nature neuroscience

Article

https://doi.org/10.1038/s41593-025-02147-2

Trans-ancestry genome-wide analyses of bipolar disorder in East Asian and European populations improve genetic discovery

Received: 25 June 2024

Accepted: 21 October 2025

Published online: 25 November 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Genome-wide association studies (GWASs) of bipolar disorder (BD) have predominantly included individuals of European (EUR) ancestry, underrep resenting non-EUR populations and limiting insight into disease mechanisms. Here we performed a GWAS of BD in Han Chinese individuals (5,164 cases and 13,460 controls) and conducted comparative and integrative analyses with independent East Asian (EAS, 4,479 cases and 75,725 controls) and EUR (59,287 cases and 781,022 controls) cohorts from the PGC4 GWAS. Our GWAS in EAS ancestry identified two genome-wide significant risk loci, including variants at the major histocomp atibility complex (MHC) class II region. Incorporating EAS data into trans-ancestry GWAS revealed 93 significant loci (23 novel). Heritability enrichment analyses implicated a variety of neuronal cell types. Multidimensional post-GWAS prioritization identified 39 high-confidence risk genes, of which 15 were differentially expressed in the brains of patients with BD, 12 modulated BD-relevant behaviors in mice and 18 are pharmacologically tractable. This work advances understanding of the biological underpinnings of BD and provides direction for future research in underrepresented populations.

Bipolar disorder (BD) is characterized by recurrent episodes of mania or hypomania and depression. The global prevalence of BD is approximately 2%[1], and the approximate lifetime risk of BD in relatives of a bipolar proband is 5–10% in first-degree relatives and 40–70% among monozygotic co-twins[2], suggesting a strong genetic predisposition.

Recent genome-wide association studies (GWASs) have uncovered hundreds of BD susceptibility loci through meta-analyses of predominantly European (EUR) ancestry cohorts[3–5]. For example, PGC4 GWAS included 67,948 cases (ascertained from clinical and community samples) and 867,710 controls[3], and the effective sample size ($N_{eff} = 4 \times N_{case} \times N_{control}/(N_{case} + N_{control})$) was 252,054. Among these samples, 87.4% derived from EUR ancestry, 3.6% from African, 6.7% from East Asian (EAS) and 1.2% from Latino, highlighting a pronounced ancestral disparity. Although the core biological mechanisms underlying BD are presumed to be conserved across ancestries, variations in allele frequencies and linkage disequilibrium (LD) patterns across continental populations could influence the detectability and magnitude of risk variants[6,7]. Multi-ancestry GWASs improve fine-mapping resolution by leveraging heterogeneous LD structures, facilitating the identification of robust disease-associated variants[8–10]. These considerations underscore the critical need to increase ancestral diversity in BD GWASs.

To address the underrepresentation of Han Chinese in BD genetic studies, we previously conducted a GWAS of 1,822 cases and 4,650 controls[11]. In the present study, we substantially expand this effort by including more Han Chinese individuals, yielding, to our knowledge, the largest-to-date GWAS cohort in this population (5,164 BD cases and 13,460 controls). We then perform comparative and integrative analyses, incorporating independent EAS and EUR cohorts from PGC4 GWAS[3]. Through these analyses, we identify novel risk loci, cell type

e-mail: yirufang@aliyun.com; zhufeng1982@xjtu.edu.cn; xiaoxiao2@mail.kiz.ac.cn; limingkiz@mail.kiz.ac.cn

https://doi.org/10.1038/s41593-025-02147-2

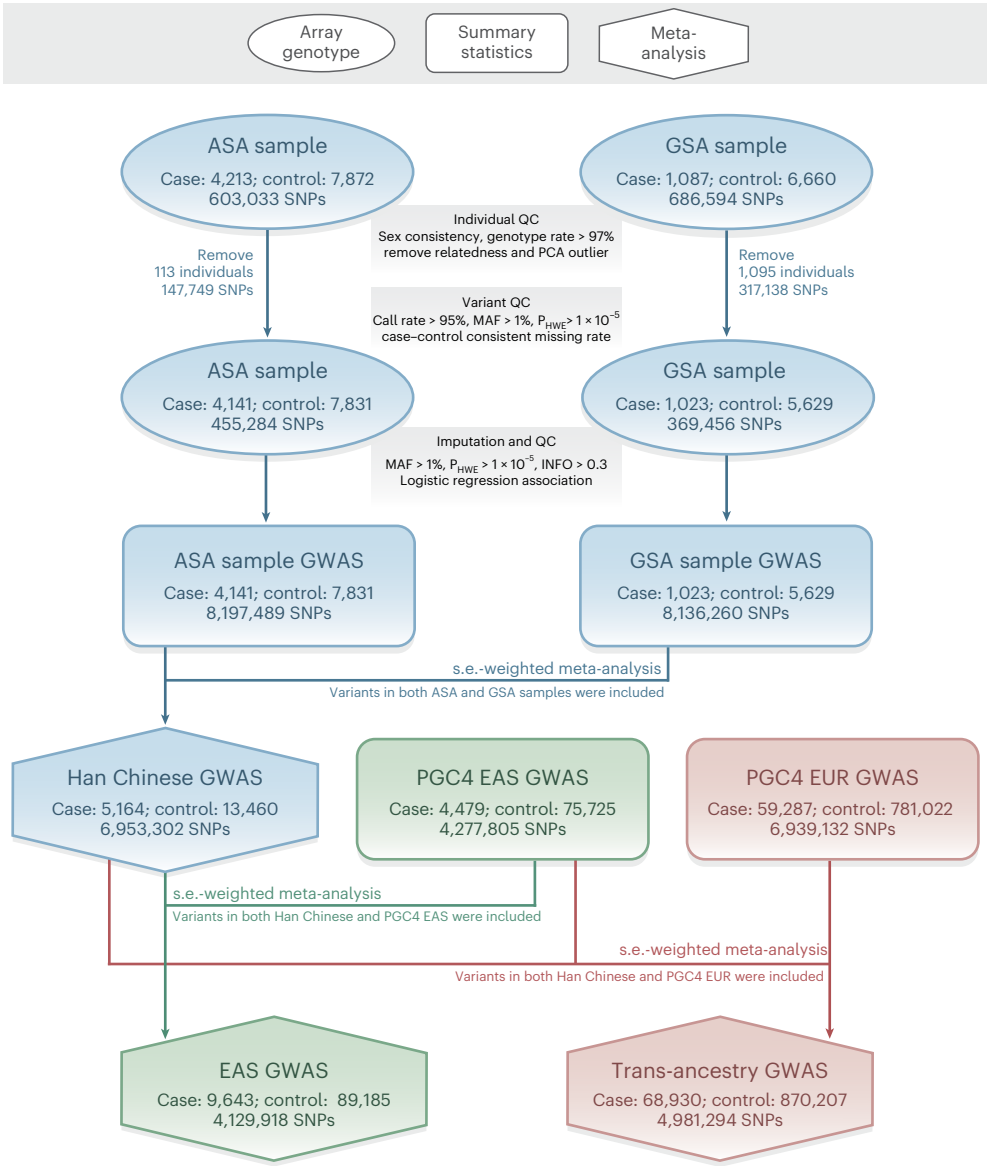

Fig. 1 | Schematic overview of data sources and analytic steps. Our study included two independent Han Chinese GWAS cohorts (ASA sample and GSA sample), an EAS GWAS summary statistic (PGC4 EAS GWAS) and an EUR GWAS summary statistic (PGC4 EUR GWAS). Quality control (QC) process and association analyses were first performed separately in two Han Chinese GWAS cohorts, with results meta-analyzed to generate the Han Chinese GWAS. The Han Chinese GWAS was then meta-analyzed with PGC4 EAS GWAS into the EAS GWAS. Finally, the Han Chinese GWAS, the PGC4 EAS GWAS as well as the PGC4 EUR GWAS were further incorporated into the trans-ancestry GWAS. PCA, principal component analysis.

enrichments and high-confidence risk genes, providing critical insights into the genetic architecture of BD.

Results

BD GWAS in Han Chinese populations

An overview of the study design and analytical procedures is shown in Fig. 1. In brief, Han Chinese individuals were genotyped using either the Illumina ASA chip (hereafter referred to as the 'ASA sample') or the GSA chip ('GSA sample'), followed by stringent variant-level and individual-level quality control as well as imputation in each sample respectively using the same pipeline (Methods), and 6,953,302 autosomal biallelic single-nucleotide polymorphisms (SNPs) common to both samples were retained for association testing and meta-analysis (referred to as the 'Han Chinese GWAS'). A final set of 5,164 clinical BD cases (mean age ± s.d. = 33.51 ± 12.72 years; 45.43% male) and 13,460 controls (27.62 ± 6.16 years; 55.46% male) was included in the Han Chinese GWAS (Table 1), and principal component analysis confirmed negligible population stratification (Supplementary Fig. 1). The GWAS showed minimal inflation ($\lambda = 1.10$; $\lambda_{1,000} = 1.01$), and LD score regression (LDSC) analysis revealed no evidence of residual confounding (intercept = 1.00 ± 0.01; attenuation ratio = 0.03 ± 0.06)[12]. The SNP-based heritability (h^2_{SNP}) was estimated to be 0.24 ± 0.024 on the liability scale assuming a 2% BD prevalence. Manhattan and quantile–quantile (Q–Q) plots are presented in Fig. 2 (top) and Supplementary Fig. 2, respectively.

At a threshold of suggestive significance ($P \leq 5.00 \times 10^{-6}$), we identified 696 SNPs across 24 loci (Supplementary Table 1), with three reaching genome-wide significance (GWS; $P < 5.00 \times 10^{-8}$): 6p21.32 within the major histocompatibility complex (MHC) region (rs9271346, $P = 4.68 \times 10^{-10}$, odds ratio (OR) = 1.21, minor allele frequency (MAF) = 0.21), 8q21.12 (rs76178353, $P = 2.21 \times 10^{-8}$, OR = 1.41, MAF = 0.04) and 20p13 (rs2422679, $P = 4.88 \times 10^{-8}$, OR = 1.42, MAF = 0.05) (Supplementary Fig. 3). However, the 8q21.12 and 20p13 loci each contained only one GWS SNP, both of which presented effect sizes (ORs ≈ 1.4) substantially larger than those typically reported in previous

https://doi.org/10.1038/s41593-025-02147-2

Table 1 | Demographics of BD cohorts

Ancestry	EAS cohorts						EUR cohorts	
Sample name	Han Chinese ASA		Han Chinese GSA		PGC4 EAS GWAS		PGC4 EUR GWAS	
	Case	Control	Case	Control	Case	Control	Case	Control
Total sample size	4,141	7,831	1,023	5,629	4,479	75,725	59,287	781,022
Sex (male%)	44.83%	55.22%	47.90%	56.22%	-		-	
Age (mean±s.d.)	11–86 (32.88±10.27)	11–70 (28.15±6.47)	11–88 (34.00±14.64)	12–68 (27.19±6.04)	-		-	
Patients with BD-I (%)	65.51%	-	68.52%	-	-		42.53%	-
Patients with family history (%)	22.52%	-	30.60%	-	-		-	
Geographical distribution	23.6% Western, 24.2% Middle, 34.2% Southeast, 18.01% Northeast of China	26.45% Western, 15.89% Middle, 32.13% Southeast, 25.41% Northeast of China	43.21% Western, 13.29% Middle, 43.5% Southeast of China	18.12% Western, 16.72% Middle, 32.44% Southeast, 32.72% Northeast of China	3 clinical cohorts (see details in PGC4 GWAS)		51 clinical cohorts and 17 community cohorts (see details in PGC4 GWAS)	
Genotyping platform	Illumina ASA chip		Illumina GSA chip		OMEX, KoreanChip		14 types of chips (see details in PGC4 GWAS)	
Data type	Raw genotype array data				Summary statistics			
Ascertainment	Clinical (DSM-IV)				Clinical (DSM-IV-TR)		Clinical + community (DSM-IV, ICD-9 or ICD-10; semi-structured or structured interviews, clinician-administered checklists, medical record review, registries or questionnaire data)	

DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision; ICD-9, International Classification of Diseases, Ninth Revision; ICD-10, International Classification of Diseases, Tenth Revision.

