nature human behaviour

SUPPLEMENTARY INFORMATION

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

In the format provided by the authors and unedited.

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]

[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

December 19, 2018

Online Supplement:

Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits

Supplementary Methods

MTAG Moment Conditions. Here we examine the connection between the MTAG model and Genomic SEM. MTAG builds onto the LDSC framework, where *K* phenotypes and *M* SNPs are measured in *N* individuals, and modeled according to the equation:

$$\phi_{i,k} = x_{i,j}\beta_{j,k} + \epsilon_{i,k}, \qquad (1.1)$$

where $\phi_{i,k}$ is the score for person *i* on phenotype *k, x* is the standardized genotype for person *i* on SNP *j,* $\beta_{j,k}$ the true genotype effect size for SNP$_j$ on phenotype *k,* and $\epsilon_{i,k}$ is the residual for person *i* on phenotype *k*. Written in matrix form, we have:

$$\Phi = \mathrm{XB} + \mathrm{E}, \qquad (1.2)$$

where Φ is an $N \times K$ matrix of scores for person *i* on phenotype *k, X* is an $N \times M$ matrix of standardized genotypes for person *i* on SNP *j,* B is an $M \times K$ matrix of true genotype effect sizes for SNP *j* on phenotype *k,* and $\mathrm{E}_{i,k}$ is an $N \times K$ matrix of residuals for person *i* on phenotype *k*.

In this framework, LDSC is used to model $\beta_{j,k}$ as phenotype-specific random effects, varying over SNPs, with $\mathrm{E}(\beta_{j,k}) = 0$ and $\mathrm{cov}(\beta_{j,k}) = \Omega$. The diagonal elements of Ω contain the average heritability explained per SNP (h_k^2 / M ; alternately referred to as genetic variance explained per SNP, i.e., σ_k^2 / M), and the off diagonal elements of Ω contain the genetic covariances between phenotypes on a per-SNP scale (σ/M, where σ is the genetic covariance between pairs of phenotypes). In other words Ω is equivalent to $\frac{1}{M} S_{LDSC}$, where S_{LDSC} is the genetic covariance matrix estimated with LDSC that is used in Genomic SEM.

By drawing on multivariate GWAS summary statistics from K genetically correlated phenotypes, MTAG attempts to obtain an estimate of the effect size $\beta_{j,t}$ for SNP *j* on target phenotype *t* that is more precise than the univariate GWAS estimate, $\hat{\beta}_{j,t}$, of this effect. In the notation of Turley et al. (2018)[11], the MTAG moment condition specifies:

$$\mathrm{E}\left(\hat{\beta}_{j,k} - \frac{\omega_{kt}}{\omega_{tt}} \beta_{j,t} \right) = 0 \qquad (1.3)$$

where $\hat{\beta}_{j,k}$ is the GWAS estimate for the regression effect of SNP *j* on phenotype *k*, ω_{kt} is the (*k*,*t*)th element of Ω (i.e., elements drawn from the *t*th column of Ω), and ω_{tt} is the (*t*,*t*)th element of Ω. In other words, ω_{kt} is the LDSC-estimate of per-SNP scaled genetic covariances between each phenotype and the target phenotype and ω_{tt} is the LDSC estimate of the per-SNP genetic variance (i.e. per-SNP heritability) of the target phenotype.

We can rewrite the MTAG moment condition in notation that is more germane to Genomic SEM. We write the GWAS estimate for the regression effect of SNP *j* on phenotype *k* as $\beta_{GWAS\,j,k}$. We write the LDSC estimate of per-SNP scaled genetic covariance between phenotype *k* and target phenotype *t* as $\sigma_{k,t}/M$, and we write the LDSC estimate of per-SNP genetic variance in target phenotype *t* as σ_t^2 / M. Finally, we write the effect size for SNP *j* on target phenotype *t* that MTAG attempts to estimate as $\beta_{MTAG\,j,k}$. Under this notation, the MTAG moment condition takes the form:

$$\beta_{GWAS\,j,k} - \frac{\sigma_{k,t} / M}{\sigma_t^2 / M} \beta_{MTAG\,j,t} = 0 \quad . \qquad (1.4)$$

Cancelling M from the numerator and denominator of the quotient and rearranging yields:

$$\beta_{GWAS\,j,k} = \frac{\sigma_{k,t}}{\sigma_t^2} \beta_{MTAG\,j,t} \quad . \qquad (1.5)$$

Standard covariance algebra holds that the covariance between variables x and y divided by the variance of x is equivalent to the unstandardized regression effect of x on y. We therefore obtain that the LDSC-derived genetic covariance between *k* and *t* divided by the LDSC-derived genetic variance of *t* is equivalent to an LDSC-inferred structural regression effect of the genetic component of phenotype *t* on the genetic component of phenotype *k*, which we label $\beta_{LDSC\,t,k}$. The moment condition therefore reduces to:

$$\beta_{GWAS\,j,k} = \beta_{LDSC\,t,k} \beta_{MTAG\,j,t} \quad . \qquad (1.6)$$

Solving for $\beta_{MTAG\,j,t}$, we obtain:

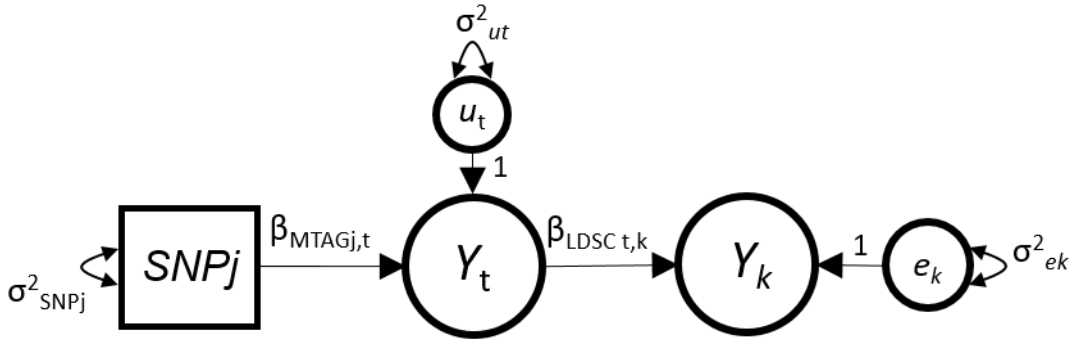
$$\beta_{MTAG\,j,t} = \frac{\beta_{GWAS\,j,k}}{\beta_{LDSC\,t,k}} \quad . \qquad (1.7)$$

Genomic SEM Covariance Expectations. We can specify a model within Genomic SEM that satisfies these same moment conditions as MTAG. We write a model in which the genetic component Y_k of each phenotype k, is regressed on the genetic component Y_t of t, and Y_t is regressed on SNP j:

$$Y_k = \beta_{\text{LDSC t},k} \times Y_t + e_k, \quad (2.1)$$

$$Y_t = \beta_{\text{MTAG}\,j,t} \times \text{SNP}_j + u_t, \quad (2.2)$$

or in path diagram form (for two phenotypes, t and k) as:



This model produces the following expectations with respect to the GWAS-estimated covariance between SNP j and phenotype k:

$$\sigma_{GWAS\,j,k} = \sigma^2_{SNPj} \beta_{MTAG\,j,t} \beta_{LDSC\,t,k}, \quad (2.3)$$

which rearranging yields:

$$\frac{\sigma_{GWAS\,j,k}}{\sigma^2_{SNPj}} = \beta_{MTAG\,j,t} \beta_{LDSC\,t,k}. \quad (2.4)$$

As the covariance between SNP_j and phenotype k divided by the variance of SNP_j is equal to the regression effect of SNP_j on phenotype k, we have:

$$\beta_{GWAS\,j,k} = \beta_{MTAG\,j,t} \beta_{LDSC\,t,k}. \quad (2.5)$$

Solving for $\beta_{MTAG\,j,t}$, we obtain:

$$\beta_{MTAG\,j,t} = \frac{\beta_{GWAS\,j,k}}{\beta_{LDSC\,t,k}}, \quad (2.6)$$

which is the same equality obtained when solving for $\beta_{MTAG\,j,t}$ from the MTAG moment condition, in equation 1.7 above.

Optimization. Both the MTAG moment condition and the specific Genomic SEM model specified to satisfy the MTAG moment condition yields:

$$\beta_{MTAG\,j,t} = \frac{\beta_{GWAS\,j,k}}{\beta_{LDSC\,t,k}} . \quad (3.1)$$

As there are K phenotypes, including the target phenotype, this yields a system of equations that is overidentified, in the sense that there are more knowns than free parameters. In, for example, the two-phenotype circumstance (1 target phenotype, *t*, and one supporting phenotype, *s*), the free parameter $\beta_{MTAG\,j,t}$ is equivalent to two separate terms:

$$\beta_{MTAG\,j,t} = \frac{\beta_{GWAS\,j,s}}{\beta_{LDSC\,t,s}} , \quad (3.2)$$

and

$$\beta_{MTAG\,j,t} = \frac{\beta_{GWAS\,j,t}}{\beta_{LDSC\,t,t}} = \frac{\beta_{GWAS\,j,t}}{1} = \beta_{GWAS\,j,t} . \quad (3.3)$$

In both MTAG and Genomic SEM, free parameters are estimated by minimizing a fit function. The MTAG fit function minimizes the weighted squared discrepancies between the MTAG-implied GWAS estimates and the univariate GWAS estimates for all K phenotypes. In the notation of Turley et al. (2018),[11] this is written as:

$$\left(\hat{\beta}_j - \frac{\omega_t}{\omega_{tt}}\beta_{j,t}\right)' W \left(\hat{\beta}_j - \frac{\omega_t}{\omega_{tt}}\beta_{j,t}\right) , \quad (3.4)$$

where *W* is a weight matrix, $\hat{\beta}_j$ is the vector of betas for the effect of SNP_j on phenotype *k* estimated from univariate GWAS, and $\frac{\omega_t}{\omega_{tt}}\beta_{j,t}$ is the vector of SNP-phenotype betas that is implied by MTAG (by multiplying the MTAG estimate for the GWAS effect on target trait *t* by the LDSC-inferred structural regression effect of the genetic component of phenotype *t* on the genetic component of phenotype *k*).

This takes the same form as the WLS fit function in Genomic SEM, which minimizes the weighted squared discrepancies between all elements in the full genetic covariance matrix (combined both from elements derived from LDSC and elements derived directly from GWAS estimates) and those implied by the specified model, according to:

$$\left(s-\sigma(\theta)\right)' W \left(s-\sigma(\theta)\right), \qquad (3.5)$$

where W is a weight matrix, s is the half-vectorized empirical genetic covariance matrix (S), and $\sigma(\theta)$ is the half-vectorized model-implied genetic covariance matrix ($\Sigma(\theta)$).

The weight matrices, W, in MTAG and Genomic SEM are very similar. In Genomic SEM, W is the inverse of the diagonal matrix D_S that contains the diagonal elements of V_S on its diagonal, where V_S is the sampling covariance matrix of all of the elements in S, and SEs of parameter estimates θ are obtained via sandwich estimation using the full V_S matrix. In MTAG, W is the inverse of a similar sampling covariance matrix of the MTAG-implied GWAS estimates. In the notation of Turley et al. (2018),[11] this matrix is formed as:

$$\left(\Omega - \frac{\omega_t \omega_t'}{\omega_{tt}} + \Sigma_j\right)^{-1}, \qquad (3.6)$$

where Ω is the LDSC-derived genetic covariance matrix among the phenotypes on a per-SNP scale, ω_t is the vector of estimates from column t of Ω that contains the genetic covariances between each phenotype and the target phenotype, ω_{tt} is the per-SNP genetic variance (per-SNP heritability) of the target phenotype, and Σ_j is the sampling covariance of the univariate GWAS effects, which is equivalent (after transformation) to the elements of the portion of the V_s matrix from Genomic SEM that contains sampling covariances of the GWAS effects (V_{SNP}) obtained from the cross-trait LDSC intercepts. The term $\frac{\omega_t \omega_t'}{\omega_{tt}}$ reduces to a matrix of genetic variances and covariances among the phenotypes mediated by their structural regressions on the genetic component of phenotype t, such that $\Omega - \frac{\omega_t \omega_t'}{\omega_{tt}}$ represents per-SNP scaled residual genetic covariances among the phenotypes after controlling for genetic variance in target phenotype t. The addition of these residual genetic covariances to the sampling covariance of the univariate GWAS effects in constructing the MTAG W matrix results in the fit function downweighting

the contribution of GWAS estimates for supporting phenotypes that have lower genetic correlations with the target phenotype.

Supplementary Results

ML Estimation. 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 reducing standard errors of the Genomic SEM parameter estimates, which may boost power for SNP discovery and increase polygenic prediction. However, because the cells in the V_S matrix (that index the precision of cells in the *S* matrix) are contingent upon the sample sizes for the contributing univariate GWASs, WLS may produce a solution that is dominated by the patterns of association involving the better powered GWASs, and contain substantial local misfit in cells of *S* that are informed by lower powered GWASs. In other words, WLS relative to ML may more heavily prioritize minimizing sampling variance of the parameter estimates in the so-called variance bias tradeoff.[37] We expect that this will only occur when the model is overidentified (i.e., $df > 0$), such that exact fit cannot be obtained, and that divergence in WLS and ML estimates will be most pronounced when there is lower sample overlap and the contributing univariate GWASs differ substantially in power.

In the case of our Genomic SEM formulation of GWIS, the model was just identified ($df = 0$) and results from ML were highly consistent with those from WLS (Supplementary Fig. 23). For anthropometric traits, results were also highly similar across ML and WLS, with ML estimation also confirming two latent factors with a modest genetic correlation ($r_g = .21$, $SE = .05$, $p < .001$). In the case of psychiatric traits, we use summary statistics characterized by discrepant sample sizes and low levels of sample overlap for which the expectation is potentially divergent WLS and ML estimates. Indeed, WLS and ML findings were discrepant, with MDD loading strongest on the *p*-factor with ML estimation (Supplementary Fig. 24), but SCZ loading strongest on the *p*-factor with WLS estimation. The follow-up models used to calculate model fit failed to converge for ML estimation of both the *p*-factor and anthropometric traits. For neuroticism, results were highly consistent across WLS and ML estimation—as would be expected giving almost entirely overlapping univariate samples—revealing a common neuroticism factor with strong loadings for all indicators and good model fit ($\chi^2[54] = 4959.08$, AIC = 5007.08, CFI = .891, SRMR = .116; Supplementary Fig. 24).

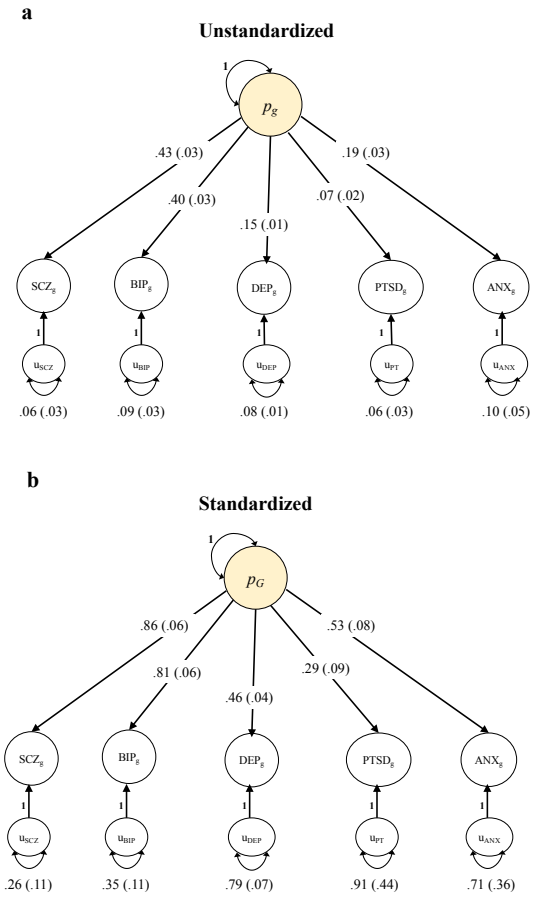
For SNP effect models estimated using ML, there was minimal enrichment of effects for the *p*-factor, but effects were similar to WLS for neuroticism (Supplementary Fig. 25 for QQ plot). More specifically, there were no lead SNPs identified for the *p*-factor with ML estimation and 105 lead SNPs identified for neuroticism with ML estimation. For estimates of Q_{SNP}, there were 63 independent hits for the *p*-factor and 63 independent hits for neuroticism. Inspection of Q_{SNP} estimates for the *p*-factor

indicated that these results were largely driven by SNPs that were highly significant for schizophrenia, but not the other indicators (Table S19). As expected based on higher sample overlap and less discrepant sample sizes for neuroticism compared to the *p*-factor, the association between *p*-values for ML and WLS were higher for neuroticism (r = .94) than for the *p*-factor (r = .15; Supplementary Fig. 26). However, the association between Q_{SNP} estimates was high for both the *p*-factor (r = .77) and neuroticism (r = .99; Supplementary Fig. 26). Biological annotation of ML-based results conducted using DEPICT revealed all null findings for the *p*-factor, and 6 prioritized genes, no gene sets, and 23 tissues for neuroticism. Loci identified for ML estimation of neuroticism and Q_{SNP} estimates for the *p*-factor and neuroticism were expressed in the nervous system (Supplementary Fig. S27).

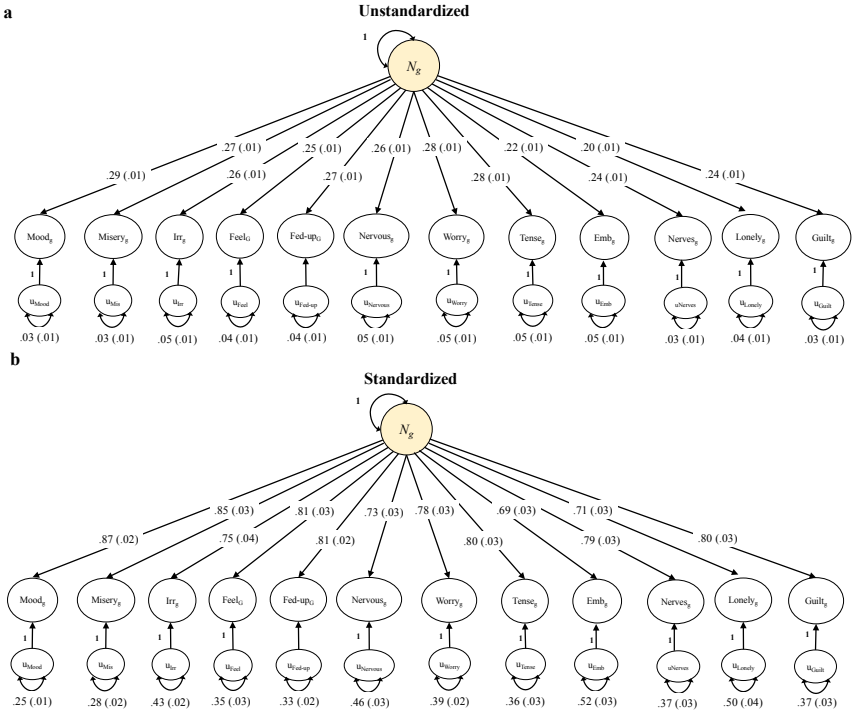
Model Comparisons: Neuroticism Example. As an example of how to use Genomic SEM to do model comparisons we examined different factor structures that might be fit to the 12 neuroticism items from UK Biobank. As a starting point, we performed an Exploratory Factor Analysis (EFA) in the fa R package using the oblimin rotation for a two-, three-, and four-factor solution. A follow-up CFA (Supplementary Fig. 5) within Genomic SEM was specified based on the EFA parameter estimates (standardized loadings > .4 were retained) for the two- and three-factor solutions, but not the four-factor solution as the fourth factor was defined only by the tense and irritability items (Table S2). The two-factor solution ($\chi^2[53] = 2758.18$, AIC = 2808.18, CFI = .940, SRMR = .077) and three-factor solution ($\chi^2[51] = 1879.31$, AIC = 1933.31, CFI = .959, SRMR = .057) both provided excellent fit to the data. For both solutions, the factors were highly correlated ($r_g \geq .67$). As these were not nested models, they could not be compared using χ^2 difference tests.

