Multivariate genome-wide analysis of education, socioeconomic status and brain phenotype

Frank R. Wendt, Gita A. Pathak, Todd Lencz, John H. Krystal, Joel Gelernter and Renato Polimanti

Socioeconomic status (SES) and education (EDU) are phenotypically associated with psychiatric disorders and behaviours. It remains unclear how these associations influence genetic risk for psychopathology, psychosocial factors and EDU and/or SES individually. Using information from >1 million individuals, we conditioned the genetic risk for psychiatric disorders, personality traits, brain imaging phenotypes and externalizing behaviours with genome-wide data for EDU/SES. Accounting for EDU/SES significantly affected the observed heritability of psychiatric traits, ranging from 2.44% \( h^2 \) decrease for bipolar disorder to 14.2% \( h^2 \) decrease for Tourette syndrome. Neuroticism \( h^2 \) significantly increased by 20.23% after conditioning with SES. After EDU/SES conditioning, neuronal cell types were identified for risky behaviour (excitatory), major depression (inhibitory), schizophrenia (excitatory and \( \gamma \)-aminobutyric acid (GABA) mediated) and bipolar disorder (excitatory). Conditioning with EDU/SES also revealed unidirectional causality between brain morphology, psychopathology and psychosocial factors. Our results indicate that genetic discoveries related to psychopathology and psychosocial factors may be limited by genetic overlap with EDU/SES.

EDU and SES are risk or protective factors for traits related to mental health and disease\(^a\). Social position has been repeatedly correlated with mood, anxiety and substance use-related disorders, while EDU phenotypes such as educational attainment, math ability and fluid intelligence are overall protective factors for development of neurological and psychiatric conditions\(^a\). Though highly correlated, the specific EDU and/or SES phenotypes used in epidemiological studies clearly account in part for the incidence of numerous health outcomes, including self-reported health, chronic conditions and overall mortality\(^a\). It is therefore imperative to understand how EDU and SES phenotypes influence what we understand about human health and disease.

Genome-wide association studies (GWAS) are powerful hypothesis-generating investigations for detecting risk loci with respect to phenotypes of interest. Their widespread use has led to risk locus discovery underlying thousands of phenotypes across the spectrum of human health, including mental and physical traits, personality, anthropometric measures, intelligence and behaviours\(^a\). An observation generated from large-scale GWAS is the widespread presence of pleiotropy; a single single-nucleotide polymorphism (SNP) (or a set of SNPs) may have a range of relatively small effects on multiple similar or disparate phenotypes. On a genome-wide scale, these pleiotropic effects, detected using GWAS summary data, may be used to determine genetic correlations between phenotypes to putatively identify genetic underpinnings of trait pairs\(^a\).

The EDU phenotypes of educational attainment and cognitive performance have relatively high SNP heritability: the phenotypic variance explained by genetic information was 40% (ref. \(^a\)) and 21.5% (ref. \(^a\)), respectively. SES is defined as the social standing or class of an individual or group, often measured as a combination of education, income and occupation\(^a\). SES phenotypes such as household income and Townsend deprivation index (a measure of SES based on whether individuals own their home, their employment status, their access to a vehicle and whether or not individuals share living accommodation with others) are significantly heritable and show strong genetic correlation with EDU traits\(^a\). Additionally, there is pleiotropy of genetic risks between EDU/SES and a range of psychopathologies and psychosocial factors (for example, psychiatric disorders, personality traits, internalizing and externalizing behaviours, social science outcomes and brain imaging phenotypes)\(^a\).

The epidemiological observations of high genetic correlations between genetic risk for EDU/SES, psychopathology and psychosocial factors\(^a\) raise two critical questions: (1) How might genetic variants with strong effects on EDU/SES affect our understanding of the overall genetic risk for psychopathology and psychosocial factors? and (2) Is there evidence that genetic variants associated with psychopathology and psychosocial factors affect our understanding of the overall genetic risk for EDU/SES? The goal of this study is to investigate how the shared genetic effects between the general categories of EDU, SES, psychopathology and psychosocial factors influence genetic risk for individual phenotypes within each of these classes.

There are several ways to approach these questions, such as polygenic risk scoring (PRS) or multi-trait analysis of GWAS (MTAG). PRS\(^a\) is an analytic approach by which a person’s genetic information is used to derive a numerical description of their risk to develop a disorder or display a certain trait\(^a\). PRS is a tempting approach to answer our question, but PRS using psychopathology and psychosocial factors to predict the same or different phenotypes from an independent dataset often explains very little variance in the outcome phenotype\(^a\). MTAG analyses the GWAS of several traits with the goal of boosting statistical power and the detection of genetic signal across those traits. MTAG adjusts per-SNP effect estimates and association \( P \)-values using the strength of the
genetic correlation between phenotypes\textsuperscript{7}. Genetic correlations between EDU/SES and related phenotypes have, however, demonstrable biases from environmental confounders. If genetic correlations involving EDU and SES proxy phenotypes are significantly upwardly biased, an MTAG adjustment of summary statistics may inappropriately correct (that is, bias) the summary statistics used for this study. To disentangle the complex genetic overlaps between EDU/SES, psychopathology and psychosocial factors, we therefore used multi-trait conditioning and joint analysis (mtCOJO), which generates conditioned GWAS summary statistics for each phenotype of interest after correcting for the per-SNP effects of another phenotype\textsuperscript{14}. The mtCOJO approach is not based on genetic correlation, but leverages the causal relationship between trait pairs inferred by Mendelian randomization (MR). For our phenotypes of interest, mtCOJO is an advantageous approach that, in theory, is independent of the effects of environmental confounders\textsuperscript{15,20}. MT detects causal inferences between trait pairs using non-modifiable risk factors (SNPs) under the assumption that (1) SNPs are associated with an exposure variable, (2) SNPs are associated with an outcome variable only through the exposure and (3) SNPs are not associated with confounders of the relationship between exposure and outcome. Because SNPs are non-modifiable, environmental confounders of the relationship between SNP, exposure and outcome should not influence MT estimates\textsuperscript{16,20}.

We used the mtCOJO approach to condition psychopathology and psychosocial factors with the per-SNP effects of EDU and SES phenotypes and investigate their underlying biology in more than 125 GWAS at multiple levels: (1) risk locus detection, (2) heritability ($h^2$), (3) gene-set enrichment, (4) tissue transcriptomic profile enrichment, (5) cell type transcriptomic profile enrichment, (6) phenotype relationships via structural equation modelling (SEM) and genetic correlation and (7) latent genetically causal relationships (see flow diagram in Supplementary Fig. 1). Our findings identify several cell types and phenotype relationships that were masked by the shared genetic aetiology between psychopathology, psychosocial factors and EDU/SES. Furthermore, we demonstrate that the same multi-level analyses of EDU and SES are largely robust to the effects of shared genetic aetiology with psychopathology and psychosocial factors.

**Results**

An overview of all the analytic approaches applied and their results is shown in Supplementary Fig. 1.

**Trait inclusion.** The genetic correlations ($r_g$) between EDU (educational attainment, cognitive performance, highest math class and self-rated math ability), SES (household income and Townsend deprivation index) and psychopathology and psychosocial factors (psychiatric disorders, personality traits, externalizing behaviours, social science outcomes and brain imaging phenotypes) were estimated using the linkage disequilibrium score regression (LDSC) method (Fig. 1, Supplementary Fig. 2 and Supplementary Tables 1–4\textsuperscript{21}). The brain imaging phenotype selection is described in the Supplemental Material. Note that sample overlap between EDU/SES, psychopathology and psychosocial factors was accounted for in conditioning experiments via incorporation of the sampling covariance estimated from GWAS summary statistics\textsuperscript{6,18,22}.

