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

Travis T. Mallard[1,2,19,*], Richard Karlsson Linnér[3,4], Andrew D. Grotzinger[1], Sandra Sanchez-Roige[5,6], Jakob Seidlitz[7,8], Aysu Okbay[3], Ronald de Vlaming[3], S. Fleur W. Meddens[9],

Bipolar Disorder Working Group of the Psychiatric Genomics Consortium,

Abraham A. Palmer[5,10], Lea K. Davis[6,11,12], Elliot M. Tucker-Drob[1,13], Kenneth S. Kendler[14,15], Matthew C. Keller[16,17], Philipp D. Koellinger[3,18], K. Paige Harden[1,13,18,*]

[1]Department of Psychology, University of Texas at Austin, Austin, TX, USA

[2]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[3]Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[4]Autism and Developmental Medicine Institute, Geisinger, Lewisburg, PA, USA

[5]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

[6]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

[7]Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[8]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

[9]Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands

[10]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

*Correspondence: tmallard@mgh.harvard.edu (T.T.M.), harden@utexas.edu (K.P.H.).

[11]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

[12]Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

[13]Population Research Center, University of Texas at Austin, Austin, TX, USA

[14]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

[15]Department of Psychiatry, Medical College of Virginia/Virginia Commonwealth University, Richmond, VA, USA

[16]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA

[17]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA

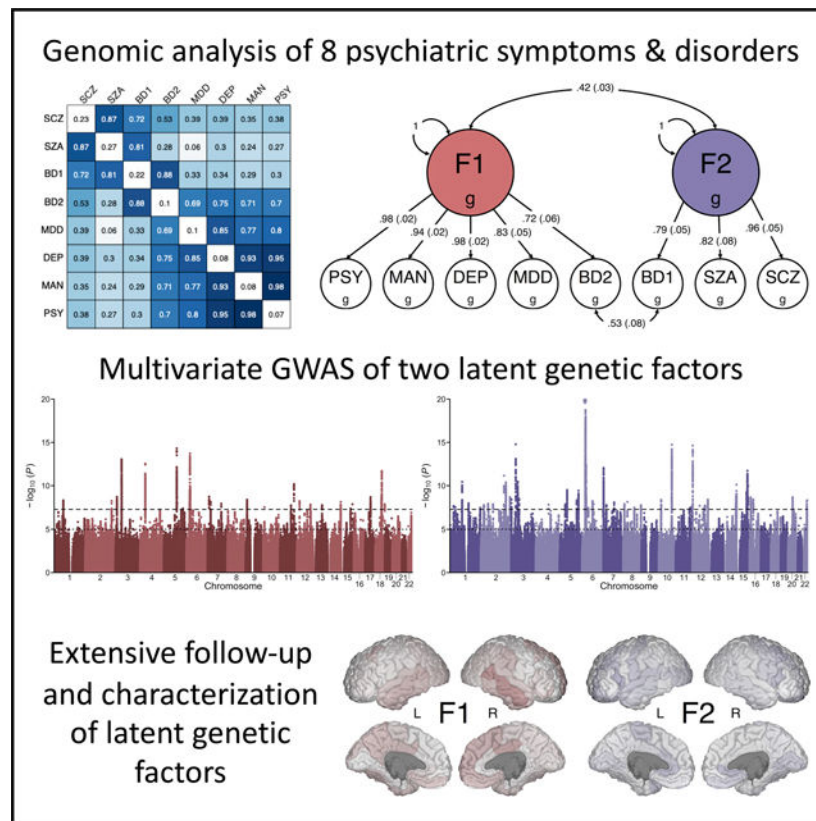[18]Senior author

[19]Lead contact

SUMMARY

Understanding which biological pathways are specific versus general across diagnostic categories and levels of symptom severity is critical to improving nosology and treatment of psychopathology. Here, we combine transdiagnostic and dimensional approaches to genetic discovery for the first time, conducting a novel multivariate genome-wide association study of eight psychiatric symptoms and disorders broadly related to mood disturbance and psychosis. We identify two transdiagnostic genetic liabilities that distinguish between common forms of psychopathology versus rarer forms of serious mental illness. Biological annotation revealed divergent genetic architectures that differentially implicated prenatal neurodevelopment and neuronal function and regulation. These findings inform psychiatric nosology and biological models of psychopathology, as they suggest that the severity of mood and psychotic symptoms present in serious mental illness may reflect a difference in kind rather than merely in degree.

In brief

Mallard et al. identified two transdiagnostic dimensions of genetic risk in a large genome-wide association study of psychiatric phenotypes related to mood disturbance and psychosis. While these genetic risk factors were modestly correlated, they were largely unique from one another and differed in their relationships with health and wellbeing.

Graphical Abstract

Genomic analysis of 8 psychiatric symptoms & disorders

Multivariate GWAS of two latent genetic factors

Extensive follow-up and characterization of latent genetic factors

INTRODUCTION

Psychiatric disorders are one of the leading causes of global disease burden, affecting more than 25% of the world's population at some point during their lifetime.[1] Twin- and family-based studies have established that a substantial portion of individual differences in liability to psychiatric disorders is attributable to genetic variation.[2] Genome-wide association studies (GWASs) have identified numerous genetic loci that have replicable associations with severe and debilitating psychiatric disorders, including schizophrenia,[3] bipolar disorder,[4] and major depressive disorder.[5]

GWASs have also identified a substantial degree of genetic overlap across psychiatric disorders, finding high genetic covariances and many pleiotropic loci.[6,7] This genetic overlap complicates efforts to identify causes, consequences, and treatments that are specific to any individual psychiatric disorder.[8] In response to these challenges, transdiagnostic approaches in psychiatry aim to identify biological systems that are perturbed across many forms of illness.[9,10] Transdiagnostic research may yield new therapeutic targets with broad utility as well as inform nosological classification and stratification of at-risk populations.

Concurrent with the emergence of transdiagnostic research, efforts to identify disorder-specific genetic loci have turned toward studying self-report measures in population-based cohorts,[11–13] as case-control study designs require diagnostic schedules that can be slow and costly. If valid, this dimensional approach in non-clinical samples has the potential to

accelerate genetic discovery via dramatic increases in sample size, as self-report survey measures of psychiatric symptoms can be administered at scale to large, genotyped population-based samples, such as UK Biobank.[14,15] However, while this approach may be valid for some common forms of psychopathology,[16] it is unknown whether the biology that influences normative variation in subthreshold symptoms also underlies rarer psychiatric conditions, such as those characterized by mania and/or psychosis.

Here, we combine transdiagnostic and dimensional research approaches to genetic discovery in two important ways. First, we use a novel combination of Bayesian item response theory and linear mixed models to perform GWASs of depressive (DEP), manic (MAN), and psychotic (PSY) symptoms on more than 250,000 individuals in the UK Biobank. This approach has been shown to yield higher heritability estimates than singleitem measures or simple composite measures.[17,18] Second, we used genomic structural equation modeling (Genomic SEM)[19] to characterize the shared genetic architecture among these three symptom dimensions and five additional psychiatric disorders: major depressive disorder (MDD), bipolar II disorder (BD2), bipolar I disorder (BD1), schizoaffective disorder (SZA), and schizophrenia (SCZ).

These analyses yield insight into three conceptual questions in the biological study of psychopathology. First, how do the genetic bases of mood and psychotic symptoms compare and contrast with the genetic bases of psychiatric diagnoses that are characterized by those symptoms? Second, to what extent can the shared genetic architecture of these symptoms and disorders be summarized with a single dimension of liability, a "p factor," as has been previously proposed on the basis of phenotypic analyses?[20] Third, how are dimensions of transdiagnostic liability similar and dissimilar in their genetic architecture, underlying biology, and associations with other aspects of human wellbeing and disease?

RESULTS

Novel loci associated with lifetime endorsement of mood and psychotic symptoms

We used a combination of Bayesian item response theory and linear mixed models to conduct univariate GWASs for self-reported measures of lifetime depression, mania, and psychosis from 252,252 individuals in the UK Biobank. We observed substantial inflation of the median test statistic for all three phenotypes, and the linkage disequilibrium (LD) score regression intercepts and attenuation ratios suggest that test-statistic inflation is primarily due to polygenic signal rather than bias (Figure 1; Table 1). After applying a standard clumping algorithm via FUMA ($r^2 = 0.1$, 250 kb merge window), we identified 23 independent loci associated with lifetime depressive, manic, and/or psychotic symptoms (Tables S1A–S1C). Nine of these loci were significantly associated with two or more phenotypes, and six loci were associated with all three psychiatric phenotypes.

The identified risk loci span 12 chromosomes and include variants tagging the major histocompatibility complex region on chromosome 6 as well a well-known inversion polymorphism on chromosome 17 previously associated with several psychiatric phenotypes.[22] Many of these risk loci replicated previous findings from GWASs of psychopathology or were in high LD with previous hits for phenotypes including

neuroticism[23] (e.g., rs7111031, rs10503002, rs4245154), broadly defined depression[11] (e.g., rs9586, rs191800971, rs7111031), and schizophrenia[24] (e.g., rs4245154, rs4702). However, several loci contained lead SNPs that were new GWAS signals altogether, identifying new regions of the genome that confer risk for psychopathology, such as rs4722389, rs7324564, and rs570217967.

Moreover, our gene-based association analyses performed via MAGMA identified 124 genes associated with at least one of the psychiatric symptoms (depression, mania, or psychosis), 31 of which were associated with all three. For all phenotypes, we observed enriched expression in brain tissue as well as an enriched signal for brain-related gene sets. We report detailed biological annotation (e.g., gene mapping, gene set enrichment, tissue enrichment) for each of these GWASs in Tables S1D–S1L.

Two transdiagnostic genetic liabilities underlie mood and psychotic psychopathology

To describe the genomic relationships among psychiatric symptoms and disorders commonly characterized by depression, mania, and/or psychosis[4,5,21] (see Table 1 for overview of study phenotypes), we first used bivariate LD score regression to estimate genetic correlations between all pairs of psychiatric phenotypes. While we observed very large positive genetic correlations among the three psychiatric symptoms (mean $r_g = 0.95$, SEM = 0.02), we observed more modest genetic correlations for the five psychiatric disorders (mean $r_g = 0.55$, SEM = 0.09). We found that schizophrenia, schizoaffective disorder, and bipolar I were highly correlated with one another ($r_{gSCZ-SZA} = 0.87$ [SE = 0.13], $r_{gSCZ-BD1} = 0.72$ [SE = 0.03], $r_{gSZA-BD1} = 0.81$ [SE = 0.12]), but these disorders generally had markedly smaller genetic correlations with bipolar II and major depressive disorder ($r_{gSCZ-BD2} = 0.53$ [SE = 0.03], $r_{gSCZ-MDD} = 0.39$ [SE = 0.04], $r_{gSZA-BD2} = 0.28$ [SE = 0.21], $r_{gSZA-MDD} = 0.06$ [SE = 0.12], $r_{gBD1-MDD} = 0.33$ [SE = 0.04]). Bipolar I and bipolar II were highly correlated, though ($r_{gBD1-BD2} = 0.88$ [SE = 0.11]). Interestingly, we found that bipolar II and major depressive disorder were also highly correlated with each other ($r_{gBD2-MDD} = 0.69$ [SE = 0.13]) as well as with all psychiatric symptoms ($r_{gBD2-DEP} = 0.75$ [SE = 0.11], $r_{gBD2-MAN} = 0.71$ [SE = 0.11], $r_{gBD2-PSY} = 0.70$ [SE = 0.11], $r_{gMDD-DEP} = 0.85$ [SE = 0.03], $r_{gMDD-MAN} = 0.77$ [SE = 0.03], $r_{gMDD-PSY} = 0.80$ [SE = 0.04]). Notably, many of these genetic correlations differed from the phenotypic correlations observed in UK Biobank (Figure S1).

After applying a hierarchical-clustering algorithm to the genetic-correlation matrix, we found two distinct clusters of psychiatric phenotypes (Figure 2A). The first cluster comprised the three psychiatric symptoms, major depressive disorder, and bipolar II, and the second cluster comprised bipolar I, schizoaffective disorder, and schizophrenia. We then conducted an exploratory factor analysis (EFA) of the genetic covariance matrix, which produced results that were consistent with the groupings suggested by the hierarchical clustering algorithm. The correlated two-factor model with approximate simple structure suggested that phenotypes principally loaded onto one of two latent genetic factors with negligible cross-loadings (Figure 2B). Combined, these two correlated latent factors explained 81.3% of the total SNP-based genetic variances across phenotypes.