There were 69 SNPs identified as significantly heterogenous for the common factor of neuroticism, indicating that these particular SNPs may be operating through factors defined by a smaller subset of items. In order to investigate this possibility, multivariate GWAS analyses were conducted for these 69 Q_{SNP} hits using the two- and three-factor solutions identified above. The SNP was specified to predict all factors in each model. Of these 69 SNPs, 28 and 20 were genome-wide significant for Q_{SNP} for the two- and three-factor solutions, respectively (Table S3). For the two-factor solution, 6 SNPs had a genome-wide significant effect on the first factor and 4 SNPs were significant for the second factor. For the three-factor solution, 5 SNPs were significant for the first factor, 1 was significant for the second factor, and 9 were significant for the third factor. Taken together, these results indicate that a proportion of the SNPs identified as significantly heterogenous for the single factor solution may have large effects on individual factors defined by a smaller subset of the neuroticism items. Indeed, plots of item-level effects for SNPs identified as significant for one of the factors indicate high levels of consistency within, but not across, factors (Supplementary Fig. 6). For SNPs that continued to be significant for Q_{SNP} for even

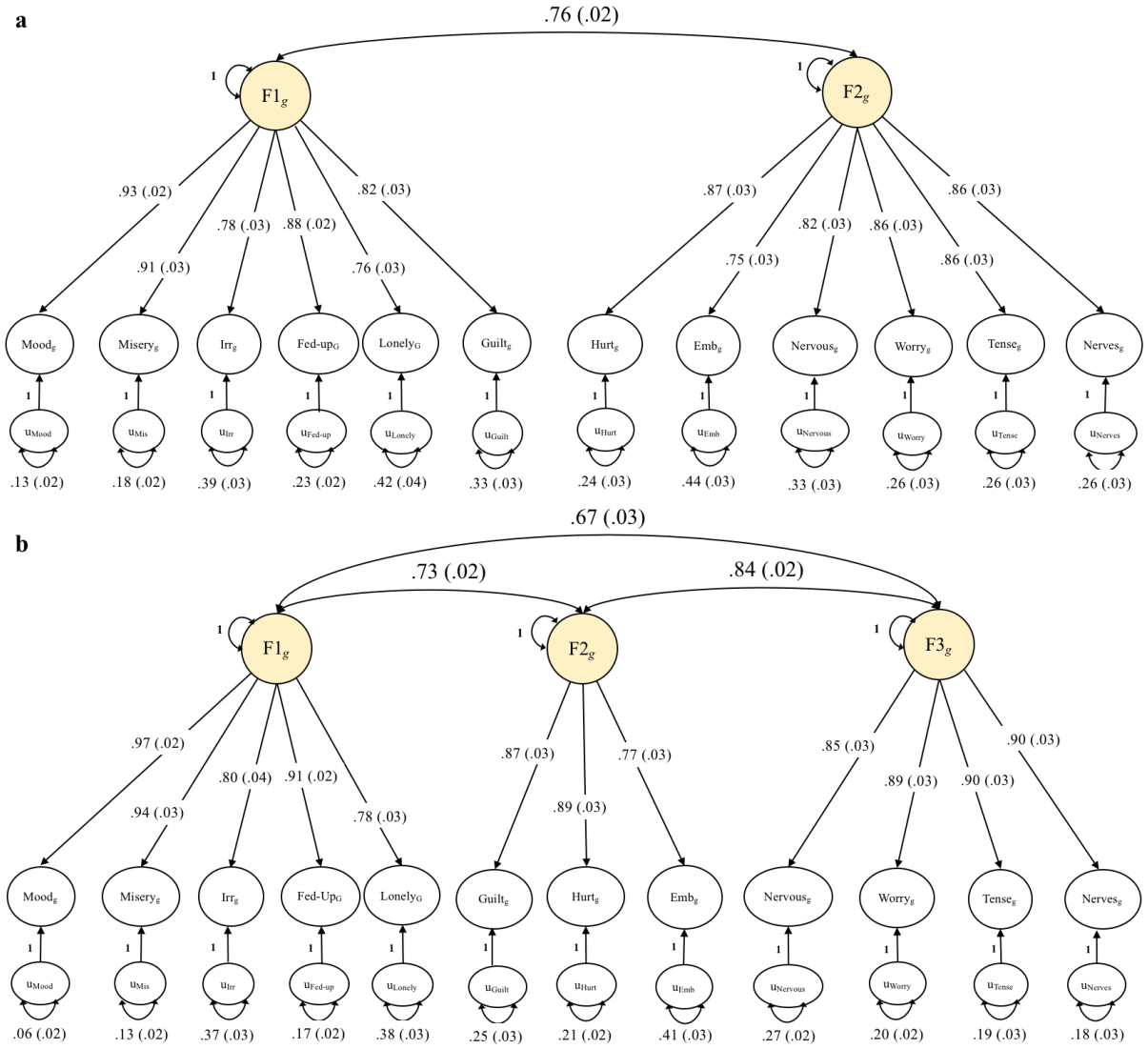
the three-factor solution, the effect may be even finer grained, with outlying effects on individual items. The iterative process outlined here of beginning with a common factor, and following up on SNPs identified as having high degrees of heterogeneity in more nuanced models, can be used to bin SNPs into categories of decreasing pleiotropic effects within a set of genetically correlated traits (Supplementary Fig. 6).



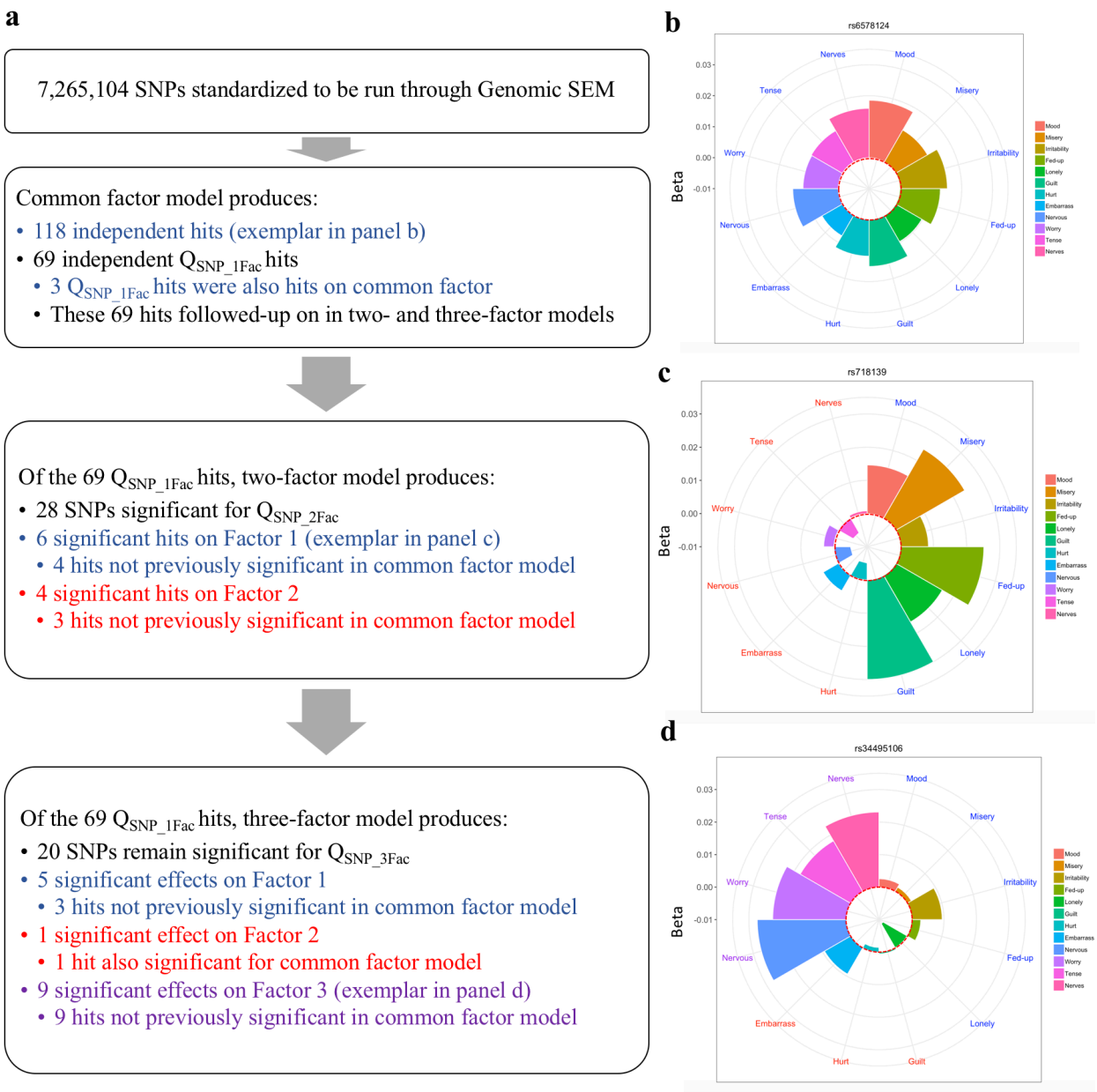
Supplementary Figure 1. Confirmatory factor analysis of genetic *p*-factor with Genomic SEM. Confirmatory factor analyses (CFA) were used to construct a genetically defined *p*-factor for unstandardized (panel a) and standardized estimates (panel b) using WLS estimation. *SE*s are shown in parentheses. The genetic covariance matrix (unstandardized) or genetic correlation matrix (standardized) and associated sampling covariance matrix were used as input for Genomic SEM. Indicators are presented as circle to reflect the fact that these are unobserved heritability estimates from LDSC. SCZ = schizophrenia; BIP = bipolar disorder; DEP = major depressive disorder; PTSD = post-traumatic stress disorder; ANX = anxiety.



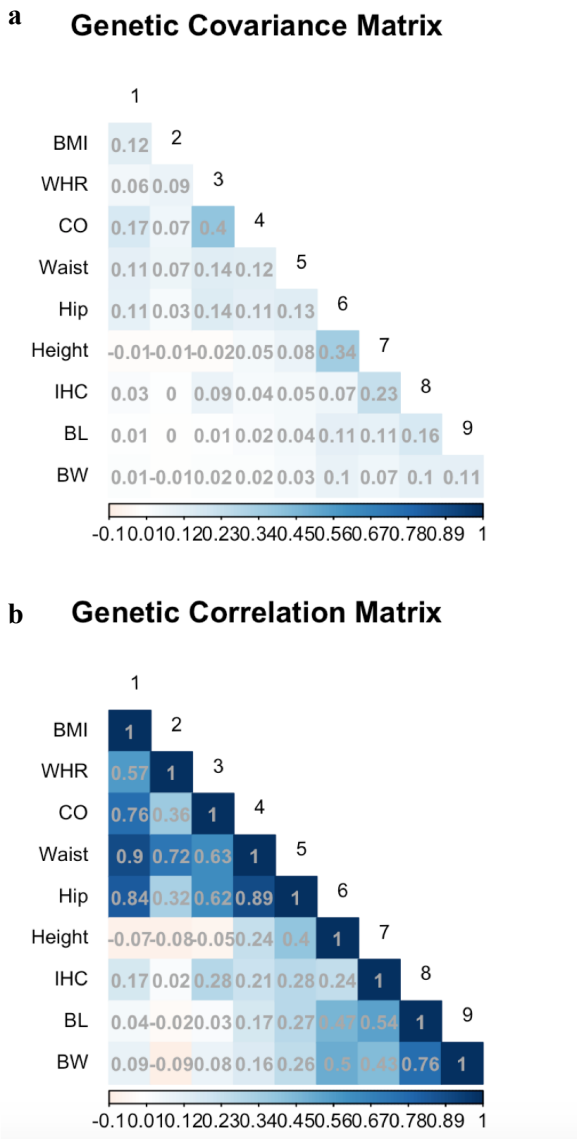
Supplementary Figure 2. Confirmatory factor analysis of a genetic factor of neuroticism with Genomic SEM. Confirmatory factor analyses (CFA) were used to construct a genetically defined neuroticism factor for unstandardized (panel a) and standardized estimates (panel b) using WLS estimation. *SE*s are shown in parentheses. The genetic covariance matrix (unstandardized) or genetic correlation matrix (standardized) and associated sampling covariance matrix were used as input for Genomic SEM. Irr = irritability; Feel = sensitivity/hurt feelings; fed-up = fed-up feelings; emb = worry too long after embarrassment.



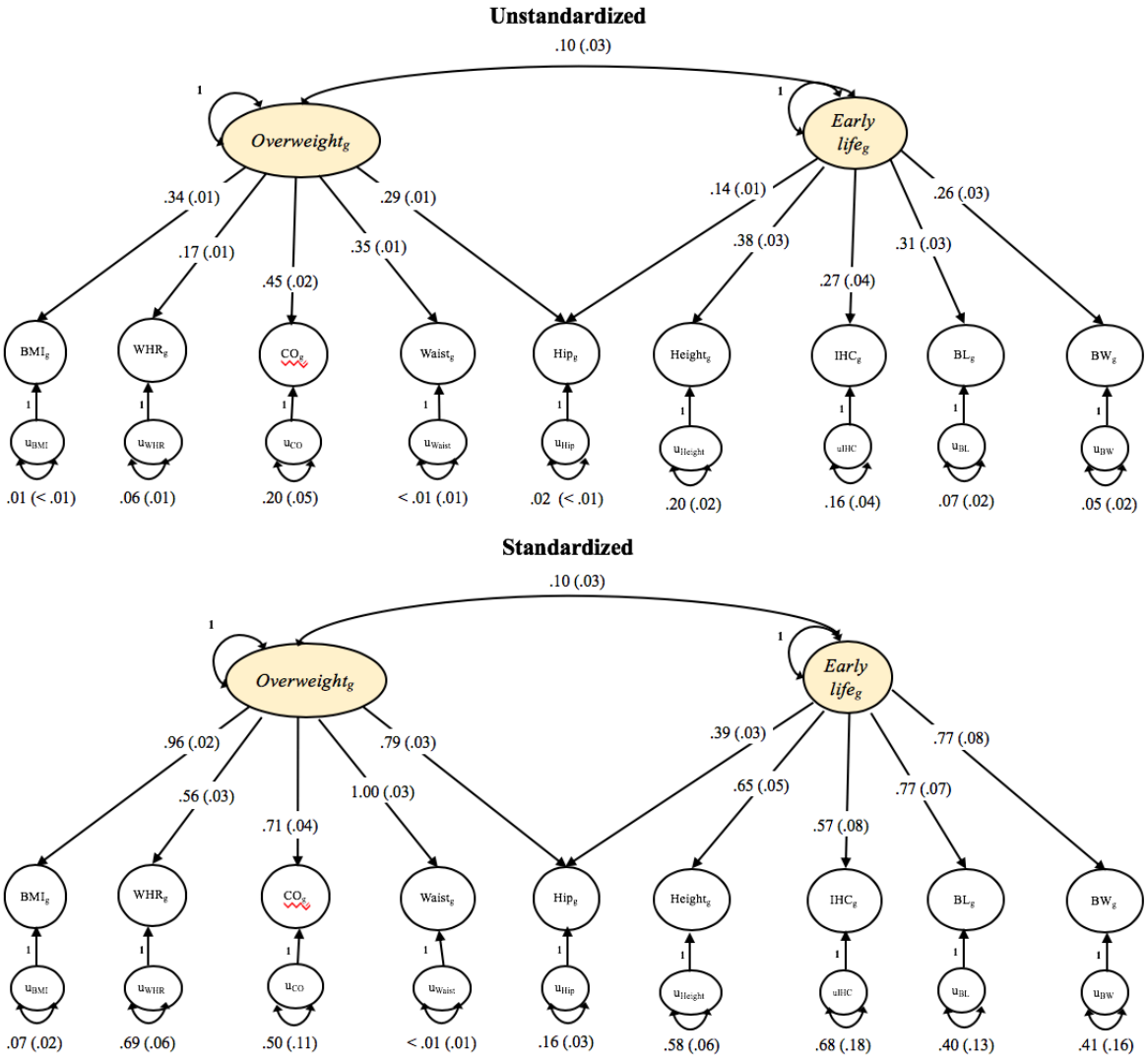
Supplementary Figure 3. Confirmatory factor analysis of two and three-factor models of neuroticism. Confirmatory factor analyses (CFA) based on initial EFAs were used to construct a two-factor (panel a) and three-factor (panel b) solution using WLS estimation. The displayed ordering of the variables is maintained across the factor solutions for comparative purposes. Standardized values are reported along with *SE*s in parentheses. The genetic correlation matrix (standardized) and associated sampling covariance matrix were used as input for Genomic SEM. Irr = irritability; Feel = sensitivity/hurt feelings; fed-up = fed-up feelings; emb = worry too long after embarrassment.



