Title

A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment

Journal

Nature Neuroscience, 22(3)

ISSN

1097-6256

Authors

Schork, Andrew J
Won, Hyejung
Appadurai, Vivek
et al.

Publication Date

2019-03-01

DOI

10.1038/s41593-018-0320-0

Peer reviewed

nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-018-0320-0

A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment

Andrew J. Schork[1,2], Hyejung Won[3,4,5,6,7], Vivek Appadurai[1,2], Ron Nudel[1,2], Mike Gandal[3,4,5], Olivier Delaneau[8,9,10], Malene Revsbech Christiansen[11], David M. Hougaard[2,12], Marie Bækved-Hansen[2,12], Jonas Bybjerg-Grauholm[2,12], Marianne Giørtz Pedersen[2,13,14], Esben Agerbo[2,13,14], Carsten Bøcker Pedersen[2,13,14], Benjamin M. Neale[15,16,17], Mark J. Daly[15,16,17], Naomi R. Wray[18,19], Merete Nordentoft[2,20,21], Ole Mors[2,22], Anders D. Børglum[2,23,24], Preben Bo Mortensen[2,13,14,24], Alfonso Buil[1,2], Wesley K. Thompson[1,2,25], Daniel H. Geschwind[3,4,5,26] and Thomas Werge[1,2,21]*

There is mounting evidence that seemingly diverse psychiatric disorders share genetic etiology, but the biological substrates mediating this overlap are not well characterized. Here we leverage the unique Integrative Psychiatric Research Consortium (iPSYCH) study, a nationally representative cohort ascertained through clinical psychiatric diagnoses indicated in Danish national health registers. We confirm previous reports of individual and cross-disorder single-nucleotide polymorphism heritability for major psychiatric disorders and perform a cross-disorder genome-wide association study. We identify four novel genome-wide significant loci encompassing variants predicted to regulate genes expressed in radial glia and interneurons in the developing neocortex during mid-gestation. This epoch is supported by partitioning cross-disorder single-nucleotide polymorphism heritability, which is enriched at regulatory chromatin active during fetal neurodevelopment. These findings suggest that dysregulation of genes that direct neurodevelopment by common genetic variants may result in general liability for many later psychiatric outcomes.

Since the first attempts to establish a consistent nosology in psychiatry[1] there has been an appreciation for the difficulties in delineating patient populations[2–4]. Without objectively discriminative biomarkers, diagnoses are made through the integration of subjective presentations, including patient experience and behavioral observations[2–4]. When research diagnostic systems are used in clinically ascertained case-control studies, diagnoses may be designated hierarchically, incorporate perceived severity, censor milder episodes and/or select for archetypical presentations; a process that may obscure complicating co-morbid or pre-morbid episodes[4,5]. In reality, psychiatric patients may present more heterogeneously, sharing features and blurring boundaries among disorders and between disorders and typical behavior[2,4,6]. To some extent this blurring is thought to extend to etiological features, with certain environmental risks predisposing to diverse outcomes[2,7,8], non-specific efficacy of drug treatments implying shared neurochemical pathologies[9] and reports of extensive overlap in genetic risk factors[3,4].

The large genetic component to susceptibility for psychiatric outcomes is established from consistently moderate to high

[1]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark. [2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Copenhagen, Denmark. [3]Department of Neurology, University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [5]Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [6]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [7]UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC, USA. [8]Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland. [9]Swiss Institute of Bioinformatics (SIB), University of Geneva, Geneva, Switzerland. [10]Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland. [11]DTU Bioinformatics, Technical University of Denmark, Lyngby, Denmark. [12]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [13]NCRR - National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [14]Centre for Integrated Register-based Research (CIRRAU), Aarhus University, Aarhus, Denmark. [15]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, USA. [16]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [17]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [18]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [19]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [20]Copenhagen Mental Health Center, Mental Health Services Capital Region of Denmark Copenhagen, Copenhagen, Denmark. [21]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [22]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. [23]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. [24]Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark. [25]Division of Biostatistics, Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA. [26]Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. *e-mail: Thomas.Werge@regionh.dk

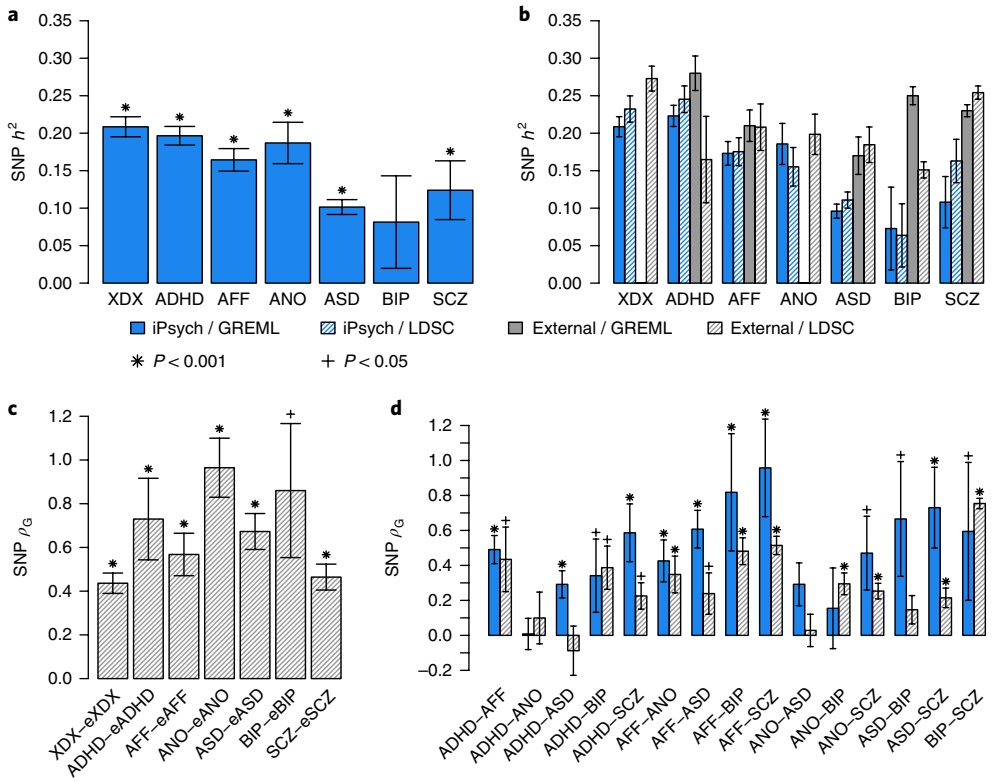

Fig. 1 | SNP heritability and genetic correlation estimates for iPSYCH indications. a, Liability scale GREML SNP heritability estimates of iPSYCH indications according to Danish population lifetime risk. Significance determined by likelihood ratio test (LRT), one-tailed *P* value. Sample sizes and statistics are provided in Supplementary Table 3. **b**, Liability scale SNP heritability estimates according to typical lifetime risk estimates for each disorder are estimated in iPSYCH and taken from external studies, using both GREML and LDSC. Sample sizes and statistics are provided in Supplementary Table 4. **c**, LDSC SNP-based genetic correlations between iPSYCH and external studies for the same disorders. Significance determined by two-sided *z*-test. Sample sizes and statistics are provided in Supplementary Table 5. **d**, SNP-based genetic correlations among iPSYCH indications estimated via GREML and among external studies estimated via LDSC. For GREML, significance was determined by one-sided LRT; sample sizes and statistics are provided in Supplementary Table 6. For LDSC, significance was determined by two-sided *z*-test; sample sizes and statistics are provided in Supplementary Table 8. All error bars denote estimate standard errors of estimates. Bar color denotes data source and shading denotes estimation method. Star symbols denote significance after Bonferroni correction for 44 variance component estimates ($P < 0.001$). Cross symbols denote nominal significance ($P < 0.05$). OTH, other indications not falling within individual disorder categories.

heritability estimates[10] and this contribution appears shared broadly among disorders. Initial observations of familial co-aggregation and genetic correlations for bipolar disorder (BIP) and schizophrenia (SCZ)[11] have been extended to include other mood disorders, autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD)[5,12,13]. Recent single-nucleotide polymorphism (SNP)-based investigations of shared polygenetic risk provide further support in patient populations with similar symptom profiles, such as SCZ and BIP[14,15] or major depressive disorders (MDD)[14–16] but are also extending this overlap to disorders with more dissimilar clinical profiles such as SCZ and ADHD[15,17,18], anorexia (ANO)[15] or ASD[14,19]. An emerging hypothesis is that at least a portion of genetic risk for psychiatric disorders is shared or perhaps non-specific with respect to outcomes.

Plausible substrates for such non-specific susceptibility are emerging from molecular genetic studies. The pleiotropic effects of large-effect copy number variants[5,20] (CNVs) such as 16p11.2 (ref. [21]), 22q11.2 (ref. [22]) and NRXN1 (ref. [23]) suggest insults to neurodevelopment and synaptic function may underlie some of the shared risk. The few studies of common variants directly investigating shared etiology[5] have implicated genes involved in calcium channel neurobiology. A recent integrative transcriptional study surmounted the challenges posed by the reliance on clinical phenotypes and the absence of biomarkers by using post-mortem gene expression profiles across multiple disorders to identify gene sets related to neuronal and astrocytic functions as shared molecular intermediaries[24]. Although plausible hypotheses are emerging, comparatively little has been done to characterize common variants that may have non-specific effects, especially in more representative patient populations.

Here we leverage the unique iPSYCH case-cohort study[25]. iPSYCH is composed of one of the largest single population samples of genotyped psychiatric patients in the world and a representative, random sample from the same national birth cohort. iPSYCH has the unique advantage of ascertaining essentially all major psychiatric patients from a single population and diagnostic schema, uniformly ascertained according to care provided under the same public healthcare system. By comparison, the majority of psychiatric disorder cohorts are ascertained from clinics which may enrich for prevalent (under treatment at the time of ascertainment) cases diagnosed by different schemas and in different populations, confounding, obscuring or otherwise diminishing cross-disorder inferences.

