^{-1}(\theta))D_k$, where D_k is the duplication matrix of order k, in place of D_S. Thus, the difference between maximum likelihood and WLS estimation can be construed as a difference in weight matrices only. A comparison between maximum likelihood and WLS results can be found in the Supplementary Results (see also Supplementary Figs. 23–27 and Supplementary Table 19).

WLS estimation more heavily prioritizes reducing misfit in those cells in the S matrix that are estimated with greater precision. This has the desirable property of potentially decreasing sampling variance of the genomic SEM parameter estimates, which may boost power for SNP discovery and increase polygenic prediction. However, because the precision of cells in the S matrix is contingent on the sample sizes for the contributing univariate GWASs, WLS may produce a solution that is dominated by the patterns of association involving the most well-powered GWASs, and contain substantial local misfit in cells of S that are informed by lower-powered GWASs. In other words, WLS relative to maximum likelihood may more heavily prioritize minimizing sampling variance of the parameter estimates in the so-called variance bias tradeoff[48]. We expect that this will only occur when the model is overidentified (that is, d.f. > 0), such that exact fit cannot be obtained, and that divergence in WLS and maximum likelihood estimates will be most pronounced when there is lower sample overlap and the contributing univariate GWASs differ substantially in power. Maximum likelihood estimation may be preferred when the goal is to most evenly weight the contribution of the univariate sample statistics.

Both WLS and maximum likelihood fit functions will produce consistent estimates of the model parameters when the model is true[47]. However, the 'naive' standard errors and fit statistic produced in stage 2 estimation will be incorrect, because neither estimator uses the full V_S matrix in estimation. Thus, robust corrections must be applied to produce consistent estimates of standard errors and test statistics. The correct sampling covariance matrix of the stage 2, genomic SEM parameter estimates (that is, V_θ) can be obtained using a sandwich correction[13,47]:

$$V_\theta = (\hat{\Delta}' \Gamma^{-1} \hat{\Delta})^{-1} \hat{\Delta}' \Gamma^{-1} V_S \Gamma^{-1} \hat{\Delta} (\hat{\Delta}' \Gamma^{-1} \hat{\Delta})^{-1}$$

where $\hat{\Delta} = \frac{\partial(\theta)}{\partial \theta'}\Big|_{\hat{\theta}}$ is the matrix of model derivatives evaluated at the parameter estimates, Γ is the naive stage 2 weight matrix that takes its form depending on the estimation method used (WLS or maximum likelihood), and V_S is the sampling covariance matrix of S obtained using multivariable LDSC.

It may not always be possible to obtain the full sampling covariance matrix, V_S. For example, for highly sensitive data, only the matrix S and the standard errors

of its elements may be available (that is, the diagonal of V_S). However, we note that when there is low sample overlap across the GWASs for each phenotype, off-diagonal elements of the sampling covariance matrix are small and pragmatically ignorable. Moreover, in other contexts with complete sample overlap, standard error inflation of the SEM parameters estimated using diagonally weighted versions of WLS has been estimated to be less than 8%[9] without robustness corrections, and 0% with robustness corrections[47].

Standardization and scaling of summary statistics for multivariate GWASs. Typically, GWAS summary statistics for quantitative phenotypes are not reported in terms of covariances, but are reported as ordinary least squared (OLS) unstandardized regression coefficients, with the phenotypes standardized before analyses (that is, the coefficients are standardized with respect to the outcome, but not the predictor). To transform the partially standardized regression coefficient ($b_{SNP,P}$) of a SNP effect on phenotype P to a covariance, we multiply by the variance of scores on the SNP. The variance (σ^2_{SNP}) of scores (0, 1, 2) of a biallelic autosomal SNP is estimated as $2pq$, assuming Hardy–Weinberg equilibrium, where p is the minor allele frequency (MAF) and $q = 1 - \text{MAF}$, with the MAF typically obtained from a reference sample. As the latent genetic factors estimated in LDSC are scaled relative to unit-variance-scaled phenotypes (by virtue of the SNP heritability estimates being placed on the diagonal of S), no further scaling is needed to transform this SNP phenotype covariance into a SNP genotype covariance.

When OLS regression coefficients and standard errors are provided from an analysis in which the phenotype has not been standardized before analyses, or only Z statistics or P values (for which Z statistics can be readily obtained) are provided, the partially standardized regression coefficients and their standard errors can be obtained as $Z = \frac{b^*_{SNP,P}}{\text{s.e.}_{b^*_{SNP,P}}}$, $b_{SNP,P} = \frac{Z}{\sqrt{N\sigma^2_{SNP}}}$ and $\text{s.e.}_{b_{SNP,P}} = \frac{b_{SNP,P}}{Z}$, where $b^*_{SNP,P}$ is equal to the regression coefficient for the OLS GWAS of the unstandardized phenotype. These derived partially standardized coefficients are then transformed into covariances by multiplying by the variance of scores on the SNP, as above.

When the GWAS summary statistics are reported for logistic regressions of liabilities for categorical outcomes (for example, case/control status) on the SNP, the logistic regression coefficients can be transformed into covariances as above, by multiplying by the SNP variances. However, it is appropriate to further transform the coefficients and their standard errors such that they are scaled relative to unit-variance-scaled liability. This can be achieved by dividing by $\sqrt{\sigma^2_{SNP} \times b\text{logit}^2_{SNP,P} + \frac{\pi^2}{3}}$, as a logistic regression model implies a residual variance of $\frac{\pi^2}{3}$. If GWAS summary statistics are reported for odds ratios, they can be transformed to logistic regression coefficients by taking their natural logarithm. Standard errors for the logistic regression coefficient are obtained as $\text{s.e.}_{OR}/\text{OR}$, where OR is the odds ratio. The derived logistic coefficients and their standard errors should further be transformed such that they are scaled relative to unit-variance-scaled phenotypes, as above. Note that when the outcomes are categorical, the liability scale heritabilities and genetic covariances from multivariable LDSC (and not what are referred to as the 'observed scale' heritabilities and genetic covariances) should be used to populate the S matrix. This has the desirable property of both modelling the continuous scale of risk in the population and providing estimates that are independent of the observed prevalence of the categorical outcomes.