BD GWASs[3,4]. Despite both SNPs passing rigorous post-imputation quality control, their low MAFs and sparse local LD raise concerns about potential imputation artifacts or effect size inflation[13]. We thus advise cautious interpretation of these associations until independent replication is achieved.

BD GWAS meta-analysis and polygenic prediction in EAS populations

We evaluated the genetic architecture similarities between Han Chinese GWAS (5,164 cases and 13,460 controls) and PGC4 EAS GWAS (4,479 cases and 75,725 controls)[3] through LDSC analysis[14], and the two datasets exhibited a strong genetic correlation ($r_g = 0.63$, s.e. = 0.11, $P = 1.91 \times 10^{-8}$). To analyze the polygenic architecture of BD, we next performed polygenic risk score (PRS) analysis using the PGC4 EAS GWAS as the training dataset and the Han Chinese cohort as the target dataset. The training dataset significantly predicted BD status in the target dataset, with SNPs explaining over 0.8% of the variance in BD (liability-scaled Nagelkerke pseudo R^2) assuming a 1–5% prevalence ($P = 9.04 \times 10^{-24}$; Supplementary Table 2).

Meta-analysis combining both GWAS datasets (hereafter referred to as the 'EAS GWAS', including 9,643 cases and 89,185 controls and 4,129,918 autosomal biallelic SNPs) yielded well-calibrated results ($\lambda = 1.13$; $\lambda_{1,000} = 1.01$) with minimal stratification (LDSC intercept = 1.01 ± 0.01; attenuation ratio = 0.08 ± 0.06) and an h^2_{SNP} of 0.13 ± 0.013 assuming a 2% BD prevalence. Manhattan and Q–Q plots for the EAS GWAS are presented in Fig. 2 (top) and Supplementary Fig. 4. The EAS GWAS identified 205 SNPs spanning 22 loci showing suggestive significance (Supplementary Table 3) and two loci reaching the GWS level, including 7p22.3 (rs61409925, $P = 4.32 \times 10^{-9}$, OR = 1.19; Supplementary Fig. 5a) encompassing the known BD risk gene *MAD1L1* (refs. 4,15) and 6p21.32 within the MHC region (rs4713555, $P = 8.67 \times 10^{-12}$, OR = 1.16; Supplementary Fig. 5b). Two SNPs within 6p21.32, rs9271346 and rs4713555, were both significant in the Han Chinese BD GWAS (rs9271346, $P = 4.68 \times 10^{-10}$, OR = 1.21; rs4713555, $P = 5.20 \times 10^{-9}$, OR = 1.17) and were in substantial LD ($r^2 = 0.13$; $D' = 0.98$). Although rs9271346 was not covered in the PGC4 EAS GWAS, rs4713555 showed nominal significance ($P = 3.91 \times 10^{-4}$, OR = 1.15) and similar OR compared to the Han Chinese GWAS. Therefore, both GWASs supported the correlation between the 6p21.32 locus and BD, although the lead SNPs were different, likely due to array content variation.

Because the 6p21.32 lead variant is located at the MHC class II region (between *HLA-DRB1* and *HLA-DQA1*), we inferred whether this association arose from human leukocyte antigen (HLA) alleles. We imputed HLA alleles on the basis of the HAN-MHC reference panel of Han Chinese individuals ($N = 20{,}635$)[16] using SNP2HLA[17] and found that the GWS SNPs at this locus were correlated with known HLA alleles, such as HLA_DRB1_07 (highest $r^2 = 0.91$; Supplementary Table 4). Although we were unable to elucidate the LD relationship between the EAS GWS SNPs and *C4* alleles due to the lack of reference panel for structural forms of the *C4A*/*C4B* locus in EAS, we observed that the GWS region at 6p21.32 in the EAS GWAS was approximately 480 kilobases (kb) away from the *C4* genes. In addition, these EAS GWS SNPs are located far (approximately 6.11 megabases (Mb)) from the reported BD risk variants at MHC in EUR[3], and the EAS GWS MHC locus was not significant in the PGC4 EUR GWAS (Supplementary Fig. 6).

Genetic correlation and polygenic overlap between BD and other traits in EAS

Using LDSC analysis[14], we quantified genetic correlations (r_g) between BD and 39 phenotypes in EAS. We used publicly available GWAS summary statistics of psychiatric, cognitive, hematological, anthropometric and metabolic traits as well as non-neurological diseases (Supplementary Table 5). Strongest correlations emerged regarding schizophrenia ($r_g = 0.65 \pm 0.05$, $P = 1.80 \times 10^{-32}$) and depression ($r_g = 0.60 \pm 0.13$, $P = 3.58 \times 10^{-6}$), and significant positive associations were also observed for coronary artery disease (CAD; $r_g = 0.16 \pm 0.05$, $P = 2.00 \times 10^{-3}$), educational attainment ($r_g = 0.14 \pm 0.06$, $P = 0.02$) and blood aspartate aminotransferase (AST) levels ($r_g = 0.12 \pm 0.05$, $P = 0.01$)

https://doi.org/10.1038/s41593-025-02147-2

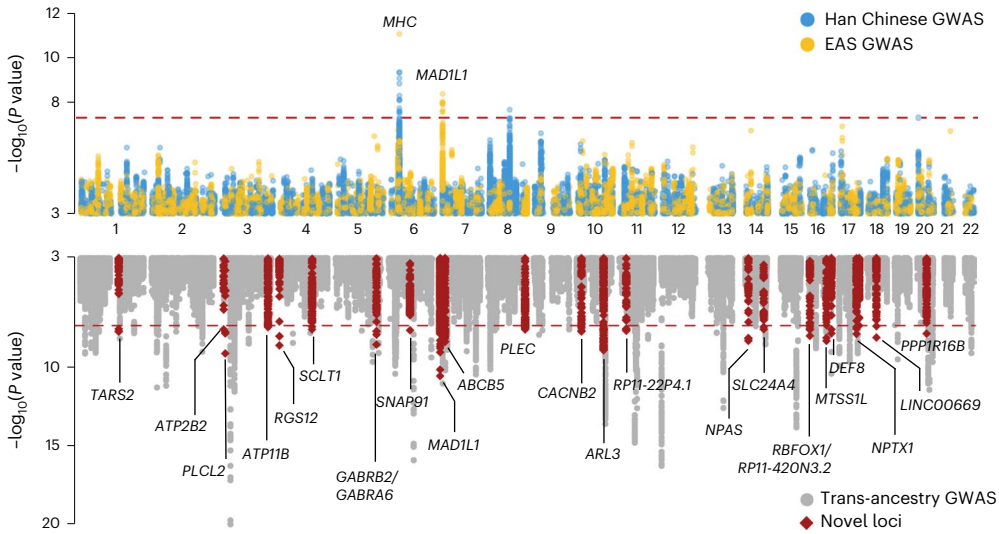

Fig. 2 | Miami plot of Han Chinese GWAS, EAS GWAS and trans-ancestry GWAS. Top: Manhattan plot presents both Han Chinese GWAS (5,164 cases and 13,460 controls) with one GWS locus and EAS GWAS (9,643 cases and 89,185 controls) with two GWS loci. Genes or regions where the lead SNPs located at or close to are labeled. Bottom: Manhattan plot for trans-ancestry GWAS (68,930 cases and 870,207 controls) with 93 GWS loci. Twenty-three loci in trans-ancestry GWAS that have not been identified in PGC4 EUR GWAS are highlighted by red diamond-shaped points. Genes where the lead SNPs located at or close to are labeled. The *x* axes show the genomic position (chromosomes 1–22 according to GRCh37). The *y* axes show the $-\log_{10}P$ values for association with BD. *P* values are two-sided and generated from inverse variance-weighted meta-analyses. The GWS threshold after Bonferroni correction ($P < 5.00 \times 10^{-8} = 0.05/1.00 \times 10^{6}$) is represented as red dashed lines.

(Supplementary Fig. 7a). Three phenotypes exhibiting negative correlations with BD were weight ($r_g = -0.07 \pm 0.04$, $P = 0.04$), blood triglyceride level ($r_g = -0.11 \pm 0.05$, $P = 0.02$) and blood creatinine level ($r_g = -0.11 \pm 0.05$, $P = 0.03$) (Supplementary Fig. 7a). Most of the above genetic correlations observed in EAS were in line with the results in EUR (Supplementary Table 5)[3,4].

We then used MiXeR[18] to evaluate the polygenic overlap between BD and those genetically correlated traits. Univariate estimates of polygenicity found that approximately 6,370 variants influenced BD, and approximately 9,010 affected schizophrenia, 2,100 for depression, 750 for CAD, 12,290 for educational attainment, 680 for blood AST level, 6,380 for weight, 160 for blood triglyceride level and 1,600 for blood creatinine level. Bivariate Gaussian mixture models were then applied to estimate the number and proportion of shared variants between BD and those traits and revealed extensive polygenic overlap (Supplementary Fig. 7b and Supplementary Table 6). In brief, 95.8% (6,100 out of 6,370) and 93.8% (5,980 out of 6,370) BD-associated variants co-influenced educational attainment and schizophrenia, respectively, and 87.6% (1,840 out of 2,100) depression-associated variants, 81.4% (610 out of 750) CAD-associated variants, 73.3% (500 out of 680) AST-associated variants and 87.2% (140 out of 160) triglyceride-associated variants were shared with BD. Directional effect concordance between BD and those traits was high for shared variants involving schizophrenia (84.6%), depression (98.4%) and CAD (83.8%). However, only 55.6% of the shared variants between BD and educational attainment exhibited concordant effects, mirroring findings from the EUR population (55% shared variants showed concordant effects)[4].

Similarities and differences in BD-related genetic effects across ancestries

Leveraging EAS GWAS (9,643 cases and 89,185 controls) and PGC4 EUR GWAS (59,287 cases and 781,022 controls), we observed substantial genetic correlations between EAS and EUR using Popcorn (version 1.0)[19], with a significant genetic impact correlation ($\rho_{gi} = 0.85$, s.e. = 0.07, 95% confidence interval: 0.72–0.98, $P_{less\_than\_one} = 0.02$ and $P_{larger\_than\_zero} = 1.28 \times 10^{-37}$) and a genetic effect correlation ($\rho_{ge} = 0.86$, s.e. = 0.07, 95% confidence interval: 0.72–0.99, $P_{less\_than\_one} = 0.04$ and $P_{larger\_than\_zero} = 3.77 \times 10^{-36}$).