Finally, we formally modeled the genetic covariance matrix via confirmatory factor analysis (CFA). We based our model on the EFA results, which consisted of two correlated latent factors, F1 and F2. F1 can be conceptualized as capturing common psychopathology related to mood disturbance (including self-reported depressive, psychotic, and manic symptoms, as well as bipolar II and major depressive disorder), while F2 can be conceptualized as capturing rarer forms of serious mental illness (bipolar I, schizoaffective disorder, and schizophrenia). We did not estimate any cross-loadings. Instead, we estimated correlated residuals between bipolar I and bipolar II, as inspection of the genetic correlation matrix suggested a unique relationship between these disorders. The path diagram for this model is presented in Figure 2C.

We compared the correlated factor model with a common factor model, where all phenotypes were indicators of a single latent factor (i.e., a p factor) (Section 1.1 of supplemental methods; Figure S2). Briefly, we found that the common factor model showed suboptimal fit to the data, while the correlated factors model with correlated residuals for BD1 and BD2 showed excellent fit (Figure S3). Fit indices from the CFA indicated that the correlated factors model closely approximated the observed genetic covariance matrix ($\chi^2(18) = 496.16$, Akaike information criterion = 532.16, comparative fit index = 0.99, standardized root mean square residual = 0.06). That is, the patterns of covariance among the eight psychiatric phenotypes were most parsimoniously represented by two transdiagnostic latent factors at the genetic level, which were correlated only modestly ($r_g = 0.42$; SE = 0.03). This is a notable divergence from the factor structure frequently observed at the phenotypic level, including that seen with similar phenotypes in the UK Biobank (Section 1.1 of supplemental methods). However, we note that a direct comparison of phenotypic and genetic factor structure cannot be performed in the UK Biobank due to an insufficient number of clinical cases.

Transdiagnostic factors have markedly divergent genetic architectures

We then conducted a multivariate GWAS of the two latent genetic factors, F1 ($N_{eff} = 377,518$) and F2 ($N_{eff} = 51,276$). The results of these analyses are summarized in Table 1 and Figure 3. Briefly, we observed substantial inflation of the median test statistic for both F1 ($\lambda_{GC} = 1.44$, mean $\chi^2 = 1.53$) and F2 ($\lambda_{GC} = 1.46$, mean $\chi^2 = 1.61$), which is indicative of a robust polygenic signal for both factors (Figure S4). The LD score regression intercepts and attenuation ratios for F1 (intercept = 1.05, SE = 0.01; ratio = 0.10, SE = 0.02) and F2 (intercept = 1.02, SE = 0.01; ratio = 0.04, SE = 0.02) suggest that test-statistic inflation is primarily due to polygenic signal rather than bias.

After applying a standard clumping algorithm, we identified 26 and 59 independent loci associated with F1 and F2, respectively (Table 2; Tables S2A and S2B). Only five loci were associated with both factors. While many of these genomic regions have been previously identified in either the constituent GWASs or related studies, several contain novel discoveries. For example, several loci associated with F1 contain lead SNPs that have not been previously associated with psychopathology, such as rs13153844 (p = 2.09 × 10⁻⁹, nearest gene = *PSMC1P5*), rs1551765 (p = 3.89 × 10⁻⁸, nearest gene = *GRIA1*), rs147584788 (p = 1.08 × 10⁻⁸, nearest gene = AC003088.1), and rs8035987 (p = 3.94

$\times 10^{-8}$, nearest gene = *SIN3A*). Several loci associated with F2 also contain lead SNPs that were also novel risk variants for psychopathology, including rs2953329 (p = 3.27×10^{-8}, nearest gene = *AKT3*), rs10199182 (p = 1.56×10^{-8}, nearest gene = AC068490.2), rs9463650 (p = 3.34×10^{-8}, nearest gene = *RPS17P5*), rs11603014 (p = 2.32×10^{-8}, nearest gene = *RP11–890B15.2*), rs10777957 (p = 1.79×10^{-8}, nearest gene = *ANKS1B*), and rs11064837 (p = 2.43×10^{-8}, nearest gene = *RP11–768F21.1*).

Tests of heterogeneity suggested that the majority of observed SNP effects operate via the latent factors (i.e., associated SNPs primarily had consistent, pleiotropic effects on the constituent phenotypes). Indeed, Q_{SNP} tests identified no heterogeneous loci for F1 and only three heterogeneous loci for F2 with lead SNPs rs11696888 on chromosome 20 (Q_{SNP} p = 2.04×10^{-8}; nearest gene = *STAU1*), rs1990042 on chromosome 7 (Q_{SNP} p = 2.10×10^{-8}, nearest gene = AC004854.1), and rs3764002 on chromosome 12 (Q_{SNP} p = 2.70×10^{-8}; nearest gene = *WSCD2*). Interestingly, the heterogeneous locus with lead SNP rs11696888 also contains rs200005157, which is a four base pair insertion/deletion that was previously identified as a locus with divergent effects on bipolar disorder and schizophrenia.[21] Fine mapping conducted by Ruderfer and colleagues identified *CSE1L* as a plausible causal gene with divergent effects for bipolar disorder and schizophrenia on chromosome 20.

Transdiagnostic factors are related to different aspects of neurobiology

To characterize the effects of variants associated with the transdiagnostic factors of psychopathology, we used FUMA to conduct a series of gene-mapping analyses. Specifically, we used positional mapping to align lead SNPs to genes based on genomic location, expression quantitative trait loci (eQTL) mapping to match *cis*-eQTL SNPs to genes whose expression they affect, and chromatin-interaction mapping to link SNPs to genes on the basis of three-dimensional DNA-DNA interactions. These three methods linked the associated SNPs for F1 and F2 to a combined 287 and 570 putative risk genes, respectively (Tables S2C–S2J). We then used MAGMA to conduct gene-based association analyses, which identified 115 and 243 genes associated with F1 and F2 (Tables S2K and S2L). Finally, we used S-Multi-Xcan to identify 50 and 91 genes associated with differential expression levels in brain tissue for F1 and F2, respectively (Tables S2M and S2N).

Collectively, these five approaches mapped a total of 332 genes to F1 and 710 genes to F2 (Figure 3). Only 159 of these putative risk genes were related to both factors. Focusing on genes implicated across all five methods, we found 11 genes robustly linked to F1 and 16 to F2 (Figure 3). Two genes (*PGBD1* and *XRCC3*) were related to both factors across all analyses. This limited overlap in mapped genes underscores the unique genetic architecture of each factor.

The modest genetic correlation between F1 and F2 (r_g = 0.42, SE = 0.03) implies that the majority of SNP-based genetic variance in each factor is unique from the other. To further characterize the shared and unique genetic architecture of these factors, we used HESS to estimate the local genetic covariance for 1,698 contiguous, similarly sized partitions across the genome. We found that approximately 27% of the genome explains 80% of the total genetic covariance between F1 and F2 and that only 15 genomic partitions share a significant local genetic correlation after correcting for multiple comparisons (Figure 4A).

Gene set enrichment and gene property (i.e., tissue expression) analyses further suggest that the genetic architectures of F1 and F2 are divergent at more granular levels of analysis, converging only at higher levels. While results from gene set enrichment analyses broadly implicated neurodevelopmental and neurobiological pathways for both factors, the specific molecular functions, cellular components, and biological processes tended to differ (Figure 4B; Tables S3A and S3B). For example, gene sets related to neurons were enriched for F1 and F2, but gene sets for specific parts of neurons were differentially enriched (e.g., the axon for F1 versus the somatodendritic compartment for F2). Similarly, in the tissue-expression analysis, we found that the brain was broadly implicated in the pathogenesis of psychopathology, as nearly all brain-related tissues were enriched for both F1 and F2 (Tables S3C and S3D). At the level of brain tissue, the only regions with divergent effects were the substantia nigra and brainstem, which were not significantly enriched for F1 after correction for multiple comparisons. However, shortcomings of these analyses include the relatively low spatial resolution of brain-related gene expression data and the limited sample size of the underlying data.

Therefore, to gain greater insight into potential etiological relationships between psychopathology and neurobiology, we estimated genetic correlations between the transdiagnostic factors of psychopathology and 101 morphological features of the human brain. Although we generally observed negative genetic correlations with cortical and subcortical features (i.e., greater risk for psychopathology was associated with smaller volumes across the brain) and positive with ventricular features (i.e., greater risk for psychopathology was associated with larger ventricular volumes), specific estimates between morphological features and F1 and F2 showed relatively little concordance (Figure 4C). After correcting for multiple comparisons, only the genetic correlation between F1 and the right middle temporal gyrus remained statistically significant ($r_g = -0.15$, SE = 0.04, p = 3.98×10^{-4})(Tables S3E and S3F).

We then used data from the Allen Human Brain Atlas to identify genes with transcriptomic profiles that were spatially similar to the neuroimaging genetic correlation maps for F1 and F2 (Tables S3G and S3H). Notably, these transcriptomically prioritized gene sets for F1 and F2 were entirely disjointed from one another and differentially expressed in pre-and postnatal cortical tissue from the PsychENCODE dataset (Figure 4D). We found that the developmental expression profile of the F1 gene set most closely resembled that of postnatal inhibitory neuronal genes, while the developmental expression profile of the F2 gene set most closely resembled that of prenatal inhibitory neuronal genes[25,26] (Figure S5). We note that the F1 and F2 gene sets also resemble postnatal microglia and prenatal neural progenitor cells, respectively, although to a lesser extent.

Transdiagnostic factors are differentially associated with human health and wellbeing

To better understand how these transdiagnostic genetic liabilities may manifest above and beyond their constituent phenotypes, we conducted a series of genetic correlation and polygenic prediction analyses focused on theoretically relevant phenotypes. In the genetic correlation analyses, we evaluated the relationships between the latent factors of psychopathology and 92 phenotypes broadly related to four broad domains of human health

and wellbeing (Figure 5A; Table S4A). We found that genetic correlation estimates for F1 and F2 were moderately correlated across all broad domains ($r = 0.60$, p = 2.77×10^{-10}) as well as within each of the four domains: demography and socioeconomic status ($r = 0.55$, p = 1.17×10^{-2}), health and disease ($r = 0.42$, p = 1.17×10^{-2}), personality and risky behavior ($r = 0.60$, p = 3.05×10^{-3}), and psychopathology and cognition ($r = 0.63$, p = 1.57×10^{-2}). Generally, we found that F1 was more consistently correlated with phenotypes typically related to psychopathology than F2. This pattern was also observed in the partial genetic correlation analyses, where we found strong evidence of divergent genetic correlations after accounting for the overlap between F1 and F2 (Figure 5B; Table S4B). Indeed, partial genetic correlation estimates for F1 and F2 were negatively correlated across all domains ($r = -0.43$, p = 2.72×10^{-5}).

In the polygenic prediction analyses, we used electronic health records from the Vanderbilt University Medical Center biobank (BioVU) to evaluate the penetrance and pleiotropy of genetic risk for the transdiagnostic factors of psychopathology across 1,335 disease phenotypes, hereby referred to as "phecodes" (Figure 5C; Tables S4C and S4D). We found that polygenic scores for F1 and F2 were generally associated with all of the constituent phenotypes for both factors, but F1 was more strongly associated with mood-related phecodes, while F2 was more strongly associated with psychosis-related phecodes. Both polygenic scores for F1 and F2 shared associations with some forms of psychopathology (e.g., suicidality, posttraumatic stress disorder, substance-use disorders, and anxiety disorders) but diverged in their associations with others (e.g., personality disorders, paranoid disorders). Beyond psychopathology, F1 was more consistently associated a variety of medical phecodes, including those related to infectious diseases (e.g., viral hepatitis, HIV disease) and pervasive developmental disorders as well as diseases of the circulatory, digestive, endocrine, genitourinary, musculoskeletal, and respiratory systems.