On occasion, summary statistics will be provided from OLS GWASs of categorical outcomes (for example, case/control status). Such an analysis is sometimes referred to as a linear probability model, as it (incorrectly) assumes that the association between the predictor and the probability of being in the comparison (for example, case) group relative to the reference (for example, control) group is linear. Parameters from the linear probability model are dependent not only on the strength of the association between the SNP and the continuous underlying liability, but also on the MAF and the proportion of comparison group members (cases) in the sample. Thus, parameters from the linear probability model cannot be used directly in genomic SEM. However, particularly in the case of complex traits, for which the effect sizes for individual SNPs are small, results from the linear probability model can be used to very closely approximate logistic regression coefficients and standard errors that are amenable for use in genomic SEM[49]. This approximation can be obtained as $Z = \frac{b^{**}_{SNP,P}}{\text{s.e.}_{b^{**}_{SNP,P}}}$, $b\text{logit}^*_{SNP,P} = \frac{Z}{\sqrt{v(1-v)N\sigma^2_{SNP}}}$ and $\text{s.e.}_{b_{SNP,P}} = \frac{b\text{logit}^*_{SNP,P}}{Z}$, where $b^{**}_{SNP,P}$ is equal to the regression coefficient from the linear probability model, $b\text{logit}^*_{SNP,P}$ is the expected logistic regression coefficient that is derived from the linear probability model results, v is equal to the proportion of cases in the sample, and σ^2_{SNP} is the variance of the SNP, computed from its MAF obtained from a reference sample, as above. To scale the derived logistic coefficient such that it is scaled relative to unit-variance-scaled liability, the coefficient should be divided by $\sqrt{\sigma^2_{SNP} \times (b\text{logit}^*_{SNP,P})^2 + \frac{\pi^2}{3}}$. Lloyd-Jones et al.[49] report that in a real data analysis of UKB data, the exponentiated regression coefficient (that is, the odds ratio) obtained directly from a logistic regression-based GWAS and that derived from the linear probability model-based GWAS was nearly perfect ($R^2 > 98\%$; slope: ~1). We have verified this nearly perfect correspondence in our own simulations (Supplementary Fig. 28).

Even within samples of the same ethnicity, there are likely to be discrepancies between the MAFs of a reference sample and the sample that GWAS summary statistics were generated from. However, some summary statistics may not include allele frequencies, and using the same reference panel for standardization across phenotypes has the desirable property of maintaining consistency across summary statistics. To examine the effect of this decision, the betas for 30,000 randomly selected SNPs for the mood phenotype from UKB were standardized using either the sample or reference panel MAF. The correlation between the betas was 0.982, and a linear regression of betas, standardized using reference panel MAF, predicting standardization using sample MAFs, revealed near-perfect correspondence (slope = 1.044; intercept = -6.54×10^{-6}; Supplementary Fig. 29).

Model fit statistics. Model χ^2 is an index of the exact fit of a SEM. It indexes whether the model-implied genetic covariance matrix, $\Sigma(\theta)$, differs from the empirical genetic covariance matrix, S. Model χ^2 can also be used as a relative fit index for comparing nested models. Conventional SEM approaches to indexing model χ^2 are based on formulas that directly incorporate sample size (N). Because there is not an N that directly corresponds to the genetic covariance matrix that is modelled by genomic SEM in the same way that N typically corresponds to an observed covariance matrix, we derived a formula for estimating model χ^2 that does not require N, but instead incorporates the sampling covariance matrix of the model residuals. This is done in two steps. In step 1, the proposed model (for example, a common factor model) is estimated. In step 2, all of the step 1 estimates are fixed, and the residual covariances and residual variances of the indicators are freely estimated. Residual variances are estimated in step 2 by estimating the variances of k residual factors defined by the indicators. This provides an estimate of the discrepancy between the model-implied and observed covariance matrices, $R = S - \Sigma(\theta)$, along with the sampling covariance matrix (V_R) of R. While the discrepancy between model-implied and observed covariance matrices can be computed simply by deriving covariance expectations from the step 1 model and subtracting the observed covariance matrix, such an approach would not provide the corresponding V_R matrix necessary for the calculations below. The V_R matrix is expected to be positive semidefinite and, consequently, have no negative eigenvalues. Therefore, the V_R matrix has the following eigendecomposition:

$$V_R = \begin{pmatrix} P_1 & P_0 \end{pmatrix} \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} P_1' \\ P_o' \end{pmatrix}$$

where P_1 is a matrix of principal components (eigenvectors) of V_R, and E is a corresponding diagonal matrix consisting of non-zero eigenvalues. P_0 reflects the null space of V_R. Projecting R_i—a vector of residual covariances estimated from the step 2 model—onto P_1 and adjusting for corresponding eigenvalues, we have:

$$E^{\frac{-1}{2}} P_1' R_i \sim N(0, I_r)$$

Therefore:

$$R_i' P_1 E^{-1} P_1' R_i \sim \chi^2(r)$$

This equation produces a test statistic that is χ^2 distributed with degrees of freedom (r) equal to the difference between the number of non-redundant elements (k^*) in the empirical covariance matrix (S) and the number of freely estimated parameters in the proposed model.