We performed a genetic dissection of this naturalistic and essentially complete patient population ascertained passively from a nation-wide Danish birth cohort (1981–2005, $n = 1{,}472{,}762$) for ADHD, affective disorder (AFF), ANO, ASD, BIP or SCZ (Supplementary Tables 1 and 2; Supplementary Fig. 1). We perform a diagnosis agnostic cross-disorder genome-wide association study (XDX GWAS) comparing common variant allele frequencies

Table 1 | Genome-wide significant associations

Locus	Chr	Range (hg19)	Index SNP	Position	A1	A2	Info	OR	s.e.	*P*	Functional candidate genes
1	2q32.1	183,279,530–183,680,199	rs4322805	183,535,884	A	G	–	1.07	0.01	2.92×10^{-8}	*PDE1A, PPP1R1C*
2	3p21.31-3p21.2	46,381,000–52,161,508	3:48644636:G:A	48,644,636	A	G	0.78	1.36	0.06	3.99×10^{-08}	*AMIGO3, AMT, APEH, ARIH2, ATRIP, BSN, C3orf84, CCDC36, CCDC51, CCDC71, CDHR4, CELSR3, COL7A1, DAG1, DALRD3, DHX30, FBXW12, GMPPB, GPX1, IMPDH2, IP6K1, IP6K2, KLHDC8B, LAMB2, MON1A, MST1, NCKIPSD, NDUFAF3, NICN1, P4HTM, PFKFB4, PLXNB1, PRKAR2A, QARS, QRICH1, RBM6, RHOA, RNF123, SHISA5, SLC25A20, SLC26A6, SMARCC1, SPINK8, TCTA, TEX264, TMA7, TREX1, UCN2, UQCRC1, USP19, USP4, WDR6, ZNF589*
3	3q13.32	117,453,031–117,997,735	rs6780942	117,828,678	T	C	0.96	1.10	0.01	1.11×10^{-10}	*IGSF11*
4	10q25.1	106,372,083–107,364,513	rs12265655	106,744,534	C	T	0.99	0.92	0.01	1.47×10^{-09}	*SORC3*[a]

Four loci indexed by genome-wide significant index SNPs in the XDX GWAS ($n=46{,}008$ cases, 19,526 controls) implicate a number of candidate genes. OR, odds ratio; s.e., standard error the natural log of OR; A1, effect allele; A2, noneffect allele; Info, imputation information score. [a]SORCS3 was implicated by overlap, not functional connection.

between the entire patient ($n=46{,}008$) and population control cohorts ($n=19{,}526$). We explore the consistency of our findings in published reports and perform an independent replication study in a small, diverse sample from the same birth cohort ($n=7{,}163$). Finally, we integrate published neurobiological data to suppose that the effects of common genetic variants with non-specific associations to psychiatric outcomes are mediated through dysregulation of early neurodevelopmental processes.

Results

The genetic architecture of iPSYCH diagnostic indications. We first asked if the genetic architecture for psychiatric diagnoses in the independent, population-based iPSCYH cohort was consistent with previous reports. The SNP heritability and genetic correlations in iPSYCH are depicted in Fig. 1 and point estimates, s.e. and *P* values are listed in Supplementary Tables 3–8. SNP heritability was generally modest, ranging from 0.08 to 0.20 and s.e. were small (0.01 to 0.03), except for BIP (0.04) and SCZ (0.06). The SNP heritability for diagnosis agnostic psychiatric indications (XDX) was of similar magnitude as each individual disorder (Fig. 1a and Supplementary Table 3). We compare our estimates to those from external studies for the same disorders (Fig. 1b and Supplementary Table 4), including linkage disequilibrium score regression (LDSC) estimates in seven external GWAS (eGWAS) (Methods): attention-deficit hyperactivity disorder (eADHD), autism spectrum disorder (eASD), major depressive disorder (eAFF), eating disorders (eANO), bipolar disorder (eBIP), schizophrenia (eSCZ) and a cross-disorder (eXDX) GWAS combining portions of eADHD, eAFF, eASD, eBIP and eSCZ. When presented together, the liability scale SNP heritability estimates by genetic restricted maximum likelihood (GREML) and LDSC for the iPSYCH indications are broadly concordant with external studies (Fig. 1b) suggesting an important contribution of common SNPs to underlying liability.

LDSC estimates of genetic correlation between iPSYCH indications and eGWAS of the same disorders were generally large (0.44–0.96; Fig. 1c and Supplementary Table 5), highly significant and consistent with the few previous reports of cross-cohort genetic correlations of the same psychiatric disorder[14,16] (Supplementary Figs. 2 and 3). These data suggest that iPSYCH indications capture similar genetic effects to external studies, with expected levels of across cohort genetic heterogeneity. The moderate genetic correlation between XDX and eXDX can likely be explained by differences in the proportion of individual case groups used in the two studies.

GREML point estimates of genetic correlations for all pairs of iPSYCH indications are at least moderate (Fig. 1d, Supplementary Tables 6 and 7) with the exception of ADHD–ANO ($\rho_G=0.01$). Standard errors ranged from moderate (0.08) to large (0.39) and were largest for pairs involving the smallest samples (SCZ, 0.21–0.39; BIP, 0.17–0.39). For ADHD–ANO and ANO–BIP standard errors were large enough that we could not reject the null hypothesis, despite moderate point estimates ($P>0.05$). Each iPSYCH indication also shows at least moderate genetic correlation with the aggregate of all remaining indications (Supplementary Fig. 4). LDSC estimates of genetic correlation using eGWAS (Fig. 1d and Supplementary Table 8) trend smaller than for iPSYCH indications of the same disorders, consistent with the additional cross-cohort genetic heterogeneity (Fig. 1c and Supplementary Figs. 2 and 3). Despite this, both sets of estimates suggest broad sharing of SNP effects among nearly all pairs of disorders.

A diagnosis agnostic GWAS. Motivated by the appreciable XDX SNP heritability and prevalence of substantial genetic correlations among iPSYCH indications, we performed a diagnosis agnostic GWAS (XDX). We combined all iPSYCH psychiatric patients into a single case cohort ($n=46{,}008$). As GWAS controls ($n=19{,}526$, see Methods) we used the subset of individuals randomly selected from the Danish national biobank (iPSYCH design controls, $n=30{,}000$) with no current diagnoses and passing quality control. We identified four independent loci tagged by genome-wide significant index SNPs (referred to throughout as loci 1–4; $P<5\times10^{-8}$; Table 1 and

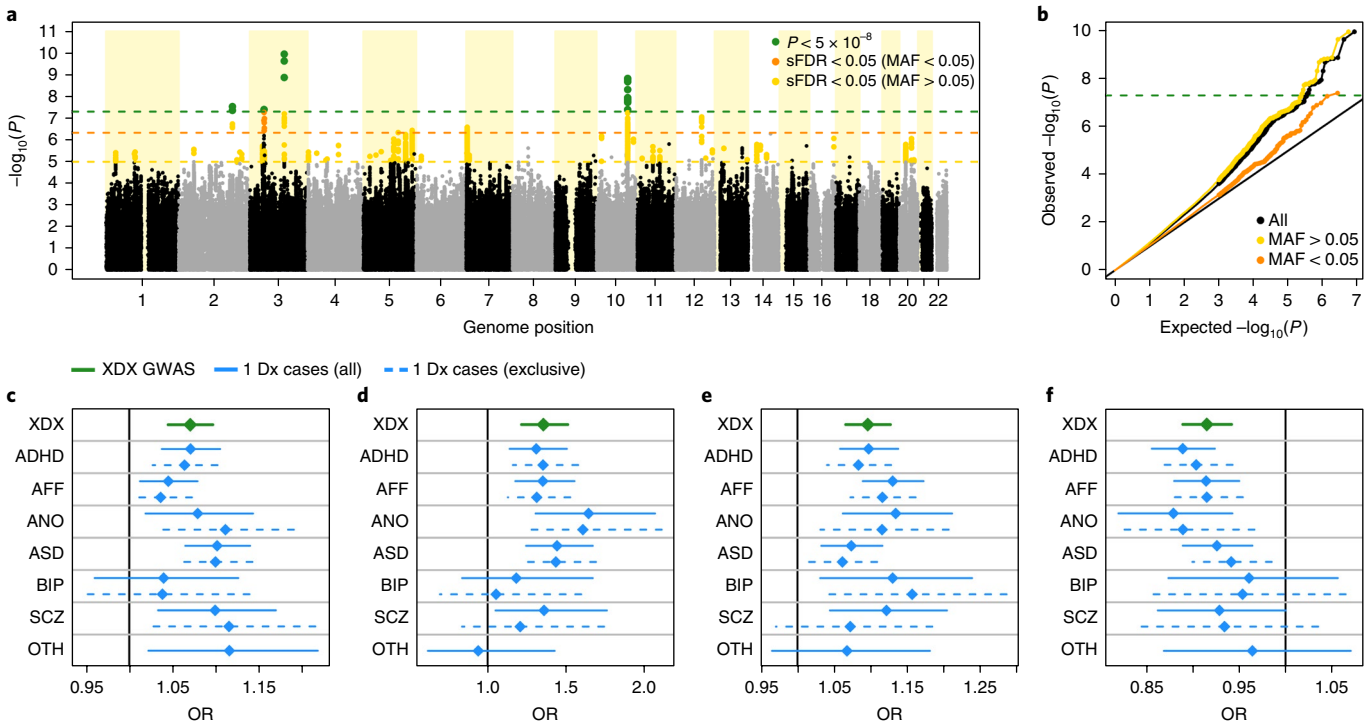

Fig. 2 | Cross-diagnosis genome-wide association study results. a, Manhattan plot depicting $-\log_{10}$ of two-tailed *P* values from case–control logistic regression association tests at 8,018,013 imputed SNP dosages in 46,008 cases and 19,526 controls identifies four genome-wide significant loci (green) and 46 suggestive loci (gold, orange). **b** Quantile–quantile (QQ) plot comparing the distribution of $-\log_{10}$ association *P* values to that expected under the global null hypothesis. **c**–**f**, For each genome-wide significant locus, 1 (**c**), 2 (**d**), 3 (**e**), and 4 (**f**), ordered as in the Manhattan plot, the odds ratio (OR) and approximate 95% confidence interval (CI) for the SNP from the XDX GWAS (green) is similar in sign and magnitude to ORs from associations with single-indication case groups (blue). Sample sizes for single-indication GWAS cohorts are described in Supplementary Table 1 and full association statistics are presented in Supplementary Tables 10 and 11. MAF, minor allele frequency; Dx, diagnosis.

