Online Supplement:

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

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* 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. 18). 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. 19), 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. 19).

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. 20 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 S17). 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. 21). However, the association between Q_{SNP} estimates was high for both the *p*-factor (r = .77) and neuroticism (r = .99; Supplementary Fig. 21). 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. S22).

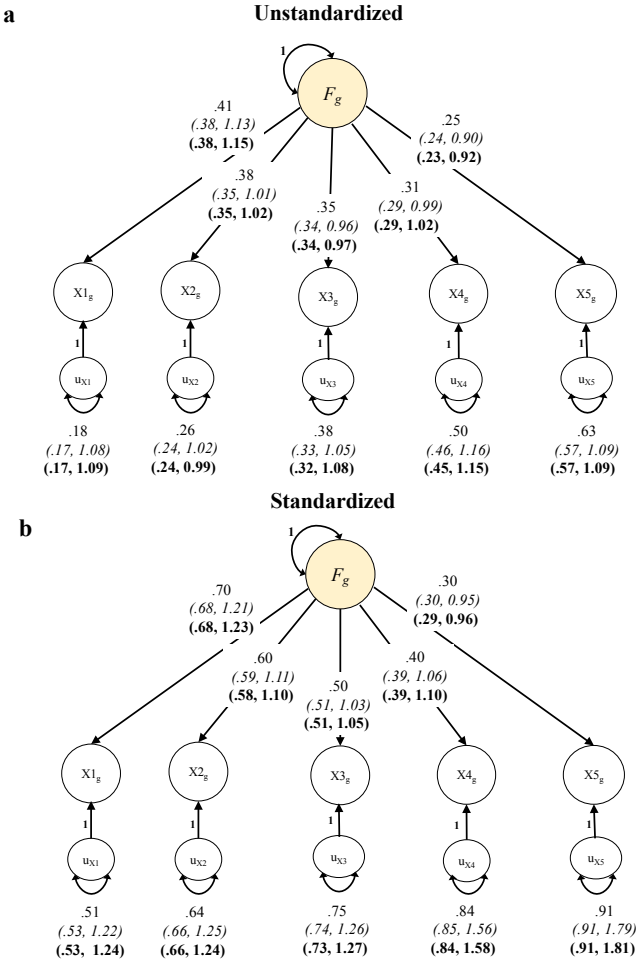

Fig. S1. Genomic SEM simulation results. Results from 100 runs of Genomic SEM using data simulated at the level of the SNPs. Results are presented for unstandardized (panel a) and standardized (panel b) estimates. Parameters outside of the parentheses indicate those provided in the generating population. 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 heritability estimates being fixed to 100%.

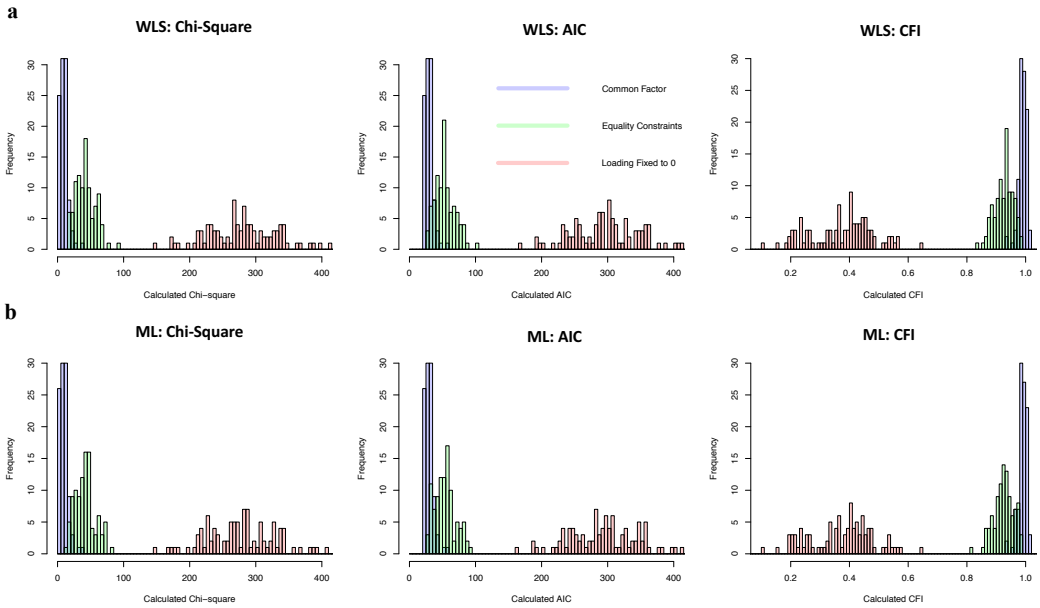

Fig. S2. Model fit indices from Genomic SEM simulations. Model fit indices were compared across the 100 runs of Genomic SEM using simulated data. Depicted in blue are model fit indices for runs specified to match the generating population (i.e., one common factor with freely estimated factor loadings). 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.

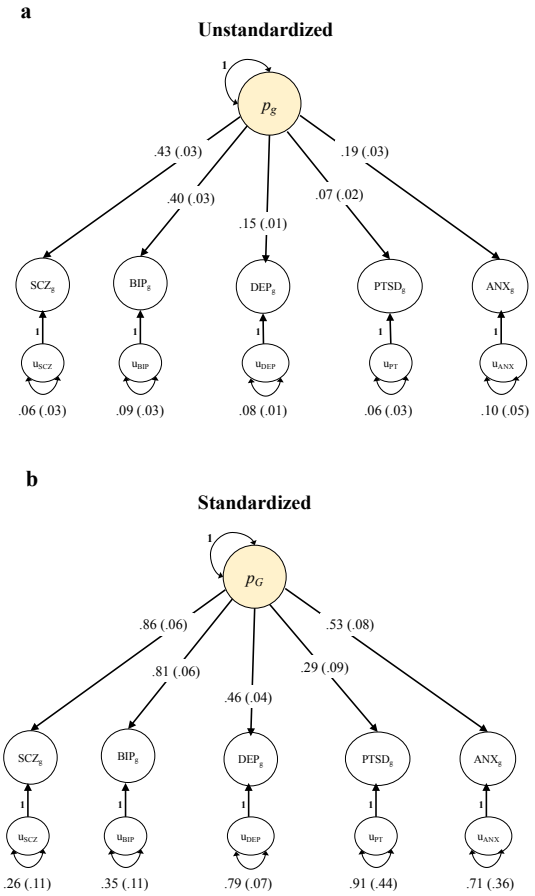

Fig. S3. Confirmatory factor analysis of genetic *p*-factor with Genomic SEM. Confirmatory factor analyses (CFA) were used to construct a genetically defined *p*-factor for unstandardized (panel a) and standardized estimates (panel b) using WLS estimation. *SE*s are shown in parentheses. The genetic 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.

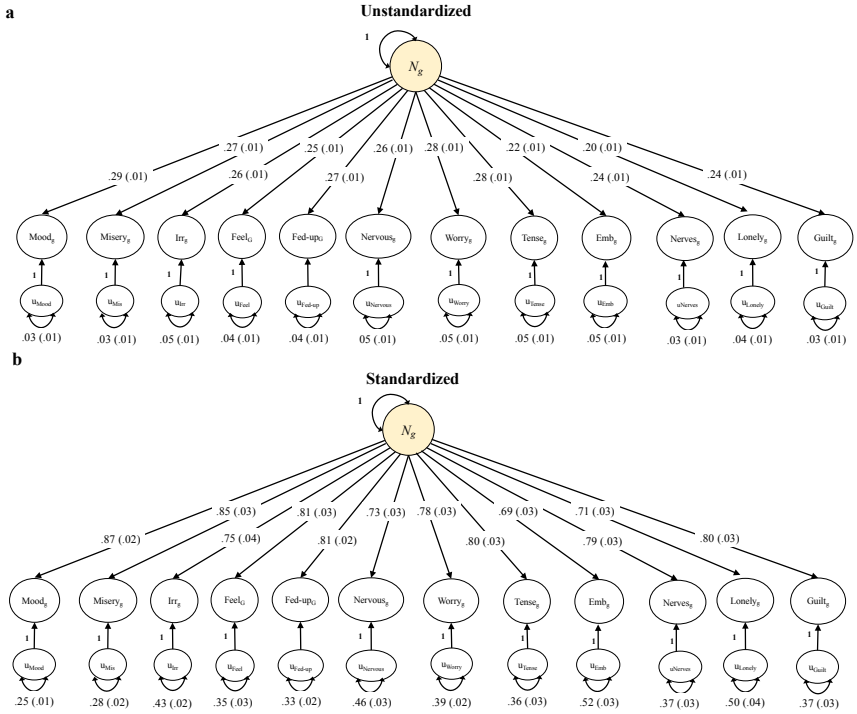

Fig. S4. Confirmatory factor analysis of a genetic factor of neuroticism with Genomic SEM. Confirmatory factor analyses (CFA) were used to construct a genetically defined neuroticism factor for unstandardized (panel a) and standardized estimates (panel b) using WLS estimation. *SE*s are shown in parentheses. 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.

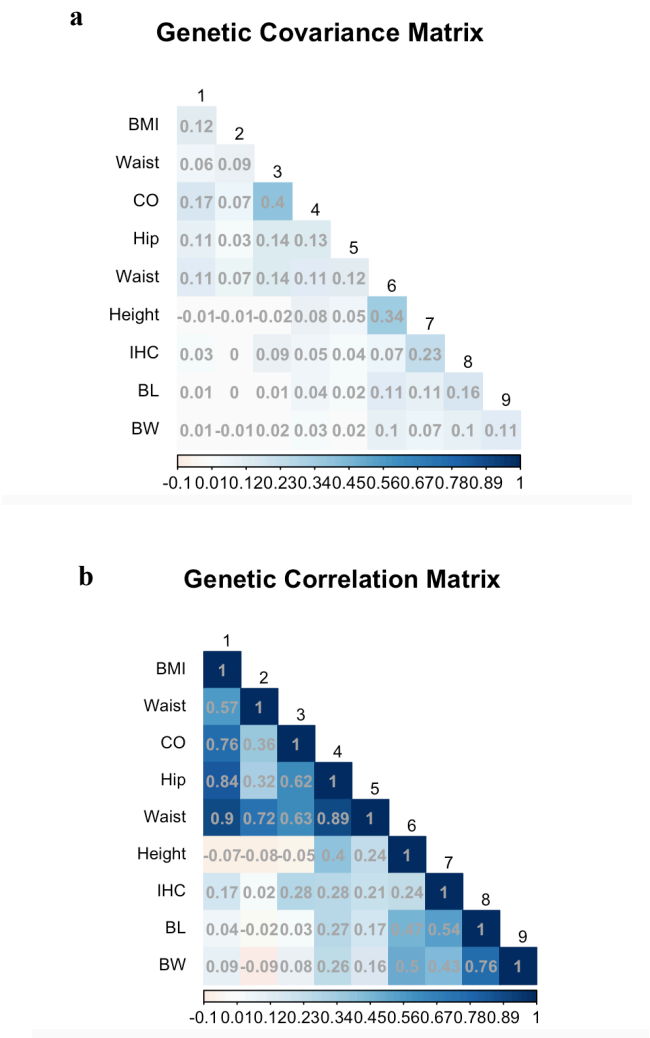

Fig. S5. Heatmap of genetic associations among anthropometric traits. Genetic covariance (panel a) and correlation (panel b) matrices with parameters estimated from multivariate LDSC. Visual inspection indicates two clusters in the upper left and lower right corner of the heatmap. BMI = body mass index; WHR = waist-hip ratio; CO = childhood obesity; IHC = infant head circumference; BL = birth length; BW = birth weight.