**Conditioning heritability and risk locus discovery.** We tested the effects of conditioning on the observed-scale heritability estimates ($h^2$) using LDSC\textsuperscript{22}. Except for major depressive disorder (MDD), anxiety and posttraumatic stress disorder (PTSD), conditioning reduced the $h^2$ for all psychiatric disorders relative to their original estimates. The decrease in $h^2$ ranged from $2.44 \pm 0.187\%$ for bipolar disorder (original $P = 3.55 \times 10^{-31}$, $h^2 = 4.39\%$, s.e. = 0.360; highest conditioned $P = 5.67 \times 10^{-40}$, $h^2 = 2.22\%$, s.e. = 0.460; lowest conditioned $P = 4.05 \times 10^{-88}$, $h^2 = 1.70\%$, s.e. = 0.440) to 29.0 ± 0.105\% for Tourette syndrome (original $P = 6.56 \times 10^{-88}$, $h^2 = 21.0\%$, s.e. = 2.52; highest conditioned $P = 2.61 \times 10^{-44}$, $h^2 = 6.72\%$, s.e. = 0.770; lowest conditioned $P = 1.27 \times 10^{-18}$, $h^2 = 6.43\%$, s.e. = 0.730; Fig. 2, Supplementary Fig. 3 and Supplementary Table 5). Tourette syndrome exhibited the largest decrease in $h^2$ after conditioning with the effects of EDU/SES phenotypes ($1.78 \times 10^{-11} < P_{\text{diff}} < 3.02 \times 10^{-41}$, mean $h^2$ decrease compared with original Tourette syndrome of 0.144, s.e. = 0.001). Conversely, two phenotypes exhibited significant increases in $h^2$ after conditioning with EDU/SES phenotypes: neuroticism ($P = 3.08 \times 10^{-29}$, highest conditioned $h^2 = 20.2\%$, s.e. = 0.630; $P = 2.35 \times 10^{-201}$, lowest conditioned $h^2 = 18.1\%$, s.e. = 0.590) and subjective well-being ($P = 8.11 \times 10^{-82}$, highest conditioned $h^2 = 3.65\%$, s.e. = 0.220; $P = 4.67 \times 10^{-52}$, lowest conditioned $h^2 = 3.34\%$, s.e. = 0.220).

Conditioning the neuroticism GWAS ($P = 3.88 \times 10^{-46}$, original $h^2 = 9.41\%$, s.e. = 0.65) with EDU/SES phenotypes revealed several novel risk loci (ranging from 59 loci (neuroticism conditioned with income) to 100 loci (neuroticism conditioned with deprivation index); Supplementary Fig. 4), increased $h^2$ ($1.94 \times 10^{-2} < P < 5.27 \times 10^{-28}$, mean $h^2$ increase 10.1\%, s.e. = 0.747; Supplementary Table 5) and confirmed known linkage disequilibrium (LD)-independent risk loci. We observed an increase in the association signal in the neuroticism GWAS, with the strongest effects observed after conditioning with SES phenotypes income (lambda genomic control (GC) 1.36; intercept 0.971, s.e. = 0.009) and deprivation index (lambda GC 1.75; intercept 0.967, s.e. = 0.009; Supplementary Fig. 5 and Supplementary Table 6). This increase was not related to an increase in the potential bias of population stratification (that is, there was no significant change in the LDSC intercept, 0.884 < $P_{\text{diff}} < 0.994$, supporting that the observation was attributable to the increased detection of valid neuroticism polygenic signals. Using a physical proximity single-SNP/single-gene based annotation of conditioned neuroticism genomic risk loci (Supplementary Tables 7–9), the top gene sets included Gene Ontology (GO) cellular component synapse ($7.51 \times 10^{-4} < P < 9.58 \times 10^{-4}$, mean enrichment 0.138, s.e. = 0.019), GO biological process long-term synaptic potentiation ($2.95 \times 10^{-3} < P < 7.43 \times 10^{-3}$, mean enrichment 0.650, s.e. = 0.042) and GO cellular component synapse part (enrichment $2.46 \times 10^{-1} < P < 0.001$, mean enrichment 0.144, s.e. = 0.017).

The significant increase in $h^2$ for GWAS of subjective well-being (original $P = 7.47 \times 10^{-36}$, $h^2 = 2.50\%$, s.e. = 0.20) uncovered a 5.7 kb genomic risk loci on chromosome 7 (minimum genome wide significant $P = 1.45 \times 10^{-4}$) which maps to the q,δ, subunit of calcium voltage-gated channel (CACNA2D1, Supplementary Tables 10–12). The protein encoded by CACNA2D1 has been implicated in familial epilepsy and intellectual disability pedigrees but to our knowledge has not been implicated in genome-wide studies of these phenotypes\textsuperscript{23,24}. The shared biology between neuroticism and subjective well-being and all other psychopathologies and psychosocial factors revealed similar results as those with EDU/SES phenotypes and are described in Supplementary Results (Supplementary Fig. 3 and Supplementary Tables 13–17).

We next considered $h^2$ estimates for psychopathology and psychosocial factors after conditioning in two additional experiments: (1) with latent factors representing EDU/SES phenotypes (excluding income, which failed to load on a latent factor; see Correlative, latent and causal relationships between psychopathology and psychosocial factors) and (2) with all EDU/SES phenotypes simultaneously. All traits exhibited a reduction in SNP $h^2$ except for extraversion ($P = 3.47 \times 10^{-41}$, $h^2 = 0.137$, s.e. = 0.010 when conditioned with all EDU/SES phenotypes and $P = 6.32 \times 10^{-4}$, $h^2 = 0.035$, s.e. = 0.010 when conditioned with latent factors; Supplementary Table 18). Though extraversion $h^2$ increased, the conditioned GWAS resulted in no genome-wide significant findings.
**Tissue-type transcriptomic profile enrichment differences.** After conditioning with GWAS of EDU/SES phenotypes (Supplementary Table 19), schizophrenia was the only trait demonstrating significant changes in tissue-specific transcriptomic profile enrichment. Compared to the unconditioned schizophrenia GWAS, all conditioned schizophrenia brain tissue Genotype-Tissue Annotation (GTEx)25 annotations, with the exception of c1 cervical spinal cord, had significantly decreased enrichments (Supplementary Fig. 6 and Supplementary Table 19). The maximum decrease was observed after conditioning schizophrenia with educational attainment (average beta decrease for all brain tissue annotations 0.038, s.e. = 0.004).

After conditioning with EDU and SES phenotypes, the cerebellum and cerebellar hemisphere GTEx annotations remained the most enriched in the schizophrenia GWAS (original cerebellum P = 1.76×10^-22, enrichment 0.080, s.e. = 0.008; original cerebellar hemisphere $P = 1.28 \times 10^{-22}$, enrichment 0.077, s.e. = 0.008; conditioned cerebellum $6.83 \times 10^{-22} < P < 5.82 \times 10^{-19}$, mean enrichment 0.047, s.e. = 0.001; conditioned cerebellar hemisphere $7.43 \times 10^{-23} < P < 6.65 \times 10^{-20}$, mean enrichment 0.047, s.e. = 0.001). After adjusting for the effects of cognitive performance and educational attainment, we uncovered enrichment of skeletal muscle tissue transcriptomic profiles in the schizophrenia GWAS (original skeletal muscle $P = 0.135$, enrichment 0.010, s.e. = 0.009; skeletal muscle conditioned with educational attainment $P = 0.032$, enrichment 0.010, s.e. = 0.006; skeletal muscle conditioned with cognitive performance $P = 0.024$, enrichment 0.011, s.e. = 0.006). Though demonstrated in early studies of schizophrenia patients26, contemporary studies are required to validate this enrichment.

**Cell-type transcriptomic profile discoveries.** Cell-type transcriptomic profile enrichments were evaluated in two ways, by: (1) assessing differences in within-dataset cell-type enrichments before and after conditioning with EDU/SES (on the basis of Multi-marker Analysis of GenoMic Annotation (MAGMA) cell-type enrichment step 1 (ref. 23)) and (2) assessing the effects of conditioning on the detection of conditionally independent proportional significance (PS) of the cell type enrichments (on the basis of MAGMA cell-type enrichment step 3). PS cell types are those whose genetic signals could be differentiated from one another. PS values ≥ 0.80 indicate independent genetic signals relative to a second cell type. We then used genes whose expression profiles define the excitatory (Ex) and inhibitory (In) cell types of PsychENCODE29 to perform gene set enrichment analyses of GO and Kyoto Encyclopedia of Genes and Genomes gene sets.