The CFI is a test of approximate model fit. CFI indexes the extent to which the proposed model fits better than a model that allows all phenotypes to be heritable, but assumes that they are genetically uncorrelated. The χ^2 statistic can be used to calculate CFI by calculating a second χ^2 statistic for a so-called independence model (that is, a model that estimates genetic variances of all phenotypes but assumes all genetic covariances to be zero, such that $\sum(\theta)$ is diagonal). CFI is calculated using the formula below[50], with $f = \chi^2 - \text{d.f.}$:

$$\frac{f(\text{independence model}) - f(\text{proposed model})}{f(\text{independence model})}$$

For the χ^2 value of the independence model, a model is estimated in step 1 that includes only the variance of the indicators and no common factor. In step 2, these variances are fixed and the covariances among the indicators and variances of k residual factors defined by the indicators are estimated and used to populate the same equation above used to calculate the proposed model χ^2. CFI values theoretically range from 0 to 1, with higher values indicating good fit. CFI values of ≥0.90 are typically considered acceptable fit, and values of ≥0.95 are typically considered good model fit[51]. When the empirical covariance matrix contains a large number of cells that are very close to 0, CFI values may be low, even when such cells are approximated well by the model.

AIC is a relative fit index that balances fit with parsimony, and can be used to compare models regardless of whether they are nested. AIC is calculated as:

$$\mathrm{AIC} = \chi^2 + 2 \times fp$$

where fp is the number of free parameters in the model[52]. Lower AIC values are considered superior.

SRMR is an index of approximate model fit that is calculated as the standardized root mean squared difference between the model-implied and observed correlations in $\Sigma(\theta)$ and S, respectively[53]. Higher SRMR values indicate a larger discrepancy between $\Sigma(\theta)$ and S. It is positively biased, with larger bias resulting when the contributing univariate GWAS samples are lower powered. SRMR values below 0.10 indicate acceptable fit, values less than 0.05 indicate good fit, and a value of 0 indicates perfect fit[54].

We recommend that model-fit indices be considered concurrently, as individual indices each have their own strengths and limitations. Model χ^2 is an index of exact fit, with lower values indicating better fit. Model χ^2 may oftentimes be statistically significant, indicating that the model-implied genetic covariance matrix significantly differs from the empirical (unrestricted) genetic covariance matrix, even when the model-implied covariance matrix very closely approximates the empirical genetic covariance matrix. Oftentimes, models that closely (albeit imperfectly) approximate the empirical genetic covariance matrix may be scientifically and inferentially useful. We thus recommend considering CFI and SRMR indices of absolute fit, even when model χ^2 is significant. We also recommend using indices of relative fit to compare competing models of the same data (that is, different models fit to genetic covariance matrices derived from the exact same summary data for the exact same phenotypes). When models are nested, their respective χ^2 values can be subtracted from one another to calculate a χ^2 difference test, with d.f. equal to the difference in d.f. between the two models. This χ^2 difference test indexes the extent to which the less complex model (that is, the model with more d.f.) approximates the empirical genetic covariance matrix significantly worse than the more complex model (that is, the model with fewer d.f.). If the χ^2 difference test is significant, the more complex model should be chosen. If the χ^2 difference test is not significant, the less complex model should be chosen, as it is more parsimonious and approximates the empirical genetic covariance matrix no worse than the more complex model. Two models are nested when the set of possible model-implied covariance matrices from one model is a subset of the set of possible model-implied covariance matrices of the second model[55]. Nesting can typically be confirmed if the less restrictive model can be derived from the more restrictive model by dropping or fixing parameters. Regardless of whether models are nested, they can be compared on CFI, SRMR and AIC, so long as the same data are being modelled.

Q_{SNP} test of heterogeneity. As with the computation of model χ^2 outlined above, Q_{SNP} is calculated using a two-step procedure. In step 1, a common pathway model is fit in which both factor loadings, the SNP effect on the common factor(s) and the residual variances of the common and unique factors are freely estimated (with one factor loading fixed to unity for factor identification and scaling). No paths representing direct effects of the SNP on the genetic components of the individual phenotypes are estimated. In step 2, a common plus independent pathways model is specified, in which the factor loadings and the SNP effect on the common factor are fixed to the values estimated in step 1, and direct effects of the SNP on individual indicators and the residual variances of each indicator are freely estimated. Supplementary Fig. 30 depicts this model, as applied to a single common factor model, with parameters that are fixed in step 2 depicted in red and those that are freely estimated in step 2 depicted in black.

Genomic SEM simulations. *Validation of summary-based model fit statistics via simulation.* A generating population with a common factor model defined by 4, 5 or 6 indicators was used to examine the null distribution of the newly derived χ^2 test statistic using a set of 1,000 simulations per model. These simulations did not include individual genotypes, and were simulated solely based on a generating factor structure. For the 6-indicator models, the standardized factor loadings in the generating population were 0.42, 0.64, 0.22, 0.59, 0.19 and 0.64. The four- and five-indicator models specified the same factor loadings, excluding the last, or last two loadings, respectively. The results indicated that the two-step procedure described above produced a test statistic equivalent to the χ^2 statistic calculated by lavaan from the raw data (Supplementary Fig. 31 and Supplementary Table 20). For a χ^2-distributed test statistic, the mean of the null sampling distribution should match the d.f. of the test. As expected, the distribution of the test statistic conformed to a χ^2 distribution with an average approaching the d.f. (Supplementary Fig. 32). Calculated CFI values were also highly consistent with those observed using the CFI statistic provided by lavaan when using raw data (Supplementary Fig. 33 and Supplementary Table 20). Calculated AIC values were not contrasted with those obtained using the lavaan package in R in the simulations below, as the software uses a formula that includes a log-likelihood estimate contingent on the provided sample size.