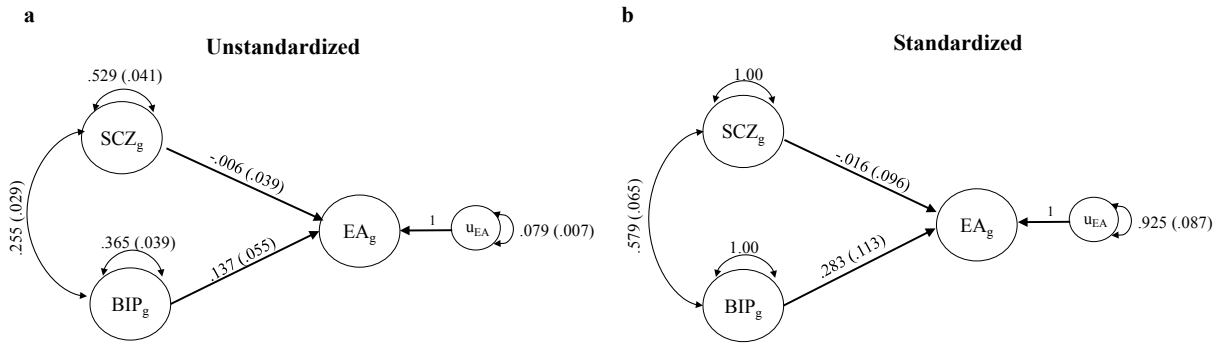
Supplementary Figure 4. Identifying SNPs with increasingly specific effects on neuroticism items. Panel a depicts the flow chart for the iterative process that can be undertaken to identify SNPs with increasingly specific effects on sets of traits. Any mention of significance is at the genome-wide level (p < 5e-8). Values inside and outside of the dotted red circle are negative and positive, respectively. Panels b-d depict polar plots for the item-level Z-statistics for exemplar SNPs identified as genome-wide significant for the one-factor solution (panel b), the first factor for the two-factor solution (panel c), and the third factor for the three-factor solution (panel d). The same coloring within item names denotes loading on the same factor. In panel b, a SNP identified as significant for a common factor shows highly consistent effects across items. In panels c and d, SNPs identified as significant for factors defined by only a subset of items show consistent effects with respect to magnitude and direction only within these factors.



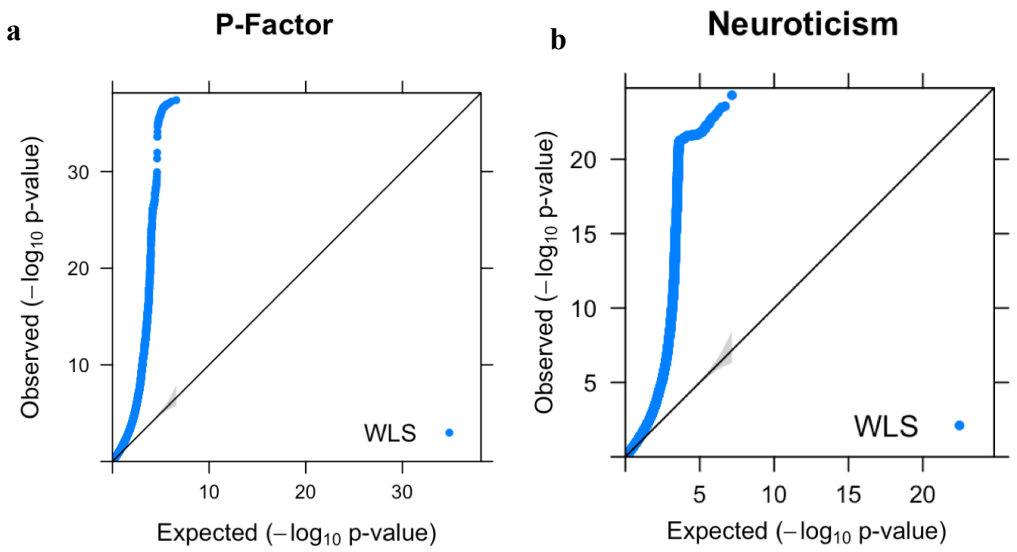
Supplementary Figure 5. Heatmap of genetic associations among anthropometric traits. Genetic covariance (panel a) and correlation (panel b) matrices with parameters estimated from multivariate LDSC. Visual inspection indicates two clusters in the upper left and lower right corner of the heatmap. BMI = body mass index; WHR = waist-hip ratio; CO = childhood obesity; IHC = infant head circumference; BL = birth length; BW = birth weight.



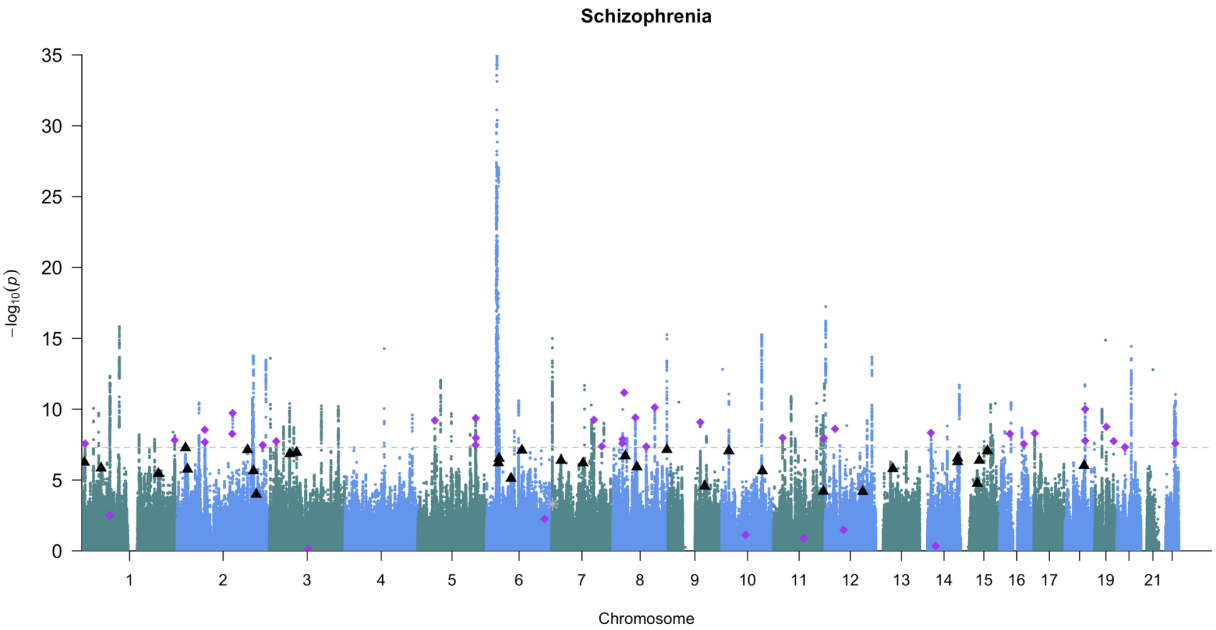
Supplementary Figure 6. Confirmatory factor analysis of multivariate genetic architecture of anthropometric traits using Genomic SEM. Confirmatory factor analyses (CFA) informed by an initial exploratory factor analysis were used to construct latent overweight and early life growth factors for unstandardized (panel a) and standardized estimates (panel b) using WLS estimation. *SE*s are shown in parentheses. The genetic covariance matrix (unstandardized) or genetic correlation matrix (standardized) and associated sampling covariance matrix were used as input for Genomic SEM. BMI = body mass index; WHR = waist-hip ratio; CO = childhood obesity; IHC = infant head circumference; BL = birth length; BW = birth weight.



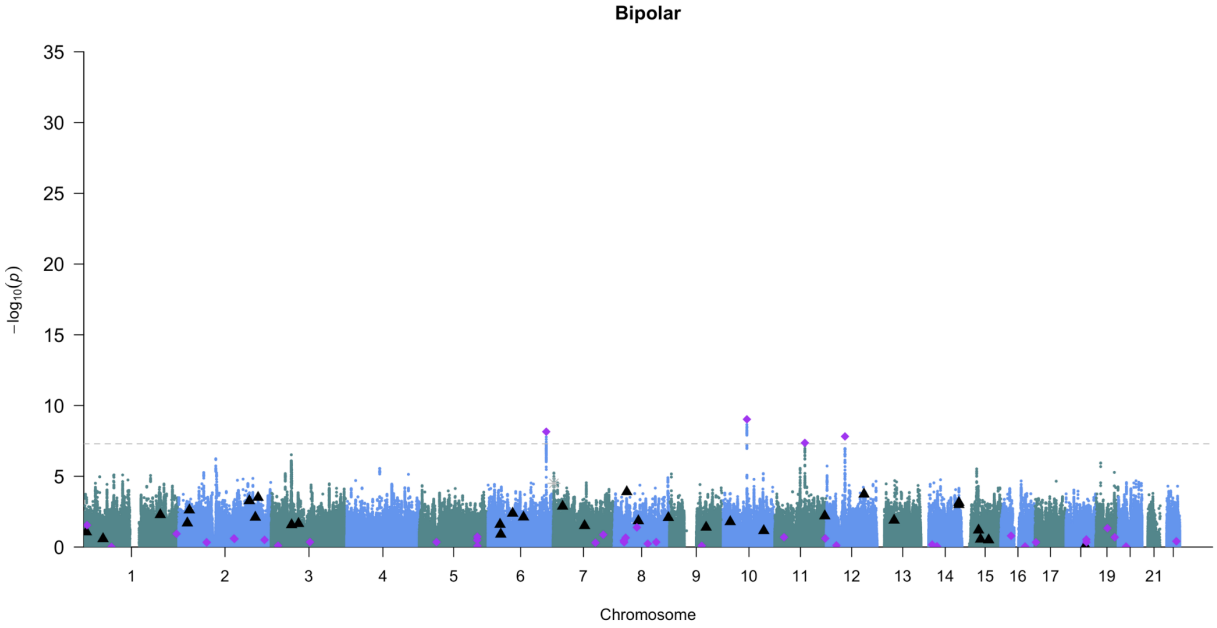
Supplementary Figure 7. Reproducing GWIS findings using Genomic SEM. Results from Genomic SEM in which the genetic component of educational achievement was simultaneously regressed on the genetic components of bipolar disorder and schizophrenia. The genetic covariance (unstandardized; panel a) and genetic correlation (standardized; panel b) matrices, and associated sampling covariance matrices, estimated from multivariate LDSC were used as input for Genomic SEM.



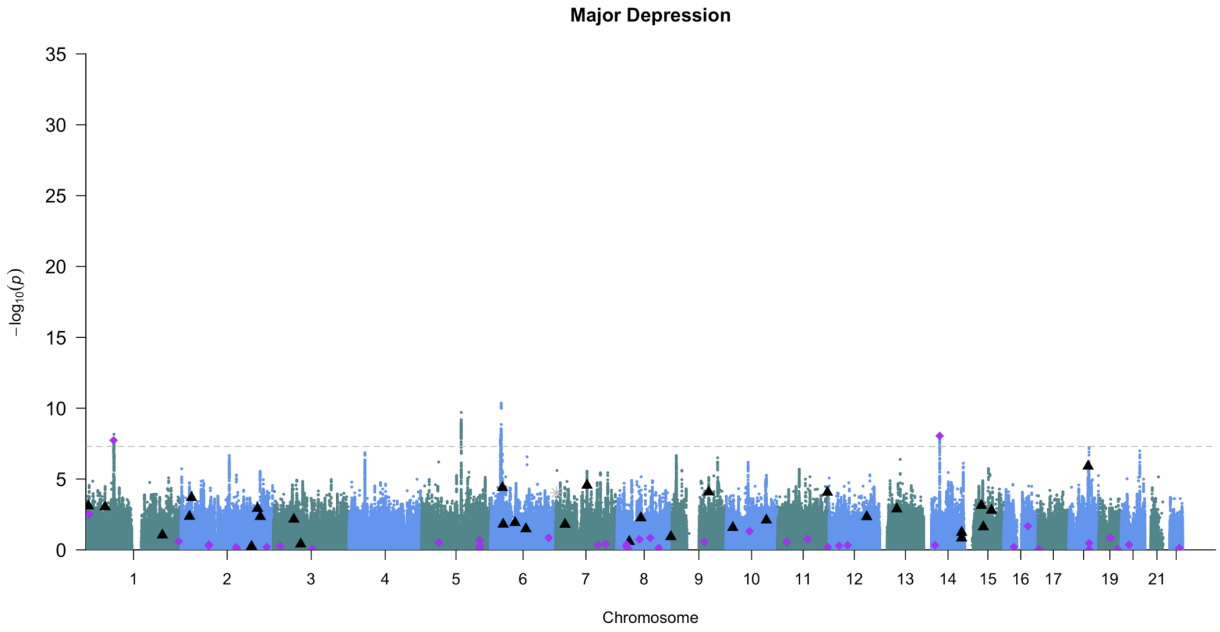
Supplementary Figure 8. Quantile-quantile plot of multivariate GWAS *p*-values for the *p*-factor and neuroticism. Expected −log10(*p*)-values are those expected under the null hypothesis. The shaded area indicates the 95% confidence interval under the null. The multivariate GWAS was conducted using Genomic SEM with WLS estimation.



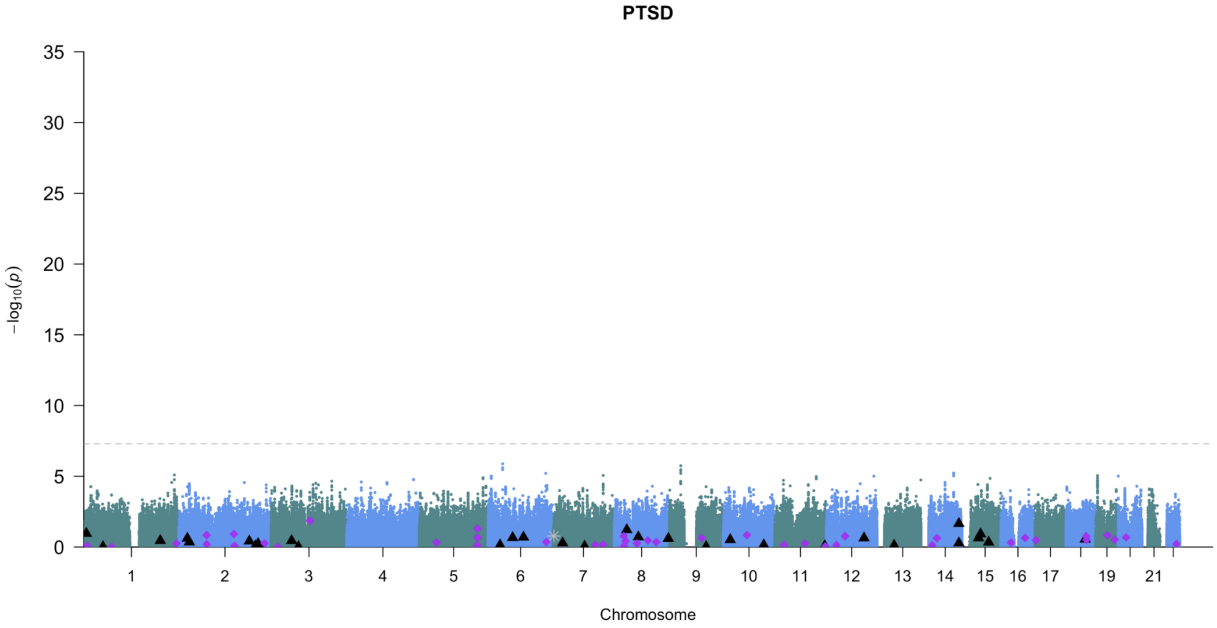
Supplementary Figure 9a. Manhattan plot for univariate Schizophrenia GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate GWAS. Purple diamonds denote independent hits for the univariate indicators that were not in LD with independent hits for the *p*-factor. Grey stars denote independent hits for Q_{SNP}.



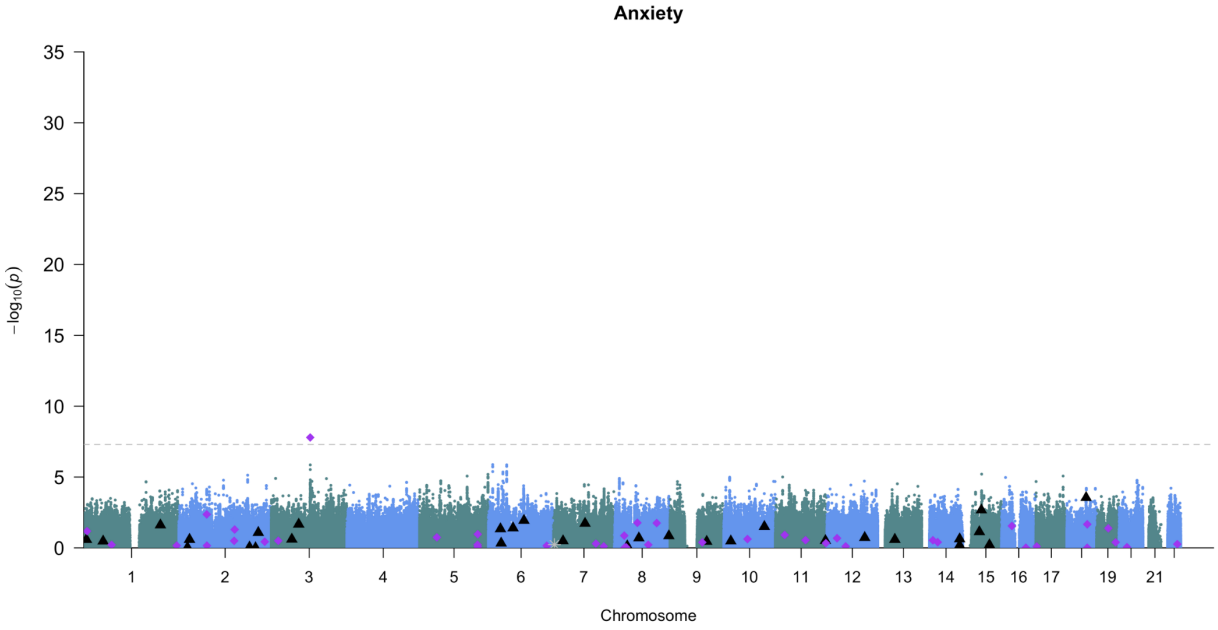
Supplementary Figure 9b. Manhattan plot for univariate Bipolar GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate GWAS. Purple diamonds denote independent hits for the univariate indicators that were not in LD with independent hits for the *p*-factor. Grey stars denote independent hits for Q_{SNP}.