There were no statistically significant differences in cell-type transcriptomic profile enrichments for psychopathology and psychosocial factors (MAGMA cell-type step 1) after conditioning with EDU/SES; however, we discovered several PS cell-type pairs not detected in the unconditioned GWAS for (1) risky behaviour (Fig. 3 and Supplementary Table 20), (2) MDD (Fig. 3 and Supplementary Table 21) and (3) schizophrenia (Supplementary Fig. 7 and Supplementary Table 22; MAGMA cell-type step 3).

In unconditioned GWAS of risky behaviour, there were no PS cell-type enrichments. After conditioning with EDU phenotypes, human cortex foetal quiescent and Ex2 were conditionally independent from one another (risky behaviour conditioned with cognitive performance Ex2: $P = 7.48 \times 10^{-4}$, $\beta = 0.035$, s.e. = 0.011, PS 1.37; foetal quiescent: $P = 0.032$, $\beta = 0.023$, s.e. = 0.012, PS 1.82; risky behaviour conditioned with educational attainment Ex2: $P = 0.001$, $\beta = 0.034$, s.e. = 0.011, PS 1.38; foetal quiescent: $P = 0.030$, $\beta = 0.024$, s.e. = 0.012, PS 1.77 (Fig. 3 and Supplementary Table 20)). The genes that define the Ex2-neuron cell type were enriched in nervous system development (GO:0007399; enrichment FDR $3.70 \times 10^{-4}$) and eye development (GO:001654; enrichment FDR $6.30 \times 10^{-4}$) gene sets.

The unconditioned MDD GWAS exhibited cell-type transcriptomic profile enrichments between adult γ-aminobutyric acid (GABA)-ergic neurons, In6b neurons and gestational week 10 (GW10) stem cells. After conditioning with self-rated math ability, the genetic signal from human midbrain neurons was conditionally independent from lateral geniculate nucleus (LGN) GABAergic neurons ($P = 0.002$, $\beta$ relative to midbrain neurons 0.041, s.e. = 0.014, PS 0.822; Fig. 3), In6b neurons ($P = 5.69 \times 10^{-4}$,
$\beta$ relative to midbrain neurons 0.517, s.e. = 0.11, PS 0.969) and In5 neurons ($P = 5.26 \times 10^{-5}$, $\beta$ relative to midbrain neurons 0.039, s.e. = 0.01, PS 0.813; Fig. 3 and Supplementary Table 21). The gene expression profiles of these cell types implicate the neurotransmitter transport (GO:0007269; enrichment FDR 0.003) and locomotory behaviour (GO:0007626; enrichment FDR 0.015) gene sets in MDD psychopathology.

**Correlative, latent and causal relationships between psychopathology and psychosocial factors.** We next evaluated relationships between phenotypes using three methods: genetic correlation, genomic SEM and latent causal variable (LCV) analysis.

Genetic correlations were assessed between all psychopathology and psychosocial factors after conditioning with EDU and SES phenotypes. Though small changes in genetic correlation magnitude were observed, psychopathology and psychosocial factor genetic correlations largely persisted (Supplementary Fig. 8 and Supplementary Table 24).

Genomic SEM was used to identify how unconditioned and conditioned psychopathology and psychosocial factors relate to

---

**Fig. 2 | Heritability ($h^2$) changes.** Each data point indicates an observed-scale $h^2$ estimate and standard error of the original (OR, grey) indicated phenotype or that same phenotype after conditioning with EDU and SES phenotypes (CP, cognitive performance; DI, deprivation index; EA, educational attainment; HM, highest math class; INC, income; MA, self-rated math ability). Two $P$-values are shown: (1) the $h^2$ estimates of each conditioned phenotype was at least nominally significant ($P < 0.05$) as demonstrated by the size of each data point; (2) solid data points indicate that the $h^2$ estimate of a conditioned phenotype was significantly different from the original $h^2$ estimate in grey. Phenotypes showing significant $h^2$ changes are shown here, while all phenotypes are presented in Supplementary Fig. 3.
a latent unobserved genetic factor connecting them (Fig. 4). In unconditioned models, exploratory factor analysis (EFA) identified a two-factor model as best suited to explain the relationships among psychopathology and psychosocial factors. In confirmatory factor analysis (CFA), these two latent factors generally highlight relationships between all psychiatric disorders and brain imaging.
Fig. 4 | Trait loading onto latent factors. Genomic SEM of psychopathology and psychosocial factors before and after conditioning with EDU and SES phenotypes. Each column shows the CFA loading value (blue shading indicates that a trait is a major contributor to the latent factor, and blue tainting indicates that a trait is a minor independent contributor to the latent factor) for each psychopathology and psychosocial factor (in the x direction) into one of two factors (F1 and F2) from EFA. Grey boxes indicate that a given trait was not predicted to load onto a given factor column. Red boxes indicate that the trait was predicted by EFA to load onto a factor but did not independently load during CFA.

phenotypes (F1) and anxiety, MDD, depressive symptoms and neuroticism (F2). The correlation between unconditioned F1 and F2 was 0.14. After conditioning with highest math class, self-rated math ability and deprivation index, the GWAS of neuroticism and MDD were no longer major contributors to the same factor. Conditioned F1 had major contributions from MDD (mean loading 0.611, s.e. = 0.005) and depressive symptoms (loading 0.538, s.e. = 0.098), while conditioned F2 had major contributions from neuroticism (loading 0.877, s.e. = 0.080) and anxiety (loading 0.658, s.e. = 0.009). Interestingly, after conditioning with the SES phenotype income, the SEM best fit converged on a single common factor between all psychopathology and psychosocial factors with major contributions from MDD (loading 0.808, s.e. = 0.068) and depressive symptoms (loading 0.831, s.e. = 0.022).

LCV analyses were used to detect causal relationships between trait pairs that are independent of the genetic correlations between them. Considering only the unconditioned psychopathology and psychosocial factors, one trait pair exhibited significant genetic causality proportion (gcp): left subcallosal cortex → obsessive compulsive disorder $\text{P} = 4.54 \times 10^{-6}$, gcp 0.167, s.e. = 0.047 (Table 1 and Supplementary Table 25 and Figs. 5 and 6). This partial causal relationship did not survive conditioning; however, 13 unique trait pairs demonstrated significant gcp after conditioning both traits with an EDU or SES phenotype (Table 1). Most notable were those causal relationships involving brain imaging phenotypes which became significant after conditioning with EDU phenotypes: (1) extraversion→left subcallosal cortex ($1.23 \times 10^{-13} < P < 1.83 \times 10^{-4}$, mean gcp 0.188, s.e. = 0.107) after conditioning with educational attainment, highest math class and self-rated math ability, (2) left subcallosal cortex→subjective well-being ($1.45 \times 10^{-3} < P < 1.16 \times 10^{-8}$, mean gcp 0.745, s.e. = 0.009) after conditioning with cognitive performance, educational attainment and highest math class and (3) left insular cortex→openness ($2.54 \times 10^{-23} < P < 3.63 \times 10^{-4}$, mean gcp 0.296, s.e. = 0.050) after conditioning with cognitive performance, highest math class and self-rated math ability. These average gcp estimates represent only Bonferroni significant relationships; however, each trait pair listed was nominally significant after conditioning with all other
EDU and SES phenotypes but not significant in the unconditioned experiment (Table 1).

The EDU/SES phenotype that revealed the most latent causal relationships between psychopathology and psychosocial factors was Townsend deprivation index. Conditioning with this phenotype revealed 7 of 13 causal relationships, most of which involved bipolar disorder or the volume of the right-ventral diencephalon (Fig. 6 and Supplementary Table 26).