Fig. 2) and another 46 loci (loci 5–50) indexed by suggestive associations (stratified false discovery rate, sFDR < 0.05; Supplementary Table 9). No locus showed genome-wide significant evidence of a second independent association after conditioning on the index SNP. The distribution of *P* values from the GWAS demonstrated moderate levels of inflation (genomic inflation factor, $\lambda_{GC} = 1.15$; Fig. 2b). A significant LDSC estimate for the XDX SNP heritability (liability scale SNP $h^2_{SNP} = 0.21$, s.e.m. = 0.01) with an intercept close to unity (1.02; s.e.m. = 0.01) suggests this inflation is likely due to polygenes rather than population stratification or cryptic relatedness.

Although the XDX GWAS is most sensitive to variants with consistent effects across diagnoses, variants strongly associated with a single indication could present as apparent cross-disorder associations. However, rather than showing skewing of the odds ratios for one or two indications, the index SNPs for loci show consistent trends across nearly all indications when association tests were performed comparing the XDX GWAS controls to each single-indication case group (Fig. 2c–f). This broad pattern holds for suggestive loci (Supplementary Tables 10 and 11). The XDX GWAS uncovered multiple loci with non-specific effects on risk across iPSYCH indications that are not driven by a single or small subset of disorders.

Independent statistical support. We sought independent statistical support for our associations from three sources. First, we queried the top associations from 4,299 GWAS aggregated in the NHGRI-EBI GWAS catalog for connections to our top 50 XDX loci (Methods; Supplementary Table 12). NHGRI had independently assigned a category to each GWAS that places psychiatric outcomes within the broader label of 'neurological disorder'. In total, our 50 loci were reported in the top hits of 22 out of 589 neurological disorder GWAS, with nearly all representing psychiatric outcomes more specifically and 21 out of 3,710 GWAS of other traits and diseases (Supplementary Table 12), a significant difference (binomial proportion test; $P = 3.48 \times 10^{-13}$). Specifically, genome-wide significant loci 2 and 4 and six of the suggestive loci found strong support in domain-relevant studies from the GWAS catalog. Locus 4 was recently reported as genome-wide significant in two meta-GWAS that include the iPSYCH data: one for MDD[16] and one for ADHD[18].

The GWAS catalog contains many studies but very few SNPs from each study, providing an incomplete picture of subtle polygenic trends. Thus, we also mapped our top 50 loci to the best proxy SNP in each of the seven eGWAS described previously (eADHD, eAFF, eANO, eASD, eBIP, eSCZ, eXDX) and, where possible, to the best proxy among the top 10,000 associations from a self-report major depressive disorder GWAS (ref. [26]; eSR–MDD; Table 2, Supplementary Table 13). We observe significant external replication ($P < 1.56 \times 10^{-3}$ correcting for four loci, each tested in eight studies) for locus 2 in the eADHD and eSCZ studies. Locus 4 significantly replicated in the eSR–MDD and eSCZ studies. Even where not individually significant, the association trends are strikingly consistent across all studies for loci 2–4. Importantly, our XDX trend is consistent for each of loci 1–4 and 35 out of 46 suggestive loci, in the eSCZ GWAS; a disorder among the least represented in iPSYCH, emphasizing the plausibility that these loci are non-specific. As a test of this concordance, we used linear regression to ask if, across the top 50 loci, the signed effects (regression statistics; *z*-scores) for proxy SNPs in the eGWAS are significantly predicted from the XDX effects for the same SNPs. We observed significant concordance ($P < 5 \times 10^{-3}$ to correct for ten concordance tests) between the XDX effects and the aggregated effects from all

Table 2 | Independent replication

	Disc.	Repl.		eADHD		eAFF		eANO		eASD		eBIP		eSR-MDD		eSCZ		eXDX	
Locus	Sign	Sign	P	Sign	P	Sign	P	Sign	P	Sign	P	Sign	P	Sign	P	Sign	P	Sign	P
1	+	+	0.26	–	0.81	–	0.63	–	0.96	–	0.56	-	0.58	NA	NA	+	0.14	–	0.52
2	+	+	0.14	+[a]	1×10^{-3}	+	0.05	+	0.09	+	0.07	+	0.01	NA	NA	+[a]	9×10^{-4}	+	3×10^{-3}
3	+	+	0.3	NA	NA	NA	NA	+	0.35	+	0.28	+	0.26	NA	NA	+	0.18	NA	NA
4	–	–	0.16	+	0.92	–	0.05	–	0.06	–	0.3	–	0.43	-[a]	1×10^{-5}	-[a]	8×10^{-5}	–	0.02

The sign of the effect of studied allele from the logistic regression association test (+, risk increasing; –, risk decreasing) from the discovery GWAS (Disc.; 46,008 cases, 19,526 controls) is presented against nine replications. The signs from logistic regression association tests in the iPSYCH replication cohort (Repl.; 4,481 cases, 2,682 controls) are concordant, although the effects are not significant (two-tailed $P > 0.05$). Looking up the effect of the four loci in the published association statistics of eight independent studies (eADHD, $n = 2{,}960$ cases, 4,519 controls; eAFF, $n = 9{,}240$ cases, 9,519 controls; eANO, $n = 3{,}495$ cases, 10,982 controls; eASD, $n = 7{,}387$ cases, 8,567 controls; eBIP, $n = 9{,}784$ cases, 30,471 controls; eSR-MDD, $n = 75{,}607$ cases, 231,747 controls; eSCZ, $n = 34{,}241$ cases, 45,604 controls; eXDX, $n = 33{,}332$ cases, 27,888 controls) shows strong concordance in direction of effect, with some evidence of strict significance for replication. Full statistics are presented in Supplementary Tables 9 (Disc.), 13 (eGWAS), and 14 (Repl.). [a]Significant after Bonferroni correction for 36 tests accounting for four loci in each of nine replication GWAS. NA, comparable SNP associations not available.

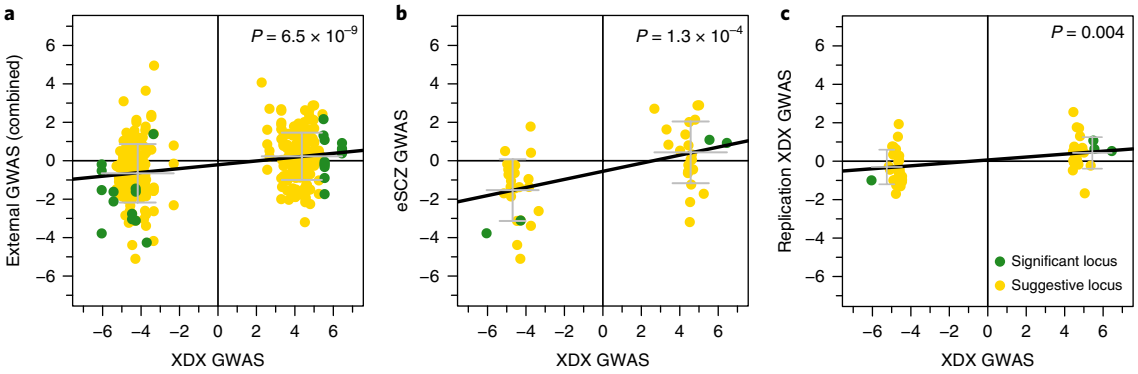

Fig. 3 | Consistency of effects in independent studies. a, iPSYCH XDX GWAS *z*-scores are associated with *z*-scores of the best proxy SNP in eADHD, eAFF, eANO, eASD, eBIP, eSCZ, eSR-MDD, and eXDX in aggregation ($n = 350$ *z*-score pairs; linear regression: $\beta = 0.10$, s.e. = 0.02; $t = 5.95$; Supplementary Table 13). **b**, iPSYCH XDX GWAS *z*-scores are associated with *z*-scores of the best proxy SNP in eSCZ only ($n = 50$ *z*-score pairs; linear regression: $\beta = 0.21$, s.e. = 0.05; $t = 4.16$). **c**, iPSYCH XDX GWAS *z*-scores are associated with iPSYCH internal replication *z*-scores ($n = 50$ *z*-score pairs; linear regression: $\beta = 0.08$, s.e. = 0.03; $t = 3.01$; Supplementary Table 14). Black lines show regression fit. Gray bars depict the mean and 1 s.d. interval of replication effects for SNPs with positive or negative effects in the XDX GWAS. Bonferroni correction for ten concordance tests, $P < 0.005$.

seven eGWAS (Fig. 3a) as well as in the eSCZ GWAS alone (Fig. 3b). Concordance was also significant for the eADHD and eXDX GWAS (Supplementary Figs. 5–11).

As a third form of external validation, we repeated the association tests for each index SNP using linear mixed models in a small, independent replication cohort. We included subjects excluded from the discovery GWAS cohort on the basis of outlying genetic ancestry (replication cohort: $n = 7{,}163$ with 4,481 cases; Supplementary Table 1; results: Table 2, Supplementary Table 14). Given the reduced sample size we did not expect, nor did we observe, any significant individual replications. We did observe, perhaps surprisingly given the diverse genetic backgrounds, a consistent trend at each of loci and for 31 of the additional 46 suggestive loci. This broad concordance was also significant by the same linear model as above ($P = 0.004$; Fig. 3c). Taken together, these analyses provide several forms of independent statistical support for the XDX loci identified here.