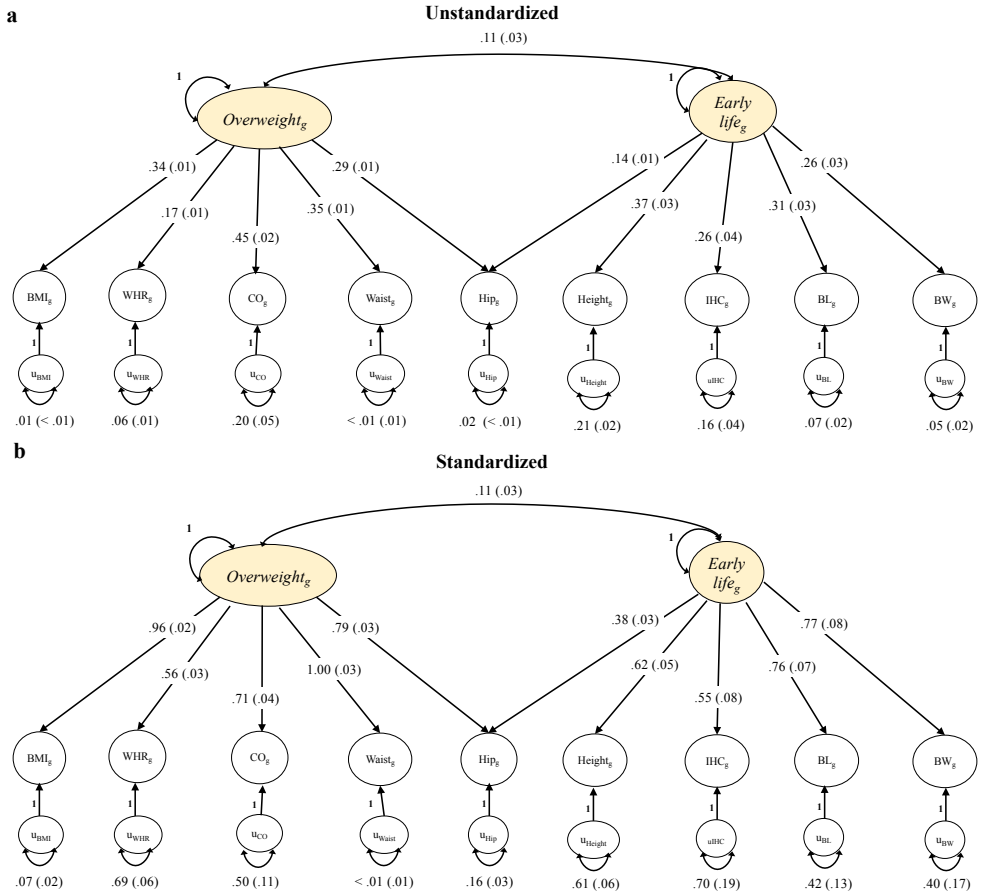

Fig S6. Confirmatory factor analysis of multivariate genetic architecture of anthropometric traits using Genomic SEM. Confirmatory factor analyses (CFA) informed by an initial exploratory factor analysis were used to construct latent overweight and early life growth factors for unstandardized (panel 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.

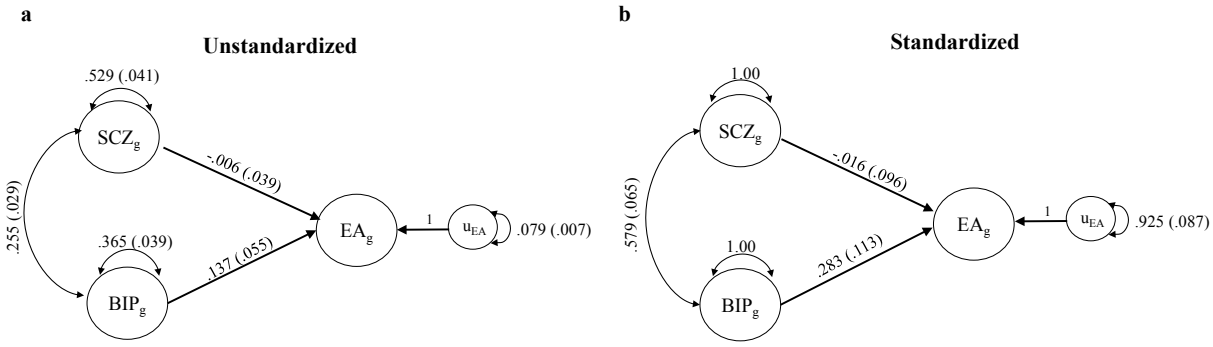

Fig. S7. Reproducing GWIS findings using Genomic SEM. Results from Genomic SEM in which the genetic component of educational achievement was simultaneously regressed on the genetic components of bipolar disorder and schizophrenia. The genetic covariance (unstandardized; panel a) and genetic correlation (standardized; panel b) matrices, and associated sampling covariance matrices, estimated from multivariate LDSC were used as input for Genomic SEM.

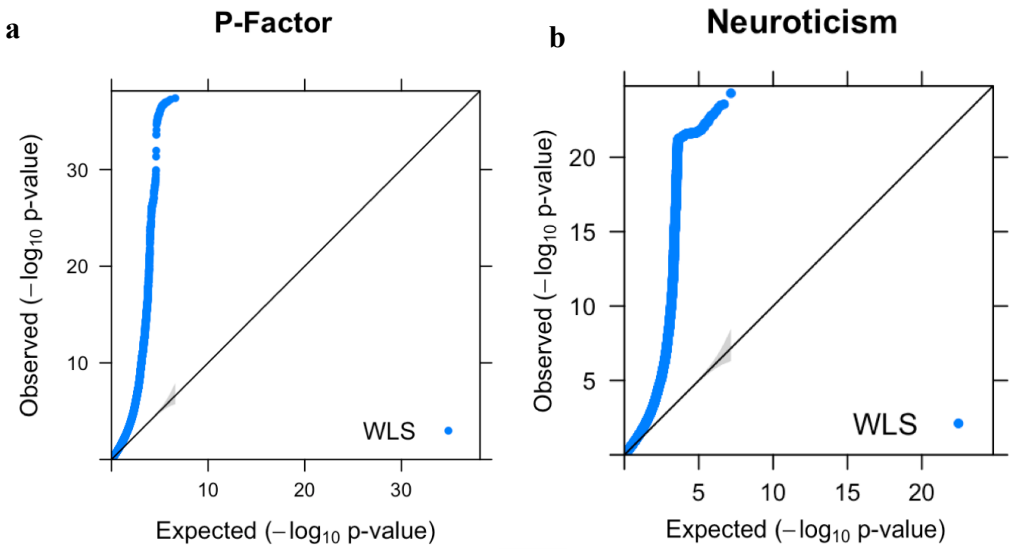

Fig. S8. Quantile-quantile plot of multivariate GWAS *p*-values for *p*-factor and neuroticism. Expected −log10(*p*)-values are those expected under the null hypothesis. The shaded area indicates the 95% confidence interval under the null. The multivariate GWAS was conducted using Genomic SEM with WLS estimation.

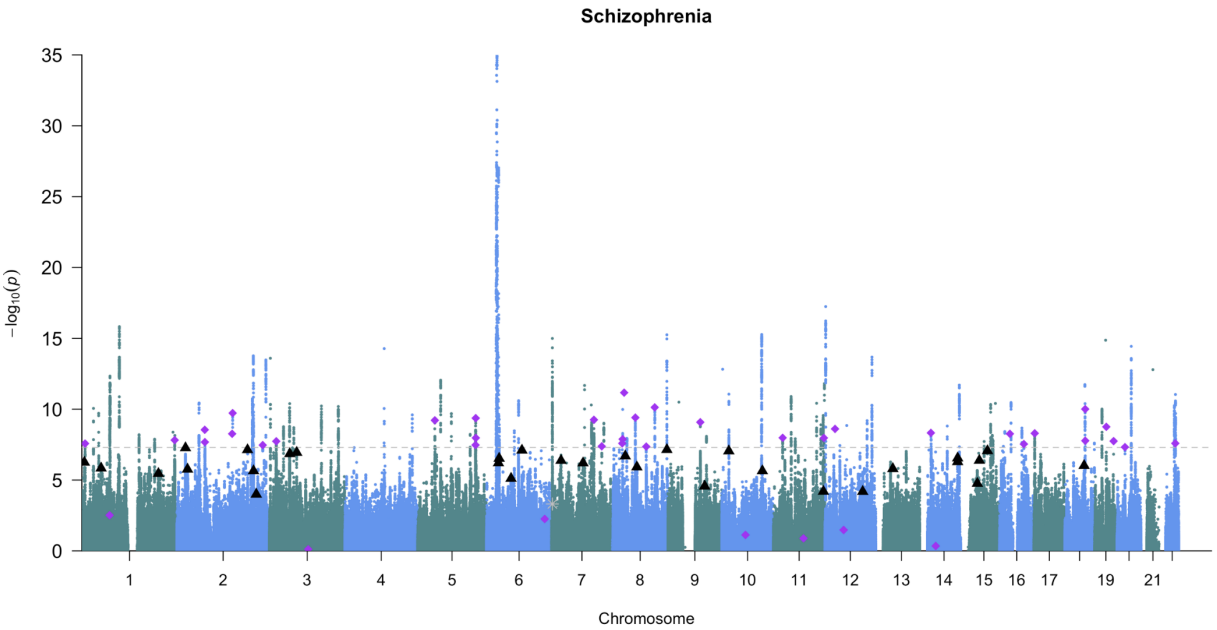

Fig. S9a. Manhattan plot for univariate Schizophrenia GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate 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}.

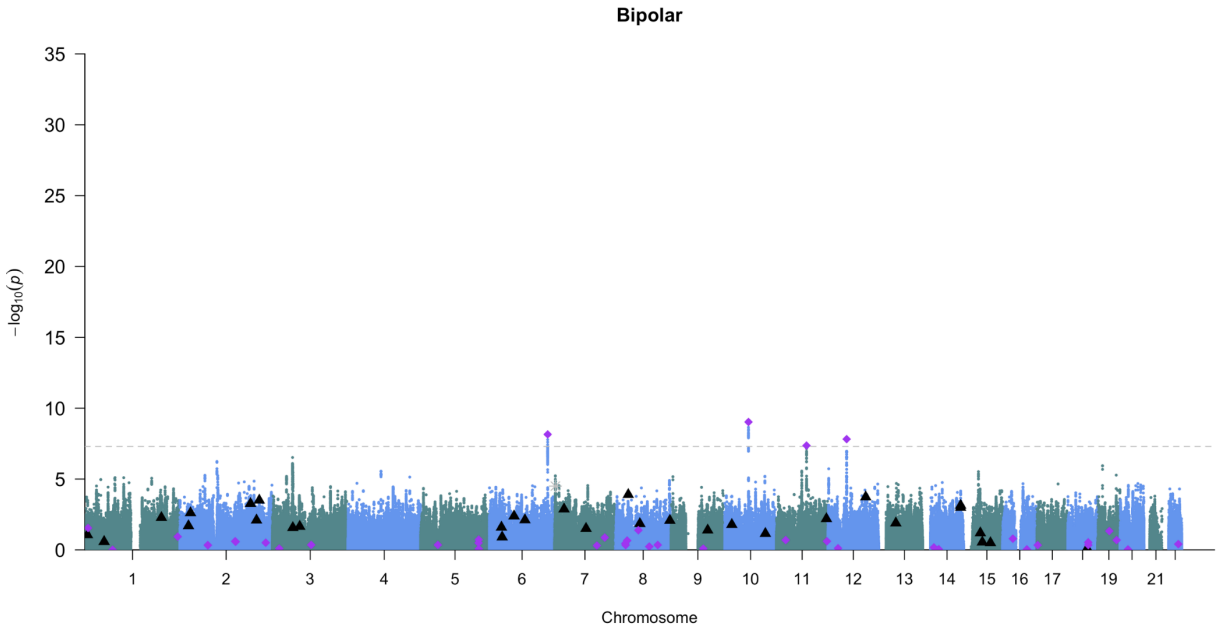

Fig. S9b. Manhattan plot for univariate Bipolar GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate 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}.