**Discussion**

EDU and SES are important influences on numerous psychopathology, psychosocial and mental disorders, but it has been difficult to determine the extent to which this is so, and the biological nature of the relationship. How much of the genetic risk for schizophrenia, for example, is caused by reduced educational attainment? Or how much of the risk for schizophrenia reflects a shared biology with the predisposition to educational attainment? These are important questions to answer if we are to understand the biology of both kinds of traits. To address this question, we conditioned one on the other, and thereby statistically removed its effects, and then asked the question: How much of the heritable risk for that trait remains? In most cases, EDU/SES accounted for some of the genetic variance in the psychopathology and psychosocial phenotype and adjusting for EDU/SES reduced the strength of the association with the heritable risk for that disorder. However, in a few cases (depression, anxiety, neuroticism, PTSD and subjective well-being), adjusting for EDU/SES either increased or did not change these associations. Below, we present a framework for interpreting the complexity of these findings.

The biology underlying psychiatric disorders was most affected by shared genetic aetiology with EDU/SES proxies, as evidenced by significant decreases in $h^2$ for all psychiatric disorder except MDD.
anxiety and PTSD when conditioned with EDU/SES. Conversely, conditioning the neuroticism and subjective well-being GWAS revealed additional risk loci that were not detected in their unconditioned GWAS. Using an independent method, Turley et al. observed similar information gain; however, we demonstrated that this information gain is due to polygenicity rather than population substructure as evidenced by stable intercept estimates no different from the unconditioned neuroticism GWAS but increasing lambda GC. That is, we believe this demonstrated biological underpinnings, as opposed to the underlying population genetics phenomena. Unlike Turley et al., we do not report comparable risk locus gain with subjective well-being.

In structural equation models of psychopathology and psychosocial factors, neuroticism and MDD originally loaded onto the same common factor. After conditioning with highest math class, self-rated math ability and deprivation index, the loadings of neuroticism and MDD separate, suggesting that their unique biology may be distinguishable. This is consistent with our observation of ameliorated genetic correlation between these two phenotypes due to conditioning. Hill et al. described two factors of neuroticism, one of which aligns more closely with anxiety and tension phenotypes and the second of which aligns more closely with worry and vulnerability phenotypes. With genomic SEM, we support these claims: neuroticism loads onto the common factor as anxiety, while neuroticism appears more similar to the Hill et al. anxiety/tension substructure here that neuroticism and MDD are highly positively genetically correlated in their unconditioned versions. On the basis of the present results, we hypothesize that conditioning these phenotypes with EDU and SES reveals their unique genetic architectures. We demonstrate that, after conditioning with EDU/SES, general neuroticism appears more similar to the Hill et al. anxiety/tension phenotype. Lastly, our genomic SEM data mirror those genetic correlation results from Hill et al., adding weight to our observed two-factor model.

Cell-type transcriptomic profile enrichments underlying the GWAS of psychopathology and psychosocial factors were robust to the effects of EDU and SES phenotypes, but we uncovered cell-type information for risky behaviour, MDD, schizophrenia and bipolar disorder, which highlight cell-specific processes. The cell types discovered in the conditioned schizophrenia GWAS overlap with those in the conditioned bipolar disorder GWAS. These findings recapitulate common therapeutic targets for these disorders. For example, inhibitory and GABAergic neuron transcriptomic profile enrichments were detected in the conditioned MDD GWAS, and these are common targets of emerging therapeutic options (for example, scopolamine, an antidepressant which blocks the M1 receptor of GABAergic interneurons in the medial prefrontal cortex; ketamine blocking the activation of somatostatin interneurons in prefrontal cortex) for MDD and depressive symptoms. Detection of overlapping cell-type information may support drug repurposing efforts in psychiatric disorders and related mental health conditions, though the effect of these detected cell types as therapeutic targets requires experimental validation.

Using genome-wide information, we uncovered putatively causal relationships between many psychopathology and psychosocial factors and brain measurements. These analyses detected well-known relationships between traits (for example, bipolar disorder, schizophrenia and MDD) but also elucidated several relationships involving brain imaging phenotypes. In particular, we identified the volume of the left subcallosal cortex as a possible mediator of the relationships between several psychopathology and psychosocial factors (for example, extraversion, subjective well-being and alcohol dependence), which in turn demonstrate potential causal relationships with mood disorders, which are commonly comorbid with alcohol dependence. The brain structural convergence detected here may indicate common disease mechanisms; however, these commonalities may be confounded by fine-grained nuances of the relationship between brain microstructure and mental health and disease. The LCV method used to identify these causal relationships does not support multivariable analyses, nor does it employ sensitivity tests to detect horizontal pleiotropy (that is, where a SNP is associated with both phenotypes through separate mechanisms) or effect-size outlier SNPs. These observations likely confound our causal inferences and require more robust testing with traditional...
| Trait 1               | Trait 2                  | Conditioning | zscore | gcp.p | gcp.pm | gcp.pse | rho.est | rho.err | h2.zscore.1 | h2.zscore.2 | Notes                   |
|----------------------|--------------------------|--------------|--------|-------|--------|---------|---------|---------|--------------|--------------|-------------------------|
| Extraversion         | Left subcallosal cortex | Original     | 0.124  | 0.901 | 0.086  | 0.522   | −4.51×10⁻⁵ | 0.038   | 12.423        | 7.772        | h2.zscore.1 too low     |
|                      |                          | Cognitive performance | 3.626  | 4.57×10⁻⁴ | 0.252 | 0.393 | 0.208 | 0.141 | 13.119       | 7.979       |                          |
|                      |                          | Educational attainment | 5.074  | 1.83×10⁻⁶ | 0.065 | 0.378 | 0.201 | 0.141 | 13.207       | 7.742       |                          |
|                      |                          | Highest math class | 8.599  | 1.23×10⁻¹³ | 0.094 | 0.386 | 0.209 | 0.141 | 12.881       | 7.706       |                          |
|                      |                          | Self-rated math ability | 6.086  | 2.20×10⁻⁸ | 0.304 | 0.326 | 0.190 | 0.130 | 11.845        | 6.692       |                          |
|                      |                          | Subjective well-being | 0.336  | 0.737 | 0.004 | 0.434   | 0.036 | 0.029 | 6.894        | 20.441      | h2.zscore.1 too low     |
|                      |                          | Cognitive performance | 6.525  | 2.92×10⁻⁹ | 0.746 | 0.168 | 0.280 | 0.091 | 7.365        | 25.776      |                          |
|                      |                          | Educational attainment | 6.227  | 1.16×10⁻⁸ | 0.736 | 0.173 | 0.269 | 0.093 | 7.129        | 25.499      |                          |
|                      |                          | Highest math class | 6.675  | 1.45×10⁻¹⁹ | 0.753 | 0.164 | 0.265 | 0.090 | 7.040        | 25.116      |                          |
|                      |                          | Self-rated math ability | 7.699  | 1.05×10⁻⁻⁹ | 0.780 | 0.148 | 0.259 | 0.100 | 6.650        | 26.939      |                          |
|                      |                          | Alcohol dependence | 0.336  | 0.737 | 0.005 | 0.434 | 0.036 | 0.029 | 6.894        | 20.441      | h2.zscore.1 too low     |
|                      |                          | Cognitive performance | 9.901  | 1.80×10⁻¹⁶ | 0.293 | 0.254 | −0.146 | 0.202 | 8.203        | 7.244       |                          |
|                      |                          | Educational attainment | 1.256  | 0.212 | 0.042 | 0.164 | −0.099 | 0.216 | 8.001        | 6.467       |                          |
|                      |                          | Highest math class | 1.008  | 0.316 | 0.005 | 0.210 | −0.087 | 0.214 | 8.148        | 6.484       | h2.zscore.1 too low     |
|                      |                          | Self-rated math ability | 3.422  | 9.05×10⁻⁴⁵ | 0.322 | 0.218 | −0.106 | 0.196 | 7.500        | 7.920       |                          |
|                      |                          | Bipolar disorder   | 0.093  | 0.926 | 0.023 | 0.560 | −0.001 | 0.022 | 7.986        | 35.945      |                          |
|                      |                          | Cognitive performance | 6.884  | 5.37×10⁻¹⁵ | 0.607 | 0.084 | 0.191 | 0.077 | 7.059        | 30.127      |                          |
|                      |                          | Educational attainment | 5.134  | 1.42×10⁻⁶ | 0.843 | 0.252 | 0.282 | 0.084 | 6.346        | 30.211      | h2.zscore.1 too low     |
|                      |                          | Highest math class | 6.201  | 1.30×10⁻⁸ | 0.834 | 0.105 | 0.227 | 0.081 | 6.412        | 29.840      | h2.zscore.1 too low     |
|                      |                          | Self-rated math ability | 5.52   | 2.72×10⁻⁷ | 0.017 | 0.051 | 0.187 | 0.073 | 8.028        | 27.169      |                          |
|                      |                          | Income              | 4.149  | 7.07×10⁻⁶ | 0.563 | 0.168 | 0.262 | 0.080 | 6.556        | 31.500      | h2.zscore.1 too low     |
|                      |                          | Deprivation index  | 0.789  | 0.431 | 0.433 | 0.333 | 0.207 | 0.078 | 5.750        | 32.739      | h2.zscore.1 too low     |
|                      |                          | Major depression    | 0.037  | 0.97  | 0.001 | 0.071 | 0.338 | 0.039 | 34.380        | 36.466      |                          |
|                      |                          | Cognitive performance | 0.999  | 0.319 | 0.086 | 0.081 | 0.329 | 0.038 | 30.433        | 34.262      |                          |
|                      |                          | Educational attainment | 1.98   | 0.233 | 0.105 | 0.085 | 0.363 | 0.038 | 30.529        | 33.800      |                          |
|                      |                          | Highest math class | 1.437  | 0.153 | 0.127 | 0.086 | 0.358 | 0.036 | 30.500        | 33.703      |                          |
|                      |                          | Self-rated math ability | 0.996  | 0.321 | 0.095 | 0.090 | 0.346 | 0.036 | 28.867        | 36.063      |                          |
|                      |                          | Income              | 1.777  | 0.078 | 0.139 | 0.078 | 0.362 | 0.036 | 31.788        | 34.070      |                          |
|                      |                          | Deprivation index  | 4.844  | 4.71×10⁻⁶ | 0.268 | 0.063 | 0.364 | 0.035 | 33.639        | 34.238      |                          |