Null distribution of Q_{SNP}. To verify that the null distribution for Q_{SNP} is χ^2 distributed, a set of simulations specified a generating population in which the direct effects of the SNP on the indicators were entirely mediated through the common factor. Each simulation included 1,000 datasets, with $N = 100{,}000$ completely overlapping participants per dataset. All simulated datasets were analysed using both WLS and maximum likelihood. We examined 3 models with $F = 1$ factor and $k = 4$, 5 or 6 phenotypes. Supplementary Table 21 presents descriptive statistics for Q_{SNP}. Using a genome-wide significance threshold, in all cases, the false discovery rate for Q_{SNP} was 0, and the power to detect a SNP effect on the common factor was 1. Both WLS and maximum likelihood estimation produced mean estimates of Q_{SNP} that were approximately equal to the d.f. of the corresponding model. Supplementary Fig. 34 depicts the null sampling distributions of Q_{SNP} estimated using WLS or maximum likelihood. Supplementary Fig. 35 plots Q_{SNP} from these two estimation methods against χ^2 distributions and against one another. These results indicate that both estimation methods produce results that are approximately χ^2 distributed.

Simulation of factor structure. To evaluate the ability of genomic SEM to capture the genetic factor structure in the generating population, the GCTA package[3] was used to generate 100 sets of 6 independent, 100% heritable phenotypes ('orthogonal genotypes') to pair with genotypic data for 39,909 randomly selected, unrelated individuals of European descent from UKB data for the 1,209,498 SNPs present in HapMap3. The generating list of causal SNPs was set to 10,000 for all 600 genotypes, with the specific list of causal variants sampled with replacement from the 1,209,498 SNPs. One of the six orthogonal genotypes per set was designated an index of the general genetic factor, and the remaining five were designated indices of domain-specific genetic factors. All of these orthogonal genotypes were scaled to $M = 0$ and s.d. $= 1$. Five new correlated genotypes were then constructed, each as the weighted linear combination of the general genetic factor and one domain-specific genetic factor. Weights for the contribution of the general genetic factor were $\lambda_{F_{g,k}} = 0.70$, 0.60, 0.50, 0.40 and 0.30 for correlated genotypes 1–5, respectively. Weights for the domain-specific factors were $\sqrt{(1-\lambda^2_{F_{g,k}})}$. Phenotypes were then each constructed as the weighted linear combination of one of the correlated genotypes and domain-specific environmental factors (randomly sampled from a normal distribution with $M = 0$ and s.d. $= 1$). Heritabilities for phenotypes 1–5 were set to $h^2_k = 35$, 40, 50, 60 and 70%, respectively, such that the weights for the genotypes were $\sqrt{h^2_k}$ and the weights for the environmental factors were $\sqrt{(1-h^2_k)}$. We chose these figures to stabilize the properties of the distributions across simulations at 100 replications with $N = \sim 39{,}000$ each. We expect that with lower SNP h^2 values, the same patterns would hold, albeit at larger sample sizes. Each of the 500 phenotypes (100 sets of 5 phenotypes) was then analysed as a univariate GWAS in PLINK[56] to produce univariate GWAS summary statistics. Our multivariable LDSC function was then used to construct 100 sets of 5×5 genetic covariance matrices (S) and associated sampling covariance matrices (V_S), and genomic SEM was used to fit a 1-factor model to each set.

Usig this procedure, we performed 100 runs of genomic SEM on raw individual-level genotype data for which we simulated multivariate phenotypic data to conform to a single genetic factor model (a latent trait that partially causes 5 observed outcomes). Across the 100 simulations, genomic SEM estimates closely matched the parameters specified in the generating population (Supplementary Fig. 36). Model standard errors also closely matched the standard deviations of parameter estimates. We also compared fit statistics (CFI, AIC and model χ^2) for the correctly specified common factor model and two deliberately misspecified models: (1) a model in which all indicators were constrained to have the same factor loading; and (2) a model for which the loading of the third indicator was set to 0. As expected, the results indicated that the common factor model matching the generating population was favoured ≥99% of the time across the model fit indices (Supplementary Fig. 37).

Simulation of partial sample overlap. To examine the effect of sample overlap on estimates obtained from genomic SEM, the GCTA package[3] was used to generate a 50% heritable, quantitative phenotype with 30,000 causal SNPs. The phenotype was paired with genetic data from 100,000 randomly selected, unrelated individuals of European descent from UKB data for 1,209,498 HapMap3 SNPs. Three sets of 60,000 participants each were created using this same phenotype, with 40,000 participants overlapping across all 3 identical phenotypes and 20,000 participants unique to each phenotype (that is, 100,000 total participants). These three subsamples were individually analysed in PLINK[56] to produce univariate GWAS summary statistics. The multivariable LDSC function was then used to construct the genetic covariance and sampling covariance matrix using the three sets of summary statistics, and genomic SEM was used to fit a one-factor model with the SNP predicting the common factor. Two key results were verified at this stage. First, we confirmed that the standardized factor loadings on the common factor were 1 for the identical phenotypes. Second, we verified that the bivariate linkage disequilibrium score intercepts used to account for sample overlap in the sampling covariance matrix were as expected. The equation for the linkage disequilibrium score bivariate intercept is[4]: $N_s\rho/\sqrt{(N_1N_2)}$, where N_s is the sample overlap, ρ is the phenotypic correlation, N_1 is the sample size of trait 1, and N_2 is the sample size of trait 2. In this simulation, we observed bivariate intercepts of 0.67, which is as expected given sample overlap of 40,000, a phenotypic correlation of 1 and sample sizes of 60,000 (that is, $40{,}000 \times 1/\sqrt{(60{,}000 \times 60{,}000)} = 0.67$). Finally, estimates

from this multivariate GWAS were compared with estimates from the univariate GWAS in PLINK for the full set of 100,000 participants. If sample overlap is not appropriately accounted for in this example, such that data are incorrectly treated as deriving from 180,000 participants (as opposed to 100,000 total participants), we would expect the *Z* statistics for the SNP effects from genomic SEM to be upwardly biased relative to those from a univariate GWAS applied directly to the single phenotype in the 100,000 participants. We observed no such bias. A linear regression of *Z* statistics from genomic SEM (from the 3 overlapping samples of 60,000 participants each) predicting univariate GWAS *Z* statistics in the complete sample (of 100,000 participants) revealed near-perfect correspondence (unstandardized slope = 1.003; intercept = −0.003).