DISCUSSION

By jointly analyzing genome-wide data for eight psychiatric disorders and symptoms in a novel multivariate framework, we identified two distinct transdiagnostic factors that distinguished common forms of psychopathology related to mood disturbance versus rare forms of serious mental illness. Together, these factors explained approximately 80% of the SNP-based genetic variance in mood and psychotic psychopathology but were themselves only moderately correlated. Extensive biological annotation of these two transdiagnostic factors revealed clear differences between their factors in their underlying genetic architecture and biology. Further follow-up analyses highlighted additional differences between the factors in their associations with human wellbeing and disease. Our results provide four critical insights into the genetic architecture of forms of psychopathology characterized by mood disturbance and psychosis.

First, we built on genomic investigations of the dimensional structure of certain forms of psychopathology, such as a large-scale study of the mood-disorder spectrum,[27] and identified two transdiagnostic factors that explain the vast majority of SNP-based genetic variance in their constituent phenotypes. Perhaps surprisingly, variation in self-reported manic and psychotic symptoms is much more closely related to common forms of mood

psychopathology (self-reported depressive symptoms, major depressive disorder, bipolar II disorder) than to psychiatric disorders characterized by severe mood disturbance and/or psychosis. These findings extend those of an earlier GWAS of psychotic experiences,[28] which also reported stronger genetic overlap with major depressive disorder than bipolar disorder or schizophrenia. Here, we also find that the factor structure at the genetic level is different than the factor structure that we observe at the phenotypic level in the UK Biobank with similar indicators. This finding contrasts with what has been called the "phenotypic null hypothesis," which states that genetic and phenotypic factor structures are expected to converge.[29] Overall, these results illustrate how diagnostic boundaries, which are known to be problematic based on widespread phenotypic comorbidity, become even fuzzier at the genetic level of analysis.

Second, our multivariate association analyses identified 80 approximately independent loci associated with one of the transdiagnostic factors. Many of these genome-wide significant loci contain novel lead SNPs and map to genes that have not been previously associated with mood or psychotic psychopathology, such as *SIN3A*, which has been reported to be a key transcriptional regulator of cortical neurodevelopment, involved in neurogenesis and corticocortical projections in the developing mammalian brain.[30] Moreover, by employing multiple gene-mapping techniques, we were also able to triangulate on novel genes associated with psychopathology, including *WDR73*, the causal gene in a rare recessive autosomal disorder characterized by severe encephalopathy, developmental delay, and neurocognitive impairment.[31] Associations such as these are particularly interesting in light of results suggesting that genes disrupted in Mendelian disorders are also dysregulated by non-coding variants in phenotypically similar traits and disorders.[32] Furthermore, we build on the results of a large GWAS of eight psychiatric disorders[33] by providing novel evidence of factor-specific pleiotropy (i.e., consistent effects across a factor's constituent indicators) via Q_{SNP} results, which also identified several novel loci with significantly heterogeneous effects for bipolar I disorder, schizoaffective disorder, and schizophrenia.

Third, our extensive biological annotation revealed a marked divergence in the biology associated with the two transdiagnostic factors. While we find that the CNS is dually implicated at a broad systems-based level (e.g., non-specific enrichment of brain tissues), the biology associated with the two factors quickly diverges at more molecular levels of investigation. Via our novel approach to gene prioritization based on spatial transcriptomics, we identified two sets of factor-specific genes with contrasting developmental expression profiles. Specifically, we found that transcriptomically prioritized genes associated with the factor broadly characterized by common mood disturbance (F1) exhibited lower expression levels during early prenatal periods, while transcriptomically prioritized genes for the factor broadly characterized by rarer forms of serious mental illness (F2) exhibited higher expression levels during early prenatal periods. Notably, both of these trajectories identify the prenatal epoch as a critical developmental period related to psychopathology, albeit in different ways. These findings coalesce with and build upon previous studies that have begun to characterize developmental expression patterns of transdiagnostic genetic liabilities.[34] Here, we found that the two observed trajectories strongly resembled those of postnatal and prenatal inhibitory neuronal genes,[25] which have been implicated in the development of mood and psychotic disorders.[35–37]

Fourth, we found that the two factors differ substantially in their associations with human wellbeing and disease. Our results expand upon recent phenome-wide association studies of genetic risk for major depressive disorder[38] and schizophrenia,[39] expanding the list of complex traits and medical phenotypes associated with mood and psychotic psychopathology. We also identified an interesting pattern of results in our genetic correlation and phenome-wide association analyses, where the factor comprising more common forms of mood disturbance (F1) had broader and often stronger negative associations with socioeconomic and health-related outcomes than the factor comprising rarer forms of serious mental illness (F2). This runs counter to associations often observed at the phenotypic level, where individuals diagnosed with more serious mental illnesses tend to face more severe impairments and consequences in these domains.[40,41] These results raise questions about the potential ascertainment biases that affect GWASs. For example, clinically ascertained samples of people with diagnosed psychiatric disorders (particularly when those disorders are rare and seriously impairing) are subject to different sources of selection, attrition, and non-response than population-based studies that utilize self-report surveys. Consider, for instance, that individuals who are homeless and incarcerated in Western countries are drastically more likely than the general population to meet diagnostic criteria for a serious mental illness,[42,43] but these socially marginalized groups are less likely to have access to adequate mental health care or be included in medical research. This selective representation of psychopathology may induce collider bias and lead to misleading estimates of genetic association.[44] Indeed, cohort-level studies have already found that educational attainment polygenic scores are positively associated with research participation, while psychopathology polygenic scores are negatively associated.[45,46]

Limitations of the study

While we have taken many steps to address potential confounds, these major findings should be interpreted in light of several limitations. First, SEM does not reveal a "ground truth" about the nature of the phenotypes included in the analysis. Instead, it is a useful statistical framework for representing complex data structures, and latent factors are most appropriately considered as convenient statistical entities that explain the (co)variances of their indicators. As such, latent genetic factors are most useful as explanatory devices when accompanied by extensive biological annotation and follow up, as done in the present study. Second, the univariate GWASs are comprised of different samples with different measurement approaches and varying levels of power. However, we have made efforts to harmonize each of the GWASs used in Genomic SEM analyses (e.g., excluding self-rated measures from diagnostic phenotypes, and vice versa), and previous examination of these concerns suggest that the genetic factor structure is not biased by sample overlap or sample-size differences.[19,33]

Third, the univariate GWASs are comprised of different cohorts that may be subject to different sources of bias that cannot be fully quantified. For example, we note that the UK Biobank is subject to a volunteer selection bias, where study participants are generally healthier than individuals who are not study participants.[47] In ideal circumstances, the phenotypic and genetic factor analyses reported here would be performed within the same participants in order to minimize confounding by differences in sample selection. However,

such deep phenotypic data are not currently available at the scale required for GWASs. Fourth, the current study focuses on forms of psychopathology that involve a wide variety of disturbances in mood and reality testing but does not comprehensively sample the full range of psychiatric disorders. These results thus complement other transdiagnostic research studies that have illuminated how schizophrenia and bipolar I disorder diverge genetically from other clinically defined disorders, such as compulsive disorders and disorders of childhood.[33]

Conclusions

In summary, we have conducted a novel multivariate GWAS of multiple symptoms and disorders spanning mood and psychotic psychopathology. This analysis identified two transdiagnostic genetic liabilities operating quite distinctly from one another. Extensive biological annotation revealed contrasting genetic architectures that implicated prenatal neurodevelopment and neuronal function and regulation in markedly different ways. Given the degree of divergence between these two factors, future research is warranted to investigate the utility and appropriateness of even broader spectra of psychopathology (e.g., the p factor[20]) as explanatory devices at the level of molecular genetics. Collectively, our results suggest that the severity of mood and psychotic symptoms evident in severe psychiatric disorders might actually reflect a difference in kind rather than merely in degree.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Any inquiries about analytical results or other information should be directed to lead contact, Travis T. Mallard (tmallard@mgh.harvard.edu).

Materials availability—No new materials were generated as part of this study.

Data and code availability—This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. This paper does not report custom code or software. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Phenotype construction in UK biobank—M*plus*[48] v8 was used to estimate person-specific thetas (*i.e.*, factor scores) for three symptom domains: depression, mania, and psychosis. As each psychiatric phenotypes was assessed by four items, thetas were estimated via a multidimensional two-parameter probit model,[61] which allowed item-level responses across measurement occasions to be combined for correlated latent variables simultaneously. Furthermore, a combination of multiple imputation and Bayesian estimation with non-informative priors was used to maximally leverage all available responses for participants to minimize the impact of missing data. See Section 1.1 of Methods S1 for further description of the phenotypic modeling.

Univariate genome-wide association analyses—BOLT-LMM[49] v2.3.2 was used to conduct GWASs in the UK Biobank for three lifetime measures of psychiatric symptoms: depression, mania, and psychosis. This approach used a linear mixed model that included a genetic relationship matrix to estimate SNP effects, which offered improved control for population stratification and maximized power by accounting for relatedness among individuals. The first 40 principal components of ancestry computed with flashPCA2 (Section 1.2 of Methods S1), sex, birth year, sex-by-birth year interactions, and batch were included as covariates. EasyQC51 was used to perform extensive quality control on the GWAS summary statistics. The main objective of the quality control was to filter out rare and low-frequency SNPs, as well as SNPs that were not imputed well. Three main filters were imposed: (i) MAF <0.005; (ii) imputation quality score <0.9; (iii) unavailable in reference panel. Additional quality control procedures and filters are further described in Section 1.3 of Methods S1. The reference panel was a combination of the 1000 Genomes phase 3 v5 and UK10K, which has been described in a previous study.[14]

Multivariate genome-wide association analyses—Genomic SEM[19] v0.0.2 was used to conduct multivariate GWAS based on eight phenotypes: depressive symptoms, manic symptoms, psychotic symptoms, major depressive disorder, bipolar II disorder, bipolar I disorder, schizoaffective disorder, and schizophrenia (see Table 1 for overview). Following identification of the confirmatory factor model that best explained the observed genetic covariances among the phenotypes, Genomic SEM was used to estimate the individual SNP effects on each latent factor in the model. Note that Genomic SEM is unbiased in the presence of varying and unknown sample overlap across the contributing GWAS samples, as the cross-trait intercepts estimated via multivariable LD score regression are used to estimate (and account for) sample overlap and phenotypic correlation.

Effective sample size (N_{eff}) for each latent factor was estimated as $N_{eff} \approx \frac{1}{m} \sum_{a}^{b} n_j$, where m is the number of SNPs in the GWAS, a is the lower MAF threshold for inclusion in the calculation (here, 10%), b the upper limit (here, 40%), and n_j is the effective sample size for SNP j, which is calculated as $(Z_j/\beta_j)^2/\sigma_j^2$. Q_{SNP} tests were used to evaluate whether SNP effects on the latent factors were driven by heterogeneous effects across constituent phenotypes. Further description of multivariate association analyses and Q_{SNP} tests is provided in Sections 2.4 and 2.5 of Methods S1, respectively.

Genetic correlations among study phenotypes—LD score regression[52] v1.0.1 was used to estimate genetic correlations between all pairwise combinations of the eight study phenotypes. Standard procedures and best practices for LD score regression were followed (*e.g.*, restricting to HapMap3[62] SNPs with a minor allele frequency 0.01). Default parameters were used for the three new GWASs of psychiatric symptoms. For the existing GWASs of psychiatric disorders, parameters (*e.g.*, sample prevalence, population prevalence) were defined as outlined in the original studies. A hierarchical clustering algorithm was applied to the final genetic correlation matrix to guide factor selection in the exploratory factor analysis. Although the original LD score regression software was used for this preliminary analysis, the multivariable version of LD score regression employed by

Genomic SEM was used for all subsequent analyses. Please note that these software produce estimates that are effectively identical.

Exploratory factor analysis—The stats R package was used to conduct an EFA of the genetic correlations among the eight study phenotypes. Specifically, the *factanal* function was used to conduct an EFA with promax rotation on the standardized S matrix derived from the multivariable version of LD score regression employed by Genomic SEM. This enabled an empirical assessment of (i) the number of latent factors that best explained the multivariate genetic architecture observed among the set of study phenotypes (*i.e.*, the number of transdiagnostic liabilities present), and (ii) how constituent phenotypes load onto separable latent factors. As suggested by the hierarchical clustering algorithm, two factors were extracted that optimally accounted for shared variation among sets of the observed variables. Results from this analysis were subsequently used to guide construction of the confirmatory factor models. A brief overview of factor analysis is provided in Section 2.2 of Methods S1.