Continued.
**Table 1 | Causal inferences detected after multiple testing correction**

| Trait 1            | Trait 2                        | Conditioning | zscore | gcp.p  | gcp.pm | gcp.pse | rho.est | rho.err | h2.zscore.1 | h2.zscore.2 | Notes |
|--------------------|--------------------------------|--------------|--------|--------|--------|---------|---------|---------|-------------|-------------|-------|
| Bipolar disorder   | Schizophrenia                  | Original     | 0.182  | 0.856  | 0.024  | 0.133   | 0.711   | 0.026   | 32.985      | 34.422      |       |
|                    |                                | Cognitive performance | 0.649  | 0.518  | 0.121  | 0.173   | 0.701   | 0.023   | 30.263      | 40.071      |       |
|                    |                                | Educational attainment | 0.619  | 0.537  | 0.107  | 0.164   | 0.697   | 0.023   | 30.312      | 39.843      |       |
|                    |                                | Highest math class | 0.755  | 0.452  | 0.138  | 0.175   | 0.707   | 0.023   | 29.911      | 40.022      |       |
|                    |                                | Self-rated math ability | 0.757  | 0.451  | 0.133  | 0.170   | 0.704   | 0.024   | 28.396      | 36.960      |       |
|                    |                                | Income        | 1.993  | 0.048  | 0.246  | 0.246   | 0.695   | 0.020   | 31.851      | 38.906      |       |
|                    |                                | Deprivation index | 6.865  | 5.88 x 10^-10 | 0.680  | 0.110  | 0.674   | 0.021   | 33.358      | 40.103      |       |
| Bipolar disorder   | Volume of right-ventral diencephalon | Original | 0.710  | 0.479  | 0.172  | 0.518   | -0.019  | 0.021   | 35.797      | 10.058      |       |
|                    |                                | Cognitive performance | 0.117  | 0.907  | 0.090  | 0.525   | -0.002  | 0.030   | 8.076       | 22.529      |       |
|                    |                                | Educational attainment | 3.828  | 2.27 x 10^-4 | 0.683  | 0.204  | -0.283  | 0.093   | 7.767       | 23.319      |       |
|                    |                                | Highest math class | 2.828  | 0.005  | 0.612  | 0.241   | -0.264  | 0.095   | 7.547       | 22.230      |       |
|                    |                                | Self-rated math ability | 4.040  | 1.06 x 10^-4 | 0.681  | 0.204  | -0.258  | 0.098   | 7.719       | 23.633      |       |
|                    |                                | Deprivation index | 9.236  | 5.07 x 10^-15 | 0.673  | 0.171  | 0.184   | 0.040   | 33.264      | 28.882      |       |
| Left insular cortex | Depressive symptoms | Original | 0.243  | 0.808  | 0.117  | 0.519   | -0.009  | 0.035   | 11.504      | 11.558      |       |
| Obsessive compulsive disorder | Anorexia nervosa | Cognitive performance | 2.957  | 0.003  | 0.064  | 0.214   | 0.514   | 0.126   | 13.034      | 11.861      |       |
|                    |                                | Educational attainment | 11.041 | 5.94 x 10^-38 | 0.805  | 0.375  | 0.479   | 0.135   | 12.701      | 10.921      |       |
|                    |                                | Highest math class | 4.252  | 4.81 x 10^-5 | 0.184  | 0.243  | 0.494   | 0.129   | 12.973      | 11.430      |       |
|                    |                                | Self-rated math ability | 2.282  | 0.024  | 0.335  | 0.304   | 0.488   | 0.122   | 14.596      | 13.977      |       |
|                    |                                | Income        | 6.585  | 2.20 x 10^-4 | 0.371  | 0.692  | 0.501   | 0.130   | 12.723      | 11.328      |       |
|                    |                                | Deprivation index | 4.691  | 8.71 x 10^-6 | 0.144  | 0.460  | 0.495   | 0.129   | 12.780      | 11.647      |       |
| Openness | Autism spectrum disorder | Cognitive performance | 0.274  | 0.784  | 0.097  | 0.544   | 0.315   | 0.111   | 14.625      | 8.038       | h2.zscore.2 too low |
|                    |                                | Educational attainment | 6.147  | 1.67 x 10^-8 | 0.761  | 0.164  | 0.286   | 0.127   | 16.717      | 6.089       | h2.zscore.2 too low |
|                    |                                | Highest math class | 5.452  | 3.66 x 10^-7 | 0.741  | 0.174  | 0.260   | 0.138   | 16.176      | 4.898       | h2.zscore.2 too low |
|                    |                                | Self-rated math ability | 5.865  | 5.96 x 10^-3 | 0.762  | 0.163  | 0.332   | 0.122   | 16.295      | 6.680       | h2.zscore.2 too low |
|                    |                                | Income        | 5.914  | 4.77 x 10^-8 | 0.777  | 0.157  | 0.351   | 0.116   | 15.955      | 7.782       | h2.zscore.2 too low |
|                    |                                | Deprivation index | 5.212  | 1.02 x 10^-6 | 0.742  | 0.174  | 0.343   | 0.122   | 16.291      | 6.466       | h2.zscore.2 too low |