We evaluated how much variation in EAS risk could be explained by EUR GWAS through PRS analysis. Because the genotype data of the PGC4 EAS GWAS were not available, we used the Han Chinese GWAS as the target dataset and the PGC4 EUR GWAS as the training dataset. At an assumed population prevalence of 1–5%, the PRS explained 2.1–3.3% ($P = 5.10 \times 10^{-89}$) of BD risk on the liability scale (Supplementary Table 7), a marked improvement over previous estimates ($R^2 = 1.27\%$ at 1% prevalence)[11]. These findings further support substantial cross-population consistency in the genetic architecture of BD.

To assess the trans-ancestral portability of BD risk loci, we systematically compared genetic effects between EAS and EUR. Among the 88 loci (lead SNPs or their proxies) identified in the PGC4 EUR GWAS, we compared their directions of allelic effects in BD cases and controls from EAS GWAS and found that 75 loci (23 achieving nominal significance, $P < 0.05$) exhibited consistent directions (Supplementary Table 8). Effect size correlation analysis of the 88 loci revealed a strong positive association between EAS and EUR ORs (Pearson $sr^2 = 0.6$, $P = 1.11 \times 10^{-9}$; Supplementary Fig. 8). To evaluate whether population differentiation influences risk allele effects of the 88 loci across different populations, we quantified genetic divergence using the fixation index (F_{ST}) and assessed heterogeneity of effect size using Cochran's *Q* test (measured by $-\log_{10}(P)$). No significant correlation emerged between F_{ST} and effect size heterogeneities (Supplementary Fig. 9), suggesting minimal impact of ancestral divergence on risk locus effects.

Population diversity improves the identification of BD-associated loci

We performed a trans-ancestry meta-analysis by combining EAS GWAS with PGC4 EUR GWAS. After quality control, 4,981,294 autosomal biallelic SNPs common to both populations were retained. The trans-ancestry GWAS revealed minimal inflation ($\lambda = 1.51$, $\lambda_{1,000} = 1.00$), with LDSC parameters (intercept = 1.05 ± 0.01, attenuation ratio = 0.07 ± 0.01) confirming polygenic architecture rather than population stratification. Manhattan and Q–Q plots of the trans-ancestry GWAS are shown in Fig. 2 (bottom) and Supplementary Fig. 10. At the

https://doi.org/10.1038/s41593-025-02147-2

Table 2 | Novel BD risk loci in trans-ancestry GWAS

CHR	BP	Lead SNP	Locus name	A1/A2	Meta-analysis		PGC4 EUR GWAS			EAS GWAS		
					P value	OR	*P* value	OR	Frequency	*P* value	OR	Frequency
1	150466925	rs12116970	*TARS2*	T/C	2.02×10^{-8}	0.95	1.06×10^{-7}	0.95	0.22	0.08	0.95	0.07
3	10525468	rs79743019	*ATP2B2*	T/C	2.82×10^{-8}	0.95	2.57×10^{-7}	0.96	0.38	9.26×10^{-3}	0.90	0.04
3	16851665	rs35992988	*PLCL2*	T/C	8.29×10^{-10}	0.94	6.45×10^{-7}	0.94	0.13	2.58×10^{-4}	0.93	0.29
3	182615560	rs59108404	*ATP11B*	G/A	3.92×10^{-8}	0.96	3.73×10^{-7}	0.96	0.67	0.04	0.96	0.36
4	3401678	rs73193388	*RGS12*	T/C	2.71×10^{-9}	1.06	7.78×10^{-7}	1.06	0.14	6.23×10^{-4}	1.08	0.14
4	129868556	rs66921209	*SCLT1*	C/T	2.81×10^{-8}	0.95	2.78×10^{-7}	0.95	0.17	0.03	0.96	0.22
5	160975332	rs3816596	*GABRB2:GABRA6*	T/C	3.06×10^{-9}	0.96	3.48×10^{-6}	0.96	0.37	0.09	0.94	0.40
6	84409255	rs217286	*SNAP91*	G/A	2.50×10^{-8}	0.96	6.01×10^{-6}	0.96	0.42	3.22×10^{-4}	0.94	0.43
7	1900208	rs4644136	*MAD1L1*	C/T	3.06×10^{-11}	1.05	5.57×10^{-8}	1.04	0.60	2.99×10^{-5}	1.08	0.45
7	20808728	rs4721948	*ABCB5*	C/T	4.51×10^{-9}	0.96	2.75×10^{-7}	0.96	0.37	1.98×10^{-3}	0.93	0.21
8	144993377	rs6992333	*PLEC*	G/A	2.81×10^{-8}	1.04	4.26×10^{-7}	1.04	0.43	0.02	1.06	0.15
10	18742254	rs4748468	*CACNB2*	A/G	2.07×10^{-8}	0.94	1.69×10^{-7}	0.94	0.11	0.04	0.94	0.08
10	104436641	rs8354	*ARL3*	T/C	1.33×10^{-9}	0.94	1.56×10^{-6}	0.94	0.09	2.04×10^{-4}	0.94	0.42
11	28614490	rs2585813	*RP11-22P4.1*	A/T	2.30×10^{-8}	1.04	5.60×10^{-7}	1.04	0.62	0.01	1.05	0.29
14	33391029	rs35985675	*NPAS3*	A/T	5.26×10^{-9}	0.95	3.59×10^{-7}	0.95	0.19	1.82×10^{-3}	0.92	0.29
14	92795912	rs4904871	*SLC24A4*	A/G	2.91×10^{-8}	0.96	1.86×10^{-7}	0.96	0.45	0.04	0.96	0.44
16	6345162	rs113039090	*RBFOX1*	G/A	1.09×10^{-8}	1.05	4.01×10^{-7}	1.05	0.23	5.74×10^{-3}	1.07	0.17
16	70703776	rs6499334	*MTSS1L*	T/C	5.26×10^{-9}	1.05	2.00×10^{-7}	1.04	0.64	4.91×10^{-3}	1.07	0.42
16	90020861	rs7195043	*DEF8*	C/T	1.58×10^{-8}	1.04	2.61×10^{-6}	1.04	0.54	3.78×10^{-5}	1.13	0.20
17	68358571	rs3848450	*CTD-2378E21.1*	A/C	1.45×10^{-8}	1.04	5.41×10^{-6}	1.04	0.29	1.89×10^{-4}	1.07	0.30
17	78448640	rs12600720	*NPTX1*	G/C	4.93×10^{-8}	1.05	2.11×10^{-6}	1.04	0.32	3.60×10^{-3}	1.08	0.36
18	36808870	rs11082129	*LINC00669*	C/T	8.67×10^{-9}	1.04	8.77×10^{-7}	1.04	0.60	1.77×10^{-3}	1.06	0.28
20	37426336	rs6071522	*PPP1R16B*	G/A	1.52×10^{-8}	0.96	1.62×10^{-7}	0.96	0.45	0.03	0.96	0.41

P values for meta-analysis and EAS GWAS are two-sided and based on inverse variance-weighted meta-analyses. CHR, chromosome; BP, base-pair position (GRCh37).

GWS threshold, we identified 2,714 significant SNPs clustered into 93 independent loci (EUR LD-based clumping: $r^2 < 0.1$ within 500 kb; Supplementary Table 9 and Supplementary Fig. 11), and there were 23 novel loci compared to the PGC4 EUR GWAS findings (Table 2). Among the 93 loci, 93.5% (87/93) presented consistent directions of effect between EAS and EUR, which further supports the trans-ancestral consistency of BD genetic architecture.

Genomic region, tissue and cell type enrichment analyses of BD-associated loci

Our analyses revealed significant enrichment of BD heritability in genomic regions with diverse functional characteristics, including chromatin accessibility states, regulatory elements and human accelerated regions (HARs) ($P_{\text{Bonferroni}} < 0.05$; Supplementary Table 10 and Supplementary Fig. 12a). Promoter and proximal enhancer elements from ENCODE candidate *cis*-regulatory elements (cCREs) showed particularly strong enrichment (promoter: OR = 6.87; proximal enhancer: OR = 8.57), highlighting the importance of regulatory elements in BD pathogenesis. HARs were similarly enriched, implicating evolutionarily divergent loci in behavioral processes potentially disrupted in BD and supporting emerging evidence of their involvement in neuropsychiatric and neurodevelopmental disorders[20]. Transcription factor binding site (TFBS) analysis revealed 40 significantly enriched transcription factors among the 137 tested candidates (Supplementary Table 10 and Supplementary Fig. 12b), with P53 demonstrating the strongest association (OR = 5.05), followed by KLF4 and GATA1, suggesting their potential regulatory roles in BD-related gene networks.

Genotype-Tissue Expression (GTEx)[21]-based tissue enrichment analysis revealed that BD association signals were most abundant in brain tissues, particularly the cerebellum and frontal cortex (Supplementary Table 11 and Supplementary Fig. 13a). We then performed a cell type enrichment analysis using datasets from two BRAIN Cell Census (BRAIN Initiative Cell Census Network (BICCN)) studies and found significant enrichment of BD heritability in excitatory and inhibitory cortical neurons (Supplementary Fig. 13b)[22]. In particular, multiple types of inhibitory neurons (In_CCK, In_INT, In_CALB2 and In_PV) showed maximal enrichment during prenatal development, whereas those of excitatory neurons (EX_L5_6_IT, EX_L4, EX_L2_3 and EX_inter) became more prominent postnatally (Fig. 3a and Supplementary Table 12). We also used single-cell expression data of 3.369 million nuclei from several subcortical brain regions, including the basal ganglia caudate, hippocampus, amygdala and midbrain substantia nigra (Fig. 3b and Supplementary Table 13)[23], and observed significant enrichment in caudal ganglionic eminence (CGE)-derived interneurons, especially in the hippocampus ($P = 7.11\times10^{-7}$ in the CA1 CA3 region; $P = 6.01\times10^{-10}$ in the dentate gyrus CA4 region) and LAMP5_LHX6_and_Chandelier neurons ($P = 2.10\times10^{-7}$ in the hippocampus CA1 CA3 region; $P = 3.00\times10^{-6}$ in the amygdala). We also observed enrichment of BD associations in some region-specific neurons, such as medium spiny neurons (MSNs) and eccentric medium spiny neurons (eMSNs) in the amygdala ($P = 3.92\times10^{-9}$ for MSNs; $P = 5.73\times10^{-7}$ for eMSNs) and caudate ($P = 3.67\times10^{-8}$ for eMSNs), midbrain-derived inhibitory neurons ($P = 8.64\times10^{-13}$), amygdala excitatory neurons ($P = 1.22\times10^{-6}$) and hippocampal dentate gyrus cells ($P = 1.19\times10^{-6}$).