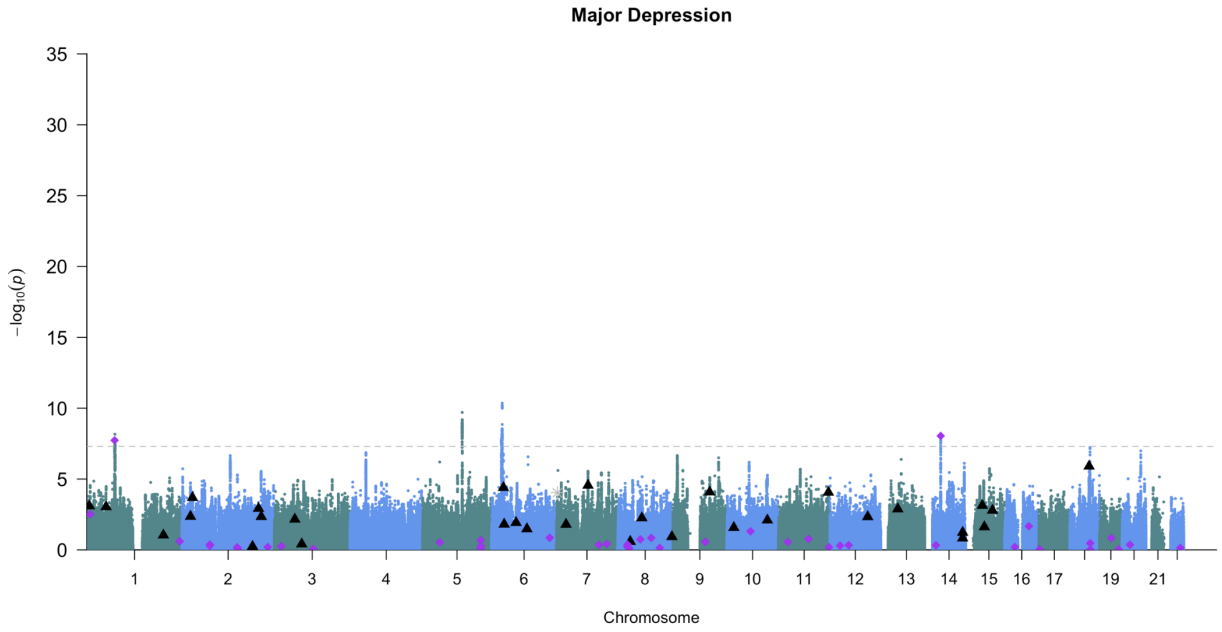

Fig. S9c. Manhattan plot for univariate Major Depression GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the 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}.

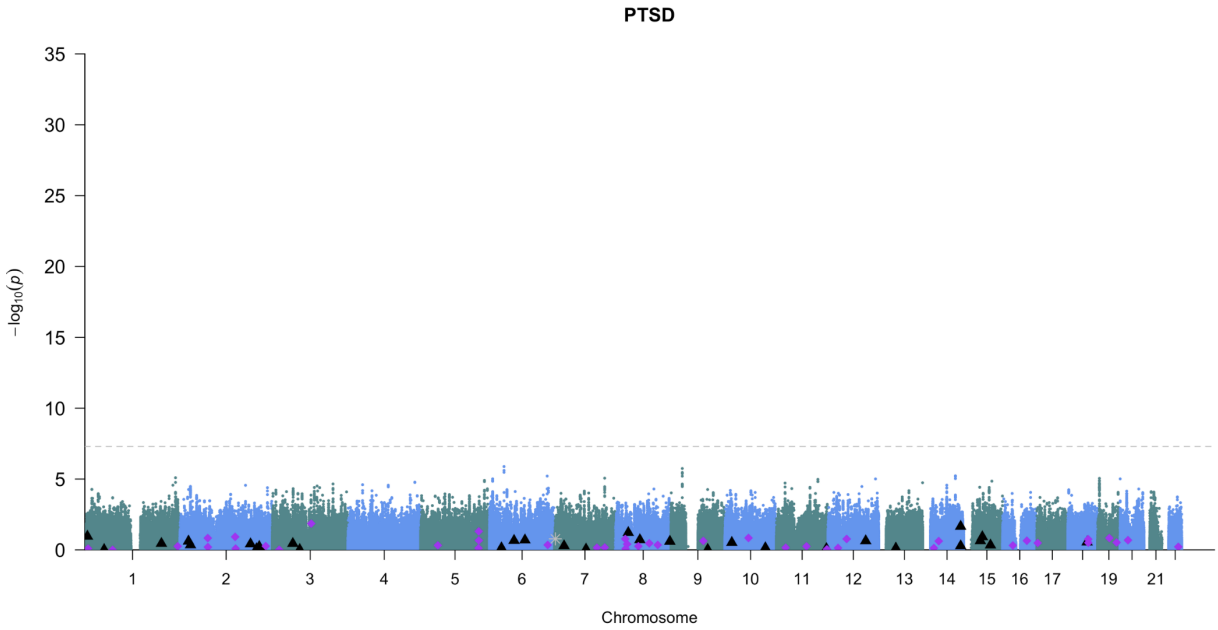

Fig. S9d. Manhattan plot for Post-Traumatic Stress Disorder GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for 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}.

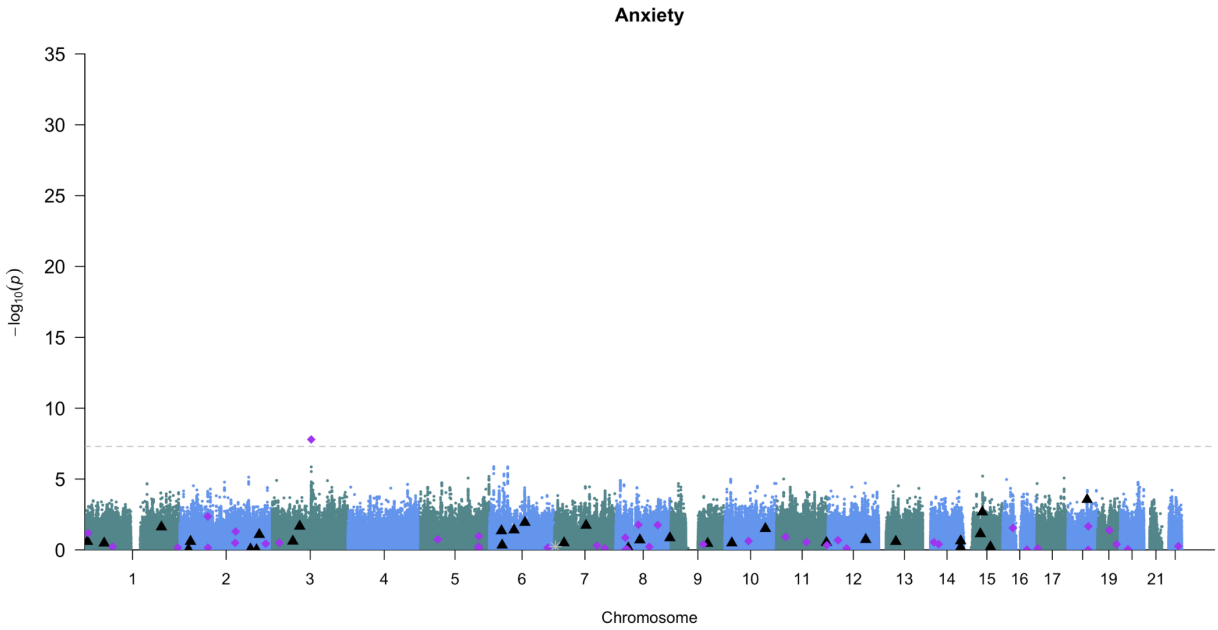

Fig. S9e. Manhattan plot for univariate Anxiety GWAS (Genomic SEM results superimposed). The gray dashed line marks the threshold for genome wide significance ($p < 5 \times 10^{-8}$). Black triangles denote independent hits for the *p*-factor that were not in LD with independent hits for the univariate 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}.

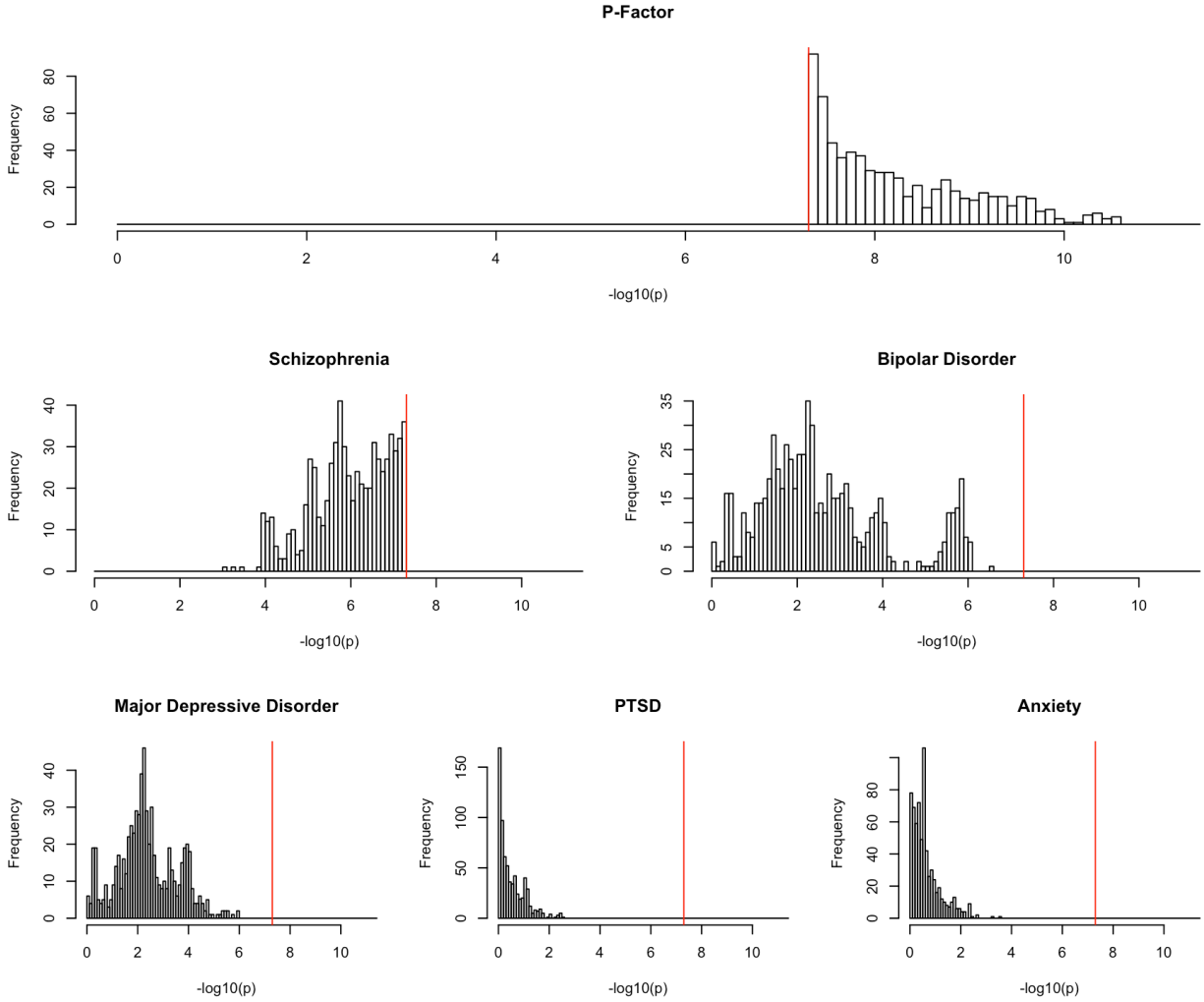

Fig. S10. Histograms of –log10 p-values for hits on *p*-factor. Histograms of –log10 p-values for the 684 non-independent SNPs that were genome wide significant for the *p*-factor using WLS estimation, but were not identified as significant in any of the individual GWASs. The vertical red line indicates genome wide significance.