Confirmatory factor analysis—Genomic SEM was used to test whether a common factor model or a correlated factors model best fit the data via CFA, where fit reflects the degree to which the specified latent variable structure adequately explains the observed covariances among the set of observed variables. Parameter estimates were derived using weighted least squares estimation. Model fit was assessed using conventional indices in structural equation modeling: the model χ^2 statistic, the Akaike information criterion (AIC), the comparative fit index (CFI), and the standardized root mean square residual (SRMR). All fit indices retain their standard interpretations within a Genomic SEM framework. However, the model χ^2 statistic is best used as a comparative measure of fit to evaluate competing models rather than a measure of statistical significance given the sensitivity of model χ^2 to sample size, which is comparatively extremely large for GWAS samples. For CFI and SRMR, values greater than .90 and less than .08, respectively, were considered reflective of good model fit.[63] Further description of structural equation modeling and confirmatory factor analysis are provided in Section 2.3 of Methods S1.

Heritability for observed and latent phenotypes—LD score regression was used to estimate the heritability of the three psychiatric symptom phenotypes, as well as the two latent genetic factors. Standard procedures and best practices for LD score regression were followed. As there is no phenotypic variance for latent genetic factors modeled in Genomic SEM, heritability is more accurately referred to as genetic variance for F1 and F2. Furthermore, as genetic variance estimates are influenced by the heritability estimates of constituent phenotypes and the metric of the latent genetic factor, we note that estimates for F1 and F2 should only be interpreted in the context of the present study.

Local heritability and genetic correlations—HESS[53] and its bivariate extension, ρ-HESS,[54] were used to estimate local genetic variance, local genetic covariance, and the proportion of the genome that contributes to the total genetic covariance for F1 and F2. For each factor, HESS was first used to estimate local genetic variance and covariance across 1,698 approximately LD-independent contiguous genomic partitions, averaging 1.5 Mb per

partition. The European samples from the 1000 Genomes Project Phase 3v5[64] (n = 503) were used as a reference panel for these analyses. Independent genomic partitions were then ranked by their absolute genetic covariance, and the percentage that accounted for 80% of the total genetic covariance between F1 and F2 was used to further quantify genetic overlap between F1 and F2.[65]

Gene mapping and identification—The FUMA[55] SNP2GENE pipeline was used to apply a standard clumping algorithm that identified associated genomic loci, lead SNPs within loci, and all independent significant SNPs within loci. The European samples from the 1000 Genomes Project Phase 3v5 (n = 503) were used as a reference panel for LD. FUMA was also used to employ an ensemble of methods to identify putative risk genes for the univariate and multivariate GWAS phenotypes. Specifically, FUMA v1.3.6c was used to conduct positional, eQTL, and chromatin interaction mapping to identify risk-conferring genes that map to genome-wide significant loci. Default parameters were used for each of these analyses. ANNOVAR annotations[66] were used for positional mapping, the Geno-type-Tissue Expression (GTEx) v8 brain dataset[67] was used as the reference tissue data for eQTL mapping, and Hi-C data from adult and fetal human brain samples[68] was used to examine enhancer-promoter and promoter-promoter chromatin interactions.

Two additional methods were employed to identify putative risk genes based on genome-wide summary statistics: and MAGMA[56] and S-PrediXcan.[57] The former was used to calculate gene-based association statistics, and the latter was used to identify functionally expressed genes via joint analysis of SNP effects and eQTL expression effects. Both methods are described in the following section.

Gene-based association and enrichment analyses—MAGMA v1.08, a bioinformatics software for gene-based biological annotation, was used to conduct gene association, gene set enrichment, and gene property analyses for all novel study phenotypes. Default MAGMA parameters were employed and standard procedures were followed for gene-based association analyses based on summary statistics. MAGMA was also used to conduct competitive gene-set enrichment and gene property analyses based on the gene-level p values produced in the association analyses. These analyses tested whether genes within an annotated set are more strongly associated with the phenotype of interest than other genes. For gene set enrichment analyses, up to 9,987 gene sets cataloged in MolSigDB v7.0 were tested, which corresponded to 7,343 biological processes, 1,001 cellular components, and 1,643 molecular functions. For the gene property analyses, 54 tissues from the GTEx v8 dataset were tested. Bonferroni-corrected thresholds of p 5.01e-6 and p 9.26e-4 were used to determine significance for gene sets and tissues, respectively.

S-PrediXcan v0.6.2 was used (i) to predict gene expression levels in brain tissues, and (ii) to test whether predicted gene expression correlated with either transdiagnostic factor. Tissue weights were computed using reference data from the GTEx v8 dataset. GWAS summary statistics for F1 and F2, the reference transcriptomic data, and covariance matrices for the SNPs within each gene model were included as input data. Thirteen brain tissues were tested: anterior cingulate cortex, amygdala, caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens basal

ganglia, putamen basal ganglia, spinal cord and substantia nigra. A Bonferonni-corrected threshold of p   8.97e-7 was established for transcriptome-wide significance, which corrected for 55,753 gene-based tests.

Genetic correlation analyses between latent factors and other complex traits —Genomic SEM was used to estimate genetic correlations and partial genetic correlations between latent factors of psychopathology and other phenotypes of interest. Specifically, genetic correlations were estimated for two broad sets of phenotypes: (i) morphological features of the human brain, and (ii) complex traits related to human health and well-being. Summary statistics for 101 neuroimaging phenotypes[69] (cortical and subcortical gray matter volumes, ventricular volumes, and global measures of brain volume) were downloaded from https://github.com/BIG-S2/GWAS. Summary statistics for 92 phenotypes broadly related to various domains of human health and well-being were downloaded from various online sources, using download links from GWAS Atlas[70] whenever possible. All summary statistics were cleaned and processed using the *munge* function of Genomic SEM, retaining all HapMap3 SNPs outside of the major histocompatibility complex regions with an allele frequency   .01. A Bonferroni correction was applied within each family of tests to adjust p values for multiple comparisons (p   4.95e-4 for neuroimaging phenotypes; p   5.43e-4 for complex traits).

Spatiotemporal transcriptomic analyses—Microarray gene expression data from the Allen Human Brain Atlas (AHBA)[71] were downloaded from https://human.brain-map.org/static/download, and subsequently aligned to the Desikan-Killiany-Tourville atlas ($N = 62$ cortical brain regions)[72] for spatial compatibility with the cortical neuroimaging phenotypes.[73] Spatial correlation coefficients (Spearman's r) were computed for each of 20,647 genes compared against the $-\log_{10}$ p values from F1 and F2. To examine the developmental trajectories of the F1 and F2 gene sets (positive *Z*-scores of AHBA correlation coefficients, p < .05), weighted gene correlation network analysis[58] was used to estimate eigengene values (i.e., gene set expression) for these gene sets in the PsychENCODE dataset, treating each factor-specific gene set as a module. These expression values were then plotted as function of time, using a non-parametric LOESS curve line-of-best-fit to characterize developmental expression trajectories for F1 and F2, which indicated that the prioritized gene sets for each transdiagnostic factor are differentially expressed in pre- and postnatal cortical tissue. Evaluation of cell-type-specific gene sets was performed as above, using available data from a recent cell-specific sequencing study in adult human brain tissue.[25]

Phenome-wide polygenic prediction—PRS-CS[59] and PLINK[60] v1.9 were used to calculate polygenic scores for the transdiagnostic latent genetic factors, F1 and F2. PRS-CS, a Bayesian polygenic prediction method, was used to apply a continuous shrinkage prior to SNP effect estimates and infer posterior SNP weights using GWAS summary statistics for F1 and F2 and an external reference panel to model LD. In the present study, PRS-CS was used to adjust weights for 1,027,871 SNPs typed on both the 1000 Genomes Project Phase 3v5 and the HapMap3 reference panels with a minor allele frequency   .01. The European samples from the 1000 Genomes Project Phase 3v5 (n = 503) were used as a reference panel

for LD. PLINK was then used to calculate polygenic scores for each individual by summing all included variants weighted by the inferred posterior effect size for the effect allele, and converting that value to a Z-score for each participant within the prediction sample.

The genotyped BioVU sample (n = 66,915) was used to test for associations between polygenic scores for F1 and F2 and a wide array of medical phenotypes. Genotyping and quality control for this sample have been described elsewhere. Case-control medical phenotypes, also referred to as "phecodes," were constructed from International Classification of Disease (ICD) diagnostic codes in participant electronic health record data. Two instances of an ICD diagnostic code were required to be present to be classified as a case for a given phecode, and 50 cases were required for a phecode to be analyzed. A total of 1,335 phecodes were included in the phenome-wide association analyses. The PheWAS R package was used to conduct phenome-wide association analyses. A logistic regression model was fit to each of 1,335 case/control phenotypes to estimate the odds of each diagnosis given the polygenic scores for F1 and F2. Sex, median age of the longitudinal electronic health record measurements, and the top 10 principal components of ancestry were included as covariates. A Bonferroni-corrected threshold of at p 3.74e-5 was established for phenome-wide significance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

CONSORTIA

The members of the Bipolar Disorder Working Group of the Psychiatric Genomics Consortium are Eli A. Stahl, Gerome Breen, Andreas J. Forstner, Andrew McQuillin, Stephan Ripke, Vassily Trubetskoy, Manuel Mattheisen, Yunpeng Wang, Jonathan R.I. Coleman, Héléna A. Gaspar, Christiaan A. de Leeuw, Stacy Steinberg, Jennifer M. Whitehead Pavlides, Maciej Trzaskowski, Enda M. Byrne, Tune H. Pers, Peter A. Holmans, Alexander L. Richards, Liam Abbott, Esben Agerbo, Huda Akil, Diego Albani, Ney Alliey-Rodriguez, Thomas D. Als, Adebayo Anjorin, Verneri Antilla, Swapnil Awasthi, Judith A. Badner, Marie Bækvad-Hansen, Jack D. Barchas, Nicholas Bass, Michael Bauer, Richard Belliveau, Sarah E. Bergen, Carsten Bøcker Pedersen, Erlend Bøen, Marco P. Boks, James Boocock, Monika Budde, William Bunney, Margit Burmeister, Jonas Bybjerg-Grauholm, William Byerley, Miquel Casas, Felecia Cerrato, Pablo Cervantes, Kimberly Chambert, Alexander W. Charney, Danfeng Chen, Claire Churchhouse, Toni-Kim Clarke, William Coryell, David W. Craig, Cristiana Cruceanu, David Curtis, Piotr M. Czerski, Anders M. Dale, Simone de Jong, Franziska Degenhardt, Jurgen Del-Favero, J. Raymond De-Paulo, Srdjan Djurovic, Amanda L. Dobbyn, Ashley Dumont, Torbjørn Elvsåshagen, Valentina Escott-Price, Chun Chieh Fan, Sascha B. Fischer, Matthew Flickinger, Tatiana M. Foroud, Liz Forty, Josef Frank, Christine Fraser, Nelson B. Freimer, Louise Frisén, Katrin Gade, Diane Gage, Julie Garnham, Claudia Giambartolomei, Marianne Giørtz Pedersen, Jaqueline Goldstein, Scott D. Gordon, Katherine Gordon-Smith, Elaine K. Green, Melissa J. Green, Tiffany A. Greenwood, Jakob Grove, Weihua Guan, José Guzman-Parra, Marian L. Hamshere, Martin Hautzinger, Urs Heilbronner, Stefan Herms, Maria Hipolito, Per Hoffmann, Dominic Holland, Laura Huckins, Stéphane Jamain, Jessica S. Johnson, Radhika Kandaswamy, Robert Karlsson, James L. Kennedy, Sarah Kittel-Schneider, James A. Knowles, Manolis Kogevinas, Anna C. Koller, Ralph Kupka, Catharina Lavebratt, Jacob Lawrence, William B. Lawson, Markus Leber, Phil H. Lee, Shawn E. Levy, Jun Z. Li, Chunyu Liu, Susanne Lucae, Anna Maaser, Donald J. MacIntyre, Pamela B. Mahon, Wolfgang Maier, Lina Martinsson, Steve McCarroll, Peter McGuffin, Melvin G. McInnis, James D. McKay, Helena Medeiros, Sarah E. Medland, Fan Meng, Lili Milani, Grant W. Montgomery, Derek W. Morris, Thomas W. Mühleisen, Niamh Mullins, Hoang Nguyen, Caroline M. Nievergelt, Annelie Nordin Adolfsson, Evaristus A. Nwulia, Claire O'Donovan, Loes M. Olde Loohuis, Anil P.S. Ori, Lilijana Oruc, Urban Ösby, Roy H. Perlis, Amy Perry, Andrea Pfennig, James B. Potash, Shaun M. Purcell, Eline J. Regeer, Andreas Reif, Céline S. Reinbold, John P. Rice, Fabio Rivas, Margarita Rivera, Panos Roussos, Douglas M. Ruderfer, Euijung Ryu, Cristina Sánchez-Mora, Alan F. Schatzberg, William A. Scheftner, Nicholas J. Schork, Cynthia Shannon Weickert, Tatyana Shehktman, Paul D. Shilling, Engilbert Sigurdsson, Claire Slaney, Olav B. Smeland, Janet L. Sobell, Christine Søholm Hansen, Anne T. Spijker, David St Clair, Michael Steffens, John S. Strauss, Fabian Streit, Jana Strohmaier, Szabolcs Szelinger, Robert C. Thompson, Thorgeir E. Thorgeirsson, Jens Treutlein, Helmut Vedder, Weiqing Wang, Stanley J. Watson, Thomas W. Weickert, Stephanie H. Witt, Simon Xi, Wei Xu, Allan H. Young, Peter Zandi, Peng Zhang, Sebastian Zöllner, Rolf Adolfsson, Ingrid Agartz, Martin Alda, Lena Backlund, Bernhard T. Baune, Frank Bellivier, Wade H. Berrettini, Joanna M. Biernacka, Douglas H.R. Blackwood, Michael Boehnke, Anders D. Børglum, Aiden Corvin, Nicholas Craddock, Mark J. Daly,