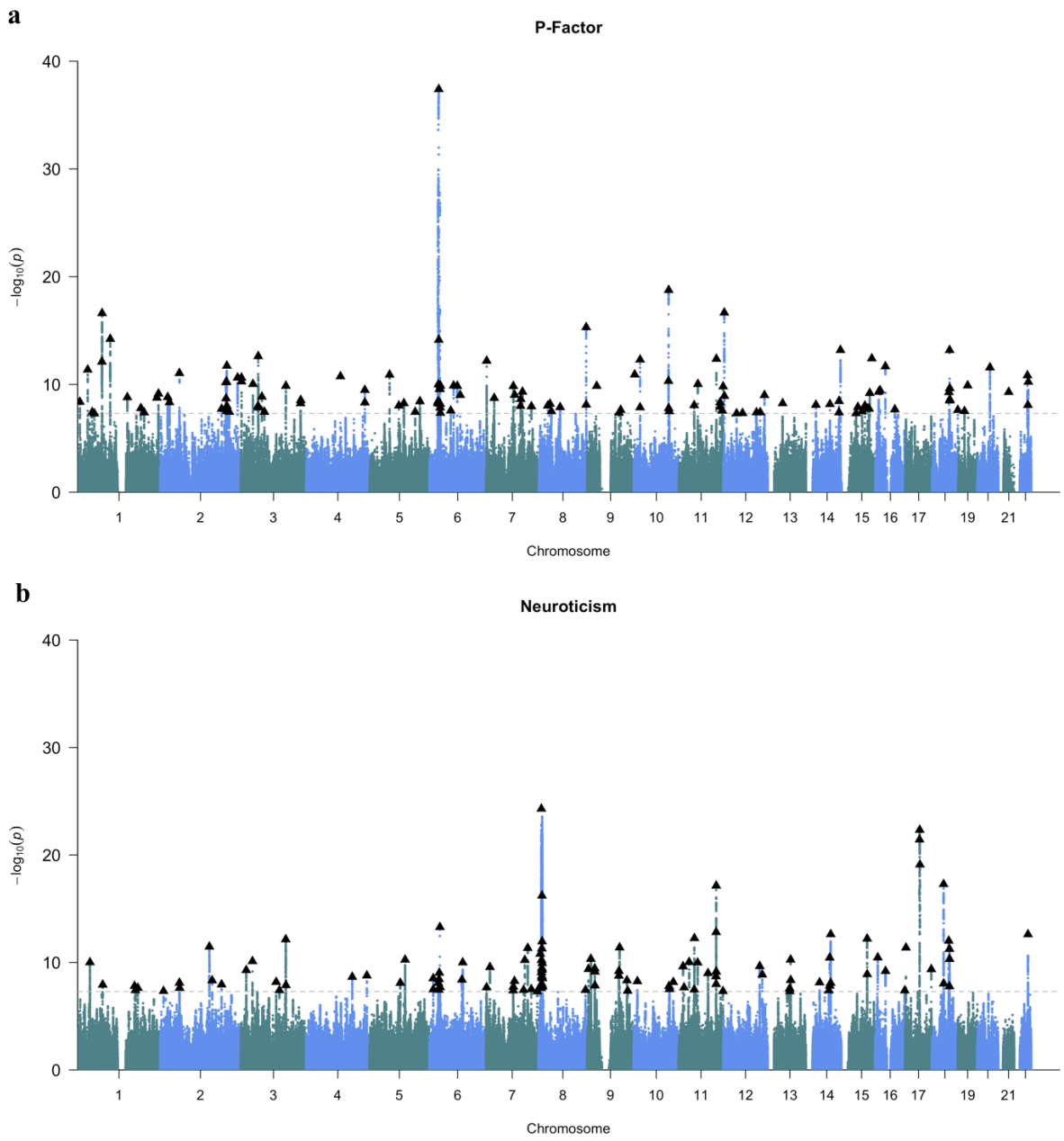
Supplementary Figure 9c. Manhattan plot for univariate Major Depression GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate GWAS. Purple diamonds denote independent hits for the univariate indicators that were not in LD with independent hits for the *p*-factor. Grey stars denote independent hits for Q_{SNP}.



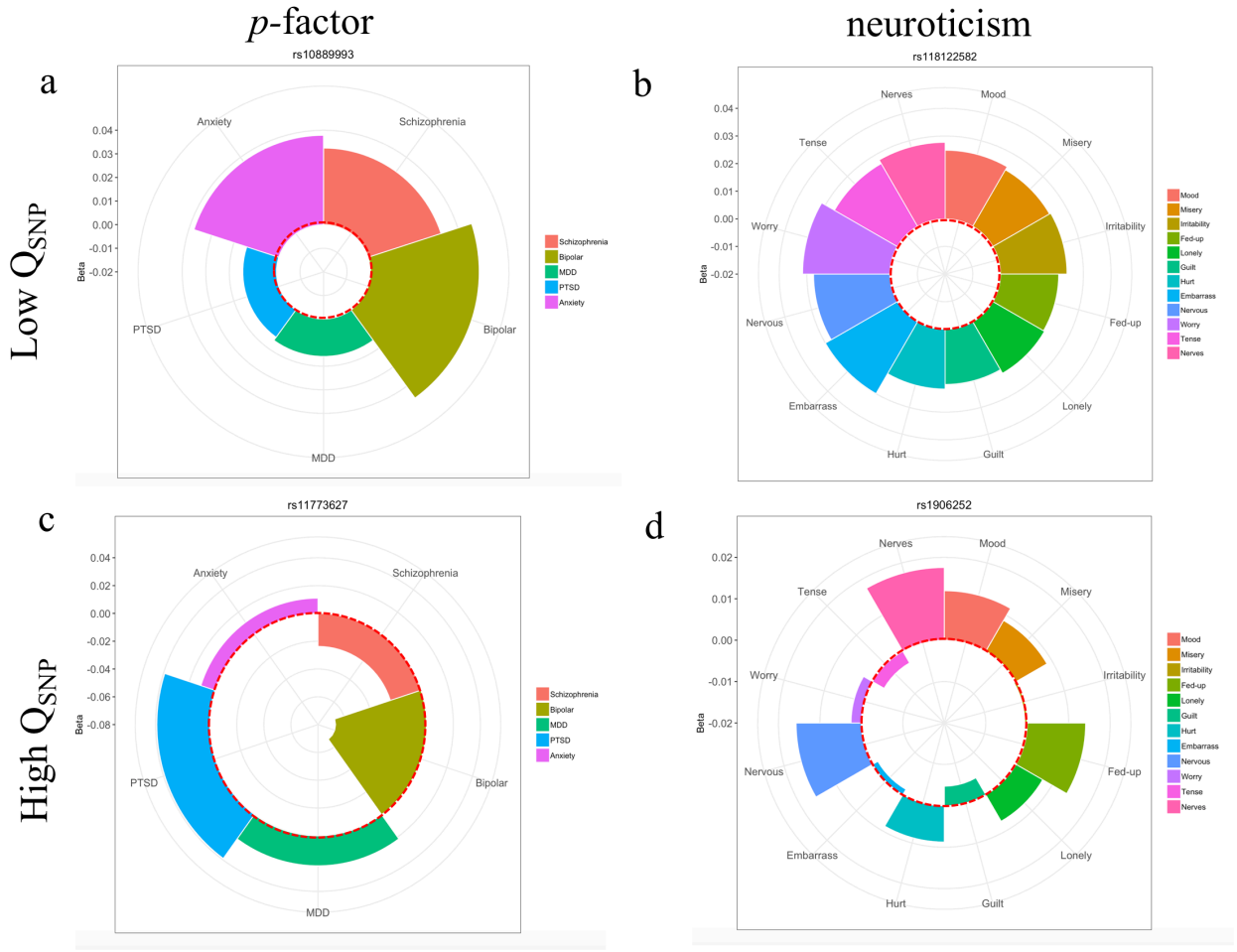
Supplementary Figure 9d. Manhattan plot for Post-Traumatic Stress Disorder GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate GWAS. Purple diamonds denote independent hits for the univariate indicators that were not in LD with independent hits for the *p*-factor. Grey stars denote independent hits for Q_{SNP}.



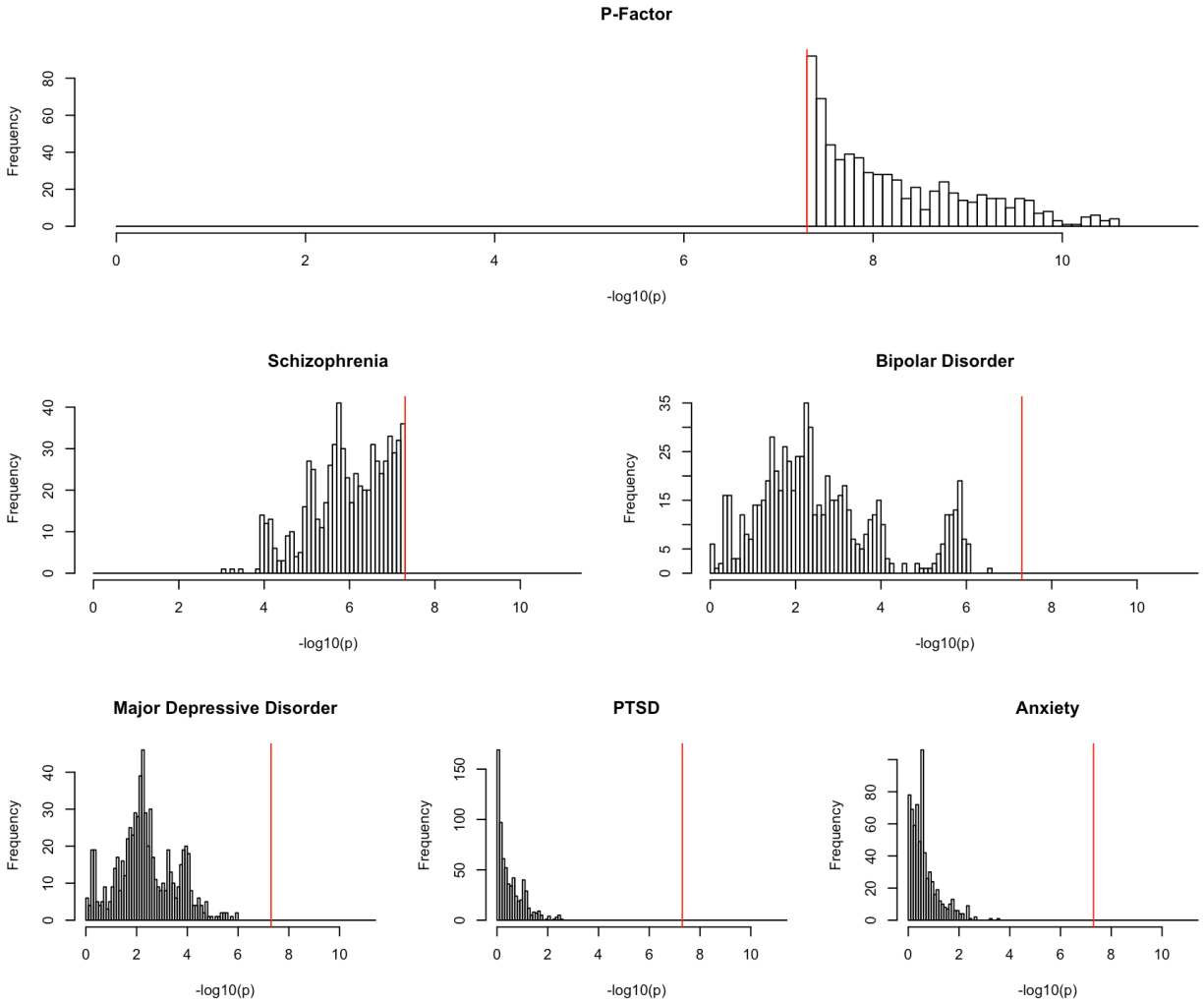
Supplementary Figure 9e. Manhattan plot for univariate Anxiety GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate GWAS. Purple diamonds denote independent hits for the univariate indicators that were not in LD with independent hits for the *p*-factor. Grey stars denote independent hits for Q_{SNP}.



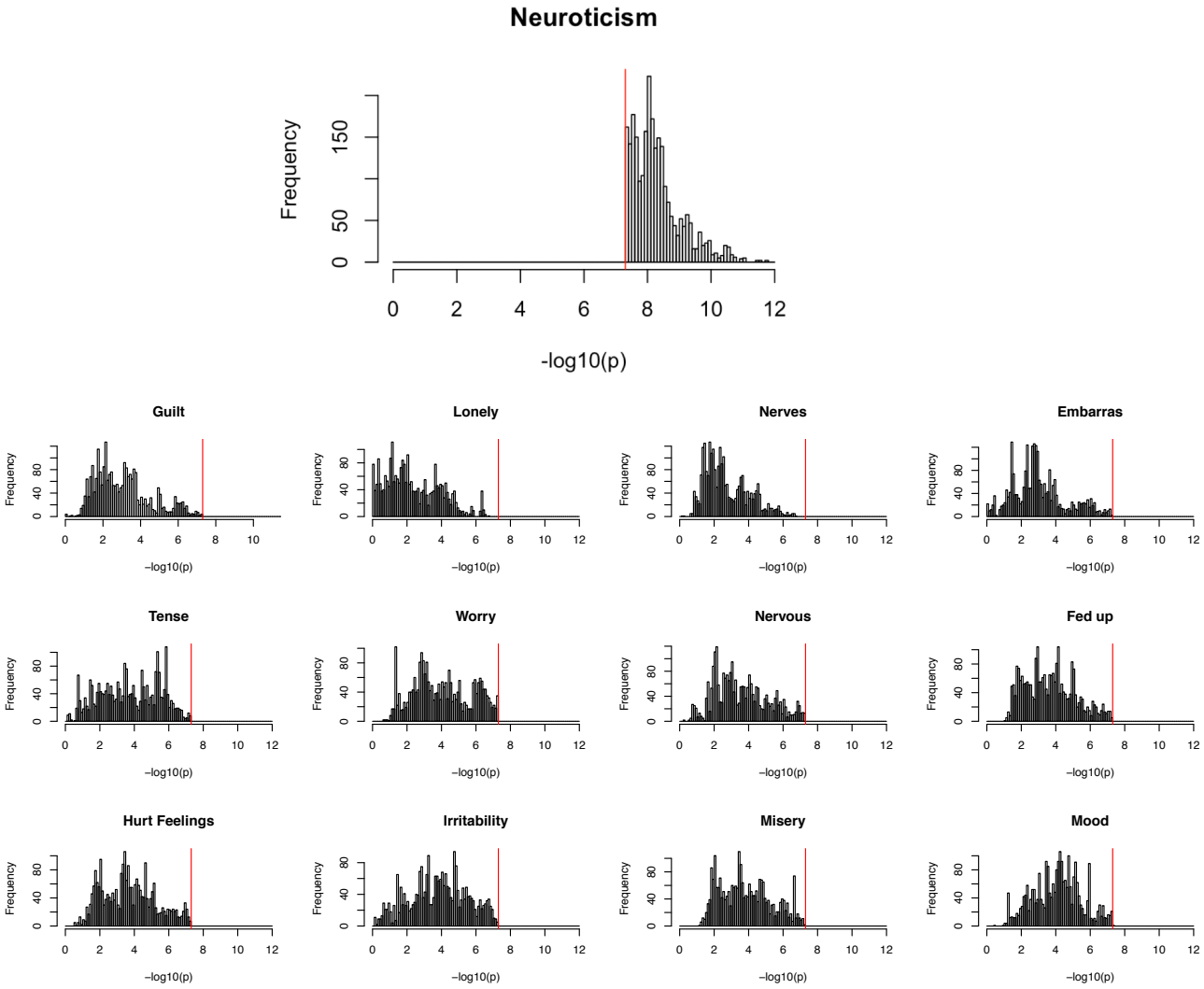
Supplementary Figure 10. Manhattan plots of hits from Genomic SEM. Genomic SEM (with WLS estimation) was used to conduct multivariate GWASs of the *p*-factor (panel a) and neuroticism (panel b). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). In both panels, black triangles denote independent hits for SNP effects from the GWAS of the general factor.



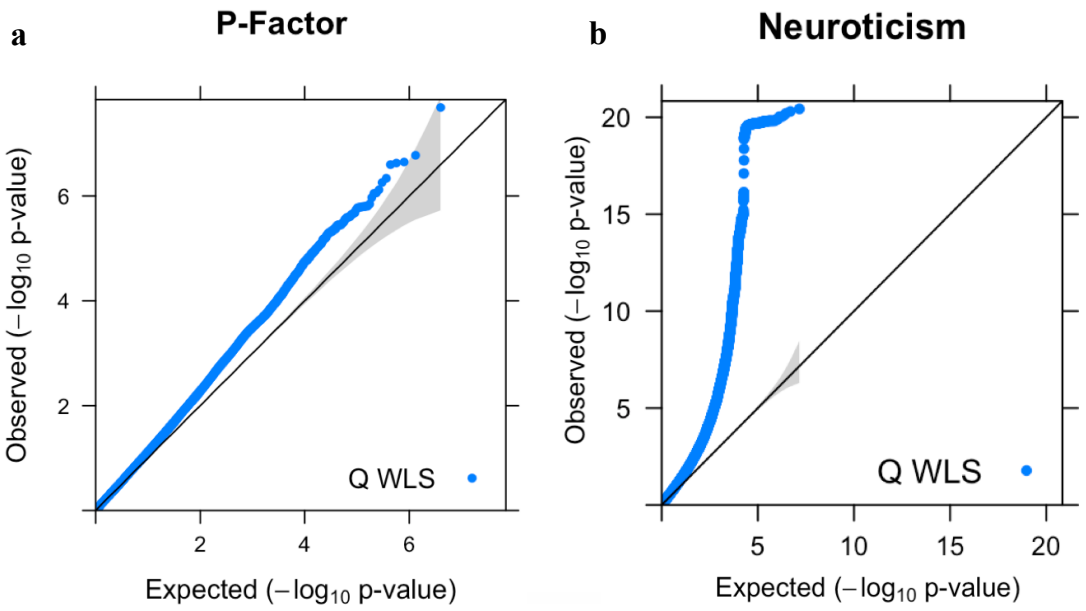
Supplementary Figure 11. Polar plots of item-level effects for genome-wide significant effects on common factors and for Q_{SNP}. All plots display the betas for the items standardized with respect to the total variance in the phenotype. Values inside and outside of the dotted red circle are negative and positive, respectively. The top panel displays item-level betas for a SNP that was genome-wide significant and produced low Q_{SNP} estimates for the *p*-factor (panel a; factor *p*-value = 7.78e-13; Q_{SNP} *p*-value = 0.57) and neuroticism (panel b; factor *p*-value = 5.06e-12; Q_{SNP} *p*-value = 0.77). As expected, the estimates in the top panel are both large in magnitude and consistent in direction across the items. The bottom panel displays item-level effects that produced genome-wide significant Q_{SNP} estimates for the *p*-factor (panel c; factor *p*-value = 5.32e-3; Q_{SNP} *p*-value = 2.02e-8) and neuroticism (panel d; factor *p*-value = 2.40e-4; Q_{SNP} *p*-value = 1.66e-14). Unlike the top panel, these SNPs are characterized by discrepant effects across the items with respect to magnitude and direction. This indicates that the Q_{SNP} test of heterogeneity is appropriately capturing discrepancy across genetic effects for the included phenotypes.