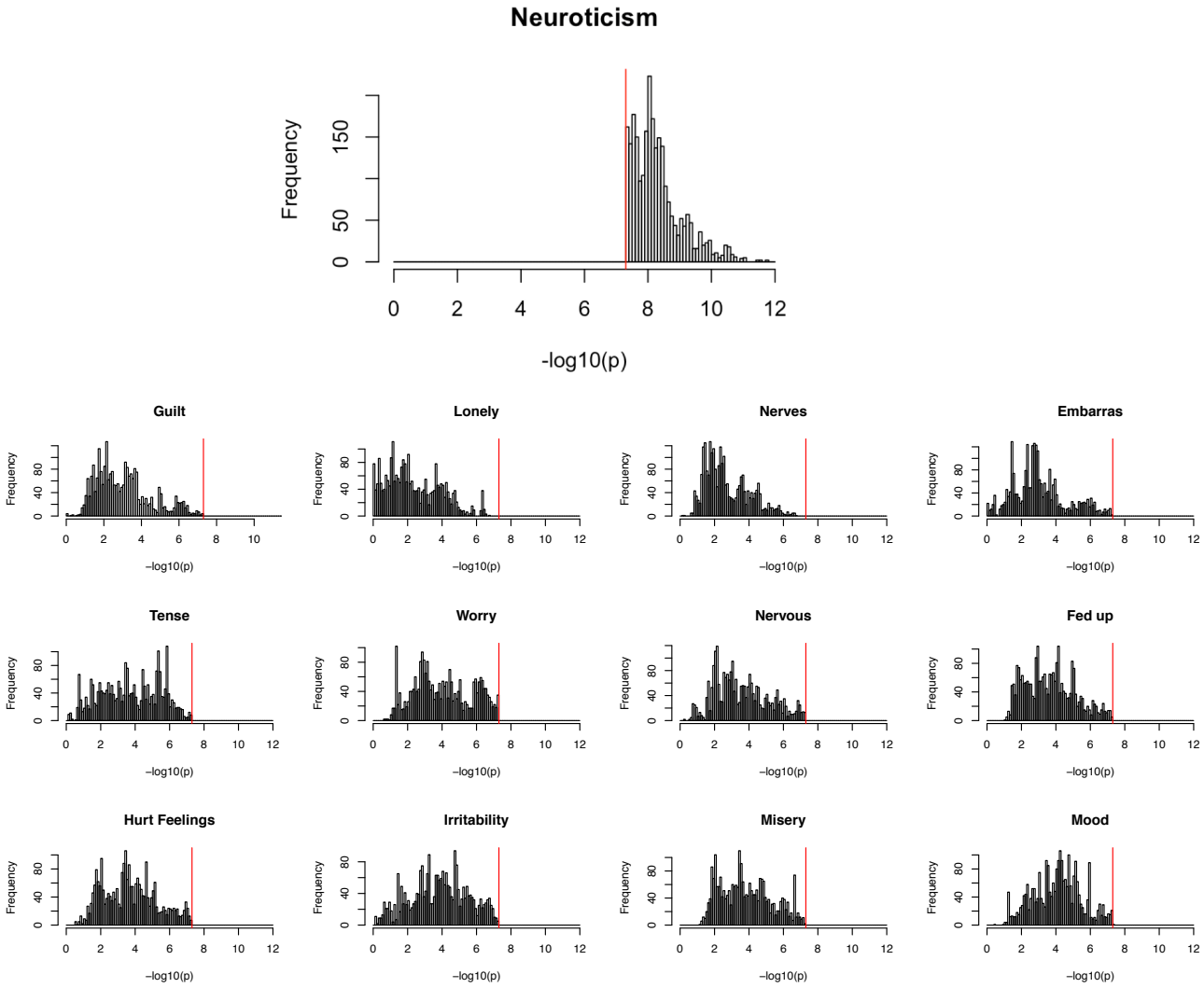

Fig. S11. Histograms of –log10 *p*-values for hits on neuroticism factor. Histograms of –log10 *p*-values for the 2,540 non-independent SNPs that were genome wide significant for neuroticism using WLS estimation, but were not identified as significant in any of the individual GWASs. The vertical red line indicates genome wide significance.

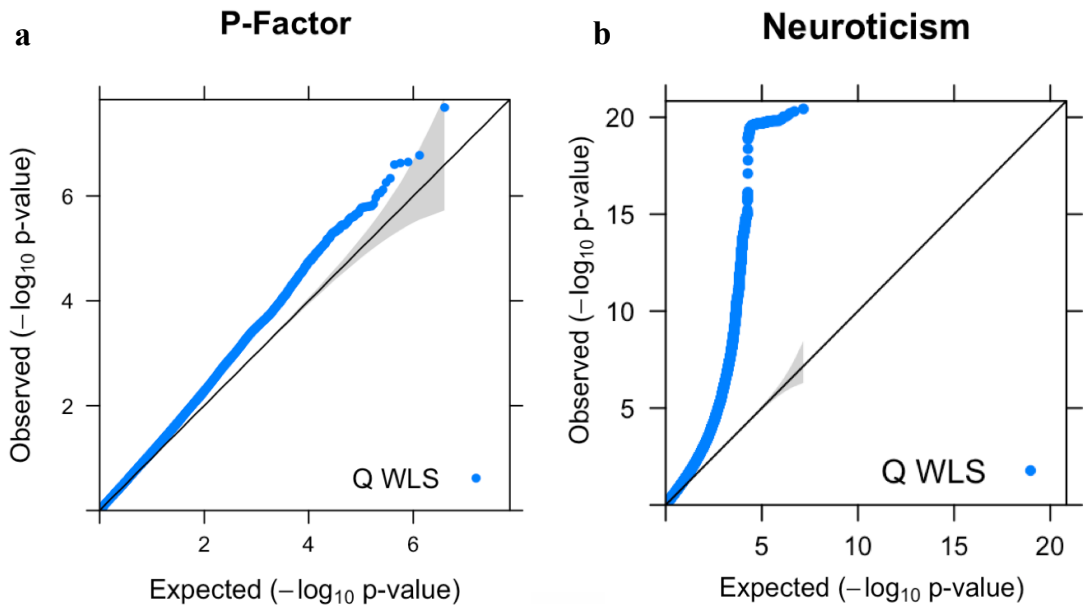

Fig. S12. Quantile-quantile plot for Q_{SNP}. Estimates are from WLS estimation for the *p*-factor (panel a) and neuroticism (panel b). Expected −log10 *p*-values are those expected under the null hypothesis. The shaded area indicates the 95% confidence interval under the null. As some Q_{SNP} estimates for neuroticism were quite large, *p*-values < 5^{-20} were set to 5^{-20}.

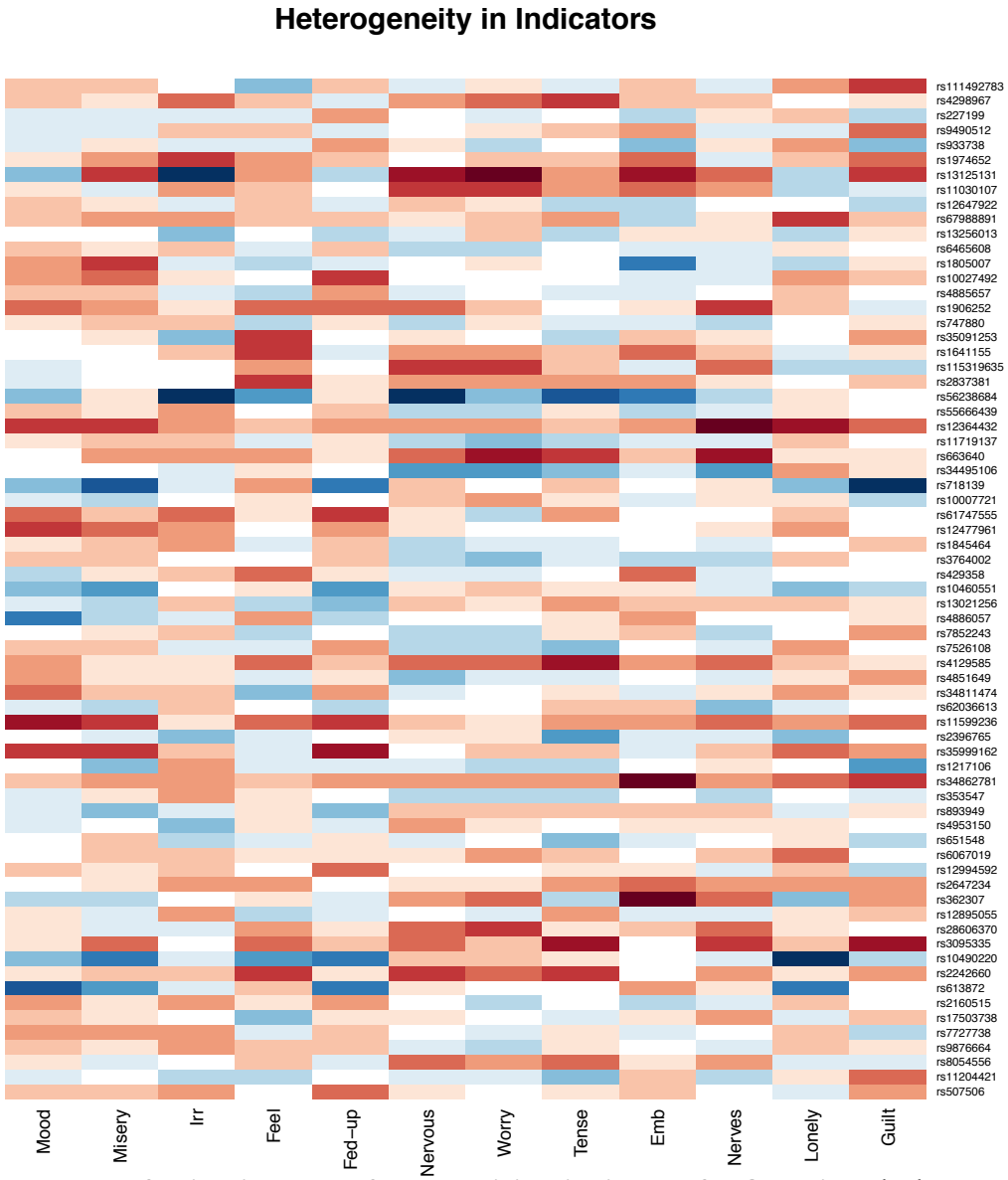

Fig. S13. Heatmap of univariate betas for neuroticism indicators for Q_{SNP} hits. The heatmap depicts univariate item-specific betas for the 69 lead SNPs for Q_{SNP} identified using WLS estimation for neuroticism. Items are on the x-axis. SNPs are on the y-axis. Cells depicted in red, white, and blue indicate negative, near zero, and positive betas, respectively. As expected, individual rows indicate substantial heterogeneity across the indicators for hits on Q_{SNP}.