Prioritization of BD risk variants and genes

We performed functional fine-mapping via Polyfun + SuSiE[24] on the 92 GWS loci identified in the trans-ancestry GWAS (excluding MHC), which

https://doi.org/10.1038/s41593-025-02147-2

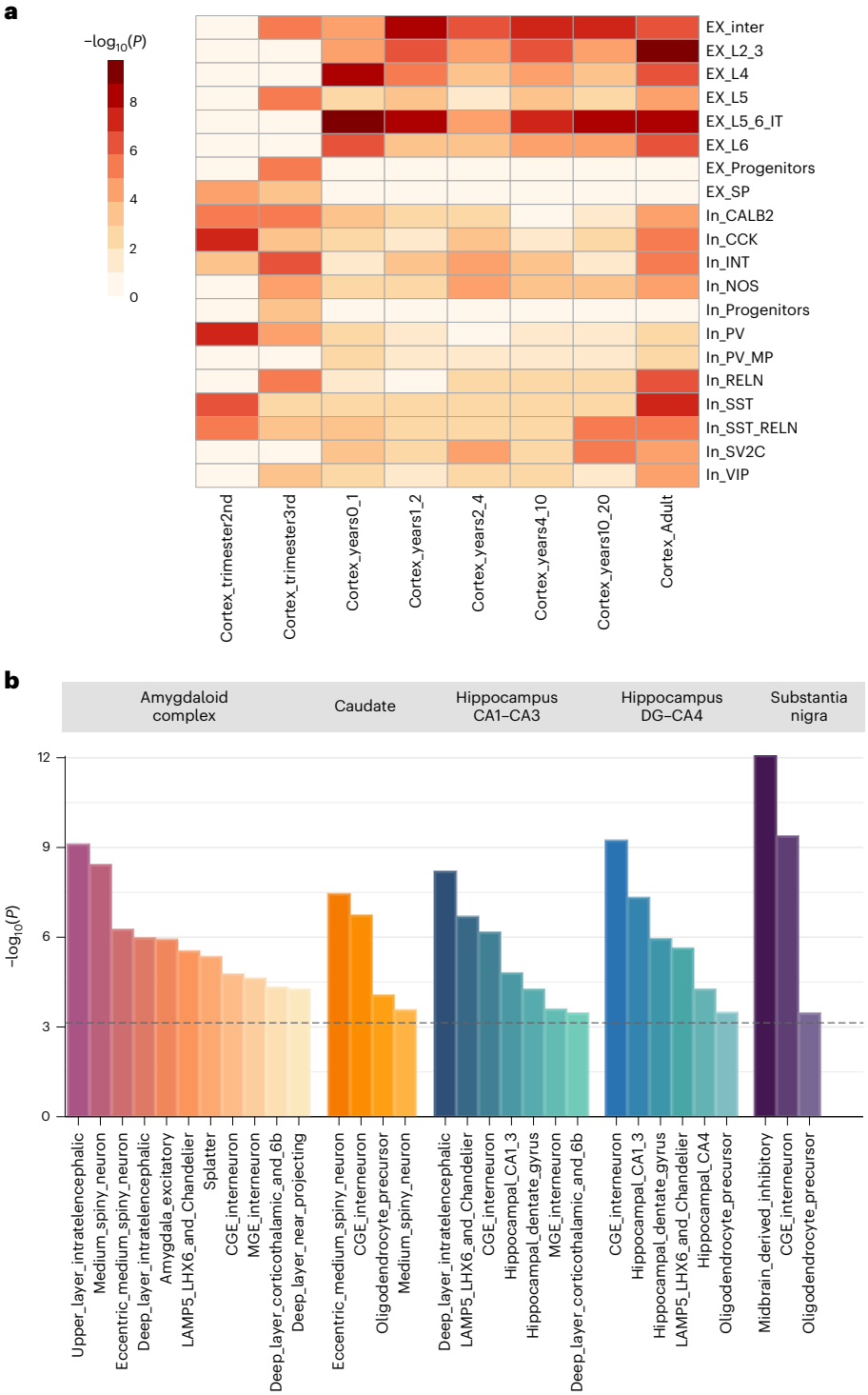

Fig. 3 | Cell type enrichment analyses for BD GWAS. a, The heatmap displays the results of enrichment analysis for cortical cell types at different developmental stages of trans-ancestry GWAS. The enrichment analysis was performed by the R package MAGMA.Celltyping (version 2.0.11) using data from Velmeshev et al.[22]. Each row represents a specific cell type, and each column represents a specific developmental period. The 'Linear' mode of the R package MAGMA. Celltyping was applied, which performed a one-side linear regression association test between cell-type-specific gene expression metric and BD gene-level associations. The color intensity represents the $-\log_{10}$ of one-sided nominal *P* values, with darker colors indicating higher significance. The significance threshold after Bonferroni correction is $P < 4.03 \times 10^{-4} = 0.05/124$. Excitatory neurons: EX_L2_3, upper-layer intratelencephalic projection neurons; EX_L5_6_IT, deep-layer intratelencephalic projection neurons; EX_L4, EX_L5 and EX_L6, extratelencephalic projection neurons in layer 4, layer 5 and layer 6; EX_inter, intermediate precursor cells; EX_Progenitors, dorsal progenitors; EX_SP, subplate neurons. Inhibitory neurons: In_INT, intermediate precursor cells; In_SST, somatostatin; In_RELN, reelin; In_SV2C, synaptic vesicle glycoprotein 2C; In_VIP, vasoactive intestinal polypeptide; In_CCK, cholecystokinin; In_SST_RELN, co-expressing SST and reelin; In_CALB2, calretinin; In_PV, parvalbumin; In_NOS, nitric acid synthase; In_PV_MP, inhibitory neurons co-expressing Pvalb with three metallopeptidases, Adamts8, Adamts15 and Mme. **b**, The bar plot shows the significant results of enrichment analysis for cell types in multiple brain dissections of trans-ancestry GWAS. The enrichment analysis was performed by the R package MAGMA.Celltyping (version 2.0.11). The *y* axis shows $-\log_{10}$ of one-sided nominal *P* values. The Bonferroni-corrected significance ($P < 9.43 \times 10^{-4} = 0.05/53$) is represented as the gray dashed line. The single-cell dataset was from Siletti et al.[23], including five brain regions. DG, dentate gyrus; MGE, medial ganglionic eminence.

https://doi.org/10.1038/s41593-025-02147-2

	Evidence count	High-confidence SNPs by fine-mapping				Lead SNP nearest	TWAS	Coloc	MAGMA	PoPS
		FUMA	SMR	Hi-C	eQTL					
WIPF3 –	9	√	√	√	√	√	√	√	√	√
STK4 –	8	√	√	√		√	√	√	√	√
TRANK1 –	8	√	√	√		√	√	√	√	√
SP4 –	7	√	√	√	√	√			√	√
CACNA1B –	6	√	√	√	√	√				√
CACNB2 –	6	√	√	√		√			√	√
HTR6 –	6	√		√	√		√	√	√	
NDFIP2 –	6	√	√	√			√	√	√	
PLEKHA8 –	6		√	√	√		√	√	√	
SYT14 –	6	√	√	√	√	√				√
SYT16 –	6	√		√		√		√	√	√
BCL11B –	5	√		√		√			√	√
DRD2 –	5	√		√		√			√	√
FURIN –	5	√	√	√			√	√		
NPTX1 –	5	√		√		√			√	√
PLCL2 –	5	√	√	√		√				√
PMVK –	5		√	√	√		√	√		
RPS6KA2 –	5	√		√		√			√	√
SHANK2 –	5	√		√		√			√	√
SHISA9 –	5	√		√		√			√	√
YWHAE –	5	√		√		√			√	√
ZBTB7B –	5	√	√	√			√	√		
ALPK3 –	4	√		√	√				√	
ARMC7 –	4	√	√	√				√		
ARTN –	4		√	√	√		√			
ATP2B2 –	4	√		√		√				√
DCLK3 –	4		√	√			√	√		
DOCK2 –	4	√				√			√	√
GABRA6 –	4	√	√			√		√		
GABRB2 –	4	√		√		√				√
KCNN3 –	4		√	√	√					√
KMT2E –	4	√		√					√	√
MACROD1 –	4	√				√			√	√
MAD1L1 –	4	√				√			√	√
MYRF –	4	√		√		√			√	
NPAS3 –	4	√		√		√				√
NT5C –	4	√	√			√			√	
PBXIP1 –	4		√	√			√	√		
PPP2R2B –	4	√		√		√				√

Fig. 4 | Gene prioritization results. A total of nine lines of evidence were used for gene prioritization, with each column representing a type of supportive evidence, and the candidate genes supported by at least four lines of evidence are displayed. Coloc, co-localization analysis; eQTL, eQTL associations across multiple brain tissues of putatively causal SNPs in the GTEx dataset; Hi-C, three-dimensional functional annotations of putatively causal SNPs in the 3DSNP website; Lead SNP nearest, nearest gene of lead SNP.

generated 117 95% credible sets of SNPs, including 59 high-confidence SNPs with posterior inclusion probability (PIP) > 0.5 and 29 SNPs having PIPs exceeding 0.95 (Supplementary Table 14).

We then employed a multi-tiered analytical framework integrating SNP-based, gene-based and similarity-based approaches to prioritize credible BD risk genes. Drawing on established methodologies[3,25], we first performed a 'discovery step' analysis leveraging genomic position, chromatin interaction (Hi-C) and expression quantitative trait locus (eQTL) data. These 59 putatively causal SNPs were annotated to 79 candidate genes by performing a SNP-to-gene mapping strategy using FUMA[26] (Supplementary Table 14), which incorporates positional and three-dimensional chromatin interaction mappings; the summary data-based Mendelian randomization (SMR)[27] of these 59 SNPs, based on cortex eQTL data from BrainMeta[28], identified an additional 36 genes (Supplementary Table 15). Therefore, the discovery step yielded 115 high-priority candidate genes. These genes were assigned a discovery score, ranging from 1 to 2.

We cross-validated these 115 candidate genes using seven complementary strategies (Supplementary Table 16). In brief, to achieve greater representation of multiple types of tissues and cells, three-dimensional functional annotations from the 3DSNP database[29] (Supplementary Table 17) and multi-tissue eQTLs from the GTEx dataset[21] (Supplementary Table 18) of the 59 putatively causal SNPs were collected. To factor in the possibility that the lead SNP of a locus was not highlighted in the fine-mapping analyses, we integrated nearest-gene mapping of lead SNPs (Supplementary Table 19). Transcriptome-wide association studies (TWASs) (Supplementary Table 20)[30] and genome-wide co-localization analyses[31] (Supplementary Table 21), which are considered important complements for SMR, were also conducted. MAGMA[32] (Supplementary Table 22), which aggregates signals across genes and detects associations driven by multiple variants, was applied. In addition, similarity-based method polygenic priority score (PoPS)[33] (Supplementary Table 23) was applied for orthogonal validations of SNP-based and gene-based methods. Each of the 115 genes received a cumulative validation score (range, 0–7) on the basis of supporting evidence across all seven validation approaches. The total score (discovery + validation) was then calculated for all genes, with higher scores reflecting greater confidence. Eventually, we identified 39 genes scored more