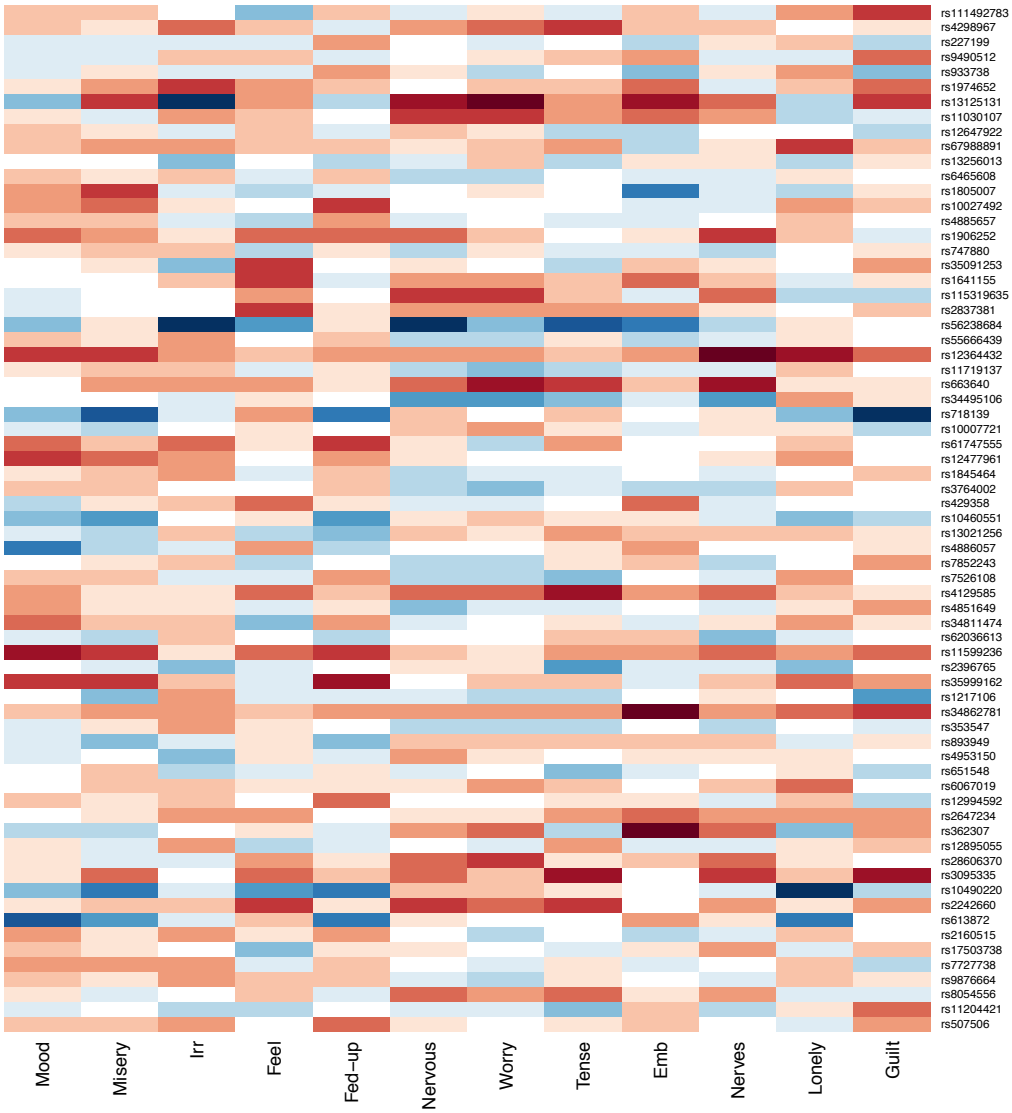
Supplementary Figure 12. Histograms of –log10 p-values for hits on *p*-factor. Histograms of –log10 p-values for the 684 non-independent SNPs that were genome wide significant for the *p*-factor using WLS estimation, but were not identified as significant in any of the individual GWASs. The vertical red line indicates genome wide significance.



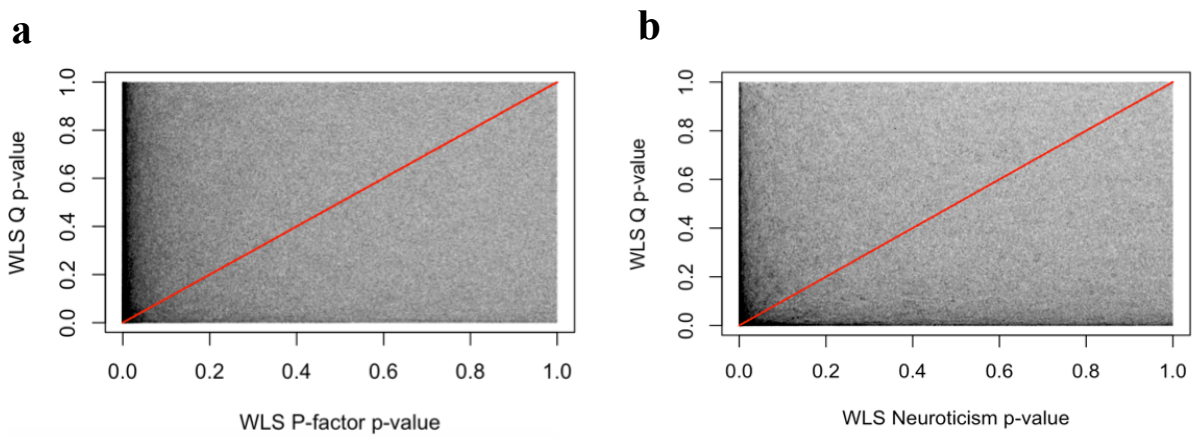
Supplementary Figure 13. Histograms of –log10 *p*-values for hits on neuroticism factor. Histograms of –log10 *p*-values for the 2,540 non-independent SNPs that were genome wide significant for neuroticism using WLS estimation, but were not identified as significant in any of the individual GWASs. The vertical red line indicates genome wide significance.



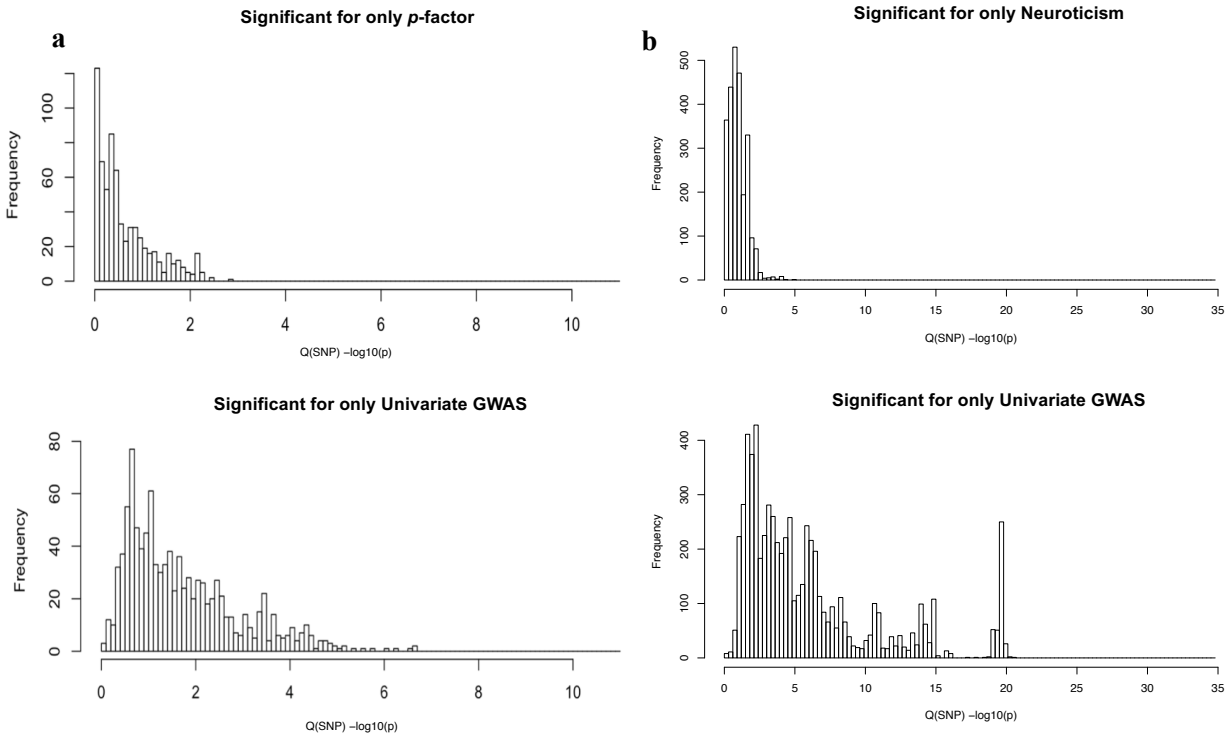
Supplementary Figure 14. Quantile-quantile plot for Q_{SNP}. Estimates are from WLS estimation for the *p*-factor (panel a) and neuroticism (panel b). Expected −log10 *p*-values are those expected under the null hypothesis. The shaded area indicates the 95% confidence interval under the null. As some Q_{SNP} estimates for neuroticism were quite large, *p*-values < 5^{-20} were set to 5^{-20}.



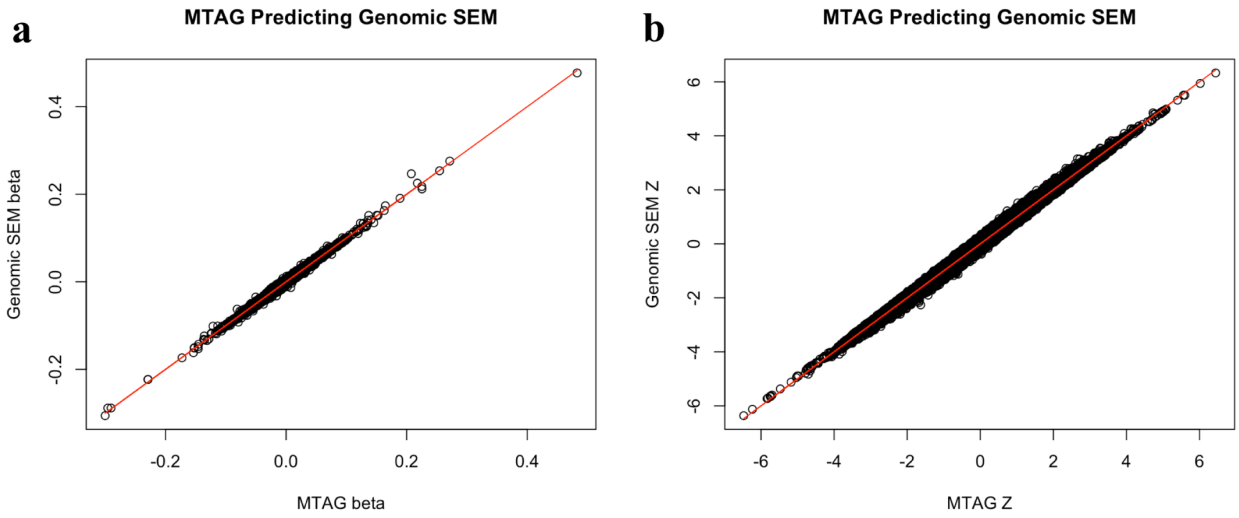
Supplementary Figure 15. Heatmap of univariate betas for neuroticism indicators for Q_{SNP} hits. The heatmap depicts univariate item-specific betas for the 69 lead SNPs for Q_{SNP} identified using WLS estimation for neuroticism. Items are on the x-axis. SNPs are on the y-axis. Cells depicted in red, white, and blue indicate negative, near zero, and positive betas, respectively. As expected, individual rows indicate substantial heterogeneity across the indicators for hits on Q_{SNP}.



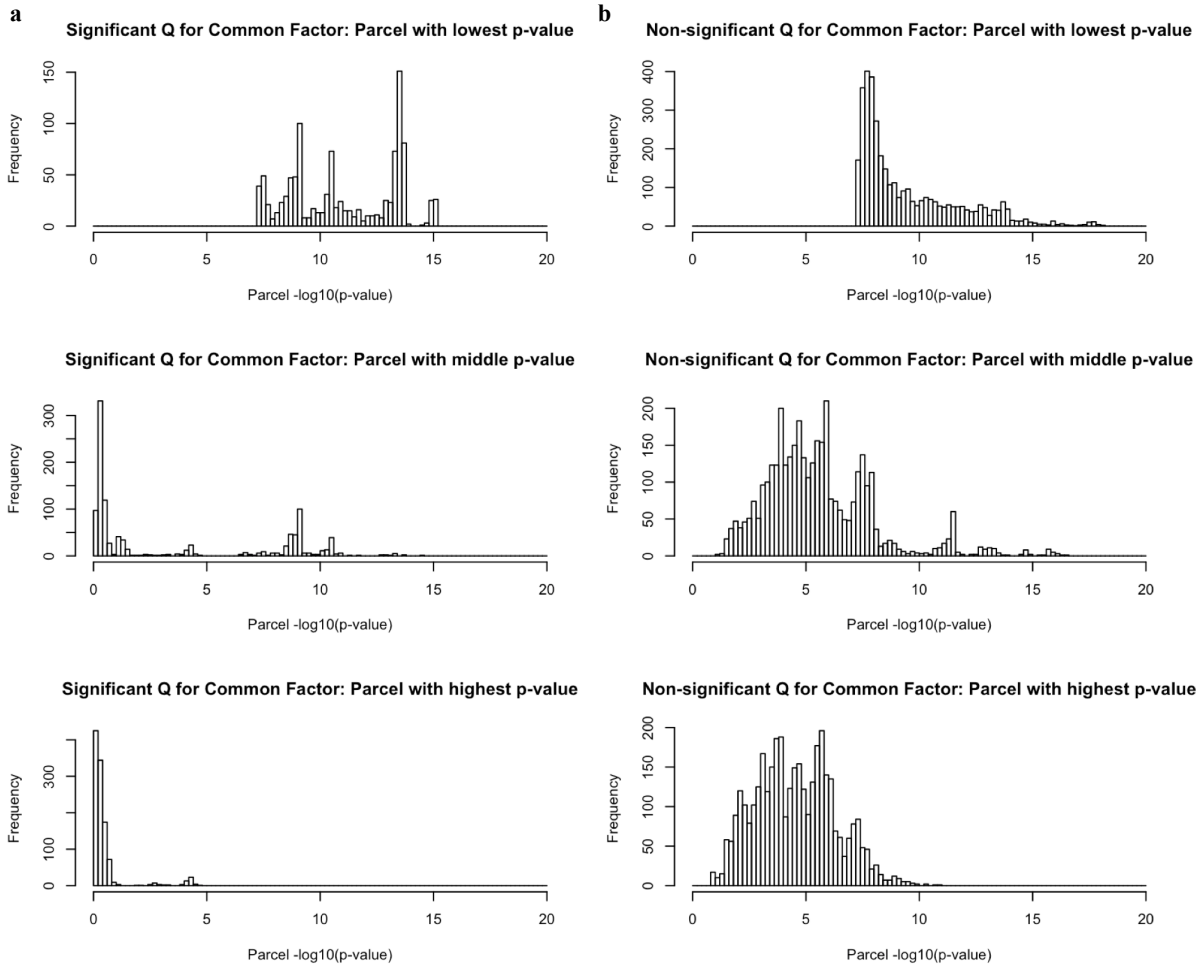
Supplementary Figure 16. Association between SNP effects on the common factor and Q_{SNP} effects. The association between the *p*-values for SNP effects on the common factor (x-axis) and the p-values for Q_{SNP} (y-axis) are plotted for WLS estimation of the *p*-factor (panel a) and neuroticism (panel b). The red line reflects the regression line for the common factor *p*-value predicting itself (i.e., a slope of 1), with dots above the line estimated as less significant for Q_{SNP}. The correlation between these two outcomes was .02 for the *p*-factor and .05 for neuroticism.



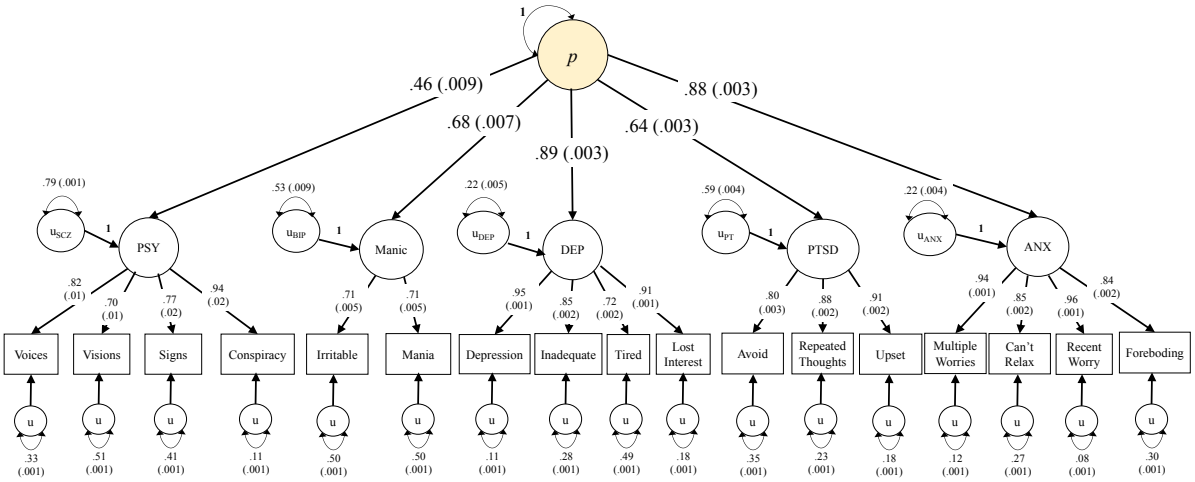
Supplementary Figure 17. Q_{SNP} –log10 *p*-values for common-factor and indicator-specific hits. Results are depicted for WLS estimation of the *p*-factor (panel a) and neuroticism (panel b). There were 684 non-independent SNPs identified as genome-wide significant for *p*-factor, but not the univariate GWAS, and 1,022 indicator-specific SNPs. For neuroticism, there were 2,540 non-independent hits specific to the common factor and 6,523 hits specific to the indicators. The average –log10 Q_{SNP} *p*-value was 0.61 for hits only on the *p*-factor and 1.81 for hits specific to the univariate indicators. For neuroticism, the average –log10 Q_{SNP} *p*-value was 0.95 for hits unique to the common factor and 5.95 for hits unique to the indicators. Thus, Q_{SNP} values were generally more significant for those SNPs not identified as significant for the common factor.