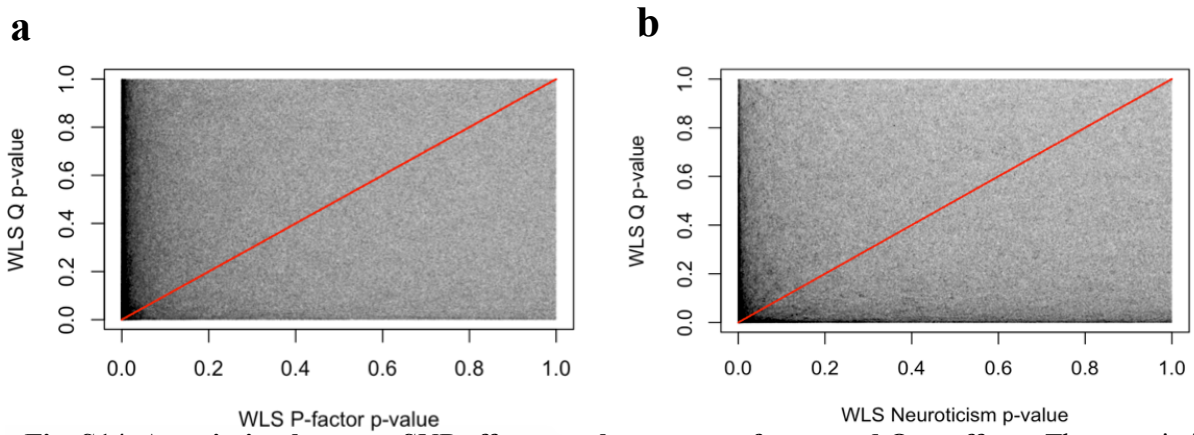

Fig. S14. Association between SNP effects on the common factor and Q_{SNP} effects. The association between the *p*-values for SNP effects on the the common factor (x-axis) and the p-values for Q_{SNP} (y-axis) are plotted for WLS estimation of the *p*-factor (panel a) and neuroticism (panel b). The 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.

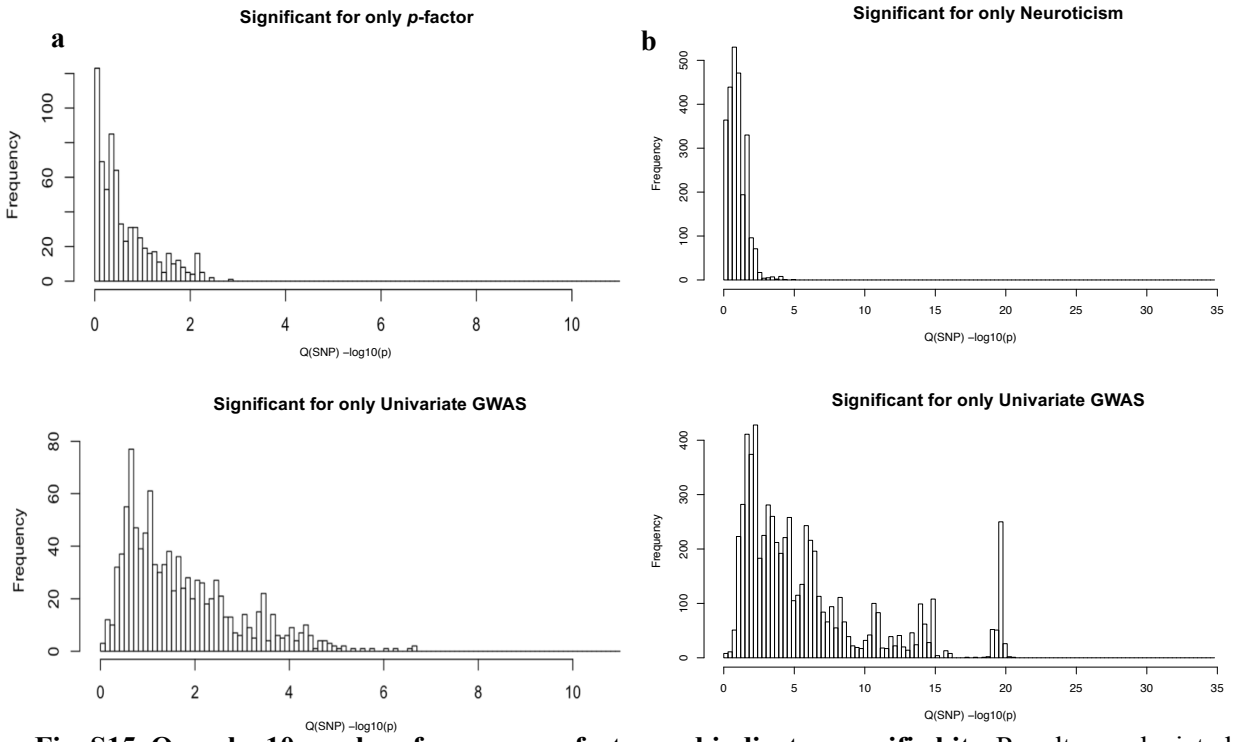

Fig. S15. Q_{SNP} –log10 *p*-values for common-factor and indicator-specific hits. Results are depicted for WLS estimation of the *p*-factor (panel a) and neuroticism (panel b). There were 684 non-independent SNPs identified as genome-wide significant for *p*-factor, but not the univariate GWAS, and 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.

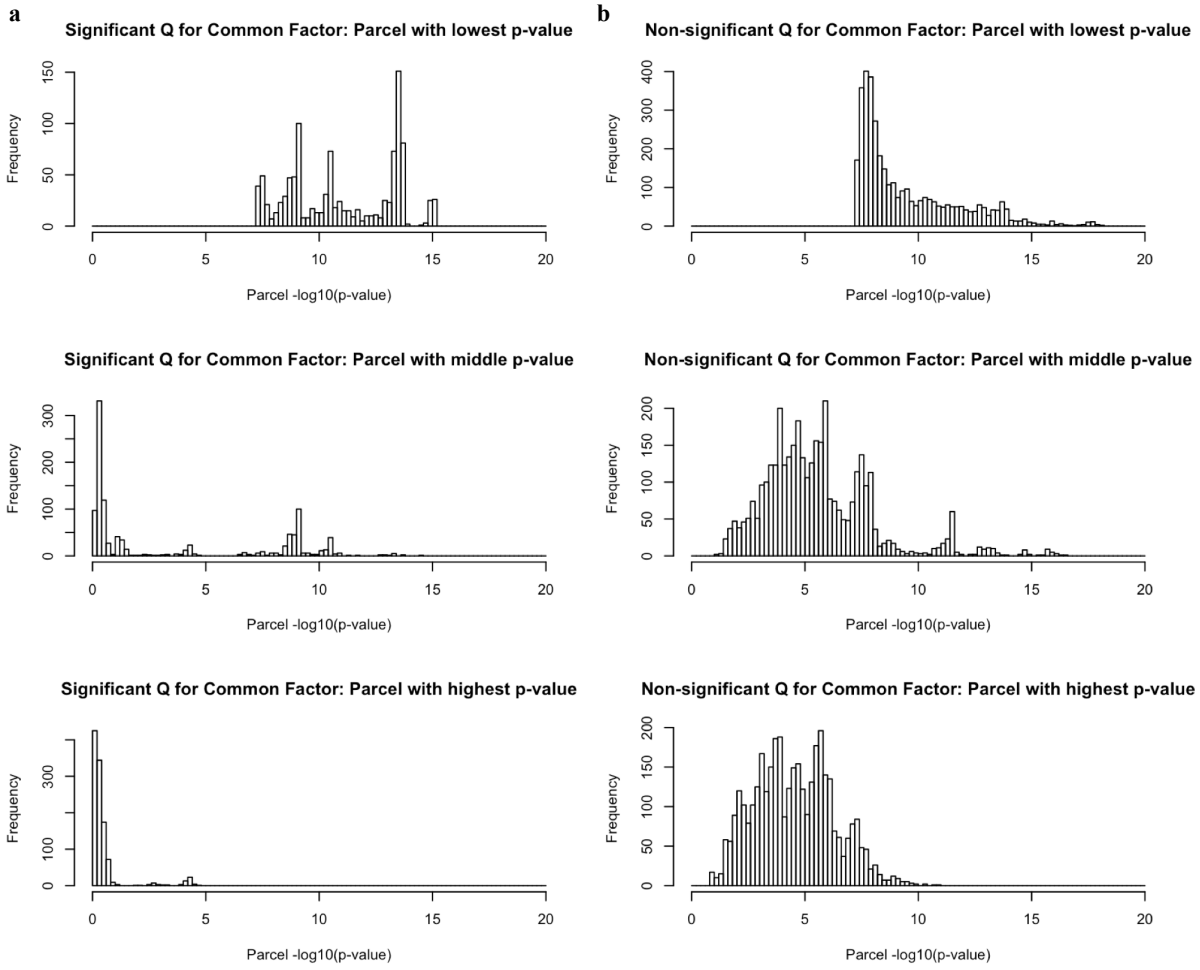

Fig. S16. Comparison of parcel *p*-values for low versus high Q_{SNP} estimates. Histograms shown for SNPs that produced genome wide significant hits for at least one of the parcels split across high (Q_{SNP} *p*-value < 5e-8; 1,090 SNPs; panel a) and low (*p* > 5e-3; 3,685 SNPs; panel b) Q_{SNP} estimates 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.

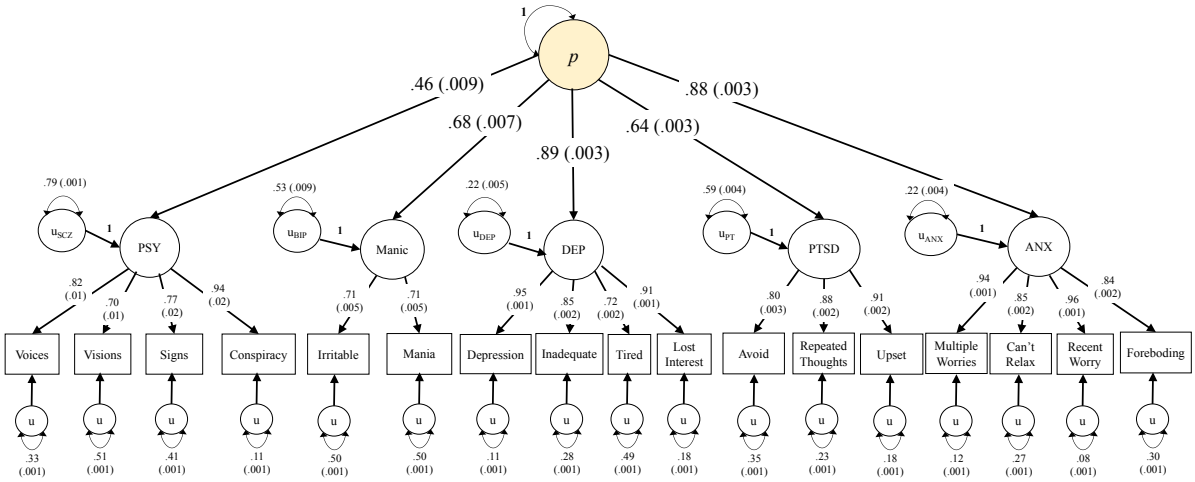

Fig. S17. UK Biobank *p*-factor. Standardized output of phenotypic *p*-factor constructed from UKB phenotypes for out of sample prediction using *p*-factor polygenic scores. PSY = psychotic experiences; DEP = depressive symptoms; PTSD = symptoms of post-traumatic stress disorder; ANX = anxious symptoms.