https://doi.org/10.1038/s41593-025-02147-2

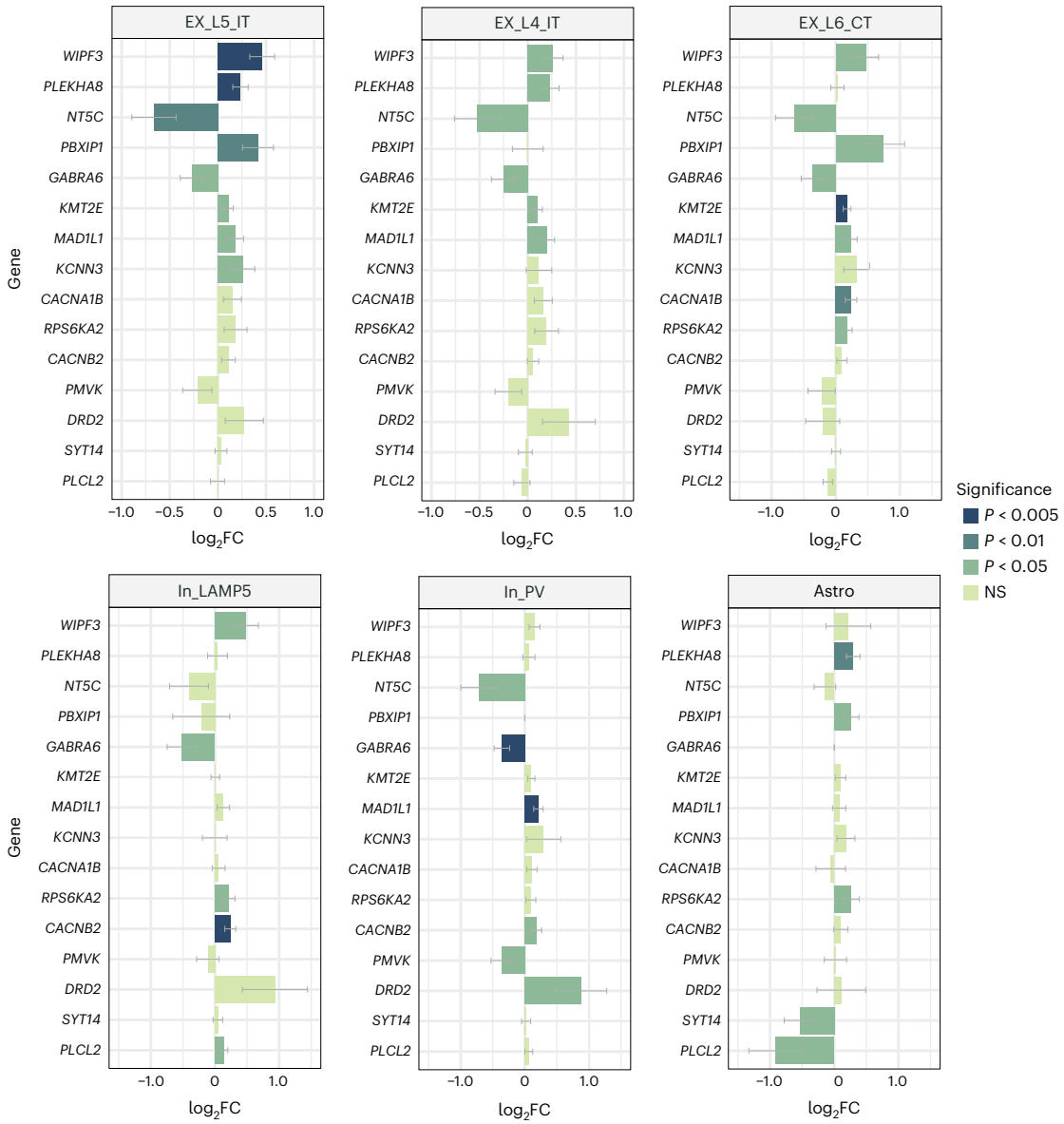

Fig. 5 | Differential expression of credible genes between BD cases and controls. Fifteen credible genes show differential expression in six cell types between BD cases ($N = 34$) and healthy controls ($N = 130$). This dataset was derived from Emani et al.[35]. Each bar shows the estimated expression differences ($\log_2$FC) between cases and controls for the credible BD risk genes; error bars indicate ±1 s.e. Statistical significance was assessed with two-sided Wald tests (reported P values are nominal and uncorrected). Different colors indicate four P value thresholds. Astro, astrocytes from DLPFC; EX_L5_IT, excitatory intratelencephalic projecting neurons from layer 5 of DLPFC; EX_L4_IT, excitatory intratelencephalic projecting neurons from layer 4 of DLPFC; EX_L6_CP, excitatory corticothalamic (extratelencephalic) from layer 6 of DLPFC; In_Lamp5, GABAergic interneurons expressing Lamp 5 from DLPFC; In_PV, GABAergic interneurons expressing parvalbumin from DLPFC. FC, fold change.

than 4 (Fig. 4), including 33 novel candidates not reported in PGC4 GWAS (which prioritized *SP4*, *CACNA1B*, *SHANK2*, *BCL11B*, *FURIN* and *SHISA9*)[3]. The gene with the most robust evidence in the present study was *WIPF3*. Despite the largely unknown function of this gene in the brain, its corresponding putatively causal SNP rs11761277 (PIP = 0.955) was predicted to increase the expression of this gene in the cortex according to the BrainMeta dataset ($P_{eQTL} = 6.72 \times 10^{-48}$)[28]; the expression of *WIPF3* was significantly upregulated in the frontal cortices of patients with BD compared to those of controls ($P = 0.0013$; data from a PsychENCODE Consortium study including 222 patients with BD and 936 controls[34]). Additionally, a recent cortical single-cell RNA sequencing analysis[35] revealed that *WIPF3* expression was increased in 34 patients with BD compared to that in 130 controls across multiple types of cells, including EX_L2_3 ($P = 4.5 \times 10^{-3}$), EX_L4 ($P = 0.01$), EX_L5_IT ($P = 3.6 \times 10^{-4}$), EX_L6_CT ($P = 0.02$), EX_L6_IT ($P = 0.01$), LAMP5 ($P = 0.01$) and In_VIP ($P = 7.3 \times 10^{-3}$).

We identified 15 of the 39 credible genes as differentially expressed in either the bulk frontal cortex or specific cellular subtypes between patients with BD and controls[34,35]. Permutation testing confirmed that these credible genes had a significantly higher likelihood of being differentially expressed than genes randomly selected from the genome ($P < 0.1$ according to the permutation test; Fig. 5 and Supplementary Table 24). For the highlighted types of cells, EX_L4_IT and EX_L5_IT refer to excitatory intratelencephalic projecting neurons from layer 4 and layer 5 of prefrontal cortex; L6 corticothalamic neurons project from layer 6 of the prefrontal cortex to various regions of thalamus; and LAMP5 neurons and parvalbumin-positive neurons (Pvalb) are inhibitory neurons from the CGE and MGE, respectively. These results corroborated those of the cellular enrichment analyses showing enrichment of BD associations in cortical LAMP5 interneurons and deep-layer intratelencephalic neurons.

Biological relevance and therapeutic implications of BD credible genes

We evaluated the behavioral phenotypes of mice lacking credible genes from the Mouse Genome Informatics (MGI) database. Among the 39 credible genes examined, multiple loci in knockout mice presented alterations across the behavioral domains implicated in BD, including emotional regulation, cognitive function, locomotor activity and social interaction (Fig. 6a and Supplementary Table 25). Specifically, *Cacna1b*, *Drd2*, *Npas3*, *Ppp2r2b*, *Shank2*, *Shisa9* and *Sp4* knockout mice presented hyperactivity in the open field test, whereas *Npas3*, *Trank1*, *Drd2* and *Cacna1b* knockouts presented anxiolytic-like phenotypes in the elevated plus maze. Exploration in a novel environment was enhanced in *Cacna1b*, *Gabrb2* and *Npas3* knockout mice, and deficits in spatial learning or memory were observed in *Sp4*, *Shank2*, *Cacna1b*, *Gabrb2* and *Npas3* knockouts, with *Sp4* and *Cacna1b* ablation particularly impairing multiple cognitive domains.

We performed functional annotation of credible genes using the SynGO database[36]. Eleven genes were annotated as synaptic components, with broad functional involvement in key synaptic processes, including membrane potential modulation, synaptic vesicle cycling and synaptic organization (Fig. 6b,c and Supplementary Table 26). Among these genes, eight (*ATP2B2*, *DRD2*, *GABRA6*, *GABRB2*, *NPTX1*, *SHANK2*, *SHISA9* and *YWHAE*) contribute to postsynaptic mechanisms or localize to postsynaptic compartments, whereas five (*ATP2B2*, *CACNA1B*, *CACNB2*, *DRD2* and *GABRB2*) have been implicated in presynaptic function.

To assess the therapeutic potential, we evaluated the druggability of BD risk genes using the Drug–Gene Interaction Database (DGIdb version 5.0)[37]. Among the 39 credible genes, 17 demonstrated interactions with 544 pharmacological agents (Supplementary Table 27), with notable drug–target interactions identified for *DRD2* (254 drugs, including 42 antipsychotics/antidepressants), *GABRB2* (87 drugs, 19 antiepileptics/anxiolytics/antidepressants) and *GABRA6* (85 drugs, 18 antiepileptics/anxiolytics/antidepressants). Gene set enrichment analysis of these drug–gene interactions showed a significant enrichment ($P < 0.001$) for targets of the anesthetic propofol and benzodiazepine-related hypnotic zopiclone (Supplementary Table 28). Complementary analysis through the Open Targets platform identified 18 credible genes with small-molecule tractability, including five (*PLEKHA8*, *ARMC7*, *FURIN*, *MACROD1* and *DCLK3*) that are not currently targeted by approved drugs (Supplementary Table 29). These findings highlight both established pharmacological pathways and novel opportunities for BD-related drug development.

Discussion

A merit of multi-ancestry GWAS is bringing advantages for variant fine-mapping by leveraging population-specific LD patterns. We present, to our knowledge, the largest Han Chinese BD cohort to date, comprising clinically diagnosed cases with an h^2_{SNP} estimate (0.24) consistent with previous reports[3]. Inclusion of EAS samples into trans-ancestry meta-analyses enabled more precise identification of credible risk variants, underscoring the value of diverse ancestry sampling for disentangling true causal variants from LD proxies. Empirical evidence demonstrated that the trans-ancestry GWAS significantly improved fine-mapping resolution: for the 69 shared GWS loci (excluding MHC) between our trans-ancestry GWAS and PGC4 EUR GWAS, the average credible set size decreased from 21.83 SNPs (spanning 108.43 kb) in EUR to 10.91 SNPs (65.79 kb) in our trans-ancestry GWAS ($P = 9.43 \times 10^{-4}$; Supplementary Table 30), representing a nearly two-fold improvement in localization resolution. Notably, although our fine-mapping approach pinpointed putatively causal SNPs for many loci, certain genomic regions (for example, 10q24.32 and 16p11.2) containing extended LD blocks remain challenging to resolve, highlighting the need for complementary approaches to characterize these architecturally complex risk loci fully.

Among the 23 novel loci compared with PGC4 EUR GWAS, four (*ATP2B2*, *MAD1L1*, *PLEC* and *CACNB2*) reside within previously reported BD-associated genomic regions[4,15]; seven (*GABRB2*, *ABCB5*, *RP11-22P4.1*, *NPAS3*, *RBFOX1*, *LINC00669* and *PPP1R16B*) were implicated in PGC4 GWAS incorporating diverse ascertainment strategies (clinical, community and self-report cohorts)[3]; and eight (*PLCL2*, *ATP11B*, *GABRB2*, *SNAP91*, *MAD1L1*, *CACNB2*, *ARL3* and *NPAS3*) demonstrated pleiotropic associations with other psychiatric disorders or related cognitive and affective traits[8,25,38–44]. These findings not only expand the catalog of BD risk loci but also reveal potential shared genetic architectures across mental health domains.