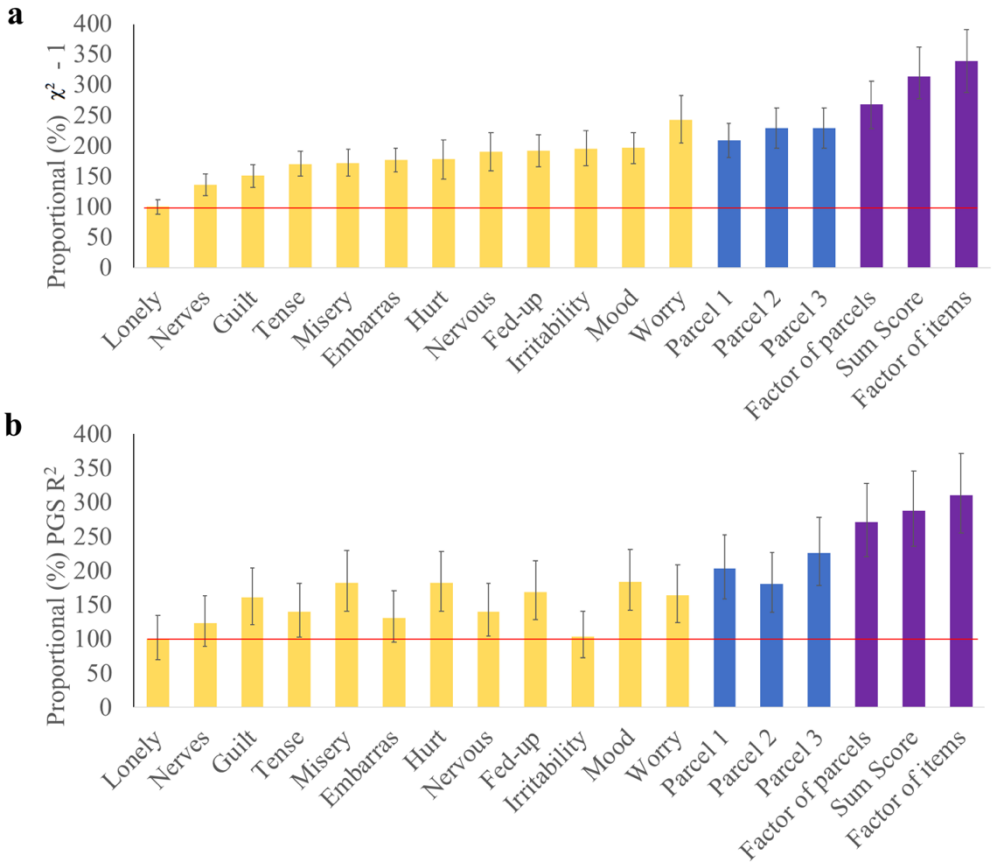
Supplementary Figure 18. MTAG predicting Genomic SEM specified as MTAG. Panel a depicts the MTAG beta predicting the Genomic SEM formulation of MTAG beta (b = .998, intercept = -1.56E-7, R^2 = .994). Panel b depicts MTAG Z-statistic predicting the Genomic SEM formulation of MTAG Z-statistic (b = .999, intercept = 2.65E-4, R^2 = .999). For both panels, the red line reflects the regression line for MTAG predicting itself (i.e., a slope of 1).



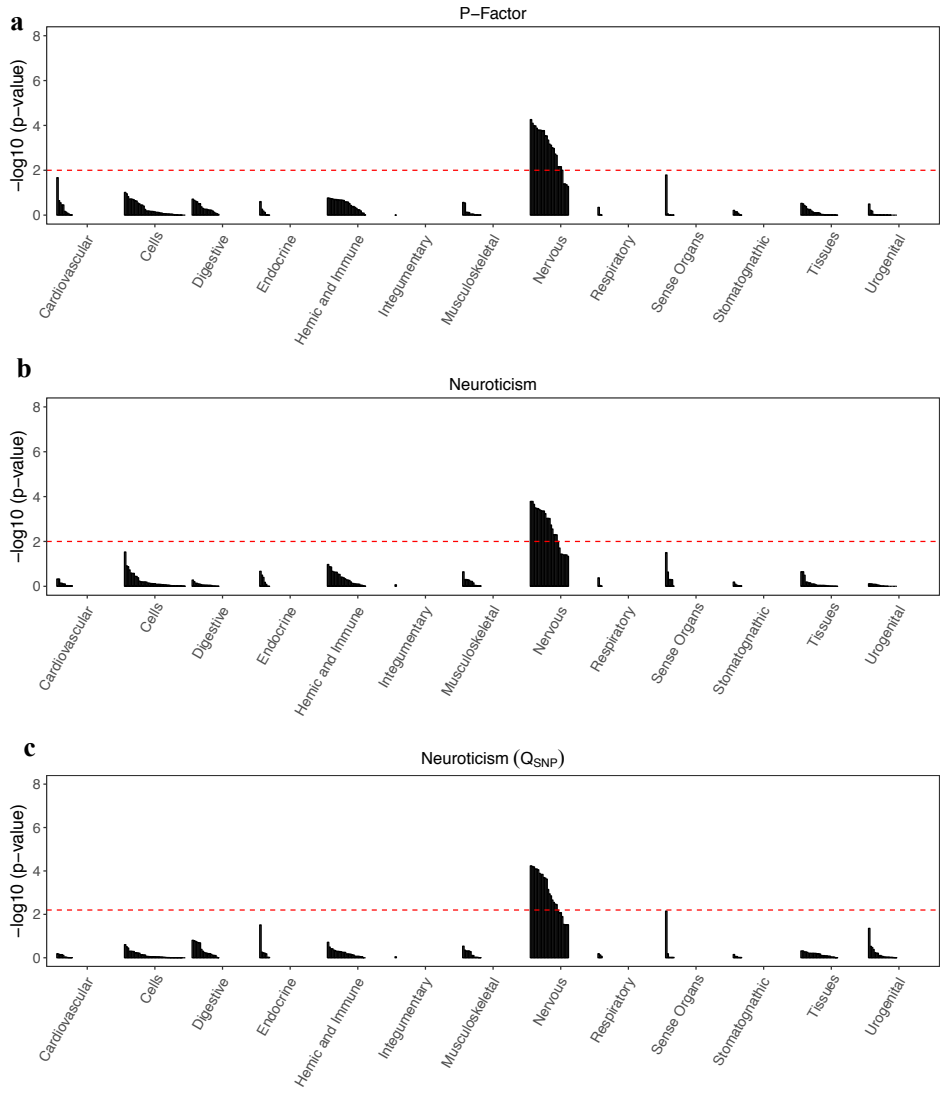
Supplementary Figure 19. Comparison of parcel *p*-values for low versus high Q_{SNP} estimates. Histograms shown for SNPs that produced genome wide significant hits for at least one of the parcels split across high (Q_{SNP} *p*-value < 5e-8; 1,090 SNPs; panel a) and low (*p* > 5e-3; 3,685 SNPs; panel b) Q_{SNP} estimates as estimated using WLS for the common neuroticism factor. For those SNPs characterized by a larger degree of heterogeneity, as indexed by Q_{SNP}, there was a corresponding heterogeneity in the *p*-values at the level of the parcel.



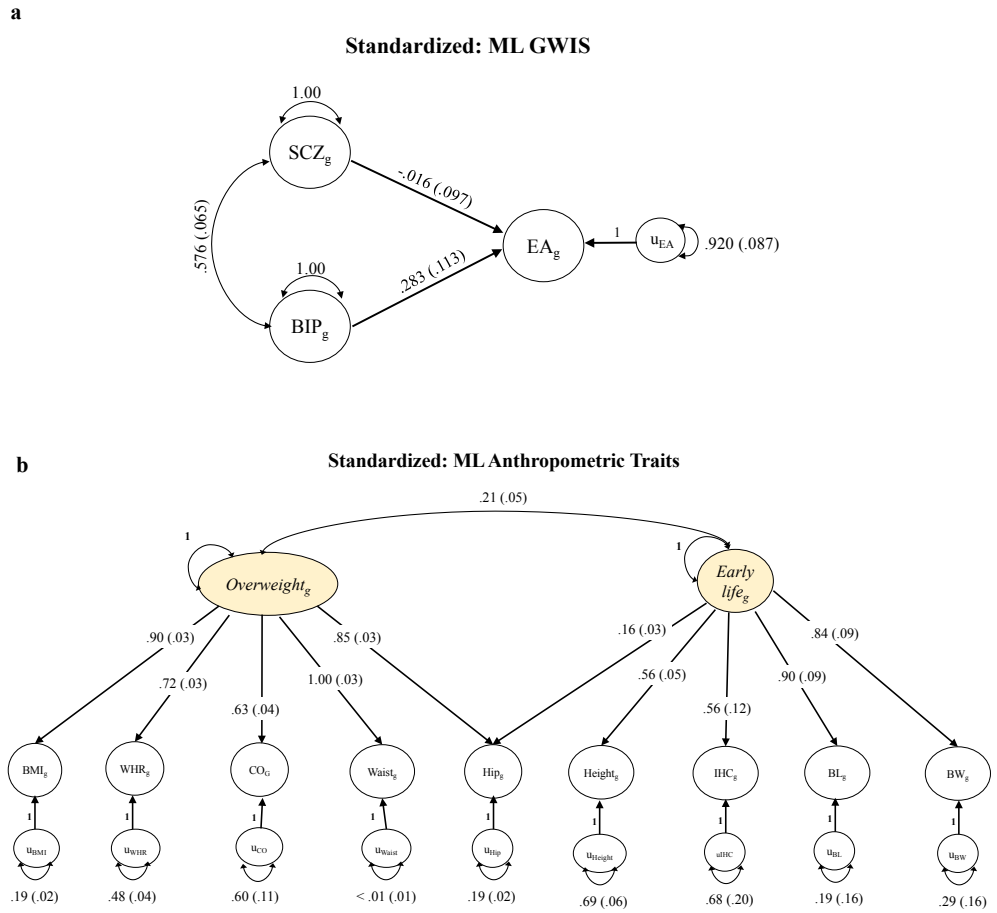
Supplementary Figure 20. UK Biobank *p*-factor. Standardized output of phenotypic *p*-factor constructed from UKB phenotypes for out of sample prediction using *p*-factor polygenic scores. PSY = psychotic experiences; DEP = depressive symptoms; PTSD = symptoms of post-traumatic stress disorder; ANX = anxious symptoms.



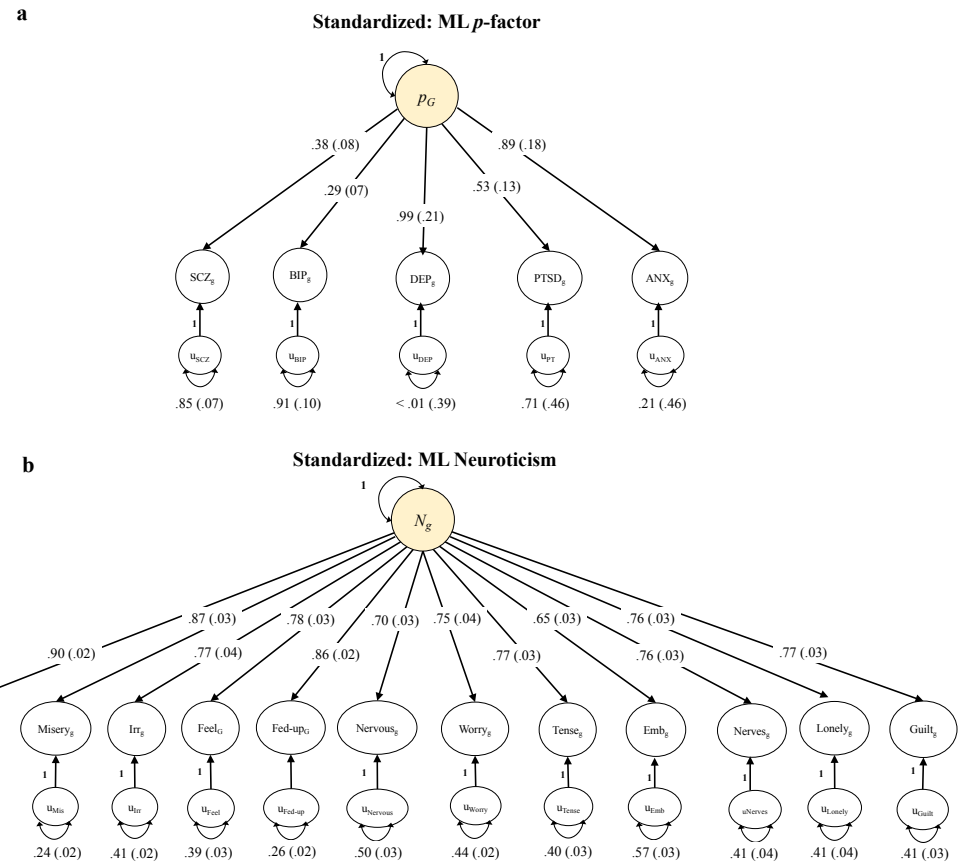
Supplementary Figure 21. Relative power and out-of-sample prediction for neuroticism. Panel **a** represents relative power of GWAS summary statistics for individual neuroticism items (yellow), parcels (blue), and factor of parcels, sum score, and factor of items (purple) from in the UKB discovery sample. Relative power is indexed by the proportion (expressed as a percentage) in the average $\chi^2 - 1$ across summary statistics relative to the lonely item (panel), which is the item with the smallest average χ^2 value. We subtract 1 because the mean of the null χ^2 distribution is equal to its degrees of freedom. Panel **b** represents relative prediction in the Generation Scotland sample for polygenic scores (PGSs) derived from GWAS sumstats for individual neuroticism items (yellow), parcels (blue), and factor of parcels, sum score, and factor of items (purple). The proportional R^2 (%) is relative to the R^2 for the lonely item PGS. PGSs were constructed using the same set of SNPs for all predictors. The summary statistics for Genomic SEM were estimated using WLS. Error bars indicate 95% confidence intervals. For both panels, the red line is drawn at 100%, to indicate distance from the lonely item baseline. The superior performance of Genomic SEM analysis of the common factor of items relative to the sum score of the items is likely, 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 GWAS summary data across phenotypes.



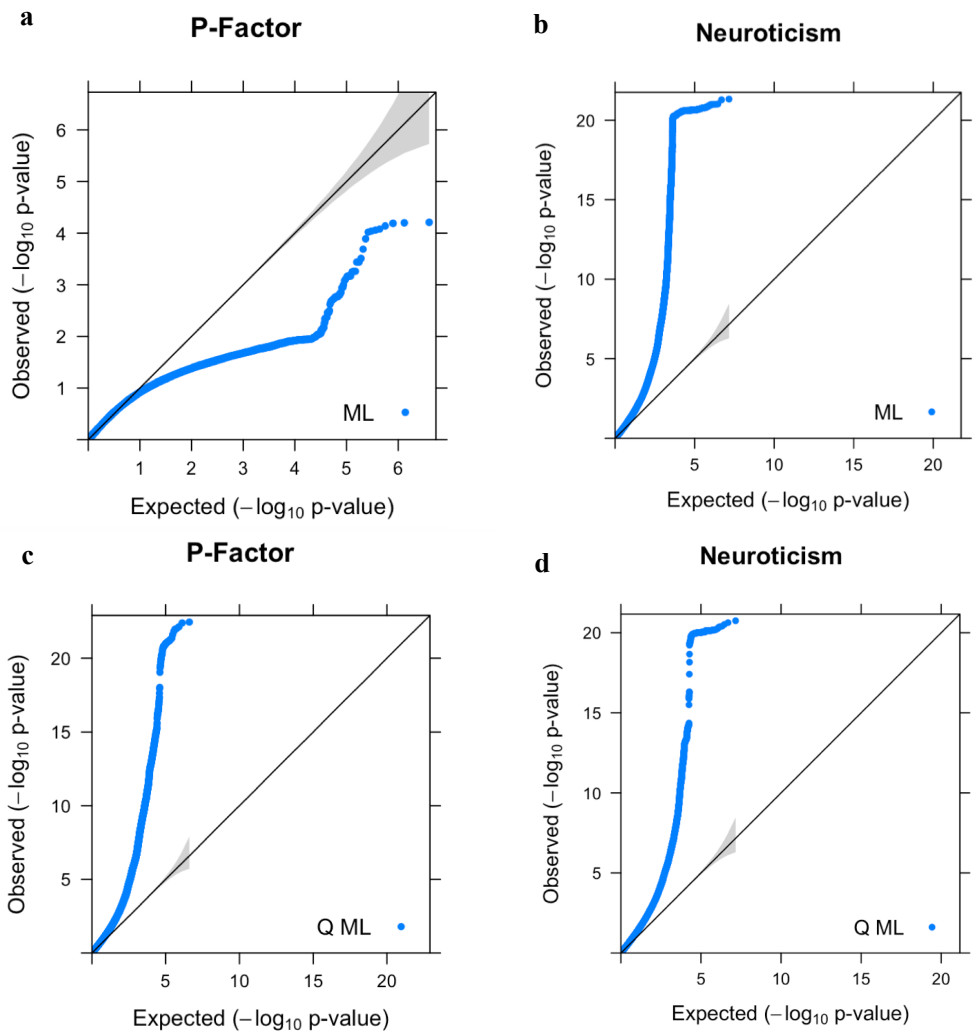
Supplementary Figure 22. Biological annotation of Genomic SEM results for *p*-factor, neuroticism, and Q_{SNP} of neuroticism. Results from tissue enrichment analyses conducted using DEPICT based on Genomic SEM results for the p-factor (panel a), neuroticism (panel b) and Q_{SNP} estimation for neuroticism (panel c) using WLS estimation. The red, dashed line indicates the false discovery rate at .05. As expected, the majority of enriched tissues were in the nervous system for both common factors and Q_{SNP} estimates.



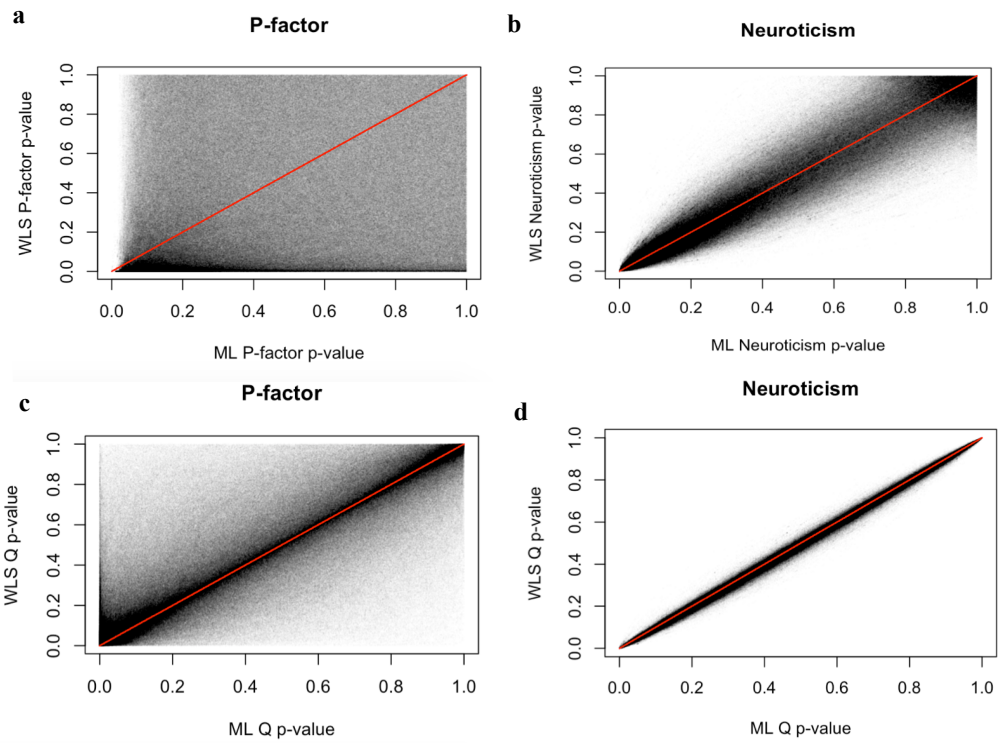
Supplementary Figure 23. ML Estimates from GWIS and anthropometric trait Genomic SEM models. Results are presented for standardized output for the multiple regression model of GWIS (panel a) and the confirmatory factor model of anthropometric traits (panel b). *SE*s are shown in parentheses. The genetic correlation matrix (standardized) and associated sampling covariance matrix were used as input for Genomic SEM. BMI = body mass index; WHR = waist-hip ratio; CO = childhood obesity; IHC = infant head circumference; BL = birth length; BW = birth weight.