Continued
### Table 1 | Causal inferences detected after multiple testing correction

| Trait 1                  | Trait 2                  | Conditioning | zscore | gcp.p  | gcp.pm | gcp.pse | rho.est | rho.err | h2.zscore.1 | h2.zscore.2 | Notes       |
|--------------------------|--------------------------|--------------|--------|--------|--------|---------|---------|---------|-------------|-------------|-------------|
| **Left insular cortex**  | Openness                 | Original     | 0.498  | 0.619  | 0.225  | 0.314   | 0.041   | 0.057   | 7.151       | 8.094       |             |
|                          | Cognitive performance    | 5.967        | 3.77×10⁻⁸ | 0.350  | 0.379   | 0.201   | 0.200   | 7.458       | 6.903       | h2.zscore.2 too low |
|                          | Educational attainment   | 9.861        | 2.20×10⁻⁶ | 0.092  | 0.381   | 0.187   | 0.221   | 7.311       | 5.574       | h2.zscore.2 too low |
|                          | Highest math class       | 8.221        | 8.04×10⁻¹³ | 0.284  | 0.456   | 0.239   | 0.184   | 7.487       | 7.298       |             |
|                          | Self-rated math ability  | 13.09        | 2.54×10⁻² | 0.252  | 0.384   | 0.198   | 0.170   | 7.583       | 7.598       |             |
| **Right insular cortex** | Volume of right-ventral diencephalon | Original     | 1.322  | 0.189  | 0.513  | 0.292   | −0.157  | 0.162   | 8.925       | 10.614      |             |
|                          | Income                   | 1.546        | 0.125  | 0.533  | 0.283   | −0.148  | 0.166   | 8.870       | 10.354      |             |
|                          | Deprivation index        | 5.975        | 3.63×10⁻⁸ | 0.478  | 0.178   | −0.131  | 0.096   | 8.735       | 30.666      |             |
| **Volume of right-ventral diencephalon** | Openness | Original     | 0.577  | 0.565  | 0.328  | 0.404   | −0.066  | 0.049   | 9.177       | 8.094       |             |
|                          | Income                   | 8.800        | 4.50×10⁻¹⁴ | 0.203  | 0.442   | 0.27    | 0.096   | 28.449      | 8.041       |             |
|                          | Deprivation index        | 0.042        | 0.966  | 0.061  | 0.302   | 0.153   | 0.17    | 9.327       | 7.409       |             |

**Column descriptions**
- **zscore**: Z score for partial genetic causality
- **gcp.p**: P; significantly positive zscore implies trait 1 partially genetically causal for trait 2
- **gcp.pm**: posterior genetic causality proportion (positive if trait 1 > trait 2)
- **gcp.pse**: posterior standard error for genetic causality proportion
- **rho.est**: estimated genetic correlation
- **rho.err**: standard error for genetic correlation estimate
- **h2.zscore.1**: Z score for trait 1 being significantly heritable
- **h2.zscore.2**: Z score for trait 2 being significantly heritable

Significant causal relationships detected between psychopathology and psychosocial factors using LCV analyses. For each significant causal estimate, all conditioned causal estimates between that phenotype pair are provided, highlighting at least nominally significant causal inferences after conditioning with EDU and SES phenotypes that could not be detected in the original unconditioned trait pair.
Mendelian randomization methods suited to accommodate weak genetic instruments (that is, those SNPs not strongly associated with either phenotype of interest, typically with association \( P \) values greater than \( 5 \times 10^{-8} \))35–37. It should be recognized that LCV and other causal inference methods indeed have reduced power with respect to highly polygenic traits, such as those studied here, leading towards downwardly biased genetic causality proportion estimates. It is therefore unlikely that estimates with high genetic causality proportions are false positives36.

Certain relationships regarding psychopathology and psychosocial factor conditioning that might have been expected were not observed in our study. Intellectual abilities are genetically correlated with autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD), and disabilities therein often co-occur with ASD and/or ADHD diagnoses16,39. According to the Diagnostic and Statistical Manual of Mental Disorders (5th edition, DSM-5), diagnosis of intellectual disability or global developmental delay must be eliminated as possible explanations of ASD symptoms prior to making a formal ASD diagnosis. We had hypothesized that, after conditioning with the effects of EDU phenotypes, these psychiatric disorders might demonstrate notable changes in their genetically predicted underlying biology, but this was not the case. This lack may suggest that ADHD and ASD diagnosis criteria robustly capture elements unique to the disorders rather than those shared with EDU/SES phenotypes. In other words, ascertainment of cases at the extreme ends of spectrum disorders16,39 reliably capture trait-specific biology with minimal phenotype confounding from shared effects with EDU and SES.

Information derived from the GWAS of EDU and SES phenotypes was generally robust to conditioning with psychopathology and psychosocial factors. When conditioned with individual psychopathology and psychosocial factors, we detected relatively few changes to the genetically predicted biology of EDU/SES phenotypes. Only when EDU/SES phenotypes were conditioned with several psychopathology and psychosocial factors did we observe changes in \( h^2 \) and genetically predicted biology. When assessing the relationship between unconditioned EDU and SES phenotypes by genomic SEM, we revealed cognitive performance and highest math class as major contributors to the factors that link EDU and SES phenotypes. When conditioned, we uncovered an independent contribution of income to a common factor with educational attainment and self-rated math ability. On the basis of recent work of Morris et al. to uncover why EDU and SES phenotypes are related to one another, and the near-perfect loading of educational attainment on the common factor, these observations point to educational attainment as a major component of the genetic and phenotypic correlations between EDU and SES.

Tissue and cell-type transcriptomic profile analyses of EDU, SES, psychopathology and psychosocial phenotypes highlighted differences in cortical and cerebellar tissue enrichment. Though not significantly decreased in all phenotypes after conditioning, the bidirectional changes in cerebellar and cortical tissue enrichment (that is, EDU/SES conditioned with psychopathology and psychosocial factors and psychopathology and psychosocial factors conditioned with EDU/SES) highlight the importance of these brain regions and their shared transcriptional regulation in human mental health and disease42. Furthermore, this observation of cerebellar and cortical tissue changes supports the common psychopathology factor (a \( p \)-factor) studied extensively in recent mental health and disease research43.

Our study has three primary limitations. First, we did not select independent genetic correlates from the psychopathology and psychosocial factors tested with which to condition the EDU and SES phenotypes. Due to high genetic correlation between psychopathology and psychosocial factors, this approach may have introduced bias in our reporting of which EDU and SES phenotypes were robust to shared genetic aetiology with all psychopathology and psychosocial factors. This potential over-conditioning likely drove our results towards the null (for example, non-significant \( h^2 \)), and therefore we have not reported gene set, tissue transcriptomic profile enrichment, cell-type transcriptomic profile enrichment or genomic SEM loadings for EDU/SES traits where there might have been over-conditioning. For this reason, our results do not imply that, for example, educational attainment is a more powerful or specific EDU phenotype than cognitive ability. Second, it has recently been demonstrated that the origin of phenotypic and genetic correlations between EDU and SES phenotypes may be driven by dynastic effects and/or assortative mating acting independently or in concert44. Dynastic effects describe a condition where offspring inherit phenotype-associated genetic risks and phenotype-associated environmental risks. Assortative mating exists when mate pairs are non-randomly chosen on the basis of certain attributes. We hypothesize that the dynastic and assortative mating events described between EDU and SES phenotypes may also appear in phenotypic and genetically correlated EDU, SES, and psychopathology and psychosocial factor pairs. Future studies will be required to describe how these evolutionary and social pressures influence the relative and causal relationships uncovered here (for example, OCD—anorexia nervosa after conditioning with the effects of educational attainment, income and deprivation index). Third, the UK Biobank is considered a generally healthy cohort not enriched for any trait or disorder of interest. To our knowledge, the brain imaging GWAS (performed on a subset of UK Biobank participants) used here were not adjusted for variables related to blood pressure, height, weight, bone mineral composition, substance-related (recreational or prescribed) or psychiatric variables. The presence of these variables in sufficient quantities among those UKB brain imaging participants could alter brain volumes, affecting the results of the genetic analyses conducted.

By conditioning psychopathology and psychosocial factors for the shared genetic aetiology with EDU and SES phenotypes, this study elucidates biological underpinnings and causal relationships between phenotypes. These biological mechanisms, cell types, tissue types, and causal trait pairs could not have been detected without adjusting the effects of EDU and SES. This study highlights how the pervasive effects of EDU and SES may mask genetically determined underlying biology of psychopathology and psychosocial factors in support of multi-phenotype analyses of GWAS to enable trait-specific discoveries.

Methods

This study was conducted using genome-wide association statistics generated by previous studies. Owing to the use of previously collected, deidentified, aggregated data, this study did not require institutional review board approval. Ethical approval had been obtained in all the original studies (Supplementary Table 1). An overview of all materials, methods and key findings from this investigation of the genetic overlap between EDU, SES, and psychopathology and psychosocial factors is shown in a flow diagram in Supplementary Fig. 1.