MTAG simulation. To evaluate the relationship between estimates from MTAG and those from a genomic SEM formulation of the MTAG model, we specified a bivariate system of heritable phenotypes, A and u. Phenotype A was constructed using the GCTA package[3], and specified to be 60% heritable and affected by a random selection of 30,000 HapMap3 SNPs. Phenotype u was constructed separately using the GCTA package, and also specified to be 60% heritable and affected by a different random selection of 30,000 HapMap3 SNPs. Both A and u were standardized ($M = 0$; s.d. = 1). Phenotype B was constructed from phenotypes A and u according to the equation $B = 0.7\,A + 0.7\,u$. This procedure resulted in 60% heritabilities for both traits A and B, with a genetic correlation of 0.7 between them. Sample sizes for phenotypes A and B were 25,000 each, with 10,000 participants contributing data for both phenotypes A and B (that is, 40% sample overlap), such that the analytic dataset was composed of 40,000 unique individuals in total. Both MTAG[11] and a genomic SEM model specified to satisfy the same moment conditions as MTAG (see Supplementary Methods) were then each run with trait A as the supporting phenotype used to boost the power for target trait B, and estimates from MTAG and genomic SEM specified to satisfy the MTAG moment conditions were compared. Results indicated near-perfect correspondence from a linear regression in which *Z* statistics from MTAG were used to predict those from a genomic SEM specified to satisfy the MTAG moment conditions (Supplementary Fig. 20; unstandardized slope = 0.999; intercept = 2.65×10^{-4}).

Quality control procedures. *LD-score regression.* For the *p*-factor, neuroticism and anthropometric traits, quality control procedures for producing the *S* and V_S matrix followed the defaults in LDSC. We recommend using these defaults for multivariable LDSC, including removing SNPs with MAF < 1%, information scores < 0.9, SNPs from the major histocompatibility complex (MHC) region, and filtering SNPs to HapMap3. Quality control procedures for the multivariable regression example mirrored those used by Nieuwboer et al.[23] for comparative purposes. More specifically, SNPs were excluded with MAFs < 0.05, as determined by the HapMap Consortium[57], and with information values less than 0.9 or greater than 1.1. SNPs were also filtered to HapMap3. The linkage disequilibrium scores used for the analyses presented were estimated from 1000 Genomes Phase 3, but restricted to HapMap3 SNPs.

Multivariate GWASs. Summary statistics are only restricted to HapMap3 SNPs for the estimation of the genetic covariance and sampling covariance matrix in linkage disequilibrium score regression, whereas all SNPs passing quality control filters are included for multivariate GWASs. To obtain summary statistics for multivariate GWASs, we recommend using quality control procedures including the removal of SNPs with MAF < 0.01 in the reference panel, and those SNPs with an imputation information (INFO) score < 0.6. MAFs were obtained for the current analyses using the 1000 Genomes Phase 3 reference panel. Using these quality control steps, 1,979,881 SNPs were present across schizophrenia, bipolar disorder, MDD, PTSD and anxiety. For neuroticism, there were 7,265,104 SNPs that were present across all phenotypes. These quality control procedures are the defaults for the processing function within the GenomicSEM package. The regression effects for the univariate indicators of the *p* factor were standardized using the procedure for logistic coefficients outlined above. Regression effects for neuroticism indicators were converted from linear probability to logistic coefficients and then standardized with respect to the variance in the outcome.

Out-of-sample prediction. *p* factor. Genomic SEM analyses that were used to produce the summary statistics for construction of PGSs for out-of-sample prediction omit the Psychiatric Genetics Consortium (PGC) MDD 2018 GWAS and schizophrenia 2018 GWAS and replace them with the PGC MDD 2013 (ref. [58]) and PGC schizophrenia 2014 (ref. [59]) GWASs to prevent overlap between discovery and target samples. This resulted in a genomic SEM-based multivariate GWAS using 930,581 SNPs. Analyses used to construct a phenotypic *p* factor for polygenic prediction in the UKB dataset were restricted to data on up to $N = 332{,}050$ European participants. The genomic SEM of the *p* factor employed case-control GWAS statistics to construct summary statistics for a general factor of liability for clinically severe levels of psychopathology as the discovery phenotype. For out-of-sample prediction, we selected a set of psychiatric symptoms (rather than diagnoses) to construct liability for general and domain-specific factors of psychiatric symptomology across the subclinical-to-clinical ranges as the target phenotypes. From the UKB dataset, we chose symptoms falling within the following domains: psychosis, mania, depression, post-traumatic stress and anxiety. We fit a confirmatory factor model (diagram shown in Supplementary Fig. 29) to the phenotypic symptom endorsements, treating them as ordered categorical variables. Analyses were run in M*plus*[60], with the target phenotypes—the *p* factor and each of the individual domains—specified as latent variables. PGS variables were specified to directly predict the latent phenotypes within the model (that is, factor score estimates were not used). To construct PGSs, we removed from both the *p*-factor and univariate summary statistics the 5 SNPs that were identified as having genome-wide significant Q_{SNP} estimates for maximum likelihood, along with SNPs that were in linkage disequilibrium with these SNPs using an r^2 threshold of 0.1 and a 500-kb window. PGSs were constructed using PRSice[61], with linkage disequilibrium clumping set to $r^2 > 0.25$ over 250-kb sliding windows. PGSs for the *p* factor were based on the WLS summary statistics produced using genomic SEM. We ran PGS analyses using a *P*-value threshold of 1.0 (that is, we used all available SNPs apart from those removed due to Q_{SNP} analyses). To maintain comparability, PGSs for the univariate summary statistics were constructed based on the same SNPs with which the PGSs for the *p* factor were constructed. In the confirmatory factor models, we included controls for age, sex, genotyping array and 40 principal components of ancestry in conjunction with the PGS predictor.