Tissue-specific partitioning of non-specific psychiatric diagnosis heritability. Next, we asked if the genetic effects constituting the XDX SNP heritability were stronger among SNPs linked to tissue-specific biological processes. We used LDSC regression for specifically expressed genes[27] (LD-SEG; Methods) to test for an enrichment of SNP heritability in each of 601 SNP sets. The SNP sets were defined by Finucane et al.[27] for proximity to genes preferentially expressed (GTEx and DEPICT sets) or chromatin marks open and active (RoadMap sets) in a variety of human tissues. None of the SNP sets defined by proximity to specifically expressed genes showed evidence for enrichment (false dicovery rate, FDR > 0.05; Supplementary Figs. 12 and 13 and Supplementary Tables 15 and 16). Among the SNP sets defined by chromatin marks (Fig. 4a and Supplementary Table 17), several related to gene regulation in fetal brain showed significant ($P < 6.88\times10^{-5}$; germinal matrix: H3K4me3, $P = 3.73\times10^{-5}$; female fetal brain: DNase, $P = 1.31\times10^{-6}$; male fetal brain: DNase, $P = 1.05\times10^{-5}$, H3K4me1, $P = 3.02\times10^{-5}$) or suggestive (FDR < 0.05; germinal matrix: H3K4me1, $P = 4.34\times10^{-4}$; female fetal brain: H3K4me1, $P = 3.38\times10^{-4}$, H3K4me3, $P = 3.45\times10^{-4}$) enrichment. No significant enrichment was observed in SNP sets related to chromatin marks in adult brain nor non-brain tissues (Supplementary Fig. 14 and Supplementary Table 18). Enrichment in the same fetal brain chromatin annotations replicated in the eXDX and eSCZ GWAS, a result described in the LDSC-SEG report[27] (Fig. 4b and Supplementary Table 19). Furthermore, iteratively removing subjects with each indication from the iPSYCH cases and re-estimating enrichment produced consistent results (Fig. 4c and Supplementary Table 20). Figures 4b,c emphasize the disorder non-specificity of these heritability enrichments. The heritability enrichments for regulatory variants active in fetal brain seen for the XDX GWAS replicate as a class in an independent cross-disorder GWAS (eXDX), in an independent psychiatric GWAS for a disorder (eSCZ) among the least represented in the iPSYCH patient population (Supplementary Fig. 1) and are not driven by any single iPSYCH indication.

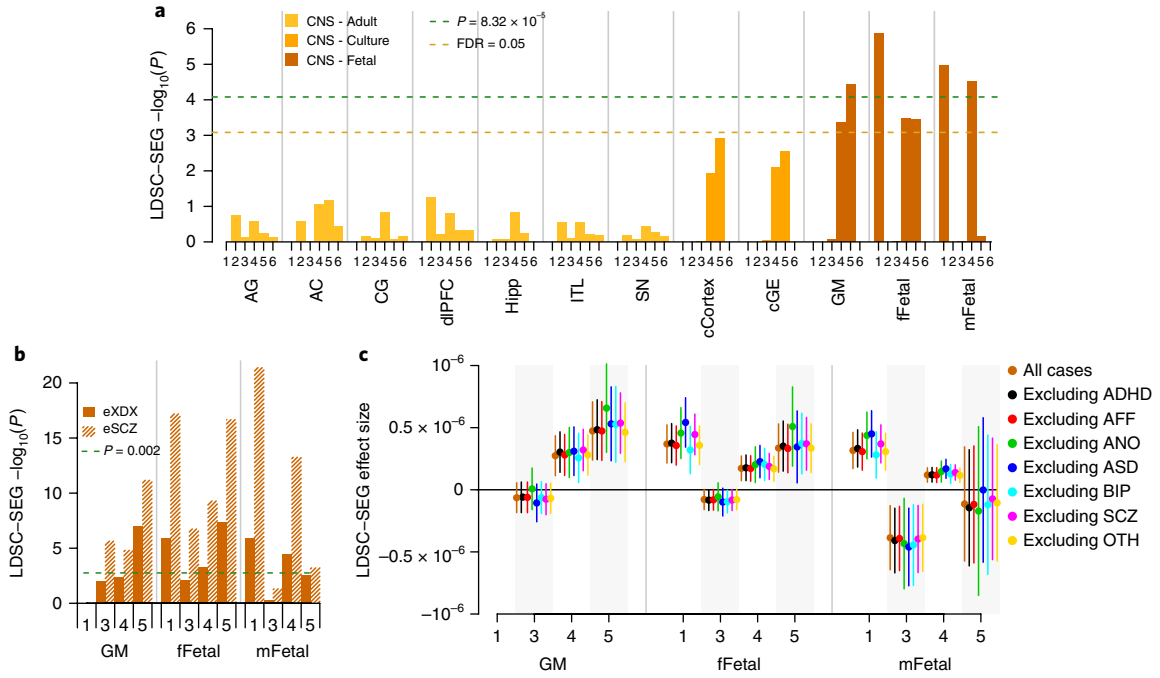

Fig. 4 | LDSC-SEG heritability partitioning. a, $-\log_{10}$ one-sided *P* values from LDSC-SEG enrichment tests for the iPSYCH XDX GWAS ($n = 46{,}008$ cases, 19,526 controls) highlight significant ($P < 8.32 \times 10^{-5}$, Bonferroni correction for 601 LDSC-SEG enrichment tests) and/or suggestive (FDR < 0.05) enrichment for heritability in SNP sets related to gene regulation in fetal brain tissue. Exact *P* values and effect sizes are presented in Supplementary Table 17. **b**, $-\log_{10}$ one-sided LDSC-SEG enrichment test *P* values for the same fetal brain SNP sets in the eSCZ ($n = 34{,}241$ cases, 45,604 controls) and eXDX ($n = 33{,}332$ cases, 27,888 controls) GWAS replicate enrichments ($P < 0.002$, Bonferroni correction for 24 replication enrichment tests). Exact *P* values and effect sizes are presented in Supplementary Table 19. **c**, The magnitude of LDSC-SEG enrichment effect estimates for fetal brain tissues in the XDX GWAS (orange) are not qualitatively different when excluding single case groups (non-orange). Gray shading denotes bunches of estimate replicates. Bars denote 95% confidence intervals for enrichment effect estimate. Exact effect sizes and sample sizes are presented in Supplementary Table 20. Axis numbers (**a**,**b**,**c**) denote SNPs sets defined by narrow peaks of 1, DNase, DNase I hypersensitivity; 2, H3K27ac, histone H3, lysine 27 acetylation; 3, H3K36me3, histone H3, lysine 36 trimethylation; 4, H3K4me, histone H3, lysine 4 methylation; 5, H3K4me3, histone H3, lysine 4 trimethylation; 6, H3K9ac, histone H3, lysine 9 acetylation. AG, adult angular gyrus; AC, adult anterior caudate; CG, adult cingulate gyrus; dlPFC, adult dorsolateral prefrontal cortex; Hipp, adult hippocampus; ITL, adult inferior temporal lobe; SN, adult substantia nigra; cCortex, cortex-derived cultured neurospheres; cGE, ganglion eminence-derived cultured neurospheres; GM, geminal matrix; fFetal, female fetal brain; mFetal, male fetal brain.

Identifying and characterizing candidate genes. We used FineMap at each of the top 50 loci (GWAS sFDR < 0.05) to identify a set of credible SNPs for each locus (Methods). To identify candidate pathways through a set of plausible candidate genes, each credible SNP was subsequently connected to target genes using chromatin interaction maps for the developing human cortical laminae (Methods). Out of 627 credible SNPs, 400 implicated one or more of 281 candidate genes (160 protein coding; Supplementary Table 21). As a set, these genes were enriched pathways including axon and dendrite development, receptor tyrosine kinases and histone regulators (Fig. 5a). Notably, histone regulators and chromatin remodelers have been recently implicated in various psychiatric disorders, including ASD and SCZ[19,28,29]. Consistent with the heritability enrichment, the candidate genes were more highly expressed during prenatal stages ($P = 1.67 \times 10^{-10}$; Fig. 5b), with peak expression during mid-gestation in the human brain (Fig. 5c). To understand what cells these genes might be functioning in, we leveraged expression profiles of single cells extracted from fetal brain tissue and tested for candidate gene enrichment (Methods). Remarkably, we observe candidates to be enriched for genes preferentially expressed in radial glia, an early form of neural stem cell and interneurons (FDR < 0.05; Fig. 5d). These results suggest that the cross-diagnosis loci are implicated in corticogenesis and establishment of neural circuits during brain development.

Specific candidate genes implicated by genome-wide significant loci include *PDE1A* and *PPP1R1C* for locus 1 (annotated local Manhattan plot, Supplementary Fig. 15). *PDE1A* encodes a Ca^{2+}/calmodulin-dependent cyclic nucleotide phosphodiesterase (PDE1s) that regulates cAMP and cGMP (ref. [30]) and can regulate L-type and T-type voltage-gated calcium channels[31]. *PPP1R1C* encodes an inhibiting regulatory subunit of protein phosphatase 1 (PP1), an enzyme involved in synaptic transmission and plasticity as part of the post-synaptic density[32] and has been shown to regulate neurite growth in cultured neurons[33]. Locus 2 is indexed by an uncommon SNP (MAF = 0.018) and covers a broad, gene-dense region implicating 53 diverse genes (Supplementary Fig. 16). Among these, *DAG1*, *QRICH1*, *RNF123* and *SMARCC1* harbor de novo risk variants for ASD and *CELSR3* for SCZ[34,35]. *IP6K2* is implicated by a well-described gene-enhancer connection (GeneHancer: GH03J048738)[36] and previous studies of SCZ[37] and ADHD[38]. Locus 3 implicates *IGSF11* (Supplementary Fig. 17), a gene encoding a neuronal adhesion molecule that binds to and stabilizes AMPA receptors regulating synapse development and plasticity[39]. Locus 4 overlaps considerably with the body of *SORC3* (Supplementary Fig. 18), which encodes a post-synaptic sorting and signaling receptor involved in aspects of neuronal functioning including synaptic plasticity[40].