Trait description and selection. Four EDU (educational attainment, highest math class, self-rated math ability and cognitive performance) and two SES phenotypes (average household income before tax and Townsend deprivation index) from the Social Science Genetic Association Consortium (SSGAC), UK Biobank (UKB) and 23andMe were used in this study to condition psychopathology and psychosocial factors. These unconditioned phenotypes were characterized on the level of heritability, tissue transcriptomic profile enrichment and cell-type transcriptomic profile enrichment (Supplementary Fig. 1). Briefly, the educational attainment phenotype describes the number of years of schooling completed per participant. Highest math class and self-rated math ability were derived from the 23andMe survey regarding participant mathematical background. For self-rated math ability, 23andMe participants were asked ‘How would you rate your mathematical ability?’, ranging from ‘very poor’ to ‘excellent’. For highest math class, 23andMe participants were asked to indicate the most advanced mathematics course they have successfully completed (excluding statistics courses) ranging from pre-algebra to coursework more advanced than vector calculus. Cognitive performance was evaluated as a standardized score of logic and reasoning questions completed.
SNPs underlying each phenotype of interest were mapped to genes within 10 kb physical proximity using FUMA. Mapped genes were assessed using the gene-set enrichment feature of FUMA, and gene ontology enrichment analysis with shinyGO.

Tissue transcriptomic profile enrichment was performed relative to the GTEx (ref. 7) v7 53 tissue types with the default 0 kb gene window.

Cell-type transcriptomic profile enrichments were performed using 11 human-specific transcriptomic profile datasets related to the brain: PsychENCODE_Devotional, PsychENCODE_Adult, Allen_Human_LGN_level1, Allen_Human_MTG_level2, DroNeC_Human_Hippocompus, GSE104276_Human_Prefrontal_cortex_all_ages, GSE104276_Human_prefrontal_cortex_per_ages, GSE67835_Human_Cortex, GSE67835_Human_Cortex_wooFetal, Linnarson_GSE101601_Human_Temporal_cortex and Linnarson_GSE63831_Human_Midbrain. Cell-type transcriptomic profiles were assessed in three ways as per Genomic SEM (appendix 1). If cell-type transcriptomic profiles within each dataset, (2) within-dataset conditionally independent cell-type transcriptomic profile enrichments and (3) across-dataset cell-type transcriptomic profile enrichments.

For analyses within datasets, step-wise conditional significance is evaluated for each cell type in a dataset against the P values for all other cell types in that same dataset. The output from these analyses identify cell types within a dataset whose transcriptomic profiles are enriched in a given GWAS independently of the signal from all other cell-type transcriptomic profiles in the same dataset.

Using within-dataset conditionally independent cell types identified above, cell-type transcriptomic profile enrichments were enriched in the GWAS of independent genetic signals for each cell type a given pair can be seen in detail online (https://fuma.ctglab.nl/tutorial#celltype) or in Watanabe et al.28.

LCVs, LCV analysis is a method for inferring genetic causal relationships between trait pairs on the basis of GWAS summary data using effect size estimates or Z scores. The LCV model assumptions are notably weaker than traditional MR assumptions. First, LCV assumes that the distribution of effect sizes for a given trait pair represents one distribution of effect sizes proportional in both traits of interest and a second distribution of SNPs that only affect the outcome trait. LCV therefore assumes that symmetry in shared genetic architectures between traits arises from the action of a latent genetic component rather than a non-genetic confounder commonly elucidated by MR. Second, the model assumes a single genetic LCV mediating trait relationships; however, in simulations of LCV power in the presence of more than one LCV, causal estimates were unlikely to be detected. There are no assumptions of parametric effect size distributions under the LCV model; however, LCV is indeed less well powered for highly polygenic traits.

LCV modelling was implemented in R using the 1000 Genomes Project phase 3 European reference panel. As recommended, GWAS summary data were filtered to include only SNPs with minor allele frequencies greater than 5% and the major allele frequency at least 0.05. As suggested by the mtCOJO analysis pipeline, this threshold is not expected to influence results from any other analysis performed herein.

Heritability and genetic correlation. The LDSC method is used for formatting GWAS summary data association data and estimating genetic correlations of a trait pair and genetic correlation between traits. LDSC assumes SNPs have not been pruned for linkage disequilibrium and that sample ascertainment of a phenotype was performed in a genetically homogeneous population. With respect to genetic correlation, phenotypes should be ascertainmented from cohorts representing similar ancestral backgrounds. Genetic correlation with LDSC is not sensitive to sample overlap.

Observed - scale h² was calculated for each original and conditioned GWAS using the LDSC method with 1000 Genomes Project European reference population.

Gene-set, tissue transcriptomic and cell-type transcriptomic profile enrichment. Original and conditioned GWAS were used as standard input for MAGMA v1.06 implemented in Func survival analysis and Annotation (FUMA) (v1.3.3c) with the following parameters: genome-wide significance P < 5 × 10⁻⁸, minor allele frequency > 0.01 and LD blocks defined at < 250 kb for LD r² > 0.6 with lead variant ⁴, ⁵.

LCVs, LCV analysis is an alternative approach for inferring genetic causal relationships between trait pairs on the basis of GWAS summary data using effect size estimates or Z scores. The LCV model assumptions are notably weaker than traditional MR assumptions. First, LCV assumes that the distribution of effect sizes for a given trait pair represents one distribution of effect sizes proportional in both traits of interest and a second distribution of SNPs that only affect the outcome trait. LCV therefore assumes that symmetry in shared genetic architectures between traits arises from the action of a latent genetic component rather than a non-genetic confounder.
and LCV estimates between conditioned and unconditioned GWAS. Note that, while much of the data generated for this study relied on one-sided tests (for example, LDSC tests of whether the $h^2$ for trait $X$ is greater than 0 and MAGMA tests of whether the transcriptomic profile of tissue $X$ is overrepresented relative to all other tissue types), $Z$ tests reported herein were used to compare the conditioned versus unconditioned versions of a trait GWAS. In other words, two sides were considered: for example, the unconditioned $h^2$ could be greater than or less than the conditioned $h^2$ for a trait. $Z$ scores for the difference between conditioned and unconditioned measurements were converted to $P$ values assuming a two-tailed distribution prior to multiple testing correction. $P$ values were corrected for multiple testing considering a false discovery rate of 5%.

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

**Data availability**

All GWAS association data and analysis materials used in this study are publicly available for download by qualified researchers. All data used to make conclusions discussed in this study are provided as Supplementary Material. Social Science Genetic Association Consortium: https://www.bsg.ac.uk/Psychiatric Genomics Consortium: https://www.med.yale.edu/cgsc/download-results.

UK Biobank: https://www.ukbiobank.ac.uk/register-apply/23andMe: https://research.23andme.com/research-innovation-collaborations/

Brain Imaging Genetics: http://big.stats.ox.ac.uk

**Code availability**

Previously developed pipelines were used to produce the results for this study. No custom code was developed.