Neuroticism. The raw total on the 12-item neuroticism subtest of the Eysenck Personality Questionnaire-Revised[62] (maximum score = 12) was used as the target phenotype for out-of-sample prediction. Both genetic and neuroticism target data were available on 19,876 European participants in the Generation Scotland cohort[63]. Neuroticism scores were residualized for age, sex and 20 principal components of ancestry before examining out-of-sample prediction. PGSs were constructed using PRSice[61], with linkage disequilibrium clumping set to $r^2 > 0.25$ over 250-kb sliding windows and using a *P*-value threshold of 1.0. PGSs for neuroticism were based on the WLS summary statistics produced using genomic SEM. Regression analyses were run using the lmekin function within the coxme package in R with a random intercept to account for nesting of individuals within families.

Clumping and biological annotation. Lead SNPs for univariate indicators and the common factors were identified using the clumping algorithm in PLINK[56]. We defined linkage disequilibrium-independent SNPs using an r^2 threshold of 0.1 and a 500-kb window using the same 1000 Genomes Phase 3 reference panel used for obtaining MAF. For chromosomes 6 and 8, an additional pruning filter was used of 1 Mb and $r^2 > 0.1$ to account for long-range linkage disequilibrium due to the MHC region and pericentric inversion, respectively. Increasing the pruning window further to 4 Mb did not influence our findings on chromosome 6 or 8. The lead SNPs identified using PLINK were entered into DEPICT. Prioritized genes, enriched gene sets and enriched tissues were identified using the standard false discovery rate of 5%.

Description of genomic SEM software. The genomic SEM software package, GenomicSEM, is written as an R package and is available through GitHub at https://github.com/MichelNivard/GenomicSEM. GenomicSEM contains several functions, including procedures for QCing and standardizing summary statistics, a function for producing genetic covariance matrices (S_{LDSC}) and their associated sampling covariance matrices ($V_{S_{LDSC}}$) using a multivariable extension of linkage disequilibrium score regression, functions for fitting genomic structural equation models to S_{LDSC} and $V_{S_{LDSC}}$, and functions for adding SNP-level data to the S_{LDSC} and $V_{S_{LDSC}}$ matrices (referred to as S_{Full} and $V_{S_{Full}}$) that are used for implementing genomic SEM for multivariate GWAS discovery. Functions include both pre-specified models (for example, a single common factor model) and user-specified models. Output includes both unstandardized and standardized solutions, along with the fit indices described above. WLS estimation is the default in the GenomicSEM package. GenomicSEM uses the lavaan Structural Equation Modeling package[64] as the primary workhorse for model specification and numerical optimization. We also provide limited support for OpenMx[65]. To run the multivariable LDSC function on 5 phenotypes takes ~15 min—a step in the analyses that only needs to be performed once. For models of multivariate genetic architecture that do not incorporate individual SNP effects, the typical run time observed for 3–15 traits is <1 s on a standard personal computer. Using parallel processing implemented in the GenomicSEM package on a 4-core/8-thread laptop, a multivariate genomic SEM GWAS with 5 indicators and ~1 million SNPs took ~8 h. The time needed to run the models will increase with increasing model complexity, and with increasing numbers of variables or SNPs. In these cases, the computing time can be greatly reduced by using a computing cluster to distribute SNP models across nodes/cores.

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

Data availability

The data that support the findings of this study are all publicly available. Links to the location of summary statistics, linkage disequilibrium scores, reference panel data and the code used to produce the current results can all be found at https://github.com/MichelNivard/GenomicSEM/wiki.

Code availability

GenomicSEM software is an R package that is available from GitHub at https://github.com/MichelNivard/GenomicSEM. The GenomicSEM R package can be installed directly at https://github.com/MichelNivard/GenomicSEM/wiki. Example GenomicSEM code, including code used to produce the results, is provided for each set of analyses at https://github.com/MichelNivard/GenomicSEM/wiki.

Received: 4 June 2018; Accepted: 22 February 2019;
Published online: 8 April 2019