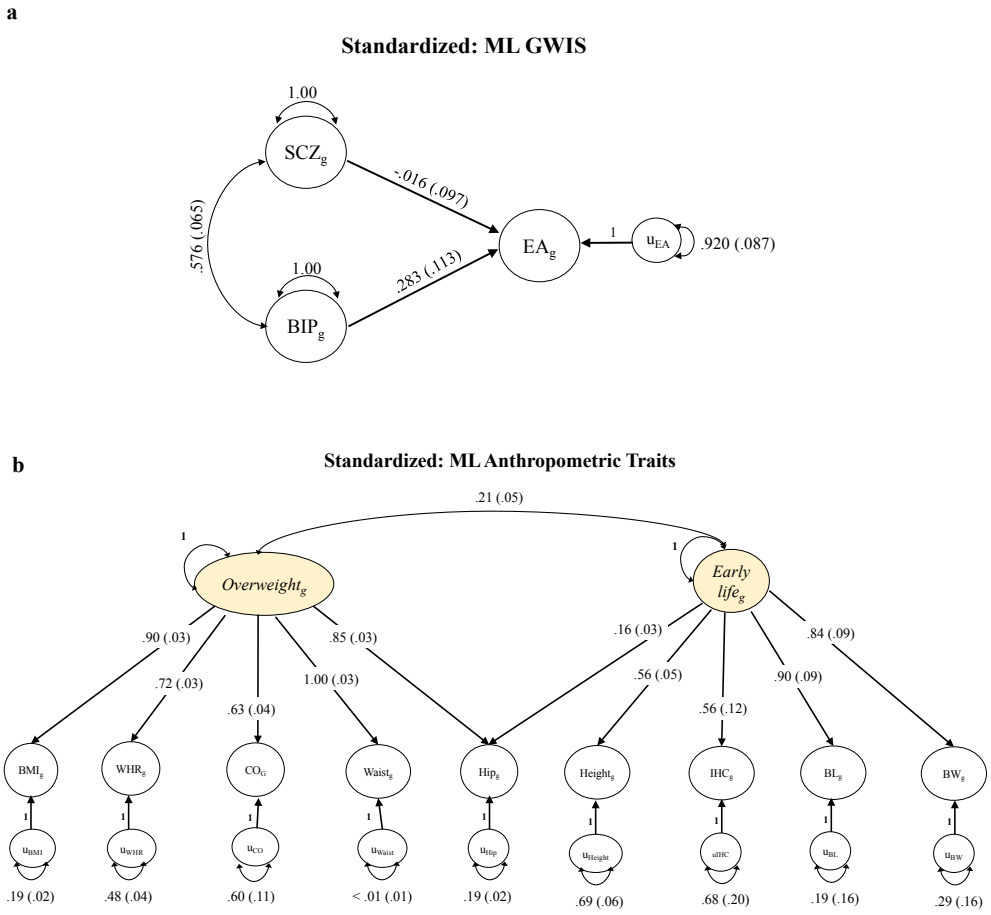

Fig S18. ML Estimates from GWIS and anthropometric trait Genomic SEM models. Results are presented for standardized output for the multiple regression model of GWIS (panel a) and the confirmatory factor model of anthropometric traits (panel b). *SE*s are shown in parentheses. The genetic correlation 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.

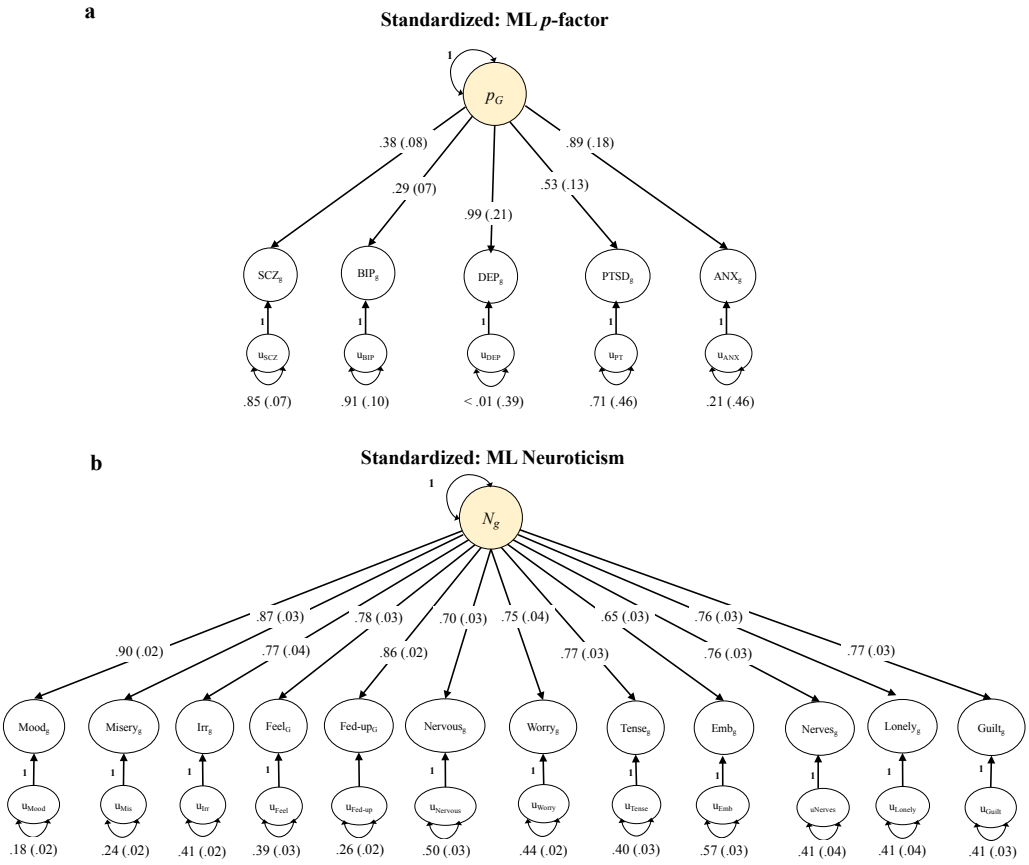

Fig S19. ML estimates for neuroticism and *p*-factor Genomic SEM models. Results are presented for standardized output for the confirmatory factor models of the *p*-factor (panel a) and neuroticism (panel b). The genetic correlation matrix (standardized) and associated sampling covariance matrix were 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.

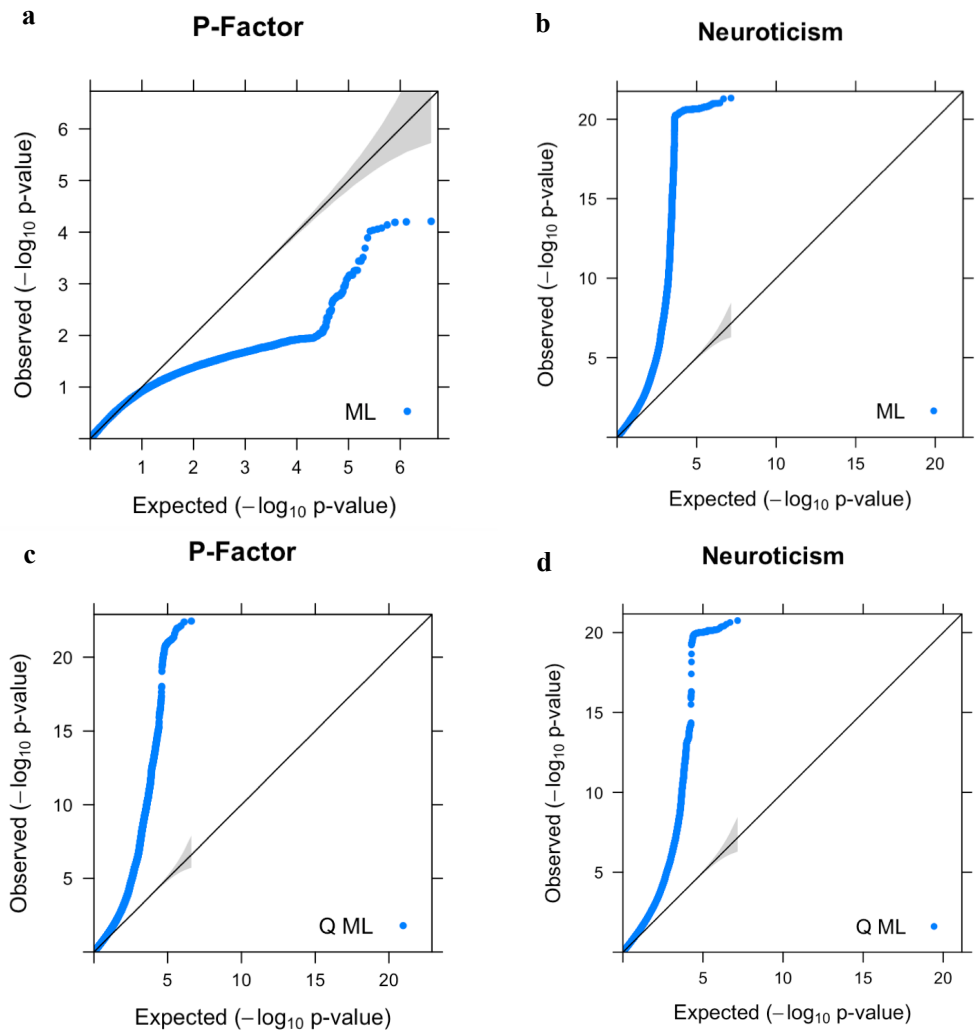

Fig S20. Quantile-quantile plot of multivariate GWAS *p*-values for *p*-factor and neuroticism (ML estimation). Estimates are from ML estimation for the *p*-factor (panel a), neuroticism (panel b), Q_{SNP} estimates for the *p*-factor (panel c), and Q_{SNP} estimates for neuroticism (panel d). Expected −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}.