**References**

1. Keyes, K. M., Platt, J., Kaufman, A. S. & McLaughlin, K. A. Association of fluid intelligence and psychiatric disorders in a population-representative sample of US adolescents. *JAMA Psychiatry* 74, 179–188 (2017).
2. McLaughlin, K. A., Costello, E. J., Leblanc, W., Sampson, N. A. & Kessler, R. C. Socioeconomic status and adolescent mental disorders. *Am. J. Public Health* 102, 1742–1750 (2012).  
3. d’Errico, A. et al. Socioeconomic indicators in epidemiologic research: a practical example from the LIFEPATH study. *PloS ONE* 12, e0178071 (2017).  
4. Baniuelo, A. et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012 (2019).  
5. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47, 1236–1241 (2015).  
6. Morris, T. T., Davies, N. M., Hemani, G. & Smith, G. D. Why are education, intelligence and health correlated? *Nat. Neurosci.* 21, 1236–1241 (2017).  
7. Morris, T. T., Davies, N. M., Hemani, G. & Smith, G. D. Why are education, socioeconomic position and intelligence genetically correlated? Preprint at bioRxiv https://doi.org/10.1101/630426 (2019).  
8. Trampush, J. W. et al. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol. Psychiatry* 22, 336–345 (2017).  
9. Marioni, R. E. et al. Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence* 44, 26–32 (2014).  
10. Hill, W. D. et al. Molecular genetic contributions to social deprivation and household income in UK biobank. *Carr. Biol.* 26, 3083–3089 (2016).  
11. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Science* 360, eaap7575 (2018).  
12. Zhang, Y., Qi, G., Park, J. H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* 50, 1318–1326 (2018).  
13. Choi, S. W. & O’Reilly, P. F. PRSice-2: polygenic risk score software for biobank-scale data. *Genicscience* https://doi.org/10.1093/genicscience/giu082 (2019).  
14. Choi, S. W., Mak, T. S. & O’Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* 15, 2759–2772 (2020).  
15. Amin, V. Educational attainment polygenic score is associated with depressive symptoms via socioeconomic status: a gene–environment–trait correlation. Preprint at bioRxiv https://doi.org/10.1101/727552 (2019).  
16. Krapohl, E. & Plomin, R. Genetic link between family socioeconomic status and children’s educational achievement estimated from genome-wide SNPs. *Mol. Psychiatry* 24, 437–443 (2016).  
17. Richards, A. L. et al. The relationship between polygenic risk scores and cognition in schizophrenia. *Schizophrenia. Bull.* 46, 336–344 (2019).  
18. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 50, 229–237 (2018).
Acknowledgements

The authors thank the research participants and employees of 23andMe Inc for making this work possible. This study was supported by the Simons Foundation Autism Research Initiative (SFARI Explorer Award 534858 (R.P.)), the American Foundation for Suicide Prevention (YIG-1-109-16 (R.P.)), the National Institutes of Health (R21 DC018098 (R.P.), R21 DA047527 (R.P.), F32 MH122058 (F.R.W) and R01 MH117646 (T.L.)) and the National Center for PTSD of the U.S. Department of Veterans Affairs. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

F.R.W. and R.P. conceived the study design; F.R.W. generated and analysed all data; F.R.W., G.A.P., T.L., J.H.K., J.G. and R.P. contributed to data interpretation; F.R.W., G.A.P. and R.P. contributed to data visualization and presentation; F.R.W. drafted the original manuscript; F.R.W., G.A.P., T.L., J.H.K., J.G. and R.P. critically evaluated and revised the manuscript.

Competing interests

J.H.K. reports compensation as the editor of Biological Psychiatry and also serves on the Scientific Advisory Boards for Bioptics Technologies, Inc., Biohaven Pharmaceuticals, BioXcel Therapeutics, Inc. (Clinical Advisory Board), Cadent Therapeutics (Clinical Advisory Board), PsychoGenics, Inc, Stanley Center for Psychiatric Research at the Broad Institute of MIT and Harvard, and the Loholca Research Corporation. He owns stock in ArRETT Neuroscience, Inc., Biohaven Pharmaceuticals, Sage Pharmaceuticals and Spring Care, Inc. and stock options in Biohaven Pharmaceuticals Medical Sciences, BlackThorn Therapeutics, Inc. and Storm Biosciences, Inc. He is co-inventor on multiple patents as listed below: (1) Seibyl J. P., Krystal J. H., Charney D. S. Dopamine and noradrenergic reuptake inhibitors in treatment of schizophrenia. US patent 5,447,948 (1995); (2) Vladimir, C., Krystal, J. H., Sanacora, G. Glutamate modulating agents in the treatment of mental disorders, US patent 8,778,979 (2014); (3) Charney D., Krystal J. H., Manji H., Matthew S., Zarate C. Intranasal administration of ketamine to treat depression US patent application 14/197,767 (2014); (4) Zarate, C., Charney, D. S., Manji, H. K., Mathew, S. J., Krystal, J. H., Department of Veterans Affairs. Methods for treating suicidal ideation, US patent application no. 14/197,767 (2014); (5) Arias A., Petrakis I., Krystal J. H. Composition and methods to treat addiction. US patent application no. 61/973/961 (2014); (6) Chekroud, A., Guergueiva, R., Krystal, J. H. Treatment selection for major depressive disorder (filing date 2016, USPTO docket number Y0087.70116US00). Provisional patent submission by Yale University: (7) Gihyun, Y., Petrakis I., Krystal, J. H. Compounds, compositions and methods for treating or preventing depression and other diseases. US provisional patent application no. 62/444,552 (filed 10 January 2017) by Yale University Office of Cooperative Research OCR 7088 US01: (8) Abdallah, C., Krystal, J. H., Durman, R., Sanacora, G. Combination therapy for treating or preventing depression or other mood diseases. US provisional patent application no. 047162-7177P1 (00754) filed on 20 August 2018 by Yale University Office of Cooperative Research OCR 7451 US01. J.G. is named as an inventor on PCT patent application 15/878,640 Genotype-guided dosing of opioid agonists, filed 24 January 2018. J.G. and R.P. are paid for their editorial work on the journal Complex Psychiatry. The other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00980-y.
Correspondence and requests for materials should be addressed to R.P.
Peer review information Primary handling editor: Charlotte Payne
Reprints and permissions information is available at www.nature.com/reprints.
Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020
Corresponding author(s): Renato Polimanti
Last updated by author(s): Sep 8, 2020

Reporting Summary

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

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a
- Confirmed

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
We did not collect data as part of this study but obtained data from the Social Science Genetic Association Consortium (SSGAC), UK Biobank (UKB), 23andMe, and Psychiatric Genomics Consortium.

Data analysis
Standard pipelines were used to produce the results for this study. No custom code was developed.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All GWAS association data and analysis materials used in this study are publicly available for download by qualified researchers. All data used to make conclusions discussed in this study are provided as Supplementary Material.
# Field-specific reporting

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

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

| Sample size | Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. |
| --- | --- |
| Data exclusions | Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established. |
| Replication | Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why. |
| Randomization | Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why. |
| Blinding | Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study. |

## Behavioural & social sciences study design

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

| Study description | We analyzed genome-wide association data for psychopathology and psychosocial factors after adjusting for effects of education and socioeconomic status. |
| --- | --- |
| Research sample | Data were accessed via the UK Biobank (UKB), Psychiatric Genomics Consortium (PGC), Social Science Genetic Association Consortium (SSGAC), and 23andMe. |
| Sampling strategy | We did not sample new subjects as part of this study. Data were made available for analysis of GWAS association data by either public data sharing (PGC, SSGAC, UKB) or approved access for qualified researchers (23andMe). |
| Data collection | We did not collect data as part of this study. ata were made available for analysis of GWAS association data by either public data sharing (PGC, SSGAC, UKB) or approved access for qualified researchers (23andMe). |
| Timing | We did not collect data as part of this study. ata were made available for analysis of GWAS association data by either public data sharing (PGC, SSGAC, UKB) or approved access for qualified researchers (23andMe). |
| Data exclusions | Exclusions were made to ensure genetically unrelated European-ancestry participants were included in each GWAS used here. |
| Non-participation | Exclusions were made to ensure genetically unrelated European-ancestry participants were included in each GWAS used here. |
| Randomization | Each GWAS used here independently covaried analyses of case-control and/or continuous phenotypes using at least age, sex, and ancestry information. No additional covariates were included in the analyses performed in this study which rely on GWAS association data. |

## Ecological, evolutionary & environmental sciences study design

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

| Study description | Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates. |
| --- | --- |
| Research sample | Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source. |
### Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

### Data collection

Describe the data collection procedure, including who recorded the data and how.

### Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken.

### Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

### Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

### Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

### Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

### Field work, collection and transport

**Field conditions**

Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

**Location**

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

**Access & import/export**

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

**Disturbance**

Describe any disturbance caused by the study and how it was minimized.

### Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

#### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ✗   | Antibodies            |
| ✓    | Eukaryotic cell lines |
| ✗   | Palaeontology and archaeology |
| ✗   | Animals and other organisms |
| ✗   | Human research participants |
| ✗   | Clinical data         |
| ✗   | Dual use research of concern |

#### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ✗   | ChIP-seq              |
| ✗   | Flow cytometry        |
| ✗   | MRI-based neuroimaging |