References

1. Lee, S. H. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
2. Bush, W. S., Oetjens, M. T. & Crawford, D. C. Unravelling the human genome–phenome relationship using phenome-wide association studies. *Nat. Rev. Genet.* **17**, 129–145 (2016).
3. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
4. ReproGen Consortium et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
5. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
6. Jansen, P. R. et al. Genome-wide analysis of insomnia (*N*=1,331,010) identifies novel loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
7. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
8. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
9. Verhulst, B., Maes, H. H. & Neale, M. C. GW-SEM: a statistical package to conduct genome-wide structural equation modeling. *Behav. Genet.* **47**, 345–359 (2017).
10. Beaumont, R. N. et al. Genome-wide association study of offspring birth weight in 86,577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27**, 742–756 (2018).
11. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
12. Cheung, M. W.-L. metaSEM: an R package for meta-analysis using structural equation modeling. *Front. Psychol.* **5**, 1521–1532 (2015).
13. Savalei, V. & Bentler, P. M. A two-stage approach to missing data: theory and application to auxiliary variables. *Struct. Equ. Modeling* **16**, 477–497 (2009).
14. Yuan, K. H. & Bentler, P. M. Robust mean and covariance structure analysis through iteratively reweighted least squares. *Psychometrika* **65**, 43–58 (2000).
15. Browne, M. W. Asymptotically distribution-free methods for the analysis of covariance structures. *Br. J. Math. Stat. Psychol.* **37**, 62–83 (1984).
16. Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F. & Botella, J. Assessing heterogeneity in meta-analysis: *Q* statistic or I^2 index? *Psychol. Methods* **11**, 193–220 (2006).
17. Caspi, A. & Moffitt, T. E. All for one and one for all: mental disorders in one dimension. *Am. J. Psychiatry* **175**, 831–844 (2018).
18. Caspi, A. et al. The *p* factor. *Clin. Psychol. Sci.* **2**, 119–137 (2013).
19. Pettersson, E., Larsson, H. & Lichtenstein, P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. *Mol. Psychiatry* **21**, 717–721 (2016).
20. Smoller, J. W. et al. Psychiatric genetics and the structure of psychopathology. *Mol. Psychiatry* **24**, 409–420 (2019).
21. Stochl, J. et al. Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychol. Med.* **45**, 1483–1493 (2015).
22. Seed, C. et al. Hail: An Open-Source Framework for Scalable Genetic Data. *Neale Lab* http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank (2017).
23. Nieuwboer, H. A., Pool, R., Dolan, C. V., Boomsma, D. I. & Nivard, M. G. GWIS: genome-wide inferred statistics for functions of multiple phenotypes. *Am. J. Hum. Genet.* **99**, 917–927 (2016).
24. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **6139**, 1467–1471 (2013).
25. Ruderfer, D. M. et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol. Psychiatry* **19**, 1017–1024 (2014).
26. Maier, R. M. et al. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat. Commun.* **9**, 989–993 (2018).
27. Van der Sluis, S., Posthuma, D. & Dolan, C. V. TATES: efficient multivariate genotype–phenotype analysis for genome-wide association studies. *PLoS Genet.* **9**, e1003235 (2013).
28. Allegrini, A. et al. Genomic prediction of cognitive traits in childhood and adolescence. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/418210v1 (2018).
29. Rhemtulla, M., Brosseau-Liard, P. É. & Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* **17**, 354–373 (2012).
30. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890–5895 (2015).
31. Li, Z. et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.* **49**, 1576–1583 (2017).
32. Hu, Y. et al. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7**, 10448–10453 (2016).
33. The Autism Spectrum Disorders Working Group et al. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* **8**, 21 (2017).
34. Hill, W. D. et al. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Mol. Psychiatry* **24**, 169–181 (2019).
35. Martin, N. G. & Eaves, L. J. The genetical analysis of covariance structure. *Heredity* **38**, 79–95 (1977).
36. Zhu, X. et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96**, 21–36 (2015).
37. Ray, D. & Boehnke, M. Methods for meta-analysis of multiple traits using GWAS summary statistics. *Genet. Epidemiol.* **42**, 134–145 (2018).
38. O'Reilly, P. F. et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One* **7**, e34861 (2012).
39. de Vlaming, R., Johannesson, M., Magnusson, P. K., Ikram, M. A. & Visscher, P. M. Equivalence of LD-score regression and individual-level-data methods. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/211821v1 (2017).
40. Lee, J. J., McGue, M., Iacono, W. G. & Chow, C. C. The accuracy of LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *Genet. Epidemiol.* **42**, 783–795 (2018).
41. Jöreskog, K. G. & Sörbom, D. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language* (Scientific Software International, 1993).
42. Boker, S. M. & McArdle, J. J. Path analysis and path diagrams. *Wiley StatsRef: Statistics Reference Online* https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06517 (2014).
43. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
44. Baselmans, B. M. et al. Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* **51**, 445–451 (2019).
45. Bates, D., Maechler, M., Davis, T. A., Oehlschlägel, J. & Riedy, R. matrix: Sparse and dense matrix classes and methods. R package version 1.2-12 (2017).
46. Flora, D. B. & Curran, P. J. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods* **9**, 466–491 (2004).
47. Savalei, V. Understanding robust corrections in structural equation modeling. *Struct. Equ. Modeling* **21**, 149–160 (2014).
48. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: lessons from machine learning. *Persp. Psychol. Sci.* **12**, 1100–1122 (2017).
49. Lloyd-Jones, L. R., Robinson, M. R., Yang, J. & Visscher, P. M. Transformation of summary statistics from linear mixed model association on all-or-none traits to odds ratio. *Genetics* **208**, 1397–1408 (2018).
50. Kenny, D. A. Measuring model fit. *David A. Kenny* http://davidakenny.net/cm/fit.htm (2014).
51. Kaplan, D. *Structural Equation Modeling: Foundations and Extensions* Vol. 10 (Sage, 2008).
52. Tanaka, J. S. Multifaceted conceptions of fit in structural equation models. In *Testing Structutal Equation Models* 10–37 (Sage, 1993).
53. Hu, L. T. & Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling* **6**, 1–55 (1999).
54. Bentler, P. M. & Hu, L. T. in *Structural Equation Modeling: Concepts, Issues, and Applications* 76–99 (SAGE Publications Inc., 1995).
55. Bentler, P. M. & Satorra, A. Testing model nesting and equivalence. *Psychol. Methods* **15**, 111–123 (2010).
56. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
57. Consortium, I. H. The international HapMap project. *Nature* **426**, 789–796 (2003).
58. Ripke, S. et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
59. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
60. Muthén, L. K. & Muthén, B. O. Mplus: The Comprehensive Modeling Program for Applied Researchers. Version 7.3. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_7.pdf (Muthén & Muthén, 2014).

61. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2014).
62. Eysenck, S. B., Eysenck, H. J. & Barrett, P. A revised version of the psychoticism scale. *Pers. Individ. Diff.* **6**, 21–29 (1985).
63. Smith, B. H. et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2012).
64. Rossel, Y. lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA) http://users.ugent.be/~yrosseel/lavaan/lavaanIntroduction.pdf (2012).
65. Neale, M. C. et al. OpenMx 2.0: extended structural equation and statistical modeling. *Psychometrika* **81**, 535–549 (2016).