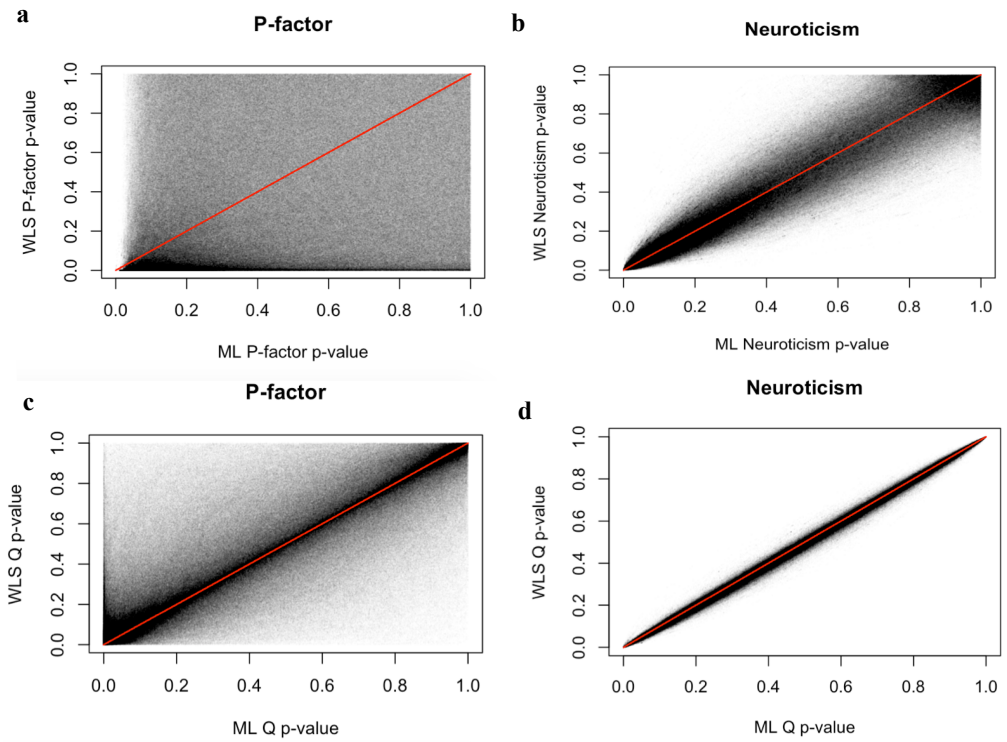

Fig. S21. Associations between ML and WLS p values. Scatter plot comparing *p*-values between WLS (y-axis) and ML (x-axis) estimation for the *p*-factor (panel a), neuroticism (panel b), Q_{SNP} for the *p*-factor (panel c), and Q_{SNP} for neuroticism (panel d). The red line reflects the regression 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.

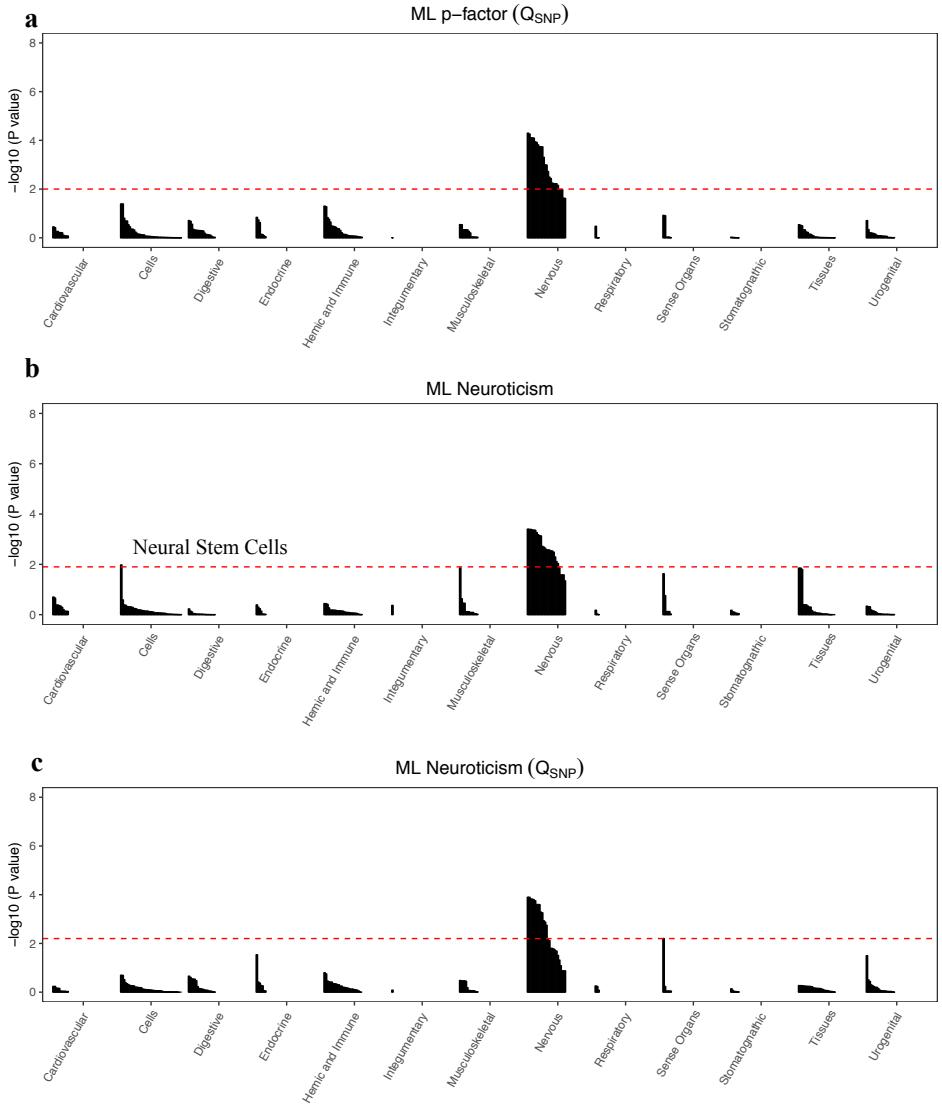

Fig. S22. Biological annotation of Q_{SNP} of *p*-factor, neuroticism, and Q_{SNP} of neuroticism. Results from tissue enrichment analyses conducted using DEPICT based on Genomic SEM results for Q_{SNP} of the *p*-factor (panel a), neuroticism (panel b) and Q_{SNP} estimation for neuroticism (panel c) 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.

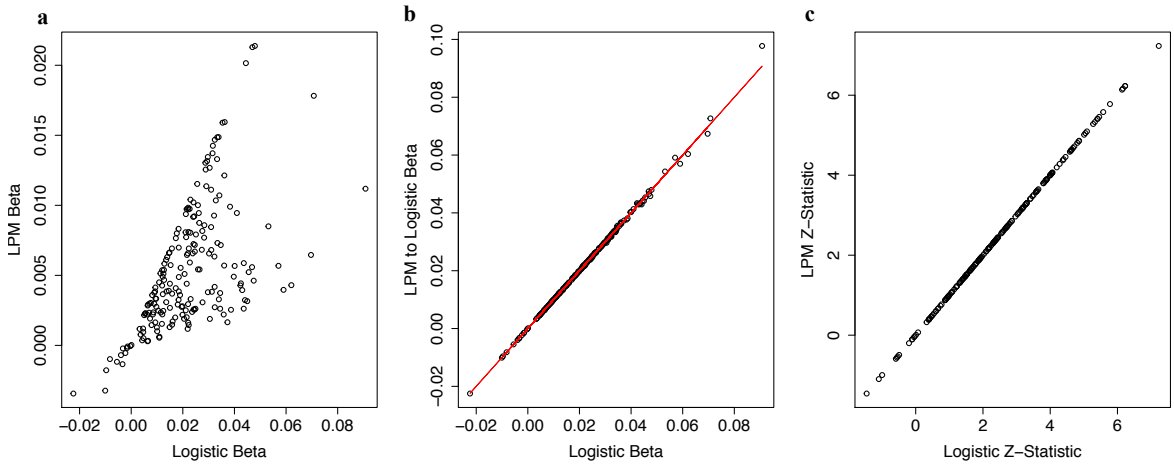

Fig S23. Associations between regression coefficients from logistic regression model and linear probability model. 200 datasets were simulated with 100,000 observations each in which a continuously distributed liability was specified to be a linear function of a biallelic autosomal SNP and a normally 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 heritabilites and genetic covariances (the liability scale estimates should still be requested).

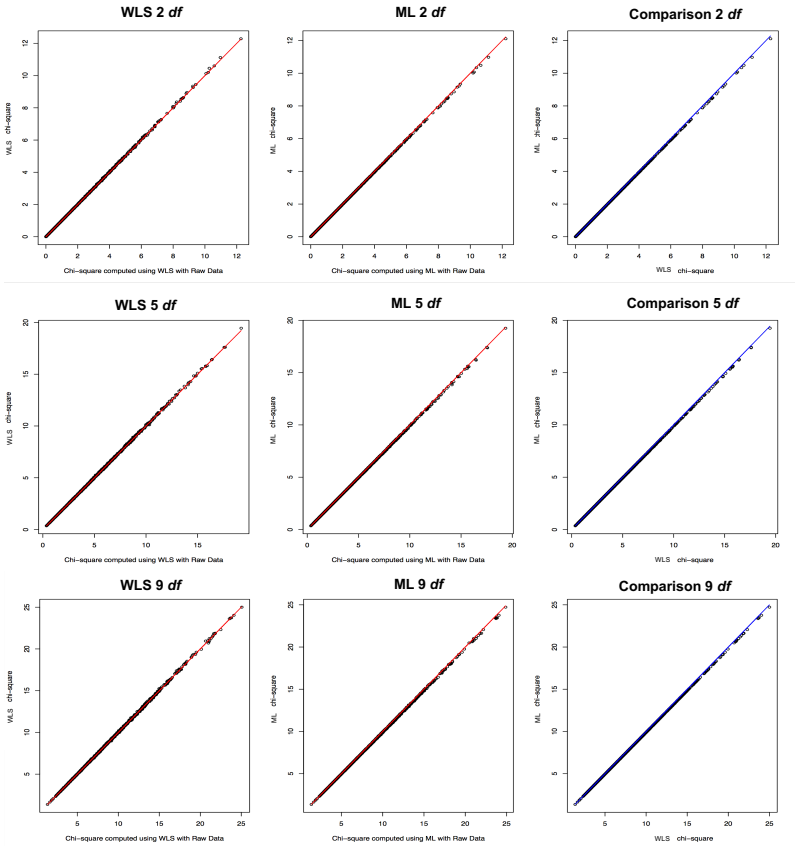

Fig. S24. Associations between model χ^2 values computed from summary data and model χ^2 values computed from raw data. Raw data-based estimates of model χ^2 were computed directly from the data using lavaan. Summary data-based estimates of model χ^2 were computed using the *S* and *V* matrices with 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.