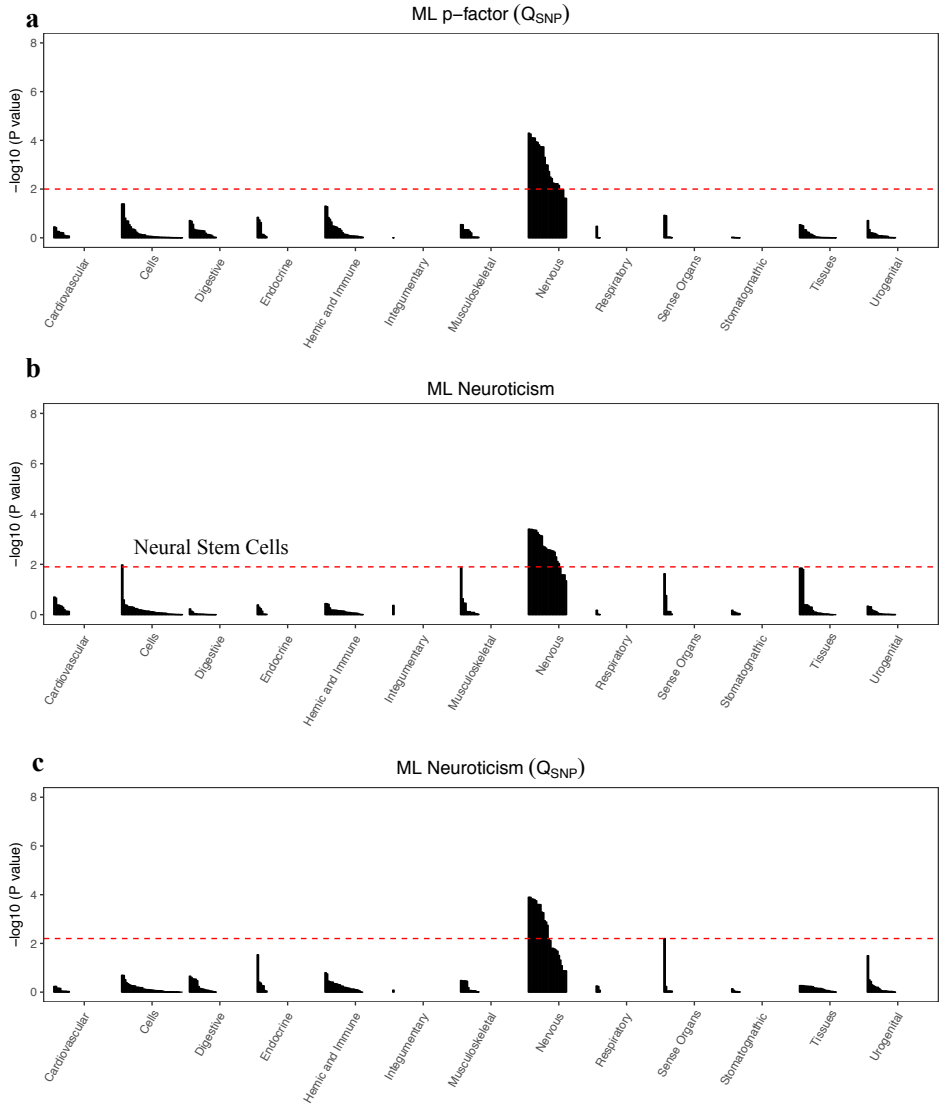
Supplementary Figure 24. ML estimates for neuroticism and *p*-factor Genomic SEM models. Results are presented for standardized output for the confirmatory factor models of the *p*-factor (panel a) and neuroticism (panel b). The genetic correlation matrix (standardized) and associated sampling covariance matrix were used as input for Genomic SEM. *SE*s are shown in parentheses. SCZ = schizophrenia; BIP = bipolar disorder; DEP = major depressive disorder; PTSD = post-traumatic stress disorder; ANX = anxiety. Irr = irritability; Feel = sensitivity/hurt feelings; fed-up = fed-up feelings; emb = worry too long after embarrassment.



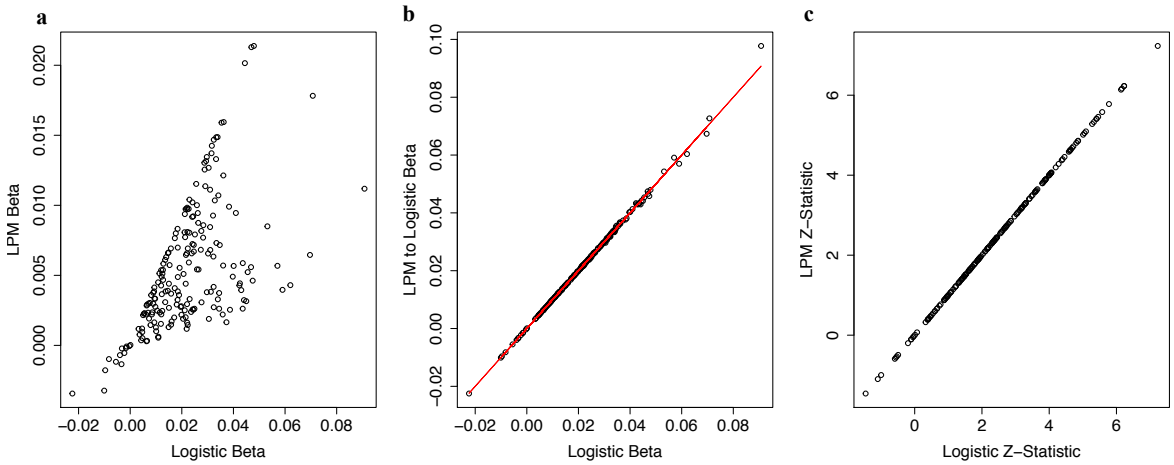
Supplementary Figure 25. Quantile-quantile plot of multivariate GWAS *p*-values for *p*-factor and neuroticism (ML estimation). Estimates are from ML estimation for the *p*-factor (panel a), neuroticism (panel b), Q_{SNP} estimates for the *p*-factor (panel c), and Q_{SNP} estimates for neuroticism (panel d). Expected −log10 *p*-values are those expected under the null hypothesis. The shaded area indicates the 95% confidence interval under the null. As some Q_{SNP} estimates for neuroticism were quite large, *p*-values < 5^{-20} were set to 5^{-20}.



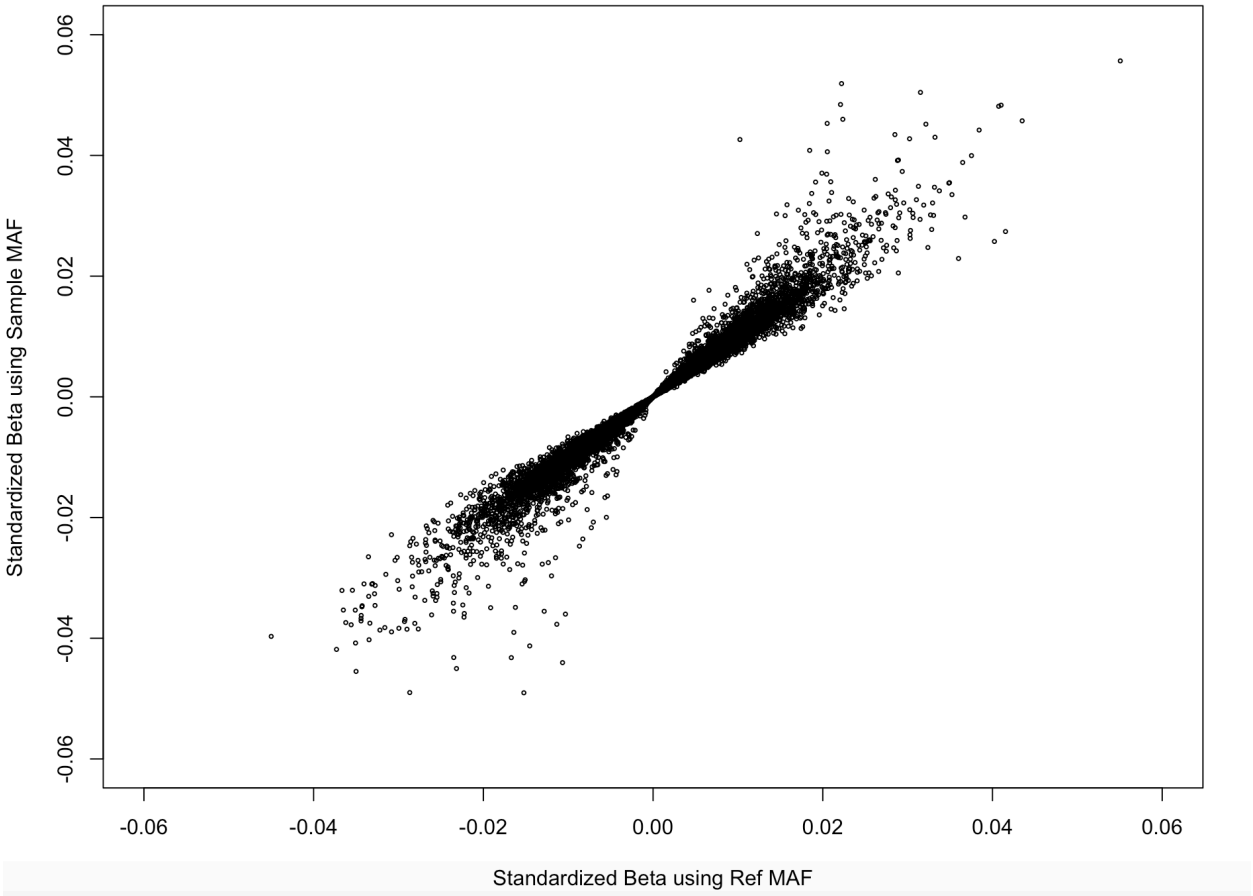
Supplementary Figure 26. Associations between ML and WLS *p*-values. Scatter plot comparing *p*-values between WLS (y-axis) and ML (x-axis) estimation for the *p*-factor (panel a), neuroticism (panel b), Q_{SNP} for the *p*-factor (panel c), and Q_{SNP} for neuroticism (panel d). The red line reflects the regression line for ML predicting itself (i.e., a slope of 1), with dots above the line estimated as less significant for WLS. The correlation between the two sets of common factor *p*-values (top panel) was .15 for the *p*-factor and .94 for neuroticism. The correlation between the two Q_{SNP} statistics (bottom panel) for neuroticism was > .99 and .77 for the *p*-factor . Thus, the rank-ordering is largely maintained across the estimation methods, but may diverge, in particular, for factor effects.



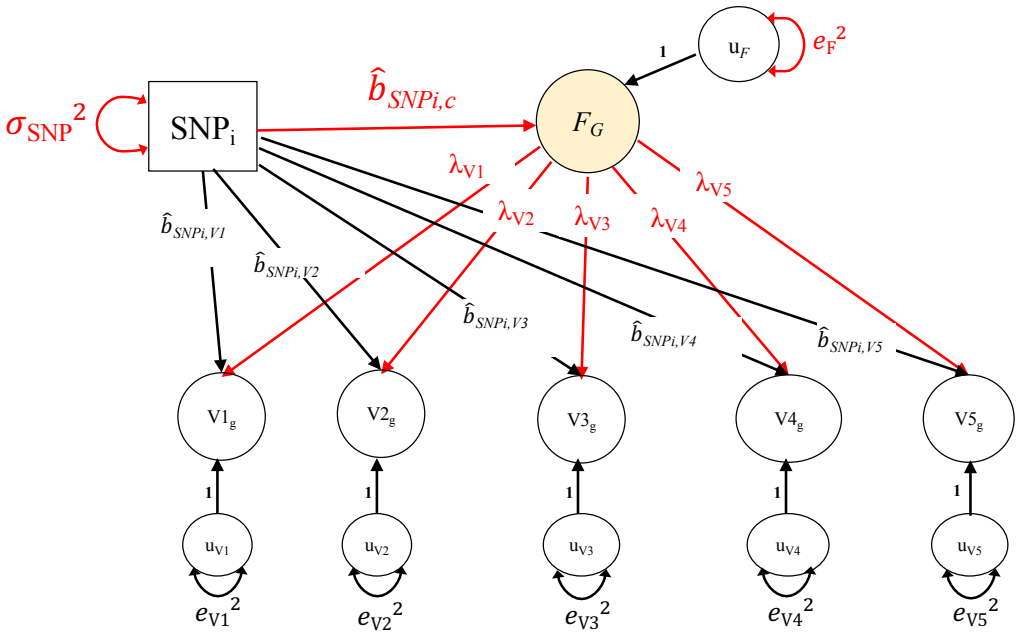
Supplementary Figure 27. Biological annotation of Q_{SNP} of *p*-factor, neuroticism, and Q_{SNP} of neuroticism using ML estimation. Results from tissue enrichment analyses conducted using DEPICT based on Genomic SEM results for Q_{SNP} of the *p*-factor (panel a), neuroticism (panel b) and Q_{SNP} estimation for neuroticism (panel c) using ML estimation. The red, dashed line indicates the false discovery rate at .05. As expected, the majority of enriched tissues were in the nervous system for both common factor and Q_{SNP} estimates.



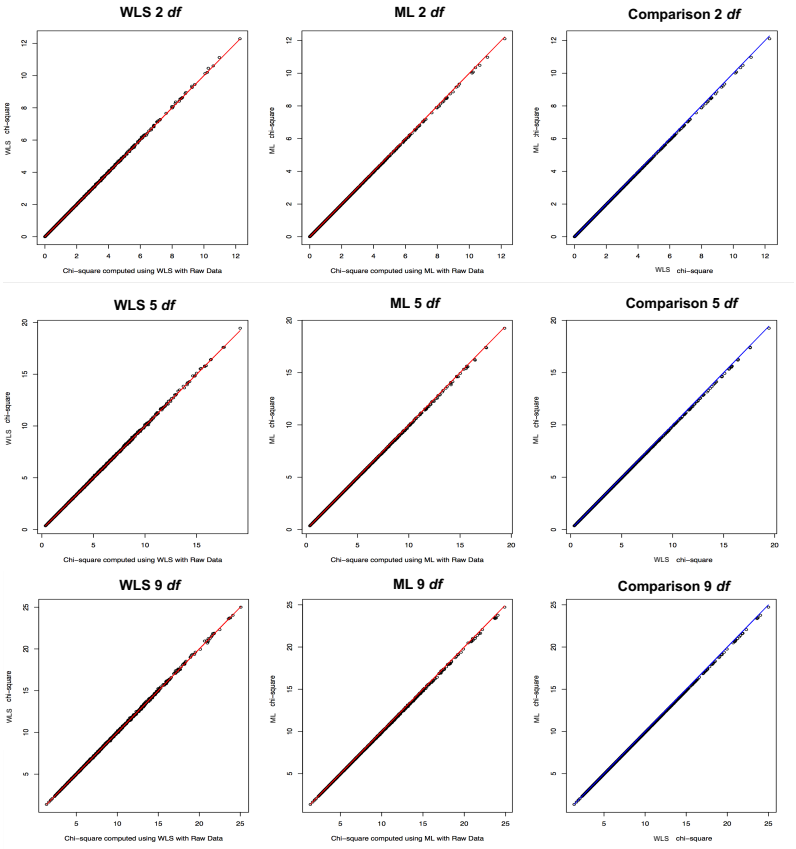
Supplementary Figure 28. Associations between regression coefficients from logistic regression model and linear probability model. 200 datasets were simulated with 100,000 observations each in which a continuously distributed liability was specified to be a linear function of a biallelic autosomal SNP and a normally distributed residual. Population effect sizes were randomly generated for each simulation, from within the range of 0 to .04 SD units per effect allele. The outcome was then dichotomized using a randomly generated threshold for each simulation within the range of -1.96 and 1.96 standard deviations from the mean of the liability distribution (i.e. the population prevalence of cases ranged between 2.5% to 97.5%). The population minor allele frequency of the SNP was randomly generated for each replication from within the range of 0 to .5. Panel a depicts the association between the betas obtained from a logistic regression of a SNP predicting the dichotomous outcome (x-axis) and from the betas obtained from a linear probability model (LPM) applied to the same data (y-axis; $r = .70$). Panel b depicts the same x-axis and the LPM output converted to logistic betas on the y-axis ($r > .99$). Panel c depicts the z-statistics (the coefficient divided by its standard error) for the logistic betas (x-axis) and LPM betas (y-axis, $r > .99$). The red lines depict the regression line (slope = 1, intercept = 0) for the logistic betas (panel b) and logistic z-statistics (panel c) predicting themselves. Thus, LPM output must be rescaled before effect sizes (i.e., regression coefficients) can be used for multivariate GWAS in Genomic SEM. However, LPM Z statistics can be used directly for LDSC to produce heritabilities and genetic covariances (the liability scale estimates should still be requested).