Acknowledgements

E.M.T.-D., K.P.H. and A.D.G. were supported by NIH grant R01HD083613. E.M.T.-D., S.J.R. and I.J.D. were supported by NIH grant R01AG054628. E.M.T.-D. and K.P.H. were each supported by Jacobs Foundation research fellowships. E.M.T.-D. and K.P.H. are members of the Population Research Center at the University of Texas at Austin, which is supported by NIH grant P2CHD042849. M.G.N. is supported by a Royal Netherlands Academy of Science Professor Award to D. I. Boomsma (PAH/6635), ZonMw grant: 'Genetics as a research tool: a natural experiment to elucidate the causal effects of social mobility on health' (pnr: 531003014) and ZonMw project: 'Can sex- and gender-specific gene expression and epigenetics explain sex-differences in disease prevalence and etiology?' (pnr: 849200011). H.F.I. is supported by the 'Aggression in children: unraveling gene–environment interplay to inform treatment and intervention strategies' (ACTION) project. ACTION receives funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement number 602768. P.D.K. and R.d.V. were supported by ERC Consolidator Grant 647648 EdGe. I.J.D., A.M.M., S.J.R., R.E.M. and W.D.H. are members of the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, which is part of the cross-council Lifelong Health and Wellbeing Initiative (MR/K026992/1). W.D.H. is supported by a grant from Age UK (Disconnected Mind Project). PGS analyses for the *p* factor were conducted under UKB dataset resource–application number 4844. PGS analyses for neuroticism were conducted using data from Generation Scotland. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). Genotyping of the Generation Scotland:Scottish Family Health Study samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland, and was funded by the Medical Research Council UK and Wellcome Trust (Wellcome Trust Strategic Award 'STratifying Resilience and Depression Longitudinally' (STRADL) reference 104036/Z/14/Z). Ethical approval for the Generation Scotland:Scottish Family Health Study was obtained from the Tayside Committee on Medical Research Ethics (on behalf of the National Health Service). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.D.G., M.R., H.F.I., M.G.N. and E.M.T.-D. developed the software. A.D.G., M.G.N. and E.M.T.-D. developed the theory underlying genomic SEM. A.D.G., M.R., R.d.V., M.G.N. and E.M.T.-D. developed the techniques and mathematical derivations. A.D.G., T.T.M., M.G.N. and E.M.T.-D. performed the simulation studies. S.J.R., R.E.M. and E.M.T.-D. performed the polygenic prediction analyses. A.D.G., M.G.N. and E.M.T.-D. wrote the manuscript. M.R., S.J.R., T.T.M., W.D.H., A.M.M., I.J.D., R.E.M., P.D.K. and K.P.H. provided feedback and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0566-x.

Correspondence and requests for materials should be addressed to A.D.G.

Corresponding author(s): Andrew Grotzinger
e-mail: agrotzin@utexas.edu

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Lead SNPs and GWAS for simulation studies were conducted using PLINK v1.9. Simulation of phenotypes with corresponding genotypes was conducting using GCTA v1.91. The construction of polygenic scores for out-of-sample prediction was conducted using PRSice v1.25. Genomic SEM was run using code we make available for download at https://github.com/MichelNivard/GenomicSEM

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

Data

Policy information about availability of data

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

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

All summary statistics used are available for public download. No raw data was used except for the construction of polygenic scores (PGSs) for out-of-sample prediction in UK Biobank and Generation Scotland. Although we are unable to make this data available ourselves, the data may be requested by others. Supplementary Table 1 provides references for all of the summary statistics used in our analyses. All summary statistics used are available for public use, and links are provide directly below.

For construction of the p-factor, summary statistics for PTSD, Anxiety, Bipolar Disorder, and Major Depressive Disorder were downloaded from: http://www.med.unc.edu/pgc/results-and-downloads. Summary statistics for Schizophrenia were downloaded from: http://walters.psycm.cf.ac.uk/.

The GWIS summary statistics for Bipolar Disorder and Schizophrenia were also downloaded from the cross-disorder section of the PGC website. The summary statistics for educational achievement are available for download from: https://www.thessgac.org/data.

Summary statistics for item-level indicators of neuroticism were downloaded from: https://docs.google.com/spreadsheets/d/1b3oGI2lUt57BcuHttWaZotQcI0-mBRPyZihz87Ms_No/edit#gid=1209628142. We use Round 1 of the Neale's Lab UKB GWAS results.

Summary statistics for early-life traits used in the factor model of anthropometric traits (birth length, birth weight, infant head circumference, childhood obesity) were obtained from: https://egg-consortium.org/.

Summary statistics used for the anthropometric traits example for BMI, waist-hip ratio, hip circumference, waist circumference, and height are available from: https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Behavioural & social sciences study design

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

Study description	Our study uses publically available summary statistics to examine a new developed method, Genomic SEM. Genomic SEM can be used to examine any form of structural equation model (e.g., factor analysis, mediation, etc.) using summary statistics. The software provides model fit statistics, and can also be used to produce summary statistics for a latent trait.
Research sample	We use only publically available summary statistics for our analyses. As Genomic SEM relies on ld-score regression (LDSC) to construct genetic covariance matrices, and LDSC requires summary statistics to be within a single ethnic population due to differences in linkage disequilibrium across populations, we use only summary statistics restricted to European populations.
Sampling strategy	This is not applicable as we use previously collected data.
Data collection	This is not applicable as we use previously collected data.
Timing	This is not applicable as we use previously collected data.
Data exclusions	We use only summary statistics from European populations due to the requirements of LDSC, as noted above.
Non-participation	This is not applicable as we use previously collected data.
Randomization	This is not applicable as we use previously collected data.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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