Developmental trajectory mapping of BD genetic risk demonstrated fetal stage-specific enrichment in inhibitory neurons versus pan-developmental enrichment in excitatory neurons, particularly deep-layer intratelencephalic projection neurons (EX_L5_6_IT). Our single-nuclei RNA sequencing analyses facilitated unprecedented cellular resolution, identifying subcortical interneuron subtypes as key enrichment targets, including hippocampal dentate gyrus–CA4 CGE interneurons and substantia nigra midbrain-derived inhibitory populations. We detected enrichment in eMSNs within the caudate and amygdala, a novel subclass indistinguishable using classic D1/D2 markers[45], mirroring findings from both PGC4 GWAS and a recent study on schizophrenia[3,46]. Hippocampal dentate gyrus neurons also showed significant enrichment, which aligns with previous reports of dentate gyrus hyperexcitability and lithium responsiveness in both BD models and patient-derived cells[47]. These results implicate disrupted cortico–subcortical circuitry in BD pathogenesis, highlighting specific interneuron–excitatory neuron interactions as potential mechanistic hubs.

We observed a counterintuitive positive genetic correlation between BD and educational attainment in EAS ($r_g = 0.14$), replicating previous findings in EUR ($r_g = 0.12$)[4]. In the epidemiological study, evidence showed that patients with BD tend to attain a greater level of education than the general population[48]. Further investigations using MiXeR suggest that the genetic relationship between BD and educational attainment is far more intricate than indicated by their overall positive association. In EAS, we found that, of the genetic variants influencing BD, 95.8% also affect educational attainment, and 55.6% have concordant effects with both traits. This mirrors findings in EUR, in which 98% of BD-related variants overlap with educational attainment, and 55% are concordant[4]. This extensive genetic overlap, coupled with longitudinal evidence of U-shaped risk associations across educational performance spectra[49,50], underscores a need for future studies dissecting subtype-specific and bidirectional relationships through refined endophenotypic approaches.

EAS GWAS confirmed the 6p21.32 MHC region association, with the lead variant rs4713555 in the HLA region, which was not significant in the PGC4 EUR GWAS. Instead of the 'lower' end of the MHC region where rs4713555 resides, the MHC locus reported in PGC4 EUR GWAS[3] is located at the other end nearing the gene *BTN3A2* and had no LD or correlations with rs4713555. In addition to HLA alleles and *BTN3A2*, the *C4* encoding genes are also pivotal in MHC. Elevated expression of *C4* genes in the brain was implicated in BD I (BD-I) and schizophrenia[51,52]. Nonetheless, PGC3 GWAS[4] imputed the five most common structural forms of the *C4A*/*C4B* locus (BS, AL, AL–BS, AL–BL and AL–AL) in EUR and reported no significant associations with BD[53]. Although the MHC is a very large genomic region showing complicated LD patterns, the observed differences on BD genetic architecture within MHC locus between ancestry does not necessarily imply different biological etiology.

The GWS variant rs174576 at 11q12.2 in the PGC4 EAS GWAS[3] was not significant in the Han Chinese GWAS. This discrepancy may reflect population-specific factors: the *FADS1*/*FADS2* locus encompassing rs174576 has undergone dietary-associated selection in EAS[54], exhibiting latitude-dependent allele frequency variation[55], suggesting potential

https://doi.org/10.1038/s41593-025-02147-2

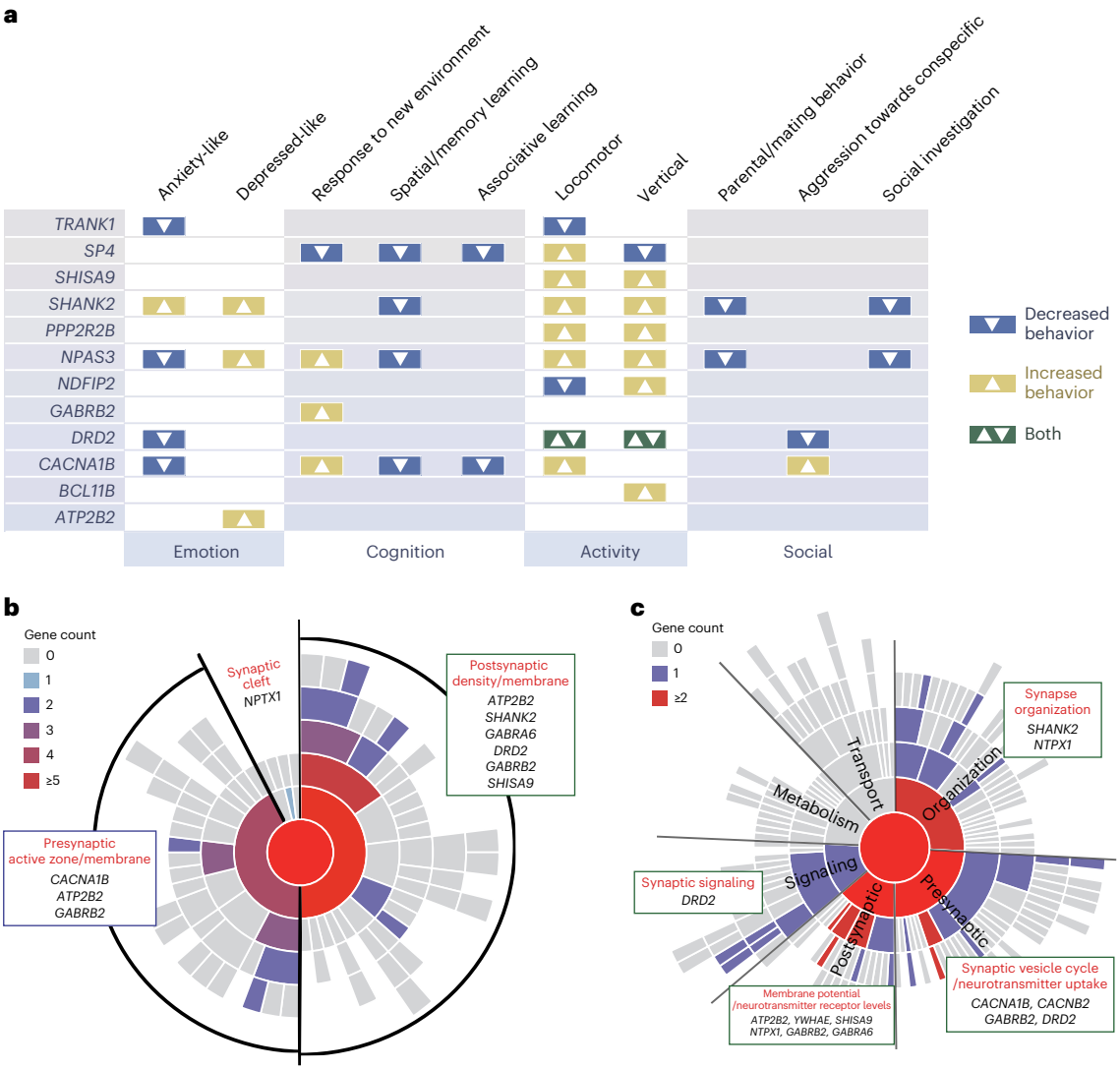

Fig. 6 | Potential biological implications of BD credible genes. a, The plot displays the behavioral phenotypes of mice lacking the credible genes in the MGI database. The phenotypes were divided into four main dimensions according to the classifications. The colors and triangle direction indicate the increased or decreased level of behaviors in mice. **b**, The sunburst plot depicts SynGO annotation[36] of BD credible genes on cellular components. The plot starts with the synapse (center) followed by presynaptic and postsynaptic locations in the first ring (divided by gray lines). The details for child terms annotation are presented in Supplementary Table 28. **c**, The sunburst plot depicts SynGO annotation[36] of BD credible genes on biological processes. The plot starts with the process in the synapse (center) followed by six dimensions (divided by gray lines) in the first ring. The details for child terms annotation are presented in Supplementary Table 28.

heterogeneity in environmental exposures across cohorts. Therefore, the heterogeneities in sample composition (for example, residences of participants or their dietary habits) may lead to inconsistencies at this locus between the Han Chinese GWAS and the PGC4 EAS GWAS. Such differences may also influence the h^2_{SNP} estimates, which were 0.239 in the Han Chinese GWAS and 0.099 in the PGC4 EAS GWAS. The lower estimate in PGC4 EAS may reflect a higher proportion of BD II cases in that cohort compared to ours, and there are also differences in the coverage of SNPs across studies, with the Han Chinese GWAS having 1.62 times more SNPs than the PGC4 EAS GWAS. In addition, of the 88 loci identified in the PGC4 EUR GWAS, 70 remained GWS in the trans-ancestry GWAS. Simulation analyses with Winner's Curse correction[56] demonstrated that the 18 non-replicated loci showed effect size patterns consistent with Winner's Curse-induced overestimation in the initial GWAS rather than representing false-positive associations (Supplementary Table 8). This observation reflects the well-documented GWAS bias wherein significant associations in discovery samples tend to have inflated effect sizes compared to their true values.

Although the current EAS GWAS remains modest compared to that of EUR cohorts, limiting the detection of population-specific signals, this work establishes a critical foundation for diversifying BD genetic research across global populations.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-025-02147-2.

References

1. Merikangas, K. R. et al. Prevalence and correlates of bipolar spectrum disorder in the World Mental Health Survey Initiative. *Arch. Gen. Psychiatry* **68**, 241–251 (2011).
2. Craddock, N. & Jones, I. Genetics of bipolar disorder. *J. Med. Genet.* **36**, 585–594 (1999).

3. O'Connell, K. S. et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature* **639**, 968–975 (2025).
4. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
5. Koromina, M. et al. Fine-mapping genomic loci refines bipolar disorder risk genes. *Nat. Neurosci.* **28**, 1393–1403 (2025).
6. McClellan, J. M. et al. An evolutionary perspective on complex neuropsychiatric disease. *Neuron* **112**, 7–24 (2024).
7. Gurdasani, D., Barroso, I., Zeggini, E. & Sandhu, M. S. Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* **20**, 520–535 (2019).
8. Lam, M. et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51**, 1670–1678 (2019).
9. Liu, Z. et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat. Genet.* **55**, 796–806 (2023).
10. Ishigaki, K. et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–1651 (2022).
11. Li, H. J. et al. Novel risk loci associated with genetic risk for bipolar disorder among Han Chinese individuals: a genome-wide association study and meta-analysis. *JAMA Psychiatry* **78**, 320–330 (2021).
12. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
13. Zhang, Z., Xiao, X., Zhou, W., Zhu, D. & Amos, C. I. False positive findings during genome-wide association studies with imputation: influence of allele frequency and imputation accuracy. *Hum. Mol. Genet.* **31**, 146–155 (2021).
14. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
15. Ikeda, M. et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol. Psychiatry* **23**, 639–647 (2018).
16. Zhou, F. et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nat. Genet.* **48**, 740–746 (2016).
17. Jia, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
18. Frei, O. et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 2417 (2019).
19. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet* **99**, 76–88 (2016).
20. Doan, R. N. et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354 (2016).
21. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
22. Velmeshev, D. et al. Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382**, eadf0834 (2023).
23. Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
24. Weissbrod, O. et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
25. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
26. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
27. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
28. Qi, T. et al. Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.* **54**, 1355–1363 (2022).
29. Quan, C., Ping, J., Lu, H., Zhou, G. & Lu, Y. 3DSNP 2.0: update and expansion of the noncoding genomic variant annotation database. *Nucleic Acids Res.* **50**, D950–D955 (2022).
30. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
31. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
32. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
33. Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).
34. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
35. Emani, P. S. et al. Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 (2024).
36. Koopmans, F. et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234 (2019).
37. Cannon, M. et al. DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res.* **52**, D1227–D1235 (2024).
38. Savage, J. E. et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
39. Okbay, A. et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).
40. Friligkou, E. et al. Gene discovery and biological insights into anxiety disorders from a large-scale multi-ancestry genome-wide association study. *Nat. Genet.* **56**, 2036–2045 (2024).
41. Watanabe, K. et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat. Genet.* **54**, 1125–1132 (2022).
42. Hatoum, A. S. et al. Genome-wide association study shows that executive functioning is influenced by GABAergic processes and is a neurocognitive genetic correlate of psychiatric disorders. *Biol. Psychiatry* **93**, 59–70 (2023).
43. Ward, J. et al. The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants, genetic correlation with psychiatric disorders, and association with gene expression and function. *Mol. Psychiatry* **25**, 3091–3099 (2020).
44. Thorp, J. G. et al. Symptom-level modelling unravels the shared genetic architecture of anxiety and depression. *Nat. Hum. Behav.* **5**, 1432–1442 (2021).
45. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).
46. Duncan, L. E. et al. Mapping the cellular etiology of schizophrenia and complex brain phenotypes. *Nat. Neurosci.* **28**, 248–258 (2025).
47. Mertens, J. et al. Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. *Nature* **527**, 95–99 (2015).