Discussion

In this study, we leveraged the uniquely designed iPSYCH study to provide an unprecedented perspective on the overlap in genetic etiology underlying major psychiatric disorders. The iPSYCH cohort

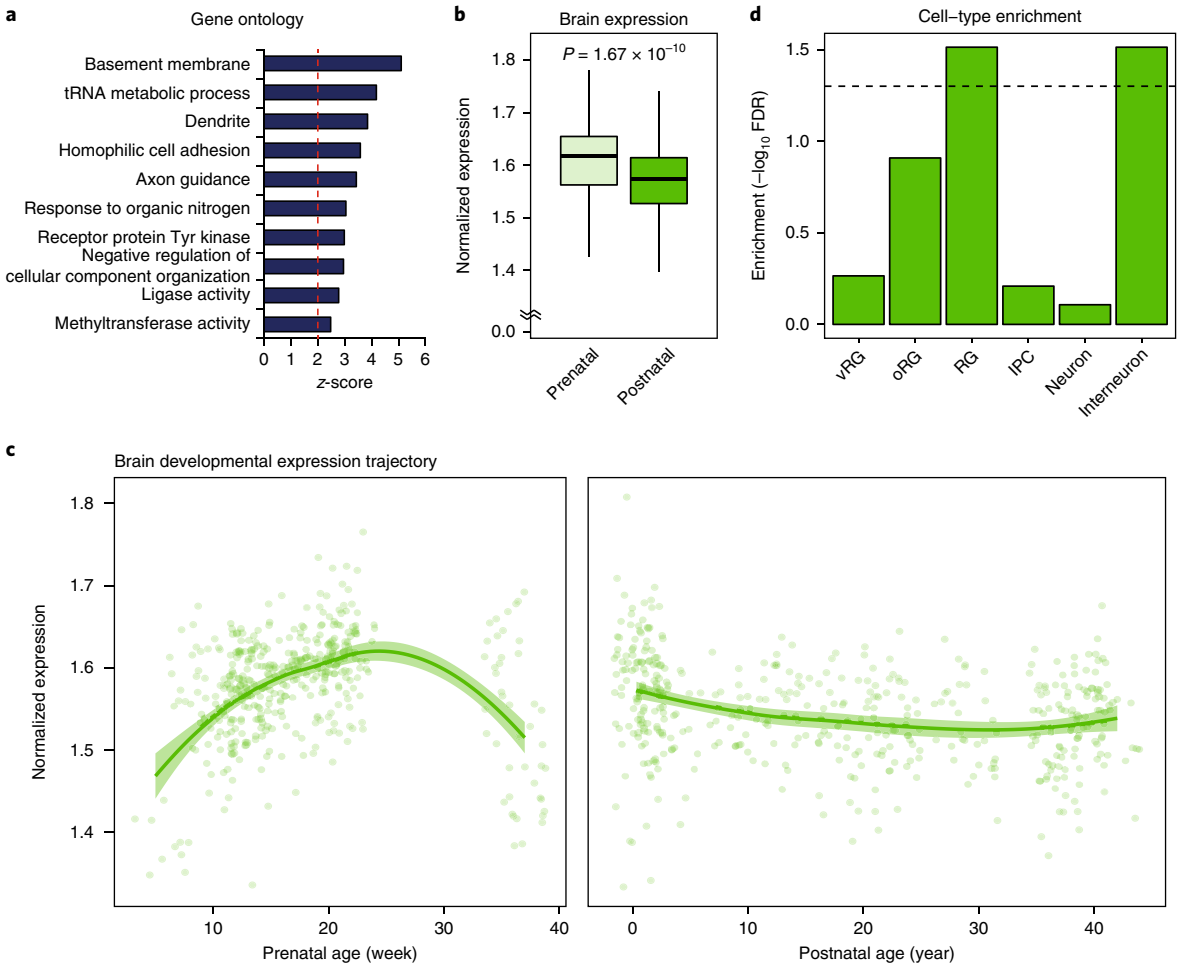

Fig. 5 | Candidate geneset enrichments for neurodevelopmental processes. a, Top ten gene ontology categories significantly enriched among the 281 XDX candidate genes. **b**, The XDX candidate genes have higher average cortical expression in the prenatal stage. *P* value calculated by two-sided *t* test. Box center is the median, box border denotes 1st–3rd quartiles (Q), the lower whisker, Q1 – 1.5× interquartile range (IQR) and the upper whisker, Q3 + 1.5× IQR. **c**, The average cortical expression trajectory for XDX candidate genes peaks during the third quarter of the prenatal period. LOESS smooth curve with shading showing 95% confidence bands; $n = 410$ and 453 for prenatal and postnatal samples, respectively (**b** and **c**). **d**, The XDX candidate genes are enriched for genes with specific expression in developing radial glia and interneurons. RG, radial glia; vRG, ventricular radial glia; oRG, outer radial glia; IPC, intermediate progenitor cells.

has the advantage of coming from a nationally representative study interrogating an essentially complete population of patients against a representative sample of diagnosis-free individuals from the same Danish birth cohort. Previous results are based on cohorts identified independently at tertiary research centers by different research groups, often in different countries, factors which could reduce true XDX overlap or shared heritability. Here, patients are ascertained against a uniform background, including genetic background, diagnostic schema, system and other cultural or sociodemographic factors, limiting the potential impact of heterogeneity that may arise when comparing independently ascertained patient cohorts.

These strengths of iPSYCH and the decades of preceding epidemiological research on mental disorders in Danish health registers make iPSYCH uniquely suited for providing robust estimates of genetic parameters. There is less uncertainty in the underlying population that our sample represents and in key parameters, such as the assumed lifetime risk for a particular diagnosis definition in the population, that define the liability scale heritability. Further, the case-cohort, register-based sampling ensures that we capture all individuals diagnosed at the time of ascertainment, while clinical ascertainment may select a subset of patients that are not a representative sample of the population of diagnosed individuals in terms of phenotypic presentation[41] or genetic architecture[42]. The effects of clinical ascertainment may be exaggerated in cohorts collected for GWAS, which may enrich for treatment resistance, archetypical presentation, or features of severity in cases and exceptional mental health in controls, as a means to boost power for single locus discovery. This kind of extreme sampling can create substantial upward bias in liability scale heritability, when not accounted for[43].

The broad trend towards larger genetic correlations estimated within the iPSYCH cohort when compared to those estimated across external studies may be expected given a component of genetic risk for psychiatric disorders that is cohort, study or population specific. The presence of within-population genetic effects is consistent with the trend for genetic correlations of the same disorder between iPSYCH and the corresponding external study to be less than 1. Importantly, we observe the magnitude of these correlations to be similar to the few previous reports[14,16], suggesting population- or cohort-specific genetic effects are not a unique feature or flaw of the register-based ascertainment in iPSYCH.

Trends towards lower heritability and higher genetic correlations have also been ascribed to increased rates of misdiagnosis[44]. We emphasize that, although each individual patient was not reassessed when they were enrolled in iPSYCH, the registered diagnoses reflect

clinical diagnoses made by trained, certified psychiatrists; small validation studies have shown them to be highly reliable (for examples, see refs. [45,46]). Further, it has been suggested[44,47] that implausibly large numbers of misdiagnosed cases are needed to explain large trends. Schork et al.[47] have undertaken a series of sensitivity analyses to provide broader context for trending differences in the genetic parameters we present here. They consider various differences in aspects of ascertainment between iPSYCH and these external studies including the accuracy of assumed lifetime risk, the age of control subjects, control subjects selected for exceptional mental health, cases selected for genetic severity and levels of misdiagnosis. Their analyses suggest that, for heritability, it is difficult to rule out sampling variability but suggest one plausible contributor to difference trends may be an overestimation due to extreme sampling[43] in external studies. Genetic correlations are more robust to sample ascertainment and we see the most likely contributor to be cross-cohort genetic heterogeneity (for example, Fig. 1c), although more work is needed to fully understand factors that generate variation in genetic correlation estimates from real data.

In general, the broad concordance of the estimated genetic parameters suggests the totality of ascertainment-related factors may be relatively modest, although assessing the true extent of between-population heterogeneity in the genetic architecture of psychiatric will require additional large population-representative cohorts and considerations for ascertainment. With the continued emergence of large population-based ascertainments, such as iPSYCH, 23andMe[26], the UKBiobank[48] and large insurance record cohorts[49], we feel this point is critically important for interpreting any differences that may emerge.

It is interesting that our functional interpretations of shared genetic risk factors, both in terms of polygenic trends and with respect to our top associations, converge on fetal neurodevelopmental processes. This epoch has been implicated by previous studies of large effect CNVs and rare or de novo variants but also by environmental exposures (for example, ref. [50]). Such convergence could suggest a critical developmental window during which part of the susceptibility to later psychiatric outcomes is defined. The overlap in timing for the action of genetic and environmental susceptibility factors may also help to carve out a hypothesis space for targeted investigations of gene–environment interactions.

While germline variants are present since the first moments of embryonic development, the chromatin they reside in and genes they may affect undergo a continual evolution of active and inactive states across developmental and life stages. We note that our top associations are potential regulators of genes with familiar functions[14,29], (post)synaptic and calcium channel biology. We also demonstrated that, in our study, these variants and their associated genes are most coherently involved in aspects of fetal neurodevelopment. If susceptibility emerges from disruptions to this familiar synapse and calcium channel biology but specifically during neuronal proliferation, migration and establishment of circuits, this could have implications for the development of interventions. To make good on the promise that GWAS can identify plausible pharmacological targets, it will become critical to consider the developmental stages during which variants induce susceptibility. It may be the case that the pathology induced by an associated germline variant occurs in one developmental stage but actionable pharmacological targets represent different downstream molecular pathologies[24]. Explicitly considering the developmental course of pathological susceptibility implied by associated germline variants is a critical next step in the translational promise of GWAS results and emphasizes the importance of partitioning risk with development in mind.

While our study has a number of strengths, it also has a few important limitations. First, we have tried to be comprehensive in our bioinformatics integration of available data resources but experimental work is needed to unequivocally confirm the functional hypotheses that arise from these genetic associations and subsequent candidate gene prioritization. Second, the relative youth of our cohort limits our ability to make strong claims about the effects in adult onset disorders. While we emphasize the evidence for replication in an external study of schizophrenia, future work should further investigate these findings across a broader spectrum of disorders. Third, although our sample is large for a single psychiatric cohort, it is still comparatively modest by GWAS meta-study standards, especially in other common, non-central nervous system disorders. As such, we cannot claim that our results suggest the only mechanisms through which common SNPs may create risk for multiple psychiatric disorders but only that our findings represent a minimal set of features that set the stage for well-informed mechanistic studies. Further, it is also likely that other susceptibility factors are more specific, contributing to the unique presentations of each disorder. Future work could take direct aim at these factors and their interactions with more general susceptibility factors, as we continue to update, partition and functionally characterize our conceptualizations of the genetic architecture underlying major psychiatric disorders.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41593-018-0320-0.