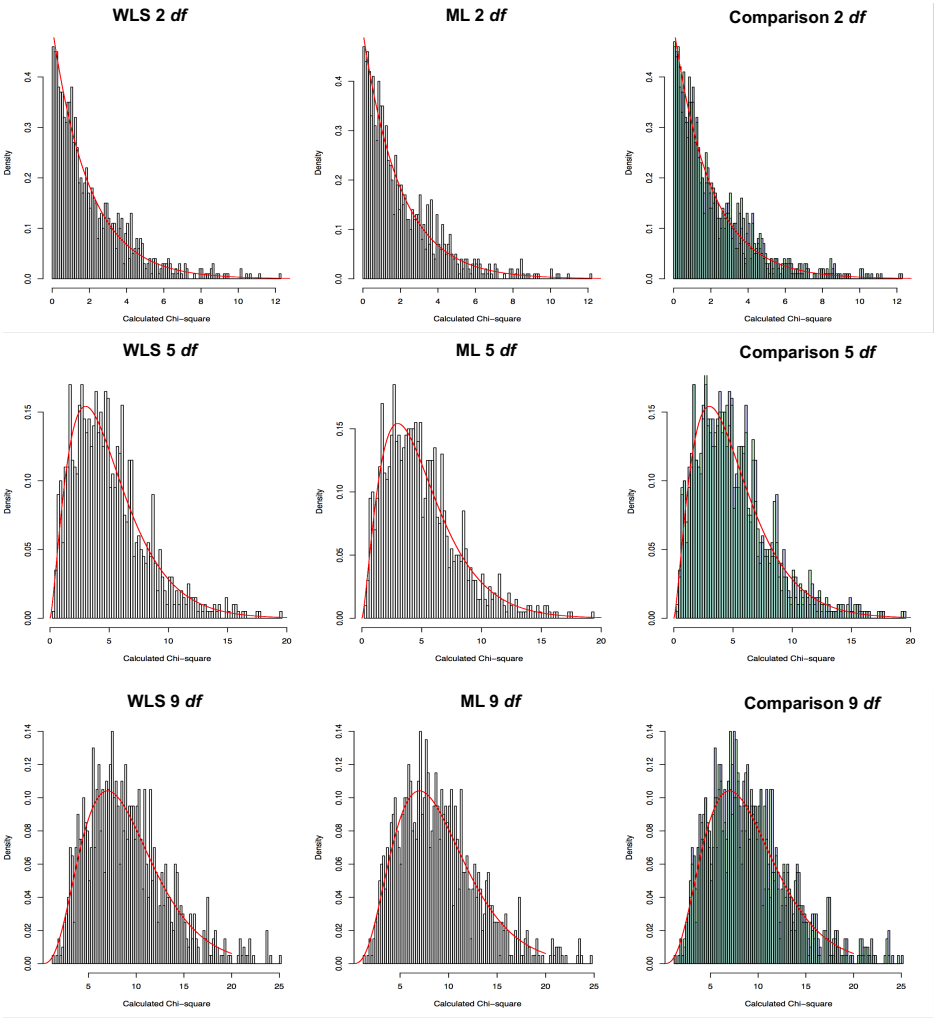
Supplementary Figure 29. Associations between betas standardized using reference panel or sample MAF. 30,000 SNPs were randomly selected from the mood UK Biobank phenotype and converted to approximate logistic regression effects and scaled relative to unit-variance scaled liability using either MAFs from a reference panel (1000 Genomes Phase 3; x-axis) or MAFs from the sample (UK Biobank; y-axis). Regardless of the MAFs used for standardization, the correspondence between the betas was very strong (r = .987, slope = 1.044, intercept = -6.54e-6). Although the use of either sample or reference MAF may be appropriate for different reasons, these results indicate that the decision will produce very similar estimates.



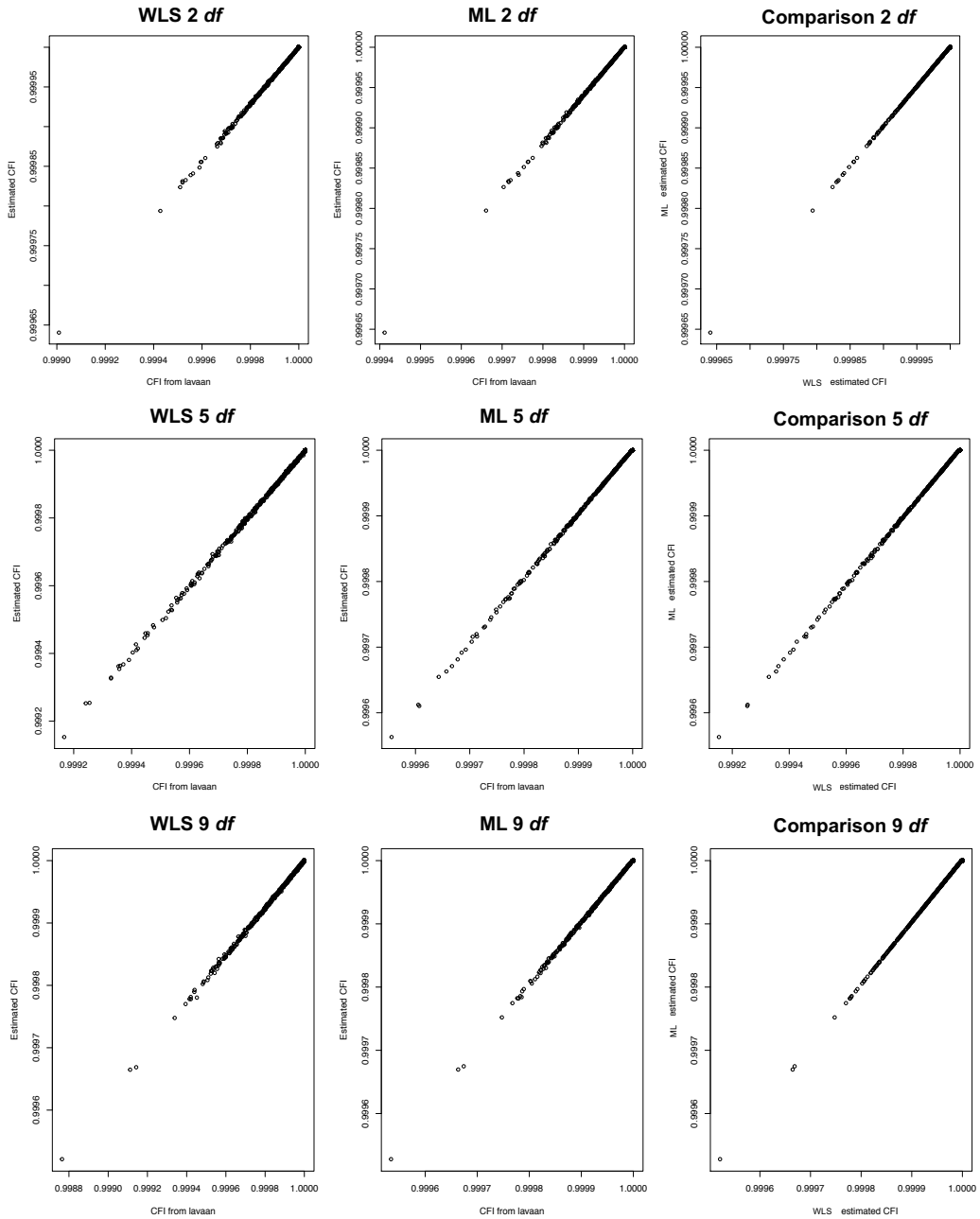
Supplementary Figure 30. Genomic SEM models for estimating Q_{SNP}. Red lines and parameters are fixed from Step 1, and black lines and parameters are freely estimated in Step 2.



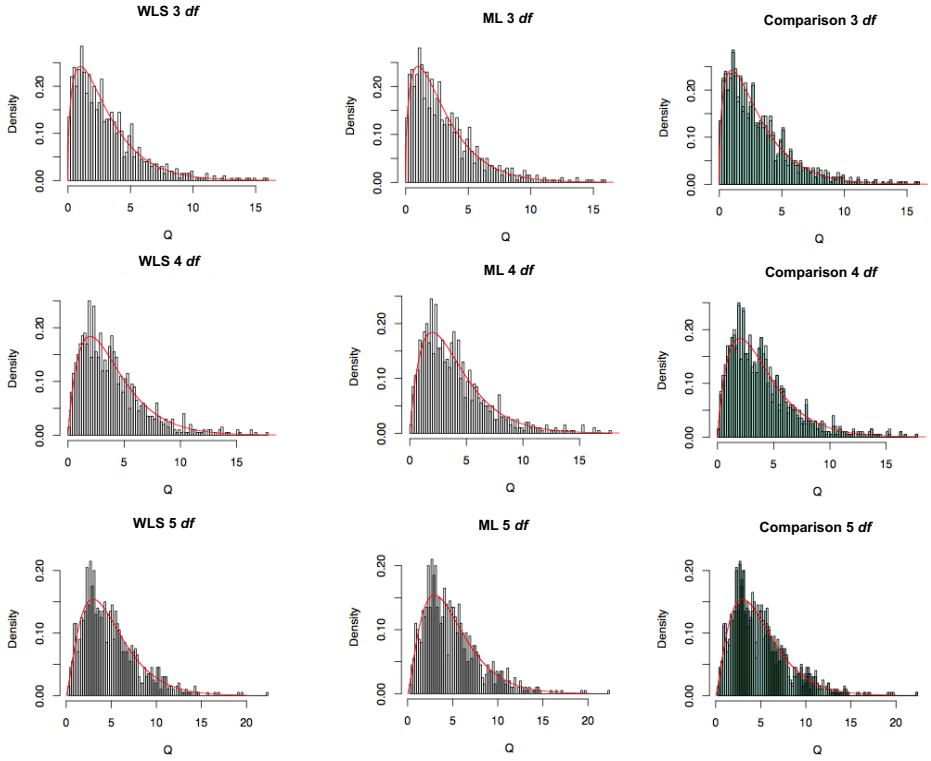
Supplementary Figure 31. Associations between model χ^2 values computed from summary data and model χ^2 values computed from raw data. Raw data-based estimates of model χ^2 were computed directly from the data using lavaan. Summary data-based estimates of model χ^2 were computed using the *S* and *V* matrices with WLS (left) and ML (right) estimation. The red line in the middle and left panel reflects the regression line for the raw data-based model χ^2 predicting itself. The blue line in the right panel reflects the regression line for the WLS χ^2 predicting itself.



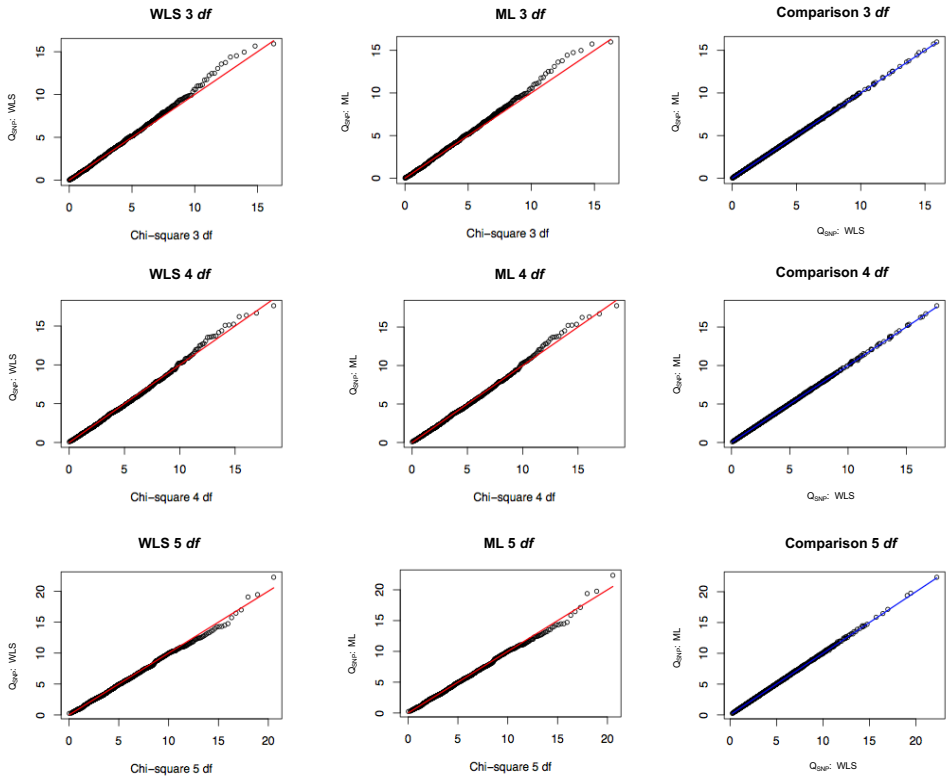
Supplementary Figure 32. Distributions of calculated and theoretical χ^2 statistics. Comparison between distribution of χ^2 values for model estimated from S and V matrices using WLS (left column) and ML (middle column) against a theoretical χ^2 distribution. The right column compares the distributions of WLS (blue bars) and ML (green bars).



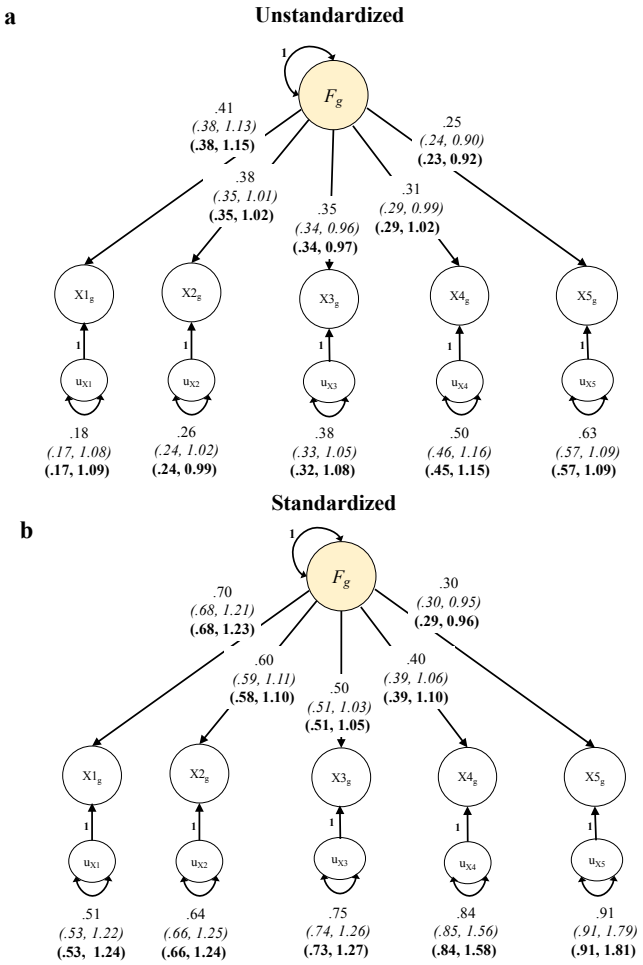
Supplementary Figure 33. Associations between CFI values derived from summary data and CFI values derived from raw data. Summary data-based estimates of CFI are depicted for models estimated using WLS (left column) and ML (middle column). We also present comparisons of the CFI from models estimated with ML and those estimated with WLS(right column). All CFI estimates were bounded at 1.



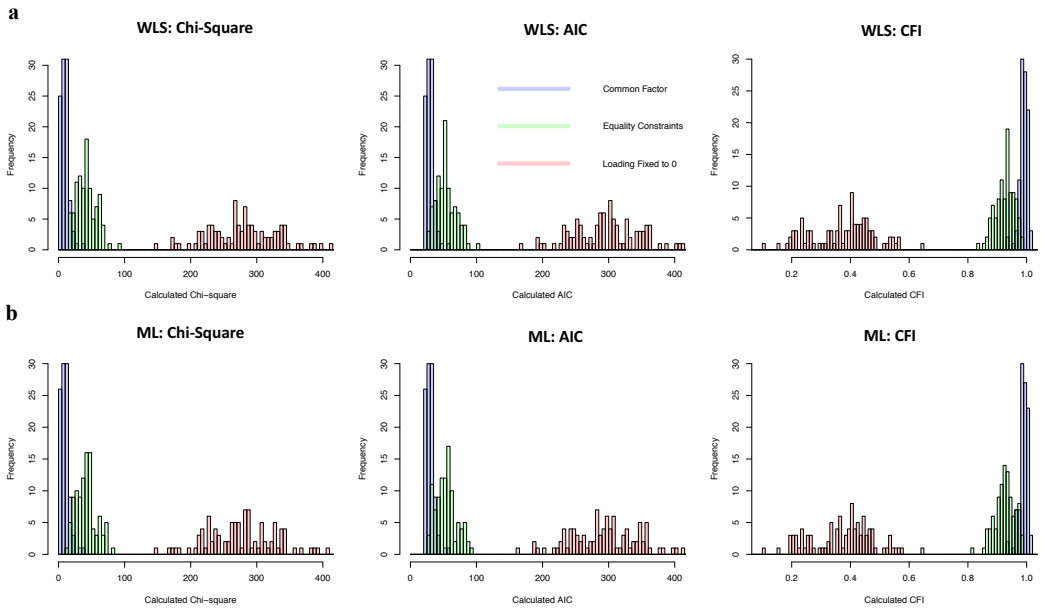
Supplementary Figure 34. Null distributions of Q_{SNP} for 1,000 simulations per model. Red lines for all panels depict the chi-square distribution with the relevant *df*. The top, middle, and bottom panels depict the sampling distributions for 3, 4, and 5 *df,* respectively. The left-most column shows estimates for WLS, the middle column estimates for ML and the right-most column overlays the WLS (depicted in light blue) and ML (light green) Q_{SNP} estimates.



Supplementary Figure 35. Associations between Q_{SNP} and theoretical χ^2 distributions. Associations shown for models estimated using WLS (left column) and ML estimates (middle column). Red lines depict the χ^2 distribution plotted against itself, with values below the line indicating under-estimated effects. The right column depicts WLS and ML plotted against one another. The blue line depicts WLS plotted against itself, with values above the line indicating Q_{SNP} estimates that were estimated as larger for ML.



Supplementary Figure 36. Genomic SEM simulation results. Results from 100 runs of Genomic SEM using data simulated at the level of the SNPs. Results are presented for unstandardized (panel a) and standardized (panel b) estimates. Parameters outside of the parentheses indicate those provided in the generating population. In parentheses, we provide for WLS (*in italics*) and ML (**in bold**) estimation the average point estimate and the ratio of the mean *SE estimate* across the 100 runs over the empirical *SE* (calculated as the standard deviation of the parameter estimates across the 100 runs). The ratio of mean and empirical *SE*s was close to 1 in all cases, although slightly above 1 (i.e., conservative) for standardized estimates of residual variance. These *SE estimates* are expected to be upwardly biased in the standardized case due to genetic variance estimates being rescaled to exactly 100%.



Supplementary Figure 37. Model fit indices from Genomic SEM simulations. Model fit indices were compared across the 100 runs of Genomic SEM using simulated data. Depicted in blue are model fit indices for runs specified to match the generating population (i.e., one common factor with freely estimated factor loadings). Depicted in green are indices for models specified to have equal factor loadings across all indicators. Depicted in red are indices for a model in which the third indicator loading was fixed to 0. Indices favored the model that matched the generating population for model chi-square, AIC, and CFI in 100% of cases, with the exception that 99 models favored the matching model for AIC with WLS estimation. Indices are presented for WLS (panel a) and ML (panel b) estimation.