48. Vreeker, A. et al. High educational performance is a distinctive feature of bipolar disorder: a study on cognition in bipolar disorder, schizophrenia patients, relatives and controls. *Psychol. Med.* **46**, 807–818 (2016).
49. MacCabe, J. H. et al. Excellent school performance at age 16 and risk of adult bipolar disorder: national cohort study. *Br. J. Psychiatry* **196**, 109–115 (2010).
50. Gale, C. R. et al. Is bipolar disorder more common in highly intelligent people? A cohort study of a million men. *Mol. Psychiatry* **18**, 190–194 (2013).
51. Sekar, A. et al. Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
52. Hörbeck, E. et al. Dissecting the impact of complement component 4A in bipolar disorder. *Brain Behav. Immun.* **116**, 150–159 (2024).
53. Kamitaki, N. et al. Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature* **582**, 577–581 (2020).
54. Chen, H. et al. Tracing Bai-Yue ancestry in aboriginal Li people on Hainan Island. *Mol. Biol. Evol.* **39**, msac210 (2022).
55. Liu, S. et al. Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history. *Cell* **175**, 347–359 (2018).
56. Forde, A., Hemani, G. & Ferguson, J. Review and further developments in statistical corrections for Winner's Curse in genetic association studies. *PLoS Genet.* **19**, e1010546 (2023).

Chu-Yi Zhang[1,2,3,43], Miao Li[1,2,3,43], Ping Sun[4,5,43], Li Hui[6,43], Yuan Gao[7,8,43], Jian-Zhong Yang[9,10,43], Nan Zhang[11], Xiaoyang Feng[12], Yong Wu[1,2], Lei Guo[13,14], Jing Yuan[15], Hong-Yan Jiang[16], Yu-Qi Cheng[16], Simeng Ma[11], Qian Gong[11], Yaoyao Sun[12], Yi Li[17,18], Na Qu[17,18], Xu-Yuan Yin[6], Lu Wang[1,2], Yongfeng Yang[19,20], Chuansheng Wang[19,20], Luxian Lv[19,20], Dongsheng Zhou[21], Xingxing Li[21], Xiaogang Chen[22,23], Chen Zhang[5], Jun Chen[5], Xueqin Song[24], Jinsong Tang[25,26], Jun Cai[5], Weixing Fan[27], Wei Tang[28], Wenxin Tang[29], Wenqiang Li[19,20,30], Xia Tang[31], Xiaoxi Zhang[31], Yan Lu[31,32], Yong-Gang Yao[1,2,3,33,34], Chuang Wang[13,14], Hon-Cheong So[34,35,36], Nakao Iwata[37], Masashi Ikeda[37,38], Takeo Saito[37], Zhongchun Liu[11,39], Shuahua Xu[31,32,40], Weihua Yue[12], GeseDNA Research Team*, Yiru Fang[41] ✉, Feng Zhu[7,8] ✉, Xiao Xiao[1,2,3,34] ✉ & Ming Li[1,2,3,34] ✉

[1]State Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [2]Yunnan Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [3]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China. [4]Qingdao Mental Health Center, Qingdao, China. [5]Clinical Research Center & Division of Mood Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [6]Suzhou Guangji Hospital, The Affiliated Guangji Hospital of Soochow University, Suzhou, China. [7]Department of Psychiatry, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China. [8]Shaanxi Provincial Key Laboratory of Biological Psychiatry, Xi'an, China. [9]Department of Psychiatry, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [10]The Key Laboratory of Mental Disorder Management in Zhejiang Province, Hangzhou, China. [11]Department of Psychiatry, Renmin Hospital of Wuhan University, Wuhan, China. [12]Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health (Peking University), National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital), Beijing, China. [13]Zhejiang Key Laboratory of Pathophysiology, Health Science Center, Ningbo University, Ningbo, China. [14]School of Basic Medical Science, Health Science Center, Ningbo University, Ningbo, China. [15]Department of Psychiatry, The Second Affiliated Hospital of Kunming Medical University, Kunming, China. [16]Department of Psychiatry, The First Affiliated Hospital of Kunming Medical University, Kunming, China. [17]Affiliated Wuhan Mental Health Center, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [18]Research Center for Psychological and Health Sciences, China University of Geosciences, Wuhan, China. [19]Henan Mental Hospital, The Second Affiliated Hospital of Henan Medical University, Xinxiang, China. [20]Henan Key Lab of Biological Psychiatry, International Joint Research Laboratory for Psychiatry and Neuroscience of Henan, Henan Medical University, Xinxiang, China. [21]Department of Psychiatry, Ningbo Kangning Hospital, Ningbo, China. [22]Department of Psychiatry, The Second Xiangya Hospital, Central South University, Changsha, China. [23]National Clinical Research Center for Mental Disorders, Changsha, China. [24]The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China. [25]Department of Psychiatry, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China. [26]Key Laboratory of Medical Neurobiology of Zhejiang Province, Hangzhou, China. [27]Jinhua Second Hospital, Jinhua, China. [28]Department of Psychiatry, The Affiliated Kangning Hospital of Wenzhou Medical University, Wenzhou, China. [29]Hangzhou Seventh People's Hospital, Hangzhou, China. [30]Collaborative Innovation Center of Prevention and Treatment of Mental Disorder, The Second Affiliated Hospital of Henan Medical University, Xinxiang, China. [31]State Key Laboratory of Genetic Engineering, Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai, China. [32]Ministry of Education Key Laboratory of Contemporary Anthropology, Fudan University, Shanghai, China. [33]National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), National Resource Center for Non-Human Primates, Yunnan Engineering Center on Brain Disease Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [34]KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [35]School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong, China. [36]Department of Psychiatry, The Chinese University of Hong Kong, Hong Kong, China. [37]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan. [38]Department of Psychiatry, Nagoya University Graduate School of Medicine, Nagoya, Japan. [39]Taikang Center for Life and Medical Sciences, Wuhan University, Wuhan, China. [40]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [41]Department of Psychiatry & Affective Disorders Center, Ruijin Hospital, Shanghai Jiao Tong University School

of Medicine, Shanghai, China. [43]These authors contributed equally: Chu-Yi Zhang, Miao Li, Ping Sun, Li Hui, Yuan Gao, Jian-Zhong Yang. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: yirufang@aliyun.com; zhufeng1982@xjtu.edu.cn; xiaoxiao2@mail.kiz.ac.cn; limingkiz@mail.kiz.ac.cn

GeseDNA Research Team

Zenan Dou[42], Shan Guan[42], Tingting Guo[42], Qinglan Liu[42], Hoyin Lo[42] & Leilei Zhang[42]

[42]Beijing Gese Technology Co., Ltd., Beijing, China.

Methods

Patients with BD and controls of Han Chinese ancestry

Each patient with BD was independently interviewed by both clinicians, and those diagnosed with the same Axis I disorder by both psychiatrists were recruited. All diagnoses were confirmed by a research psychiatrist through an extensive clinical interview and a Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Version (SCID-P). Patients were excluded if they were diagnosed with other psychiatric disorders or neurological disorders, were pregnant or were breastfeeding at the time of study. The control individuals were recruited from the Chinese mainland and were not specifically screened for psychiatric disorders by psychiatrists but had no current self-reported serious illnesses or disabilities or any personal or family history (including first-degree, second-degree and third-degree relatives) of psychiatric illness. This study protocol was approved by the institutional review board of Kunming Institute of Zoology, Chinese Academy of Sciences, and the ethics committees of all participating units. All participants were volunteers and provided written informed consent before any study-related procedures were performed.

Quality control for GWAS data

The raw genotyping data of 603,033 autosomal SNPs across 12,085 individuals (N_{case} = 4,213, $N_{control}$ = 7,872) were obtained using an Illumina ASA chip (referred to as the 'ASA sample'). Quality control analyses were performed at both the variant and individual levels using PLINK version 1.9 as described by Anderson et al.[57,58] (Supplementary Note). Eventually, a total of 455,284 autosomal biallelic SNPs across 4,141 cases (1,856 (44.83%) males and 2,285 (55.17%) females) and 7,831 controls (4,322 (55.22%) males and 3,590 (44.78%) females) remained for genotype imputation.

In addition, raw genotyping data of 686,594 autosomal SNPs in an additional 7,747 individuals (N_{case} = 1,087, $N_{control}$ = 6,660) were obtained using an Illumina GSA chip (referred to as the 'GSA sample'). After quality control (Supplementary Note), a total of 369,456 autosomal biallelic SNPs in 1,023 cases (490 (47.90%) males and 533 (52.10%) females) and 5,629 controls (3,165 (56.22%) males and 2,464 (43.78%) females) remained for genotype imputation.

Genotype imputation

Following the prephasing/imputation stepwise approach, genotyped SNPs were prephased by SHAPEIT4 (ref. 59) and then imputed using IMPUTE5 (refs. 60,61) on the basis of Phase 3 of the 1000 Genomes Project reference set[62]. A final set of 8,136,260 autosomal biallelic SNPs across 1,023 cases and 5,629 controls (GSA sample), and 8,197,489 autosomal biallelic SNPs across 4,141 cases and 7,831 controls (ASA sample), had INFO > 0.3, MAF > 1% and minor allele counts larger than 10 in cases and had Hardy–Weinberg equilibrium (HWE) $P > 1.00 \times 10^{-5}$. These SNPs were subjected to subsequent association analyses.

Statistical analysis of the GWAS samples

In the Han Chinese GWAS, the additive logistic regression of BD diagnosis on the imputed hard-called genotypes was performed separately on the ASA sample and the GSA sample using PLINK version 1.9, which involves sex and principal components associated with diagnostic status as covariates to adjust for population stratification. Summary statistics from different GWAS cohorts were combined for an inverse variance-weighted meta-analysis using PLINK version 1.9. A genome-wide significant locus was defined using the '--clump' option in PLINK to report the region around a lead variant ($P \leq 5.00 \times 10^{-8}$) with LD $r^2 < 0.1$ within a 500-kb window[58].

LDSC heritability estimation

LDSC (version 1.0.1) was used to estimate h^2_{SNP} and assess potential population stratification within Han Chinese populations[12]. The precomputed LD scores for EAS in HapMap3 were downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/. Before LDSC analysis, we removed SNPs with A/T or G/C alleles across the whole genome as well as SNPs in the extended MHC region (hg19, chr6: 25–35 Mb). We assumed that the lifetime prevalence of BD was 0.02 ('--pop-prev' flag) to estimate h^2_{SNP} on the liability scale.