Received: 13 January 2018; Accepted: 6 December 2018;
Published online: 28 January 2019

References

1. Robins, E. & Guze, S. B. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am. J. Psychiatry* **126**, 983–987 (1970).
2. Kendell, R. & Jablensky, A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am. J. Psychiatry* **160**, 4–12 (2003).
3. Krystal, J. H. & State, M. W. Psychiatric disorders: diagnosis to therapy. *Cell* **157**, 201–214 (2014).
4. O'Donovan, M. C. & Owen, M. J. The implications of the shared genetics of psychiatric disorders. *Nat. Med.* **22**, 1214–1219 (2016).
5. Doherty, J. L. & Owen, M. J. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med.* **6**, 29 (2014).
6. Widiger, T. A. & Sankis, L. M. Adult psychopathology: issues and controversies. *Annu. Rev. Psychol.* **51**, 377–404 (2000).
7. Bulik, C. M., Prescott, C. A. & Kendler, K. S. Features of childhood sexual abuse and the development of psychiatric and substance use disorders. *Br. J. Psychiatry* **179**, 444–449 (2001).
8. Brown, G. W., Harris, T. O. & Eales, M. J. Social factors and comorbidity of depressive and anxiety disorders. *Br. J. Psychiatry Suppl.* **168**, 50–57 (1996).
9. Gorman, J. M. & Kent, J. M. SSRIs and SNRIs: broad spectrum of efficacy beyond major depression. *J. Clin. Psychiatry* **60** (Suppl.4), 33–38 (1999).
10. Polderman, T. J. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
11. Lichtenstein, P. et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
12. Sullivan, P. F. et al. Family history of schizophrenia and bipolar disorder as risk factors for autism. *Arch. Gen. Psychiatry* **69**, 1099–1103 (2012).
13. Larsson, H. et al. Risk of bipolar disorder and schizophrenia in relatives of people with attention-deficit hyperactivity disorder. *Br. J. Psychiatry* **203**, 103–106 (2013).
14. Lee, S. H. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
15. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
16. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
17. Hamshere, M. L. et al. Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. *Br. J. Psychiatry* **203**, 107–111 (2013).

18. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for ADHD. *Nat.Genet.* **51**, 63–75 (2019).
19. Grove, J. et al. Common risk variants identified in autism spectrum disorder. Preprint at *bioRxiv* https://doi.org/10.1101/224774 (2017).
20. Kirov, G. et al. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol. Psychiatry* **75**, 378–385 (2014).
21. Green Snyder, L. et al. Autism spectrum disorder, developmental and psychiatric features in 16p11.2 duplication. *J. Autism. Dev. Disord.* **46**, 2734–2748 (2016).
22. Schneider, M. et al. Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 deletion syndrome. *Am. J. Psychiatry* **171**, 627–639 (2014).
23. Rujescu, D. et al. Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18**, 988–996 (2009).
24. Gandal, M. J. et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
25. Pedersen, C. B. et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
26. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
27. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
28. Parikshak, N. N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
29. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).
30. Bender, A. T. & Beavo, J. A. Cyclic nucleotide phosphodiesterases: molecular regulation to clinical use. *Pharmacol. Rev.* **58**, 488–520 (2006).
31. Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**(D1), D1074–D1082 (2018).
32. Munton, R. P., Vizi, S. & Mansuy, I. M. The role of protein phosphatase-1 in the modulation of synaptic and structural plasticity. *FEBS Lett.* **567**, 121–128 (2004).
33. Han, Q. J. et al. IPP5 inhibits neurite growth in primary sensory neurons by maintaining TGF-β/Smad signaling. *J. Cell Sci.* **126**, 542–553 (2013).
34. Kosmicki, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504–510 (2017).
35. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
36. Fishilevich, S. et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017** https://doi.org/10.1093/database/bax028 (2017).
37. Glatt, S. J. et al. Similarities and differences in peripheral blood gene-expression signatures of individuals with schizophrenia and their first-degree biological relatives. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **156B**, 869–887 (2011).
38. Dela Peña, I. et al. Common prefrontal cortical gene expression profiles between adolescent SHR/NCrl and WKY/NCrl rats which showed inattention behavior. *Behav. Brain. Res.* **291**, 268–276 (2015).
39. Jang, S. et al. Synaptic adhesion molecule IgSF11 regulates synaptic transmission and plasticity. *Nat. Neurosci.* **19**, 84–93 (2016).
40. Breiderhoff, T. et al. Sortilin-related receptor SORCS3 is a postsynaptic modulator of synaptic depression and fear extinction. *PLoS ONE* **8**, e75006 (2013).
41. Cohen, P. & Cohen, J. The clinician's illusion. *Arch. Gen. Psychiatry* **41**, 1178–1182 (1984).
42. Meier, S. M. et al. High loading of polygenic risk in cases with chronic schizophrenia. *Mol. Psychiatry* **21**, 969–974 (2016).
43. Yap, C. X. et al. Misestimation of heritability and prediction accuracy of male-pattern baldness. *Nat. Commun.* **9**, 2537 (2018).
44. Wray, N. R., Lee, S. H. & Kendler, K. S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur. J. Hum. Genet.* **20**, 668–674 (2012).
45. Jakobsen, K. D., Frederiksen, J. N., Parnas, J. & Werge, T. Diagnostic agreement of schizophrenia spectrum disorders among chronic patients with functional psychoses. *Psychopathology* **39**, 269–276 (2006).
46. Lauritsen, M. B. et al. Validity of childhood autism in the Danish Psychiatric Central Register: findings from a cohort sample born 1990-1999. *J. Autism. Dev. Disord.* **40**, 139–148 (2010).
47. Schork, A. et al. Exploring contributors to variability in estimates of SNP-heritability and genetic correlations from the iPSYCH case-cohort and published meta-studies of major psychiatric disorders. Preprint at *bioRxiv* https://doi.org/10.1101/487116 (2018).
48. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS. Med.* **12**, e1001779 (2015).
49. Wang, K., Gaitsch, H., Poon, H., Cox, N. J. & Rzhetsky, A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat. Genet.* **49**, 1319–1325 (2017).
50. Mortensen, P. B. et al. Effects of family history and place and season of birth on the risk of schizophrenia. *N. Engl. J. Med.* **340**, 603–608 (1999).

Acknowledgments

The iPSYCH Initiative is funded by the Lundbeck Foundation (grant nos. R102-A9118 and R155-2014-1724), the Mental Health Services Capital Region of Denmark, University of Copenhagen, Aarhus University and the university hospital in Aarhus. Genotyping of iPSYCH samples was supported by grants from the Lundbeck Foundation, the Stanley Foundation, the Simons Foundation (SFARI 311789) and NIMH (5U01MH094432-02). The iPSYCH Initiative use the Danish National Biobank resource that is supported by the Novo Nordisk Foundation. iPSYCH data was stored and analysed at the Computerome HPC facility (http://www.computerome.dtu.dk/) and the authors are grateful for continuous support from the HPC team led by A. Syed of DTU Bioinformatics, Technical University of Denmark. The following grants provided support for this work: NIH grant nos. R00MH113823 (H.W.), R01GM104400 (W.K.T.), 1R01MH109912 (D.G.), 1R01MH110927 (D.G.) and 1R01MH094714 (D.G.). Australian National Health and Medical Research Council grant nos. 1113400, 1078901, 1087889 (N.R.W).

Author contributions

D.G. and T.W. conceived of and supervised the study. A.J.S., H.W., D.G. and T.W. designed the analysis plan. A.J.S., V.A., A.B and W.K.T. prepared the data. A.J.S. performed the GWAS, (partitioned) SNP-(co)heritability, fine mapping and replication analyses. H.W. performed the candidate gene identification and enrichment analyses. R.N., M.G. and N.R.W. provided interpretive support. O.D. contributed imputation software and protocols. M.R.C. contributed analytic support. D.M.H., M.B.-H., J.B.-G., M.G.P., E.A., C.B.P., B.M.N., M.J.D., M.N., O.M., A.D.B., P.B.M. and T.W. designed, implemented and/or oversaw the collection and generation of the iPSYCH data. A.J.S., H.W., D.G. and T.W. wrote the manuscript. All authors discussed the results and contributed to the revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0320-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

The iPSYCH study cohort and data. The iPSYCH cohort is described in detail elsewhere[25]. Briefly, iPSYCH samples from essentially the entire population of Denmark born between 1981 and 2005 ($n = 1{,}472{,}762$). In total, 87,764 individuals were selected. A total of 30,000 design controls were selected randomly from this birth cohort, without regard for psychiatric disorders, to create a representative population sample. The remaining 57,764 design cases were ascertained for indications of clinical diagnoses recorded in the Danish Civil[51], National Patient[52] and/or Psychiatric Central Research[53] Registers describing care received for ADHD, ANO, ASD, AFF, BIP or SCZ. Where available a dried blood spot was obtained from the Danish Neonatal Screening Biobank[54]. The use of this data is according the guidelines provided by the Danish Scientific Ethics Committee, the Danish Health Data Authority, the Danish data protection agency and the Danish Neonatal Screening Biobank Steering Committee. No statistical methods were used to pre-determine the sample size but our sample sizes are consistent with the state-of-the-field[14,55].

For this study, indications were from the Danish National Patient Register[52] (complete through 2012) and Psychiatric Central Research Register[53] (PCRR; complete through 2013) and coded according to criteria previously described[56]. Indications are International Classification of Disease[57,58] (ICD) codes representing the clinical diagnosis associated with an instance of care provided at one of many psychiatric facilities throughout Denmark. Records include all admissions to inpatient psychiatric facilities since 1969 and outpatient psychiatric care received in all psychiatric hospitals, psychiatric wards and emergency rooms since 1995 (ref. [53]). Care administered by primary care or private practice physicians is not recorded into national registers. Diagnoses before 1994 were associated with codes from the ICD 8th revision[58] and converted to equivalent ICD 10th revision[57] (ICD10) codes to match the majority of indications[56]. We define the following seven cases groups as having at least one indication with the corresponding ICD10 (or equivalent ICD8) codes: cross-diagnosis (XDX: F00-99), attention-deficit/hyperactivity disorder (ADHD: F90.0), affective disorders (AFF: F32-F39), anorexia (ANO: F50.0, F50.1), autism spectrum disorders (ASD: F84.0, F84.1, F84.5, F84.8, F84.9), bipolar disorder (BIP: F30-31), and schizophrenia (SCZ: F20). As an additional case group, we consider those patients with other psychiatric indications, exclusively (OTH: F00-F99 not listed previously, only). Case status can reflect the presence of a single or multiple indications and were not censored or integrated hierarchically. The counts, prevalence and co-occurrence of indications are shown in Supplementary Tables 1 and 2 and co-occurrence is depicted in Supplementary Fig. 1.