Udo Dannlowski, Tõnu Esko, Bruno Etain, Mark Frye, Janice M. Fullerton, Elliot S. Gershon, Michael Gill, Fernando Goes, Maria Grigoroiu-Serbanescu, Joanna Hauser, David M. Hougaard, Christina M. Hultman, Ian Jones, Lisa A. Jones, RenéS. Kahn, George Kirov, Mikael Landén, Marion Leboyer, Cathryn M. Lewis, Qingqin S. Li, Jolanta Lissowska, Nicholas G. Martin, Fermin Mayoral, Susan L. McElroy, Andrew M. McIntosh, Francis J. McMahon, Ingrid Melle, Andres Metspalu, Philip B. Mitchell, Gunnar Morken, Ole Mors, Preben Bo Mortensen, Bertram Müller-Myhsok, Richard M. Myers, Benjamin M. Neale, Vishwajit Nimgaonkar, Merete Nordentoft, Markus M. Nöthen, Michael C. O'Donovan, Ketil J. Oedegaard, Michael J. Owen, Sara A. Paciga, Carlos Pato, Michele T. Pato, Danielle Posthuma, Josep Antoni Ramos-Quiroga, Marta Ribasés, Marcella Rietschel, Guy A. Rouleau, Martin Schalling, Peter R. Schofield, Thomas G. Schulze, Alessandro Serretti, Jordan W. Smoller, Hreinn Stefansson, Kari Stefansson, Eystein Stordal, Patrick F. Sullivan, Gustavo Turecki, Arne E. Vaaler, Eduard Vieta, John B. Vincent, Thomas Werge, John I. Nurnberger, Naomi R. Wray, Arianna Di Florio, Howard J. Edenberg, Sven Cichon, Roel A. Ophoff, Laura J. Scott, Ole A. Andreassen, John Kelsoe, and Pamela Sklar. Their affiliations are provided in the supplemental information.

REFERENCES

1. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulkader RS, Abdulle AM, Abebo TA, Abera SF, et al. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 390, 1211–1259. [PubMed: 28919117]

2. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, and Posthuma D (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. 47, 702–709. advance on. [PubMed: 25985137]

3. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H, et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427. [PubMed: 25056061]

4. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, Mattheisen M, Wang Y, Coleman JRI, Gaspar HA, et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. 51, 793–803. [PubMed: 31043756]

5. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, Adams MJ, Agerbo E, Air TM, Andlauer TMF, et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681. [PubMed: 29700475]

6. Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, Escott-Price V, Falcone GJ, Gormley P, Malik R, et al. (2018). Analysis of shared heritability in common disorders of the brain. Science 360, eaap8757. [PubMed: 29930110]

7. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Nguyen-Viet TA, Bowers P, Sidorenko J, Linnér RK, et al. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. Nat. Genet. 50, 1112. [PubMed: 30038396]

8. O'Donovan MC, and Owen MJ (2016). The implications of the shared genetics of psychiatric disorders. Nat. Med. 22, 1214. [PubMed: 27783064]

9. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, and Wang P (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders.

10. Cuthbert BN (2014). The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 13, 28–35. [PubMed: 24497240]

11. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, Coleman JRI, Hagenaars SP, Ward J, Wigmore EM, et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. 22,343. [PubMed: 30718901]

12. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, 23andMe Research Team, the S.U.D.W.G. of the P.G.C.; Adams MJ, Howard DM, Edenberg HJ, Davies G, Crist RC, et al. (2018). Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. Am. J. Psychiatry 176, 107–118. [PubMed: 30336701]

13. Levey DF, Gelernter J, Polimanti R, Zhou H, Cheng Z, Aslan M, Quaden R, Concato J, Radhakrishnan K, Bryois J, et al. (2020). Reproducible genetic risk loci for anxiety: results From~ 200,000 participants in the million veteran Program. Am. J. Psychiatry, appi–ajp. 10.1176/appi.ajp.2019.19030256.

14. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209. [PubMed: 30305743]

15. Sanchez-Roige S, and Palmer AA (2020). Emerging phenotyping strategies will advance our understanding of psychiatric genetics. Nat. Neurosci. 23, 475–480. 10.1038/s41593-020-0609-7. [PubMed: 32231337]

16. Plomin R, Haworth CMA, and Davis OSP (2009). Common disorders are quantitative traits. Nat. Rev. Genet. 10, 872–878. 10.1038/nrg2670.n [PubMed: 19859063]

17. van den Berg SM, Glas CAW, and Boomsma DI (2007). Variance decomposition using an IRT measurement model. Behav. Genet. 37, 604–616. 10.1007/s10519-007-9156-1. [PubMed: 17534709]

18. Cheesman R, Purves KL, Pingault J-B, Breen G, Plomin R, Rijsdij K F, Eley TC, and others. (2018). Extracting stability increases the SNP heritability of emotional problems in young people. Transl. Psychiatry 8, 223–229. 10.1038/s41398-018-0269-5. [PubMed: 30333497]

19. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, Ip HF, Marioni RE, McIntosh AM, Deary IJ, et al. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat. Hum. Behav. 3, 513–525. 10.1038/s41562-019-0566-x. [PubMed: 30962613]

20. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, Meier MH, Ramrakha S, Shalev I, Poulton R, and Moffitt TE (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders? Clin. Psychol. Sci. 2, 119–137. 10.1177/2167702613497473. [PubMed: 25360393]

21. Ruderfer DM, Ripke S, McQuillin A, Boocock J, Stahl EA, Pavlides JMW, Mullins N, Charney AW, Ori APS, Loohuis LMO, et al. (2018). Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell 173, 1705–1715. 10.1016/j.cell.2018.05.046. [PubMed: 29906448]

22. Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, Meddens SFW, Linnér RK, Rietveld CA, Derringer J, et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet. 48, 624–633. 10.1038/ng.3552. [PubMed: 27089181]

23. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, Savage JE, Hammerschlag AR, Skene NG, Muñoz-Manchado AB, et al. (2018). Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat. Genet. 50, 920–927. 10.1038/s41588-018-0151-7. [PubMed: 29942085]

24. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. 50, 381–389. 10.1038/s41588-018-0059-2. [PubMed: 29483656]

25. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, Duong TE, Gao D, Chun J, Kharchenko PV, and Zhang K (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat. Biotechnol. 36, 70–80. 10.1038/nbt.4038. [PubMed: 29227469]

26. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, Sunkin SM, Li Z, Shin Y, Zhu Y, et al. (2018). Integrative functional genomic analysis of human

brain development and neuropsychiatric risks. Science 362, eaat7615. 10.1126/science.aat7615. [PubMed: 30545854]

27. Coleman JRI, Gaspar HA, Bryois J, Byrne EM, Forstner AJ, Holmans PA, de Leeuw CA, Mattheisen M, McQuillin A, Pavlides JMW, et al. (2019). The genetics of the mood disorder spectrum: genome-wide association analyses of more than 185,000 cases and 439,000 controls. Biol. Psychiatry 88, 169–184. [PubMed: 31926635]

28. Legge SE, Jones HJ, Kendall KM, Pardiñas AF, Menzies G, Bracher-Smith M, Escott-Price V, Rees E, Davis KAS, Hotopf M, et al. (2019). Association of genetic liability to psychotic experiences with neuropsychotic disorders and traits. JAMA Psychiatry 76, 1256–1265. 10.1001/jamapsychiatry.2019.2508. [PubMed: 31553412]

29. Turkheimer E, Pettersson E, and Horn EE (2014). A phenotypic null hypothesis for the genetics of personality. Annu. Rev. Psychol. 65, 515–540. 10.1146/annurev-psych-113011-143752. [PubMed: 24050184]

30. Witteveen JS, Willemsen MH, Dombroski TCD, Van Bakel NHM, Nillesen WM, Van Hulten JA, Jansen EJR, Verkaik D, Veenstra-Knol HE, van Ravenswaaij-Arts CMA, et al. (2016). Haploinsufficiency of MeCP2-interacting transcriptional co-repressor SIN3A causes mild intellectual disability by affecting the development of cortical integrity. Nat. Genet. 48, 877–887. 10.1038/ng.3619. [PubMed: 27399968]

31. Colin E, Huynh Cong E, Mollet G, Guichet A, Gribouval O, Arrondel C, Boyer O, Daniel L, Gubler M-C, Ekinci Z, et al. (2014). Loss-offunction mutations in WDR73 are responsible for microcephaly and steroid-resistant nephrotic syndrome: galloway-Mowat syndrome. Am. J. Hum. Genet. 95, 637–648. 10.1016/j.ajhg.2014.10.011. [PubMed: 25466283]

32. Freund MK, Burch KS, Shi H, Mancuso N, Kichaev G, Garske KM, Pan DZ, Miao Z, Mohlke KL, Laakso M, et al. (2018). Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. Am. J. Hum. Genet. 103, 535–552. 10.1016/j.ajhg.2018.08.017. [PubMed: 30290150]

33. Lee PH, Anttila V, Won H, Feng Y-CA, Rosenthal J, Zhu Z, Tucker-Drob EM, Nivard MG, Grotzinger AD, Posthuma D, et al. (2019). Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179, 1469–1482. [PubMed: 31835028]

34. Schork AJ, Won H, Appadurai V, Nudel R, Gandal M, Delaneau O, Revsbech Christiansen M, Hougaard DM, Bækved-Hansen M, Bybjerg-Grauholm J, et al. (2019). A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. Nat. Neurosci. 22, 353–361. 10.1038/s41593-018-0320-0. [PubMed: 30692689]

35. Volk DW, and Lewis DA (2014). Early developmental disturbances of cortical inhibitory neurons: contribution to cognitive deficits in schizophrenia. Schizophr. Bull. 40, 952–957. 10.1093/schbul/sbu111. [PubMed: 25053651]

36. Fine R, Zhang J, and Stevens HE (2014). Prenatal stress and inhibitory neuron systems: implications for neuropsychiatric disorders. Mol. Psychiatry 19, 641–651. 10.1038/mp.2014.35. [PubMed: 24751963]

37. Duman RS, Sanacora G, and Krystal JH (2019). Altered connectivity in depression: GABA and glutamate neurotransmitter deficits and reversal by novel treatments. Neuron 102, 75–90. 10.1016/j.neuron.2019.03.013. [PubMed: 30946828]

38. Dennis J, Sealock J, Levinson RT, Farber-Eger E, Franco J, Fong S, Straub P, Hucks D, Linton MF, Song W-L, et al. (2019). Genetic risk for major depressive disorder and loneliness in gender-specific associations with coronary artery disease. Preprint at bioRxiv, 512541. 10.1038/s41380-019-0614-y.