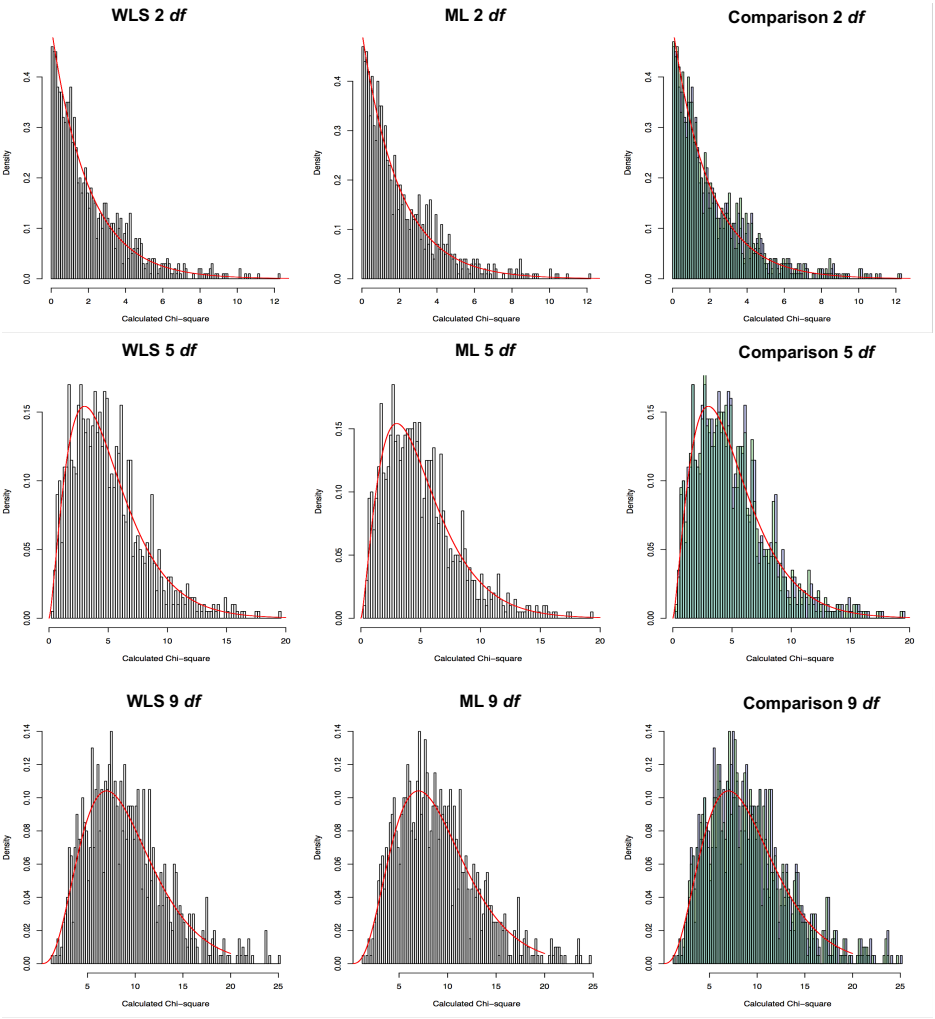

Fig. S25. Distributions of calculated and theoretical χ^2 statistics. Comparison between distribution of χ^2 values for model estimated from S and V matrices using WLS (left column) and ML (middle column) against a theoretical χ^2 distribution. The right column compares the distributions of WLS (blue bars) and ML (green bars).

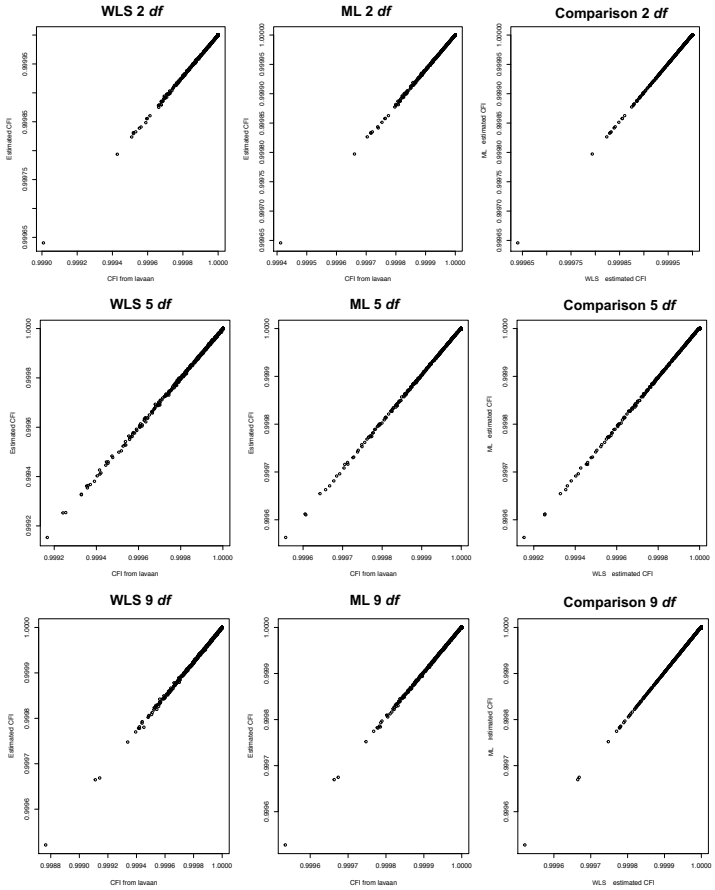

Fig. S26. Associations between CFI values derived from summary data and CFI values derived from raw data. Summary data-based estimates of CFI are depicted for models estimated using WLS (left column) and ML (middle column). We also present comparisons of the CFI from models estimated with ML and those estimated with WLS(right column). All CFI estimates were bounded at 1.

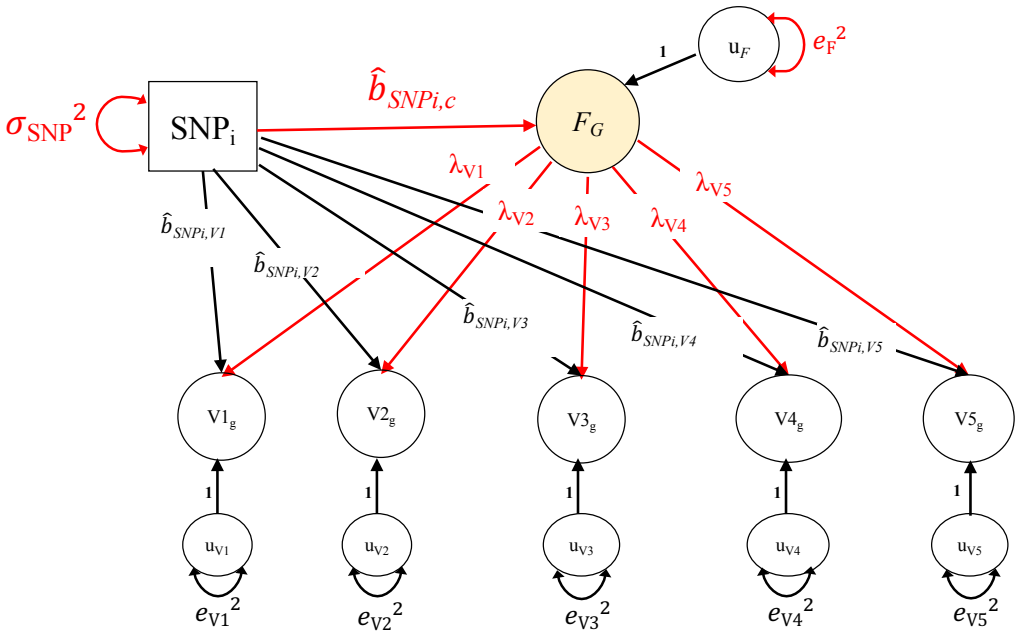

Fig. S27. Genomic SEM models for estimating Q_{SNP}. Red lines and parameters are fixed from Step 1, and black lines and parameters are freely estimated in Step 2.

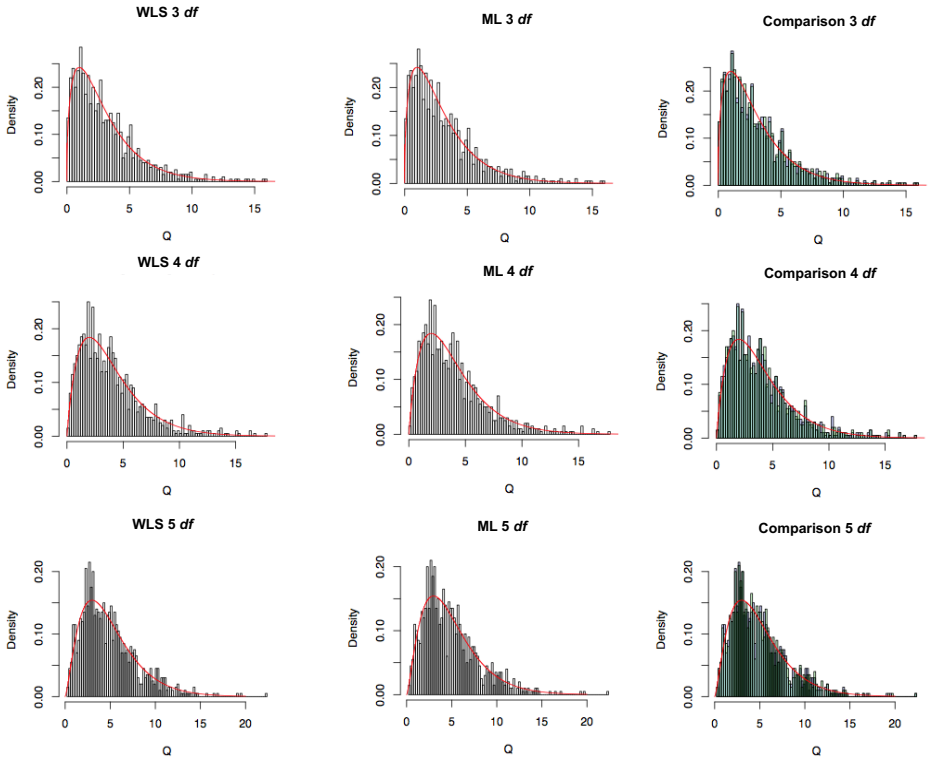

Fig. S28. Null distributions of Q_{SNP} for 1,000 simulations per model. Red lines for all panels depict the chi-square distribution with the relevant *df*. The top, middle, and bottom panels depict the sampling distributions for 3, 4, and 5 *df,* respectively. The left-most column shows estimates for 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.

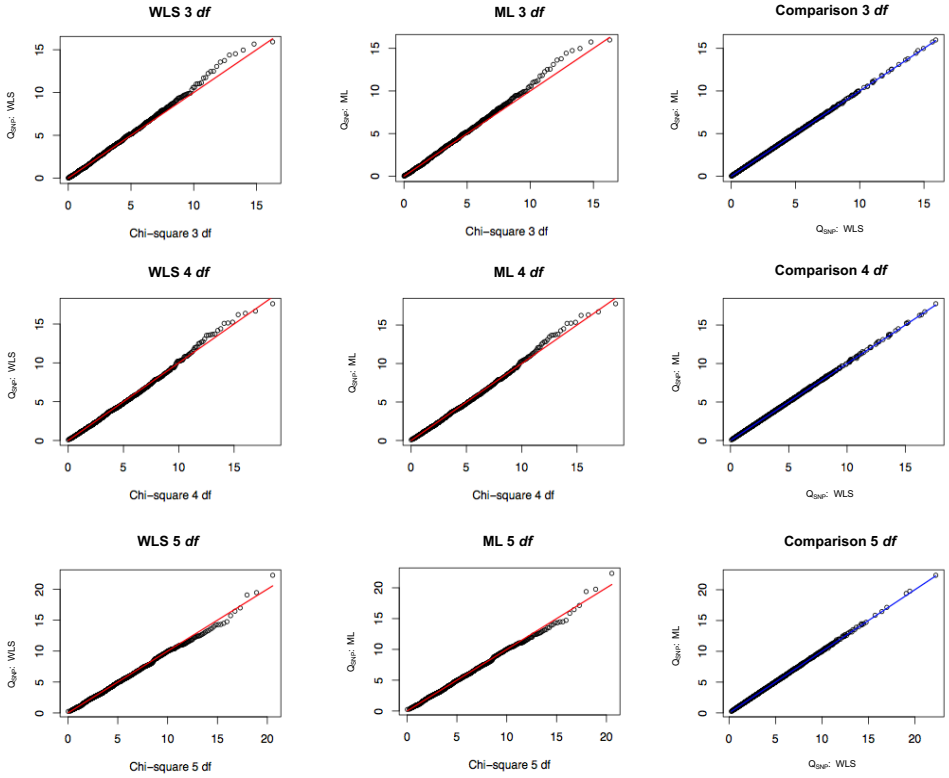

Fig. S29. Associations between Q_{SNP} and theoretical χ^2 distributions. Associations shown for models estimated using WLS (left column) and ML estimates (middle column). Red lines depict the χ^2 distribution plotted against itself, with values below the line indicating under-estimated effects. The 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.