Initial genotyping was performed at the Broad Institute with amplified DNA extracted from dried blood spots and assayed on the Inifinium PsychChip v1.0 array[25]. In total, 78,050 subjects were successfully genotyped across 25 waves at approximately 550,000 SNPs. A subset of good-quality common SNPs ($n = 246{,}369$) were phased into haplotypes in a single batch using SHAPEIT3[59] and imputed in ten batches using Impute2 (ref. [60]) with reference haplotypes from the 1,000 genomes project phase 3 (ref. [61]). Imputed additive genotype dosages and best-guess genotypes were checked for imputation quality (INFO > 0.2), deviations from Hardy–Weinberg equilibrium (HWE; $P < 1 \times 10^{-6}$), association with genotyping wave ($P < 5 \times 10^{-8}$), association with imputation batch ($P < 5 \times 10^{-8}$; Supplementary Figs. 19–23) and differing imputation quality between subjects with and without psychiatric diagnoses ($P < 1 \times 10^{-6}$) as well as censored on MAF > 0.01). In total, 8,018,013 imputed dosages and best-guess genotypes were used for analysis.

Three sets of cohorts of unrelated subjects with homogenous genetic ancestry were created by sub-setting the design cohorts, one for our primary GWAS analyses (GWAS cohorts) and two for heritability by either LDSC regression analysis (LDSC cohorts) or by GREML SNP heritability analysis (GCTA cohorts). Genetic ancestry for all cohorts was characterized using principal components analysis using smartPCA implemented in the Eigensoft package v6.0.1 (refs. [62,63]). We performed two iterations of censoring, removing subjects outlying from joint distribution of the first ten principle components defined in the subset of iPSYCH with four grandparents recorded in the Danish civil register as born in Denmark ($n = 6{,}474$ outliers removed; Supplementary Fig. 24), re-computing principle components on the remaining subjects and censoring again according to the same criteria ($n = 689$ outliers removed; Supplementary Fig. 25). Censored individuals were aggregated into a fourth ancestry diverse cohort (Replication cohort). For the GWAS and LDSC cohorts kinship was estimated using KING v1.9 (ref. [64]) and individuals were censored to ensure no pair had closer than third degree kinship ($n = 4{,}988$ removed). For the GCTA cohorts, kinship was more strictly filtered such that no pair had GCTA-based estimate greater than 0.034, the absolute value of the minimum estimated kinship ($n = 22{,}223$ removed). When possible cases were retained and the control relative was censored. All subject genotypes were flagged for abnormal sample heterozygosity, high levels of missing genotypes (>1%), sex concordance and inconsistencies among duplicate samples and those failing one or more tests were excluded ($n = 364$). In total, 65,534 subjects were retained in the GWAS and LDSC cohorts, 43,311 in the GCTA cohorts and 7,163 in the Replication cohort. A more detailed quality control protocol is available in our consortium white paper posted with our GWAS summary statistics (https://ipsych.au.dk/downloads/).

For the GWAS cohort, control subjects were defined as the subset of the design controls with no indications of any mental disorders. This cohort was used for XDX LDSC heritability, XDX–eXDX genetic correlation, XDX GWAS, single-indication odds ratios in Fig. 2c–f, and LDSC-SEG analysis. The same control definition was applied to the Replication cohort for internal replication tests. For the LDSC and GCTA cohorts, the definition of control subjects was different for each indication. Control subjects were the subset of the design controls without that indication, only, and are expected to have other indications at the population prevalence, consistent with a representative sample of the population without the considered diagnosis. These cohorts were used for single-indication GCTA and LDSC heritability and genetic correlation analyses. Unless explicitly noted, cases cohorts include all individuals with a specified indication, including those with co-morbid diagnoses.

Statistical analyses. All statistical tests were two-sided, unless specifically noted and assumptions (for example, data distribution, homoscedasticity and so on) were not formally tested. Randomization and blinding procedures do not apply to our study design.

SNP heritability and genetic correlations. SNP heritability and genetic correlations were estimated in the GCTA cohorts with the GREML approach available in GCTA v1.25.2 (refs. [65–67]). Age, gender and ten principal components were included as fixed-effects covariates. Estimates were converted to the liability scale[67] according to estimates of lifetime risk take from Pedersen et al.[56] (Supplementary Table 2). Estimation of genetic correlation between indications was performed using bivariate GREML[66,68]. For each pair of phenotypes, subjects with both indications were excluded and controls were randomly and evenly split, creating two independent case-control groups. Splitting and estimation were repeated five times for each pair and the median values were retained.

Published GCTA SNP heritability estimates for ADHD, AFF, ASD, BIP and SCZ were taken from Lee et al.[14]. GREML estimates of SNP for ANO and XDX were unavailable. GWAS statistics for eXDX[55], eADHD[69], eAFF[70], eANO[71], eASD[72] and eSCZ[73] were downloaded from the Psychiatric Genomics Consortium (PGC) repository (http://www.med.unc.edu/pgc/results-and-downloads). Statistics for eBIP (ref. [75]) were downloaded from the NHGRI-EBI GWAS catalog[75]. Linkage disequilibrium score regression (LDSC v1.0.0)[76] was used to estimate SNP heritability for these published studies and for each single iPSYCH indication. Reference LD scores and protocol were provided by the authors (https://github.com/bulik/ldsc/wiki). Genetic correlations between iPSYCH indications and published GWAS were estimated with LDSC[15] using the authors' protocols. For LDSC regression heritability and genetic correlation, single-indication iPSYCH GWAS were performed in the LDSC cohort according to the analysis approach described below. To facilitate comparisons a typical population prevalence was used for each liability scale transformation (XDX = 0.35, ADHD = 0.05, AFF = 0.15, ANO = 0.01, ASD = 0.01, BIP = 0.01, SCZ = 0.01)[14,71], including re-scaling the iPSYCH GREML estimates.

Association testing. GWAS were performed using imputed additive genotype dosages and logistic regression implemented in PLINK v1.90 (ref. [77]). The XDX GWAS included all subjects in the GWAS cohort (46,008 cases, 19,526 controls). Inflation was assed via genomic inflation factor (λ_{GC})[78] and LDSC[76]. Age, gender and ten principal components were included as fixed effects covariates. Stratified false discovery rates[79] (sFDR) were estimated according to Story's q value[80] and computed independently for common (MAF ≥ 0.05) and uncommon SNPs (0.01 < MAF < 0.05). The suggestive SNP threshold of sFDR q value < 0.05 corresponds to a P value less than 1.02×10^{-5} for common SNPs and less than 4.71×10^{-7} for uncommon SNPs. Single-indication odds ratios in Fig. 2c–f and XDX GWAS excluding each single indication used in Fig. 4 were performed to provide context for the XDX results in the GWAS cohort. For the internal replication cohort (7,163 individuals, 4,481 cases), association tests used best-guess genotypes and linear mixed models implemented in GCTA[66] accounting for relatedness and heterogeneity in genetic background with genome-wide estimates of empirical kinship. Gender and age were included as fixed-effects covariates.

Fine mapping. Region-based loci associated with independent index SNPs were defined and refined iteratively. The most significant SNP was selected and PLINK v1.90[77] (https://www.cog-genomics.org/plink/1.9/) was used to estimate pairwise r^2 LD between the index SNP and all SNPs within five megabases. All SNPs with r^2 LD > 0.1 with the index SNP were considered index-associated SNPs. Locus bounds were determined by the physical positions of the furthest index-associated SNP upstream and downstream. The process was repeated until all suggestive SNPs were labeled as index or index-associated SNPs. Loci with overlapping index-associated SNP sets were merged. Conditional analysis was performed within each locus, including the imputed dosage for the most significant SNP as a covariate and re-computing within locus association statistics. Secondary suggestive SNPs were retained as independent index SNPs and the process was repeated until no SNPs within the locus showed suggestive association. Credible SNPs were defined for each locus using FineMap v1.1 (ref. [81]) with default parameters. For each locus, FineMap input SNPs had LD $r^2 > 0.6$ with the index SNP and an association

P value less than 0.001. Using the per SNP posterior probabilities and Bayes factors provided by FineMap, we define credible sets as the the smallest collection of SNPs in each locus with a total posterior probability for containing the causal SNPs of 95% (ref. [82]), supplemented with the index SNP and any individual SNPs with $\log_{10}$(Bayes Factors) > 2.

External statistical support. The 25 June 2018 NHGRI GWAS catalog[75] contains 69,102 autosomal single SNP associations with positions that could be mapped to the hg19 reference aggregated from 4,299 GWAS described in 2,918 publications. A total 589 GWAS were labeled by NHGRI as 'neurological disease', a broad category including psychiatric outcomes; 3,710 were not. Catalog SNPs were connected to XDX loci according to an objective hierarchy. First, catalog SNPs were checked if they were an index SNP. If not, they were checked if they were a credible SNP. Remaining catalog SNPs were checked for r^2 LD > 0.6 with a credible SNP using iPSYCH genotypes. If catalog SNPs were not present in the iPSYCH genotypes, we checked for an LD connection in the 1,000 genomes project phase 3 (ref. [61]). Effect directions were aligned to the same strand by allele codes for unambiguous SNPs (A–T/C–G) and by frequency when the MAF was less than 0.40 for strand ambiguous SNPs (A–T/T–A, C–G/G–C). Enrichment for connections to neurological GWAS was tested with a binomial proportion test, although this test may not be optimally specified due to overlap among cataloged studies.