39. Zheutlin AB, Dennis J, Karlsson Linné r R., Moscati A, Restrepo N, Straub P, Ruderfer D, Castro VM, C.-Y. Chen, Ge T, et al. (2019). Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. Am. J. Psychiatry 176, 846–855. 10.1176/appi.ajp.2019.18091085. [PubMed: 31416338]

40. Lewis DA, and Levitt P (2002). Schizophrenia as a disorder of neurodevelopment. Annu. Rev. Neurosci. 25, 409–432. 10.1146/annurev.neuro.25.112701.142754. [PubMed: 12052915]

41. Laursen TM, Nordentoft M, and Mortensen PB (2014). Excess early mortality in schizophrenia. Annu. Rev. Clin. Psychol. 10, 425–448. 10.1146/annurev-clinpsy-032813-153657. [PubMed: 24313570]

42. Fazel S, and Danesh J (2002). Serious mental disorder in 23 000 prisoners: a systematic review of 62 surveys. Lancet 359, 545–550. 10.1016/S0140-6736(02)07740-1. [PubMed: 11867106]

43. Fazel S, Geddes JR, and Kushel M (2014). The health of homeless people in high-income countries: descriptive epidemiology, health consequences, and clinical and policy recommendations. Lancet 384, 1529–1540. 10.1016/S0140-6736(14)61132-6. [PubMed: 25390578]

44. Munafó MR, Tilling K, Taylor AE, Evans DM, and Davey Smith G (2018). Collider scope: when selection bias can substantially influence observed associations. Int. J. Epidemiol. 47, 226–235. 10.1093/ije/dyx206. [PubMed: 29040562]

45. Martin J, Tilling K, Hubbard L, Stergiakouli E, Thapar A, Davey Smith G, O'Donovan MC, and Zammit S (2016). Association of genetic risk for schizophrenia with nonparticipation over time in a population-based cohort study. Am. J. Epidemiol. 183, 1149–1158. 10.1093/aje/kww009. [PubMed: 27188935]

46. Taylor AE, Jones HJ, Sallis H, Smith GD, Lawlor D, Davies NM, Zammit S, Lawlor DA, Munafó MR, Davey Smith G, et al. (2018). Exploring the association of genetic factors with participation in the avon longitudinal study of parents and children. Int. J. Epidemiol. 47, 1207–1216. 10.1093/ije/dyy060. [PubMed: 29800128]

47. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, Collins R, and Allen NE (2017). Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. Am. J. Epidemiol. 186, 1026–1034. 10.1093/aje/kwx246. [PubMed: 28641372]

48. Muthén LK, and Muthén B (2016). Mplus. The Comprehensive Modelling Program for Applied Researchers: User's Guide.

49. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. 47, 284–290. 10.1038/ng.3190. [PubMed: 25642633]

50. Abraham G, Qiu Y, and Inouye M (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics.

51. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, Ferreira T, Fall T, Graff M, Justice AE, et al. (2014). Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. 9, 1192–1212. 10.1038/nprot.2014.071. [PubMed: 24762786]

52. Bulik-Sullivan BK, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, Duncan L, Perry JRB, Patterson N, Robinson EB, et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241. 10.1038/ng.3406. [PubMed: 26414676]

53. Shi H, Kichaev G, and Pasaniuc B (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. Am. J. Hum. Genet. 99, 139–153. [PubMed: 27346688]

54. Shi H, Mancuso N, Spendlove S, and Pasaniuc B (2017). Local genetic correlation gives insights into the shared genetic architecture of complex traits. Am. J. Hum. Genet. 101, 737–751. [PubMed: 29100087]

55. Watanabe K, Taskesen E, Van Bochoven A, and Posthuma D (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826–1911. 10.1038/s41467-017-01261-5. [PubMed: 29184056]

56. de Leeuw CA, Mooij JM, Heskes T, Posthuma D, Visscher P, Brown M, McCarthy M, Yang J, Manolio T, Collins F, et al. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219. 10.1371/journal.pcbi.1004219. [PubMed: 25885710]

57. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. 9, 1825–1920. 10.1038/s41467-018-03621-1. [PubMed: 29739930]

58. Langfelder P, and Horvath S (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. 10.1186/1471-2105-9-559. [PubMed: 19114008]

59. Ge T, Chen C-Y, Ni Y, Feng Y-CA, and Smoller JW (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat. Commun. 10, 1776. 10.1038/s41467-019-09718-5. [PubMed: 30992449]

60. Chang CC, Chow CC, Tellier LCA, Vattikuti S, Purcell SM, and Lee JJ (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience 4, 7. 10.1186/s13742-015-0047-8. [PubMed: 25722852]

61. Sheng Y, and Wikle CK (2008). Bayesian multidimensional IRT models with a hierarchical structure. Educ. Psychol. Meas. 68, 413–430.

62. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Dermitzakis E, Schaffner SF, Yu F, Dermitzakis E, Bonnen PE, de Bakker PIW, et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58. 10.1038/nature09298. [PubMed: 20811451]

63. Hu L, and Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equation Model. a multidisciplinary J. 6, 1–55.

64. Auton A, Abecasis GR, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. (2015). A global reference for human genetic variation. Nature 526, 68–74. 10.1038/nature15393. [PubMed: 26432245]

65. Cai N, Revez JAA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, Clarke T-K, Forstner AJJ, Grabe HJJ, Hamilton SP, et al. (2019). Minimal phenotyping yields gwas hits of reduced specificity for major depression. Preprint at BioRxiv, 440735. 10.1101/440735.

66. Wang K, Li M, and Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164. 10.1093/nar/gkq603. [PubMed: 20601685]

67. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, Kasela S, Kim-Hellmuth S, Liang Y, Oliva M, et al. (2019). The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. Preprint at BioRxiv, 787903. 10.1126/2Fscience.aaz1776.

68. Giusti-Rodriguez PMD, and Sullivan PF (2019). Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. Preprint at BioRxiv, 406330. 10.1101/406330.

69. Zhao B, Luo T, Li T, Li Y, Zhang J, Shan Y, Wang X, Yang L, Zhou F, Zhu Z, and Zhu H (2019). Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. Nat. Genet. 51, 1637–1644. 10.1038/s41588-0190516-6. [PubMed: 31676860]

70. Watanabe K, Stringer S, Frei O, Umicevic Mirkov M,deLeeuw C, Polderman TJC, van der Sluis S, Andreassen OA, Neale BM, and Posthuma D (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. 51, 1339–1348. 10.1038/s41588-019-0481-0. [PubMed: 31427789]

71. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, Van De Lagemaat LN, Smith KA, Ebbert A, Riley ZL, et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489, 391–399. 10.1038/nature11405. [PubMed: 22996553]

72. Klein A, and Tourville J (2012). 101 labeled brain images and a consistent human cortical labeling protocol. Front. Neurosci. 6, 171. 10.3389/fnins.2012.00171. [PubMed: 23227001]

73. Romero-Garcia R, Whitaker KJ, Váša F, Seidlitz J, Shinn M, Fonagy P, Dolan RJ, Jones PB, Goodyer IM, Bullmore ET, and Vertes PE (2018). Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex. Neuroimage 171, 256–267. 10.1016/j.neuroimage.2017.12.060. [PubMed: 29274746]

Highlights

- Identification of two dimensions of genetic risk in mood and psychotic psychopathology

- Pleiotropic genes broadly implicate neuronal pathways in psychopathology

- Dimensions of genetic risk differ in their associations with health and disease
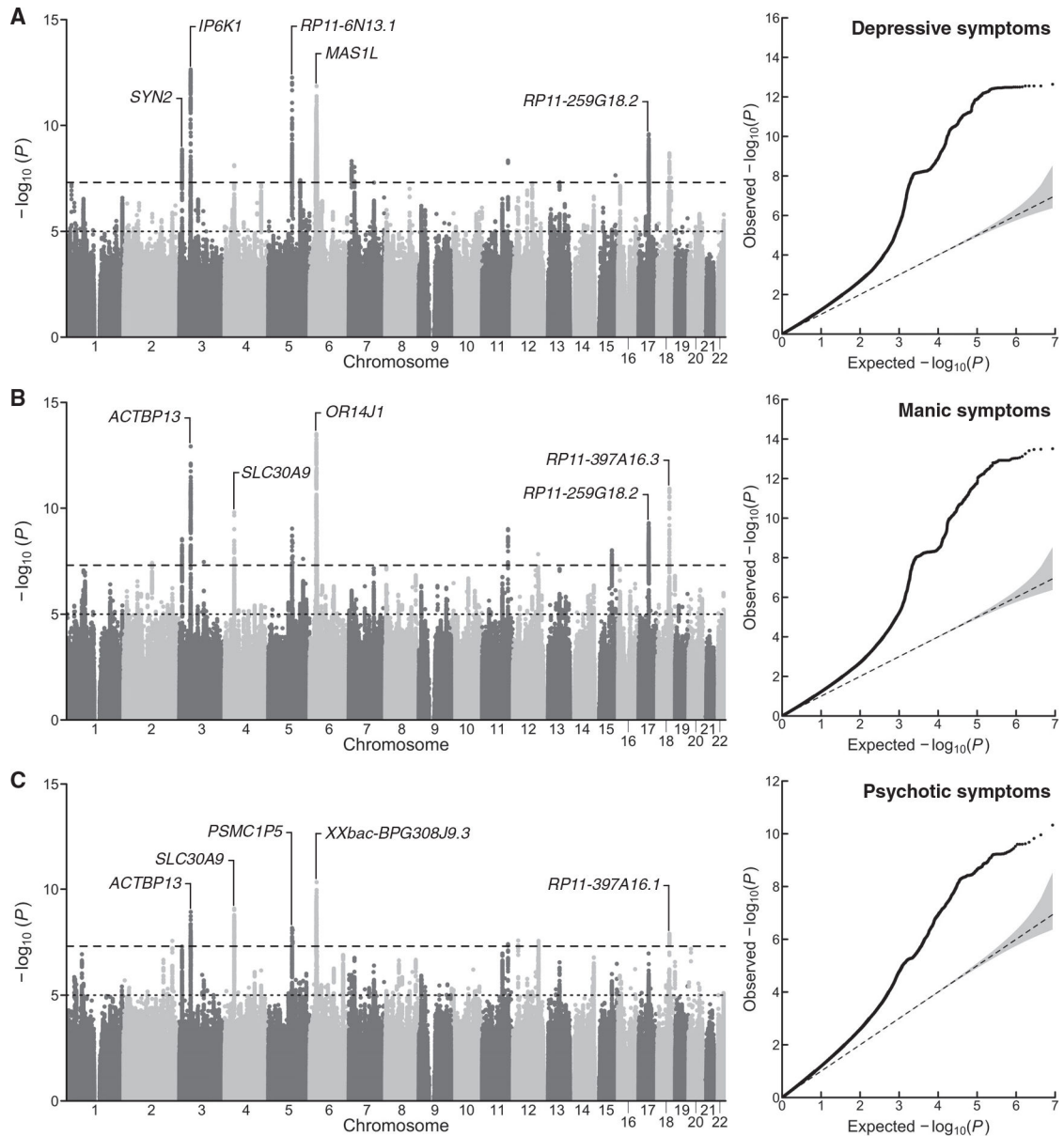
Figure 1. Univariate association results for lifetime measures of mood disturbance and psychosis