To connect each locus to the full results from the seven eGWAS described previously, we followed a similar protocol. Priority was given to index SNPs genotyped in both studies, then credible SNPs, then the strongest proxy-credible LD pair with r^2 at least 0.6 in the iPSYCH data. Only strand unambiguous SNPs were considered (A–T/C–G). For concordance tests, we considered effects of the proxy SNP in both studies to ensure we used statistics for the same variant in the case where an index or credible SNP was not directly present in the eGWAS. Concordance was estimated via a linear model predicting the external *z*-score from the XDX *z*-score, including an intercept. The same concordance test was used for results from the internal replication association tests.

LDSC-SEG. LDSC regression for specifically expressed genes (LDSC-SEG)[27] tests for enrichments in per SNP heritability among sets of SNPs defined for plausibly tissue preferential biological signatures while controlling for tissue general effects. Three sets of pre-computed annotation files (LD scores) are provided by the LDSC-SEG authors (GTEx, DEPICT and Roadmap; https://data.broadinstitute.org/alkesgroup/LDSCORE/LDSC_SEG_ldscores/biorxiv). The GTEx LD scores represent 53 variant annotations capturing preferentially expressed genes defined from human tissue RNA sequencing[83]. The DEPICT LD scores represent 152 variant annotations capturing tissue preferential gene sets defined from an amalgamation of human, mouse and rat microarray experiments[84]. The Roadmap represent 396 variant annotations constructed from data produced by the Roadmap Epigenetics project[85], namely the narrow peaks defined by Roadmap for DNase hypersensitivity, H3K27ac, H3K4me3, H3K4me1, H3K9ac and/or H3K36me3 chromatin marks in 88 cell types or primary tissues. Analytic protocols were as provided by the LDSC-SEG developers documentation (https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses). *P* values less than 8.32×10^{-5} were declared significant to correct for testing 601 sets of LD scores and suggestive significance was determined by Story's *q* value[80] (FDR < 0.05 which corresponds to $P < 8.27\times10^{-4}$). Replication LDSC-SEG partitioning used the eSCZ[73] and eXDX statistics. Internal replication enrichments used results from seven secondary GWAS, each one censoring all patients with a different indication.

Identifying candidate genes. For each locus, candidate genes were identified by functional connections between credible SNPs and plausible targets according to the union of three selection criteria as described previously[86]. First, genes containing credible SNPs that cause missense variation or nonsense mediated decay were selected (133 credible SNPs implicating 13 candidate genes). Second, genes with credible SNPs located in the promoter regions were selected (2 kilobases upstream from the transcription start site; 15 credible SNPs implicating 12 candidate genes). Finally, unannotated SNPs were assigned to genes on the basis of three-dimensional chromatin contacts defined by an interaction map for fetal brain (GEO accession number: GSE77565)[86]. Genes contacting regions containing credible SNPs were selected (252 implicating 262 candidate genes). In total, 400 credible SNPs were assigned to 281 candidate genes.

Candidate gene enrichment tests. Gene ontology enrichments were performed using GOElite[87] and ontologies from ENSMart[88] v77 against a background of all autosomal protein coding genes. Developmental expression trajectories for candidate genes were plotted using a published transcriptome atlas constructed from post-mortem cortices (GEO accession number: GSE25219)[89]. Expression values were log-transformed and centered using the mean expression values for all brain-expressed genes. Mean expression values for the 281 candidate genes were plotted across prenatal (6–37 weeks post-conception) and postnatal (4 months to 42 years) developmental stages. Developing neural cell-type enrichments were estimated using expression profiles of single cells taken from fetal cortical laminae[90]. Cell-type specific genes were selected according to a significant (FDR < 0.05) Pearson correlation between the gene and an idealized cluster marker for each cell-type, following the approach described in the data generation report[90]. Candidate gene enrichment for each set of specifically expressed genes was estimated by logistic regression and adjusting for gene length.

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

Code availability

Code and scripts available by request from authors.

Data availability

In accordance with the consent structure of iPSYCH and Danish law, individual level genotype and phenotype data are not able to be shared publicly. Cross-disorder (XDX) GWAS summary statistics are available for download (https://ipsych.au.dk/downloads/). Summary statistics from secondary GWAS of single disorders are available upon request from the corresponding author. BrainSpan RNA data are available in the GEO with the accession code GSE25219. Fetal Brain Hi-C data are available in the GEO with the accession code GSE77565.

References

51. Pedersen, C. B. The Danish civil registration system. *Scand. J. Public Health* **39**, 22–25 (2011). Suppl.
52. Lynge, E., Sandegaard, J. L. & Rebolj, M. The Danish national patient register. *Scand. J. Public Health* **39**, 30–33 (2011). Suppl.
53. Mors, O., Perto, G. P. & Mortensen, P. B. The Danish psychiatric central research register. *Scand. J. Public Health* **39**, 54–57 (2011). Suppl.
54. Nørgaard-Pedersen, B. & Hougaard, D. M. Storage policies and use of the Danish newborn screening biobank. *J. Inherit. Metab. Dis.* **30**, 530–536 (2007).
55. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
56. Pedersen, C. B. et al. A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. *JAMA Psychiatry* **71**, 573–581 (2014).
57. World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: diagnostic criteria for Research. (WHO: Geneva, 1993). http://www.who.int/iris/handle/10665/37108.
58. World Health Organization. *Klassifikation af sygdomme; Udvidet dansk-latinsk udgave af verdenssundhedsorganisationens internationale klassifikation af sygdomme. 8 revision, 1965* [Classification of diseases: Extended Danish-Latin version of the World Health Organization International Classification of Diseases, 8th revision, 1965] (Danish National Board of Health, Copenhagen, 1971).
59. O'Connell, J. et al. Haplotype estimation for biobank-scale data sets. *Nat. Genet.* **48**, 817–820 (2016).
60. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS. Genet.* **5**, e1000529 (2009).
61. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
62. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
63. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS. Genet.* **2**, e190 (2006).
64. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
65. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
66. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
67. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
68. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).
69. Neale, B. M. et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 884–897 (2010).
70. Ripke, S. et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
71. Duncan, L. et al. Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **174**, 850–858 (2017).
72. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism

spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism* **8**, 21 (2017).
73. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
74. Hou, L. et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Hum. Mol. Genet.* **25**, 3383–3394 (2016).
75. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**(D1), D896–D901 (2017).
76. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
77. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
78. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
79. Sun, L., Craiu, R. V., Paterson, A. D. & Bull, S. B. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genet. Epidemiol.* **30**, 519–530 (2006).
80. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. U.SA.* **100**, 9440–9445 (2003).
81. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
82. Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**, 497–508 (2014).
83. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
84. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
85. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
86. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
87. Zambon, A. C. et al. GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* **28**, 2209–2210 (2012).
88. Kasprzyk, A. et al. EnsMart: a generic system for fast and flexible access to biological data. *Genome Res.* **14**, 160–169 (2004).
89. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
90. Pollen, A. A. et al. Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).

nature research

Corresponding author(s): Thomas Werge

Reporting Summary

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

Statistical parameters

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

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	R v3.3.1 was used to organize subject records into appropriate input files.
Data analysis	R v3.3.1, python v2.7, Plink v1.90 beta3, GCTA v1.25.2, LDSC v1.0.0, FineMap v1.1, GOElite, Eigensoft 6.0.1, KING v1.9, SHAPEIT3, Impute2

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

Data

Policy information about availability of data

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

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

In accordance with the consent structure of iPSYCH and Danish law, individual level genotype and phenotype data are not able to be shared publicly. Cross-disorder (XDX) GWAS summary statistics are available for download (https://ipsych.au.dk/downloads/). Summary statistics from secondary GWAS of single disorders are

available upon request from the corresponding author.

Field-specific reporting

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

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

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

Behavioural & social sciences study design

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

Study description	A genome-wide associate study of lifetime psychiatric diagnoses as of 2012 in a 1981-2005 birth cohort .
Research sample	We used data from the iPSYCH case-cohort study: A random, representative sample of the Danish neonatal biobank for individuals born between 1981 and 2005, with a known mother, alive on their first birthday, who have not opted out of the Danish national register based research program and the complete population of individuals from the Danish population birth cohort meeting the same inclusion criteria and with a major psychiatric diagnosis in the national registers as of 2012. The case sample was chosen to be complete and the random population sample to be representative and of comparable size to the largest individual case cohort. Further details can be found in PMID: 28924187.
Sampling strategy	Data was ascertained according to a case-cohort design where the size of the cohort (random sample) was chosen to represent 2% of the broader population it was sampled from (population of Denmark born between 1981 and 2005). Cases were all cases in the population at the time of ascertainment. These samples are state-of-the-field for single cohort GWAS studies, representing the largest of its kind, and consistent with previous reports adequately powered for discovery.
Data collection	We used data from the iPSYCH case-cohort study: DNA was extracted from dried neonatal blood spots and amplified before genotyping on the Infinium PsychChip v1.0. Blood was collected between 4-7 days after birth and stored at -20 C until the time of ascertainment. Psychiatric diagnoses were aggregated from national registers. Demographic and social variables were aggregated from national civil registers. Further details can be found in PMID: 28924187.
Timing	All data was initially collected in 2012 and psychiatric diagnoses were later updated, complete through 2014.
Data exclusions	Among the 78,050 samples with genotype data available in the iPSYCH case-cohort study, 5,353 were excluded according to genotype and imputation quality control procedures described in great detail in Supplementary Note 2. These data exclusion criteria were determined before the study was designed or conducted.
Non-participation	All subjects in the iPSYCH case-cohort data resource were initially included. This sample was drawn only from those meeting consent requirements.
Randomization	When control samples were split into random sub-cohorts for analysis the sample() function in R was used to select subsamples of subject unique IDs.

Reporting for specific materials, systems and methods

Materials & experimental systems

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

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	Patients were recruited according to consent by non-opt out from a national research program in which very few individuals opt out. Our control sample was drawn nearly randomly from the national biobank making the potential for ascertainment effect negligible. Our case cohort is the complete population as of 2012, which is young for psychiatric disorders meaning there may be some enrichment for early onset and some yet-to-convert cases in the controls for anyone disorder. These are unlikely to affect most of our results.