(A–C) Manhattan plots and a quantile-quantile plots for (A) depressive, (B) manic, and (C) psychotic symptoms. In the Manhattan plots, the x axis refers to chromosomal position, the y axis refers to the significance on a $-\log_{10}$ scale, the horizontal dashed line denotes genome-wide significance ($p = 5 \times 10^{-8}$), and the horizontal dotted line marks suggestive significance ($p = 5 \times 10^{-5}$). In the quantile-quantile plots, the x axis refers to expected p value, while the y axis refers to the observed p value. For each plot, the nearest gene for the lead SNP in the top five genome-wide significant loci is labeled.
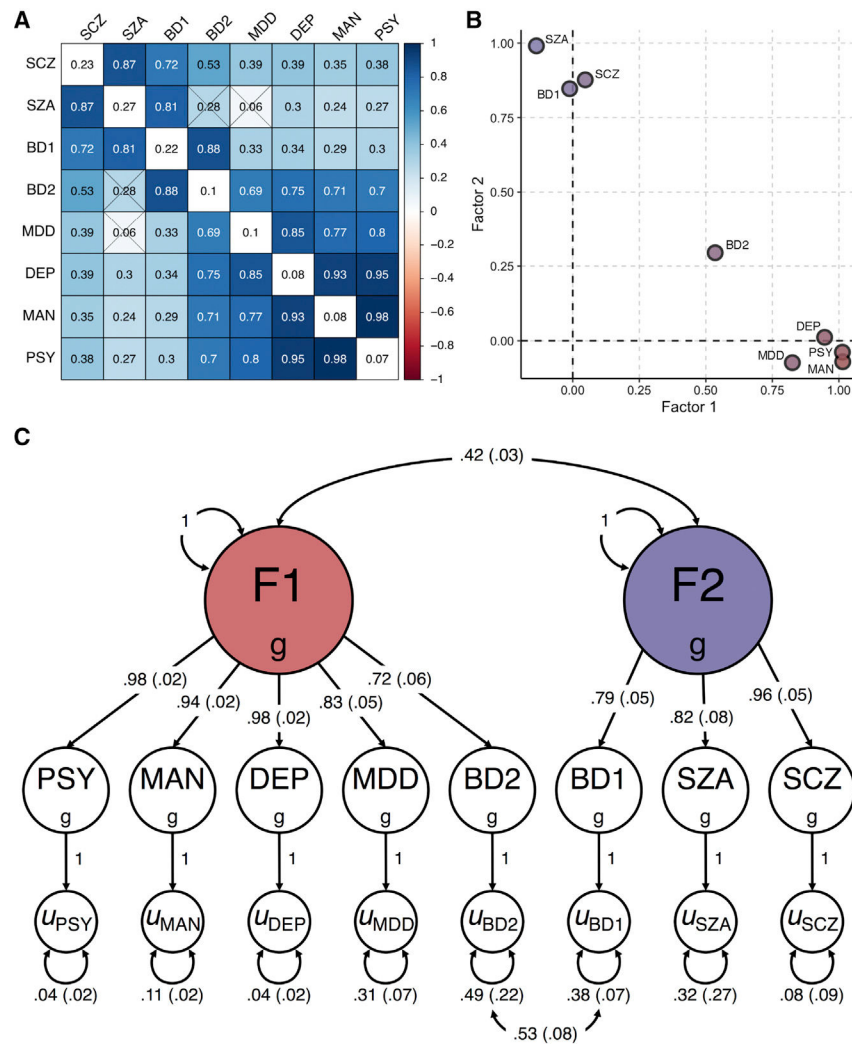
Figure 2. Relationships between eight psychiatric symptoms and disorders

(A) Matrix of bivariate genetic-correlation estimates, where the diagonal elements correspond to SNP h^2 and the off-diagonal elements correspond to genetic correlations. Estimates that are non-significant are crossed out.

(B) Scatterplot of standardized factor loadings from the exploratory factor analysis.

(C) Path diagram for the final confirmatory factor model with standardized parameter estimates.
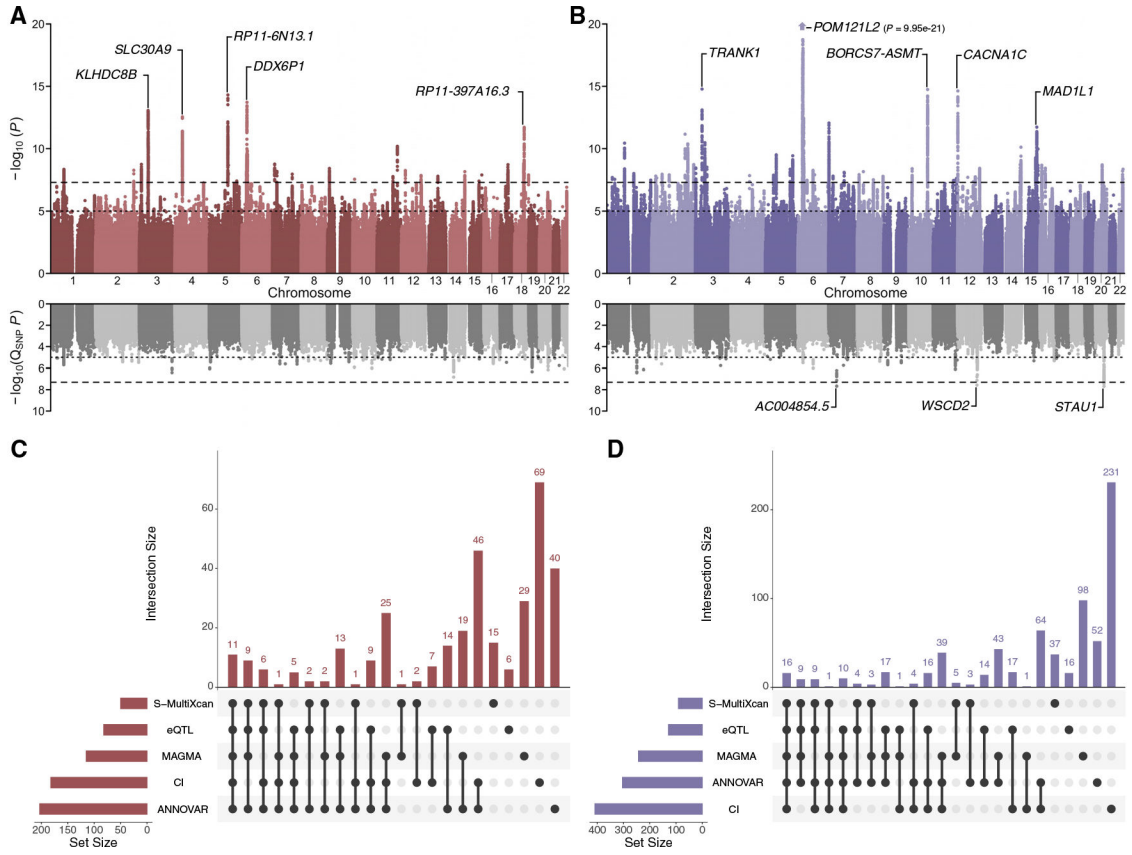
Figure 3. Multivariate association results for the two transdiagnostic latent genetic factors

(A and B) Miami plots for (A) F1 and (B) F2. The top of each Miami plot corresponds to the significance of SNP effects on each latent factor, as traditionally conveyed in a Manhattan plot, while the bottom corresponds to the significance of heterogeneity tests for SNP effects (Q_{SNP}; i.e., the degree to which SNP effects are not mediated by F1 or F2). For each plot, the x axis refers to chromosomal position, the y axis refers to the significance on a $-\log_{10}$ scale, the horizontal dashed line denotes genome-wide significance ($p = 5 \times 10^{-8}$), and the horizontal dotted line marks suggestive significance ($p = 5 \times 10^{-5}$). For each plot, the nearest gene for the lead SNP in the top five genome-wide significant loci is labeled.

(C and D) UpSet plots illustrating the intersection of the five gene-mapping methods, ranked by degree of overlap.
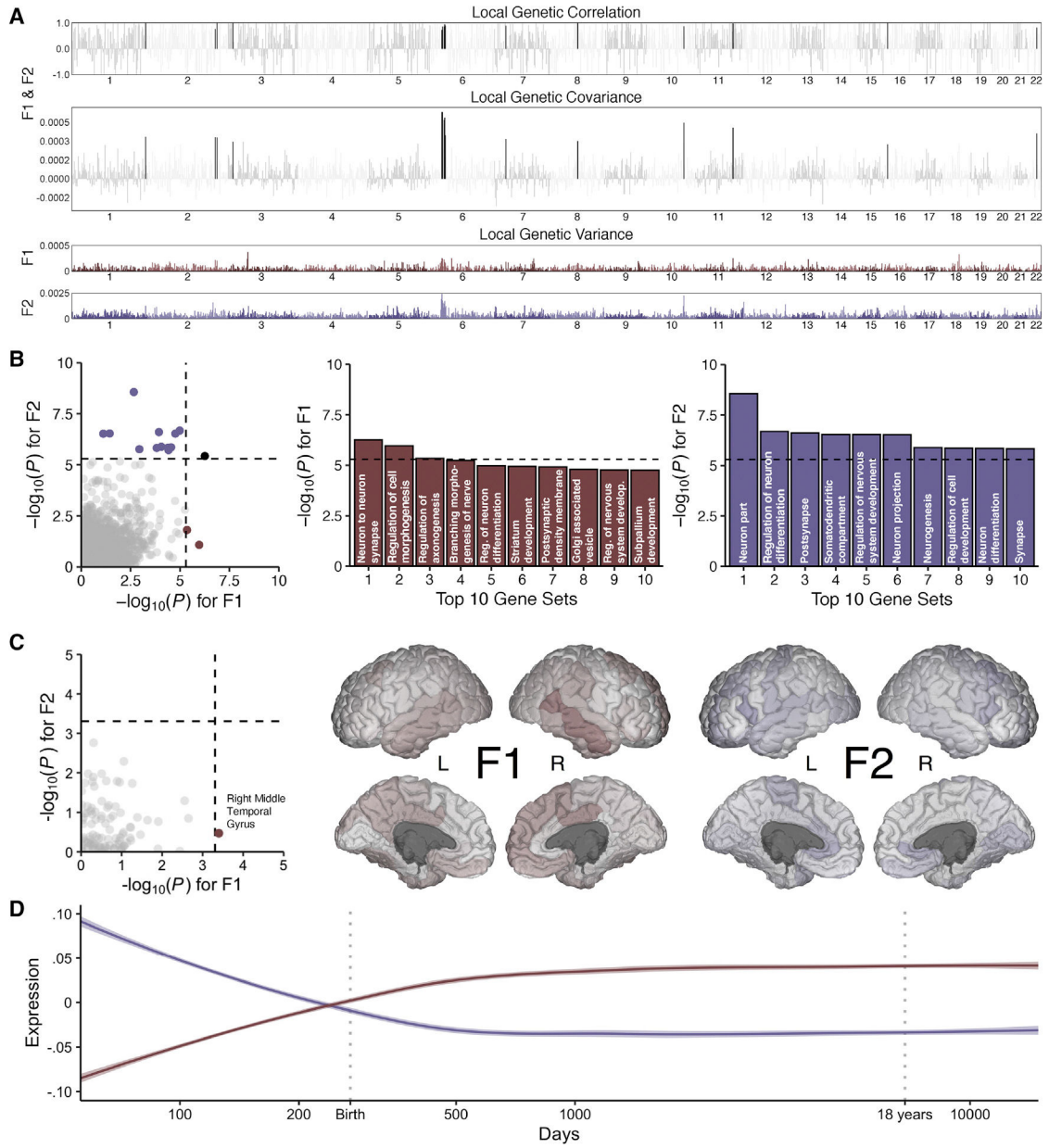
Figure 4. Biological annotation of the two transdiagnostic latent genetic factors

(A) Manhattan plots for local genetic correlation, covariance, and variance for F1 and F2. Black bars indicate significant local genetic correlation.

(B) Scatterplot of gene set enrichment results illustrating convergence and divergence across the latent genetic factors with accompanying histograms for the top 10 gene sets for each factor.

(C) Scatterplot of neuroimaging genetic-correlation results with accompanying figures where the $-\log_{10}$ p values are mapped across the cortex, as parcellated in the Desikan-Killiany-Tourville atlas.

(D) Smoothed line plots of gene set expression across developmental time in the PsychENCODE dataset for prioritized genes with transcriptomic profiles that are spatially

similar to the neuroimaging genetic correlation maps for F1 and F2 (as indexed in the Allen Human Brain Atlas). For all plots, the dashed black line corresponds to the Bonferroni-corrected significance threshold when applicable.
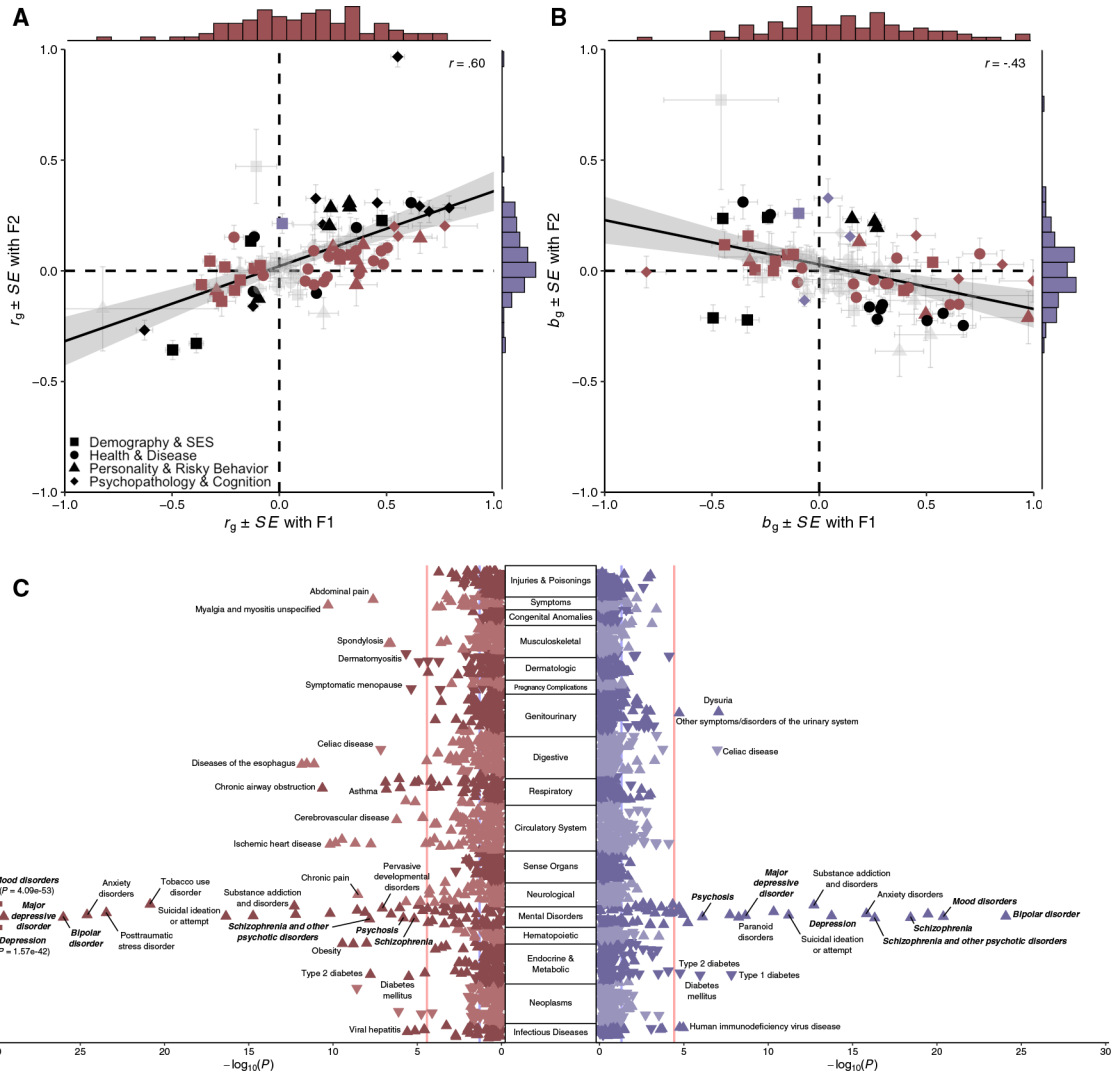
Figure 5. Genetic-correlation and phenome-wide association results for the two transdiagnostic latent genetic factors

(A) Scatterplot of genetic correlations (r_g) with marginal histograms.

(B) Scatterplot of partial genetic correlations (b_g) with marginal histograms. For both plots, phenotypes are grouped into one of four broad domains: (1) demography and socioeconomic status, (2) health and disease, (3) personality and risky behavior, and (4) psychopathology and cognition. A line of best fit (with 95% confidence interval) is fit for all 92 data points. Points are colored burgundy if significant only for F1, violet if significant only for F2, black if significant for both, and faded gray if non-significant for both. The standard errors (SEs) for point estimates are plotted for both factors.

(C) Rotated Miami plot for (left) F1 and (right) F2, where the y axis refers to the ICD-10 code category, the x axis refers to the significance on a –$\log_{10}$ scale, the vertical light red line denotes phenome-wide significance (p = 3.27 × 10^{-5}) following Bonferroni correction, and the vertical light blue line marks nominal significance (p = 0.05). The direction of the

triangle refers to the direction of effect. Phecodes closely resembling Genomic SEM model phenotypes are bolded and italicized for emphasis.

Table 1.

Summary of study phenotypes

GWAS (abbr.)	Source	N	h^2	λ_{GC}	Mean χ^2	Intercept	Ratio
Depressive symptoms (DEP)	present study	252,252	0.08	1.31	1.38	1.01	0.02
Manic symptoms (MAN)	present study	252,252	0.08	1.31	1.39	1.00	0.00
Psychotic symptoms (PSY)	present study	252,252	0.07	1.31	1.33	1.00	0.01
Major depressive D/O (MDD)	Wray et al., 2018[5]	138,884	0.10	1.19	1.20	1.00	<0
Bipolar II D/O (BD2)	Stahl et al., 2019[4]	25,576	0.10	1.07	1.08	1.03	0.42
Bipolar I D/O (BD1)	Stahl et al., 2019[4]	45,871	0.22	1.31	1.37	1.04	0.09
Schizoaffective D/O (SZA)	Stahl et al., 2019[4]	9,667	0.27	1.06	1.06	1.02	0.35
Schizophrenia (SCZ)	Ruderfer et al., 2018[21]	65,967	0.23	1.49	1.63	1.05	0.08
F1 (mood disturbance)	present study	377,518	N/A	1.44	1.53	1.05	0.10
F2 (serious mental illness)	present study	51,276	N/A	1.46	1.61	1.02	0.04

Heritability (h^2) was estimated using LD score regression. λ_{GC} refers to the median χ^2 statistic of the GWAS divided by the expected median of the χ^2 distribution with 1 degree of freedom. Mean χ^2 refers to the average χ^2 statistic of the GWAS. Intercept refers to the estimated intercept from univariate LD score regression. Ratio refers to a measure of stratification bias that is defined as (Intercept − 1)/(Mean χ^2 − 1). To harmonize measurement approaches among psychiatric disorders, summary statistics for MDD were obtained for the clinically ascertained cohorts, excluding 23andMe and UK Biobank. D/O, disorder; The effective sample size is reported for F1 and F2 (STAR Methods).

Table 2.

Lead SNPs for the top ten loci per latent factor from multivariate association analyses

Lead SNP	CHR:BP	A1	A2	MAF	Z	p	Nearest gene	Function	CADD	RDB
F1 (depressive symptoms, manic symptoms, psychotic symptoms, major depressive disorder, and bipolar II disorder)										
rs30266	5:103972357	G	A	0.32	−7.83	4.94×10^{-15}	*RP11-6N13.1*	ncRNA intronic	2.275	N/A
rs148682985	6:29288001	G	A	0.03	−7.66	1.93×10^{-14}	*DDX6P1*	intergenic	2.643	5
rs9586	3:49213637	C	T	0.02	7.46	8.80×10^{-14}	*KLHDC8B*	UTR3	11.63	N/A
rs28656217	4:42099424	T	C	0.16	7.30	2.85×10^{-13}	*SLC30A9*	intergenic	4.861	6
rs67526282	18:53471187	T	C	0.33	−7.04	1.97×10^{-12}	*RP11-397A16.3*	intergenic	7.656	6
rs7934649	11:113372671	C	T	0.36	−6.53	6.50×10^{-11}	*DRD2*	intergenic	1.426	7
rs17410557	18:50776391	T	C	0.38	−6.04	1.52×10^{-9}	*DCC*	intronic	4.502	7
rs3807866	7:12250378	G	A	0.40	−6.03	1.69×10^{-9}	*TMEM106B*	upstream	7.544	N/A
rs184262	3:12134740	A	G	0.15	6.02	1.77×10^{-9}	*SYN2*	ncRNA intronic	6.559	7
rs2696673	17:44315803	A	C	0.22	−6.01	1.89×10^{-9}	*RP11-259G18.2*	intergenic	3.3	N/A
F2 (bipolar I disorder, schizoaffective disorder, and schizophrenia)										
rs7746199	6:27261324	C	T	0.17	9.34	9.95×10^{-21}	*POM121L2*	intronic	0.879	1f
rs9834970	3:36856030	T	C	0.49	−7.96	1.67×10^{-15}	*TRANK1*	intergenic	11.17	4
rs12764899	10:104635103	G	A	0.23	7.95	1.82×10^{-15}	*C10orf32-ASMT:AS3MT*	intronic	1.133	7
rs4298967	12:2408194	A	G	0.34	7.92	2.37×10^{-15}	*CACNA1C*	intronic	10.73	5
rs6461049	7:2017445	C	T	0.44	−7.15	8.80×10^{-13}	*MAD1L1*	intronic	0.914	5
rs12902973	15:85105982	G	C	0.28	7.04	1.89×10^{-12}	*UBE2Q2P1*	ncRNA intronic	1.808	7
rs4380187	2:185811940	A	C	0.45	6.86	6.95×10^{-12}	*ZNF804A*	intergenic	2.923	7
rs2535627	3:52845105	T	C	0.50	6.63	3.36×10^{-11}	*ITIH4*	intergenic	5.307	N/A
rs1198588	1:98552832	A	T	0.23	−6.62	3.65×10^{-11}	*NFU1P2*	intergenic	1.829	3a
rs11693528	2:200736507	C	G	0.18	−6.60	4.18×10^{-11}	*AC073043.1*	ncRNA intronic	2.344	6

Results for all lead SNPs are presented in Tables S2A and S2B. Lead SNPs refer to approximately independent lead SNPs identified via FUMA. CHR:BP refers to genomic location of the lead SNP, specifically the chromosome and base pair location on that chromosome. A1 and A2 refer to the alleles for that SNP. MAF, minor allele frequency; CADD, Combined Annotation Dependent Depletion score; RDB, RegulomeDB score.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Summary statistics for multivariate GWAS of F1 & F2	This study	https://osf.io/jm3y6/
Summary statistics for individual psychiatric symptoms	This study	https://osf.io/jm3y6/
Summary statistics for individual psychiatric disorders	Psychiatric Genomics Consortium	https://www.med.unc.edu/pgc/download-results/
Individual-level data from UK Biobank	UK Biobank	https://biobank.ndph.ox.ac.uk/showcase/
Genotype reference panel	1000 Genomes Project	http://hgdownload.soe.ucsc.edu/downloads.html#human
European LD scores and weights	Broad Institute	https://alkesgroup.broadinstitute.org/LDSCORE/
Developmental postmortem brain gene expression data	PsychENCODE	http://development.psychencode.org#
Adult postmortem brain gene expression data	Allen Human Brain Atlas	https://human.brain-map.org/static/download
Software and algorithms		
Mplus	[48]	https://www.statmodel.com
BOLT-LMM	[49]	https://alkesgroup.broadinstitute.org/BOLT-LMM/downloads/
flashPCA2	[50]	https://github.com/gabraham/flashpca
EasyQC	[51]	https://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software
Genomic SEM	[19]	https://github.com/GenomicSEM/GenomicSEM
LD score regression	[52]	https://github.com/bulik/ldsc
HESS	[53]	https://github.com/huwenboshi/hess
ρ-HESS	[54]	https://github.com/huwenboshi/hess
FUMA	[55]	https://fuma.ctglab.nl
MAGMA	[56]	https://ctg.cncr.nl/software/magma
S-PrediXcan	[57]	https://github.com/hakyimlab/MetaXcan
WGCNA	[58]	https://cran.r-project.org/web/packages/WGCNA/index.html
PRS-CS	[59]	https://github.com/getian107/PRScs
PLINK	[60]	https://www.cog-genomics.org/plink/