Genetic correlations of BD with other phenotypes in EAS

LDSC (version 1.0.1) was applied to calculate genetic correlations between BD and other 39 phenotypes (Supplementary Note) within EAS using GWAS summary statistics[14]. We removed SNPs with A/T or G/C alleles across the whole genome as well as SNPs in the extended MHC region. Genotype information of EAS ancestry from the 1000 Genomes Project[62] was used as the reference panel.

Polygenic overlap analysis using MiXeR

To assess polygenic overlap between BD and genetically correlated traits in EAS, we used the bivariate causal mixture model implemented in MiXeR version 1.3 (ref. 18). Variants in the extended MHC region were removed, and the input GWAS summary statistics were preprocessed to calculate the *z*-score. The total number of variants with non-zero additive genetic effects was calculated for each trait based on the univariate Gaussian causal mixture model using a likelihood-based framework by MiXeR. The shared and unique polygenic components of trait pairs were plotted as Venn diagrams, and the Dice coefficient score (polygenic overlap measure on the 0–100% scale) was computed[63]. Genotype information of EAS ancestry from the 1000 Genomes Project[62] was used as a reference panel.

Cross-ancestry genetic effect comparison

Popcorn. The trans-ancestry genetic correlation between EAS GWAS and PGC4 EUR GWAS was estimated using Popcorn (version 1.0)[19]. The precomputed cross-population scores for EAS and EUR in the 1000 Genomes Project were downloaded from https://www.dropbox.com/sh/37n7drt7q4sjrzn/AAAa1HFeeRAE5M3YWG9Ac2Bta. SNPs with A/T or G/C alleles across the whole genome as well as SNPs in the extended MHC region were removed. The analysis was performed under all default values except one (--MAF 0.05). The trans-ancestry genetic effect correlation (ρ_{ge}) and genetic impact correlation (ρ_{gi}) represented the genetic relationships between EAS BD and EUR BD.

PRS. PRS-CS version 1.1.0 (ref. 64) was used to infer posterior SNP effect estimates in the PGC4 GWAS training dataset through a high-dimensional Bayesian regression framework. The LD reference panel, which is based on UK Biobank data, was provided by PRS-CS developers (https://github.com/getian107/PRScs). We calculated the PRS for Han Chinese target samples (5,164 cases and 13,460 controls) on the basis of the PRS-CS estimated effects using the '--score' function in PLINK version 1.9. The case–control status of the Han Chinese participants was estimated through logistic regression analysis of *z*-score transformed polygenic scores. The analysis included the same covariates as in the Han Chinese GWAS and was performed using the glm() function in R with family = binomial(logit). The variances explained by PRS (R^2) in two samples were first converted to the Nagelkerke pseudo R^2 using the fmsb package in R and then converted to the liability-scaled Nagelkerke pseudo R^2, assuming that BD incidence ranged from 1% to 5%.

Allelic effect. To estimate the consistency of the effect for each locus, we conducted a Pearson s correlation test for OR values of PGC4 EUR GWS lead SNPs or their proxies between EAS and EUR GWASs.

F_{ST}. F_{ST} was used to evaluate the population differentiation due to genetic structure. We calculated the F_{ST} using the Weir and Cockerham approach[65] with PLINK version 1.9 using genotypes of EUR (N = 503) and EAS (N = 504) individuals from the 1000 Genomes Project.

Enrichment analysis

Genomic region enrichment. Gene regulatory element annotations for cCREs were derived from ENCODE cCREs version 3 (ref. 66) and the 15-state ChromHMM model BED files from the Roadmap Epigenetics Project[67]. Genome-wide maps of TFBSs were obtained from Vorontsov et al.[68], and the list of HARs was defined based on previous studies[69–71]. We used the clumping procedure in PLINK and assessed enrichment of BD risk loci with specific genomic element regions using the LOLA R package[72]. Significant enrichment was defined as Benjamini–Hochberg values less than 0.05.

Tissue and cell type enrichment. We used FUMA version 1.6.4 (ref. 26) to perform tissue expression enrichment analyses for trans-ancestry GWAS signals with default settings based on tissue-specific gene expression levels in GTEx version 8 (ref. 21).

Single-nuclei RNA sequencing data expression profiles were obtained from two BRAIN Cell Census (BICCN) studies[22,23]. Velmeshev et al.[22] provided a comprehensive view of human cortical development across different developmental stages using a total of 709,372 nuclei and 169 brain tissue samples from 106 individuals. Another BICCN study sampled 3,369,219 nuclei from 105 different dissections of three postmortem brain donors[23]. We used the R packages MAGMA.Celltyping (version 2.0.11)[73] and MungeSumstats (version 1.6.0)[74] to test for enrichment of trans-ancestry GWAS associations within brain cell populations. The required CellTypeDataset (CTD) object formats of single-cell datasets were generated using the EWCE R package (version 1.6.0)[75] from the gene expression matrix and cell annotation metadata. For the cell type association method, the 'Linear' enrichment mode available in the celltype_associations_pipeline function was used.

Fine-mapping analysis

To identify putatively causal SNPs, we applied Polyfun + SuSiE[24], a well-established framework to improve fine-mapping accuracy through leveraging functional annotations across the genome, to carry out fine-mapping analysis of GWS loci in the trans-ancestry GWAS. We constructed a customized LD panel based on the proportions of samples from different ancestries in the trans-ancestry GWAS ($N_{eff\_EAS} = 34{,}808$, $N_{eff\_EUR} = 220{,}416$, $N_{total} = N_{eff\_EAS} + N_{eff\_EUR}$). Sample-size-weighted LD matrices ($LD_{weighted}$) for each independent GWS locus were calculated, with the LD matrices based on EAS sample (LD_{EAS}) and EUR sample (LD_{EUR}) combined following the study by Wojcik et al.[76].

$$LD_{weighted} = \frac{N_{eff\_EAS}}{N_{total}} \times LD_{EAS} + \frac{N_{eff\_EUR}}{N_{total}} \times LD_{EUR}$$

The LD_{EAS} matrices were generated based on 11,972 Han Chinese individuals genotyped using the ASA array in our study using PLINK version 1.9 '–r square'. The LD_{EUR} matrices, which were precomputed LD reference panels based on 337,000 unrelated British ancestry individuals from the UK Biobank, were downloaded from https://broad-alkesgroup-ukbb-ld.s3.amazonaws.com/UKBB_LD/. LD matrices from the two populations were integrated on the premise that the included SNPs were the same and in the same order and had consistent reference alleles in both LD matrices. The independent GWS loci (defined by PLINK version 1.9 '--clump' with LD $r^2 < 0.1$ within a 500-kb window) were used as targeted regions, and the MHC locus was excluded owing to its complex LD structure. Within each GWS locus, we assigned each SNP a PIP score, and multiple SNPs with an ordered sum of PIP scores greater than 0.95 were considered a '95% credible set'. Within each 95% credible set, a SNP with a PIP ≥ 0.50 was defined as putatively causal (which means that there is no more than one putatively causal SNP within a single credible set), and the maximum number of 95% credible sets within each GWS locus was set to five.

Gene mapping for putatively causal SNPs

FUMA annotation. Putatively causal SNPs (PIP > 0.50 in the 95% credible set) were annotated to genes using the SNP2GENE module in FUMA version 1.6.4 (https://fuma.ctglab.nl/snp2gene)[26]. The annotation considered both position-based mapping (10-kb maximum distance to genes) and promoter-anchored Hi-C loops (PsychENCODE resource[77])-based mapping strategies. According to the positional mapping method, for SNPs mapped to multiple genes with overlapping locations, all genes remained.

SMR analysis. We implemented SMR version 1.0.3 to prioritize potentially functional genes from the trans-ancestry GWAS risk loci through leveraging the BrainMeta version 2 *cis*-eQTL dataset (2,865 adult cortex transcriptomes) as our eQTL reference28. SMR tests for putatively causal or pleiotropic relationships between gene expression and BD via SNPs[78] and infers whether variants account for LD through the heterogeneity in dependent instruments (HEIDI) test[27]. Putatively causal SNPs identified from fine-mapping analysis served as instrumental QTLs, with each SNP analyzed against all probes within a ±2-Mb window (using the –extract-target-snp-probe option) and the EUR Haplotype Reference Consortium LD reference panel. To distinguish true *cis*-regulatory effects from linkages, we applied stringent dual thresholds ($FDR_{SMR} < 0.05$ for SMR significance and $P_{HEIDI} > 0.01$ to exclude pleiotropic associations).

Genes were assigned discovery scores of 1 (highlighted by FUMA or SMR) or 2 (highlighted by both methods).

Additional validation approaches in credible gene prioritization

Given the polygenic architecture of BD and the potential for spurious associations in single-method analysis, we leveraged convergent evidence across seven approaches and employed a multidimensional gene-mapping strategy to validate genes prioritized by FUMA, SMR or both methods. Each supporting line of evidence was assigned a unitary weight, and genes with a total score (discovery + validation) over 4 were classified as high-confidence candidates, ensuring stringent prioritization while mitigating false positives inherent to any single method. Below, we briefly outline these seven additional analytical strategies.

Three-dimensional functional annotations. We annotated all the putatively causal SNPs with their spatially interacting genes in the 3DSNP database (https://omic.tech/3dsnpv2/)[29], an integrative platform that incorporates multi-omics data, including single-molecule sequencing and single-cell assay for transposase-accessible chromatin using sequencing (ATAC–seq) profiles across diverse tissues and cell types, enabling identification of chromatin interactions between SNPs and their target genes.

Multi-tissue eQTL analyses. To characterize the neurobiological relevance of putatively causal SNPs, we evaluated their eQTL regulatory potentials across distinct brain regions using GTEx version 8 (https://www.gtexportal.org/)[21]. Given that BD risk variants may exhibit regulatory effects across varied regions in the brain, we analyzed eQTL associations across 13 brain regions in GTEx, and SNP–gene correlations showing consistent significance in at least two tissues (FDR < 0.05) were considered reliable.

Nearest gene mapping of the lead SNP. We mapped the lead SNPs in the GWS locus to nearest genes using the SNP2GENE module in FUMA version 1.6.4 (ref. 26). This approach assumes that the nearest gene to a GWS lead SNP is often a candidate of credible genes[33,79,80] and, hence, factors in the possibility that the lead SNP of a locus was not highlighted in the fine-mapping analyses. We implemented a 10-kb window criterion for gene assignment, capturing both intragenic variants and potential *cis*-regulatory elements in proximal intergenic regions. For SNPs mapping to multiple genes within this window (for example, in gene-dense regions), all assigned genes were retained, as complex regulatory interactions may span multiple adjacent genes in GWS loci.

TWAS. TWAS leverages eQTL datasets and trains a multivariate model that considers multiple *cis*-SNPs to impute gene expression, thereby complementing the reliance of SMR on a single putatively causal SNP[81]. For TWAS analysis, we employed FUSION software[30], which uses cortex-derived predictive weights from the PsychENCODE Consortium (N = 1,321 individuals)[34]. Predictive models (GBLUP, LASSO, Elastic Net, BSLMM and TOP1) were evaluated via five-fold cross-validation, with the top-performing model selected based on predictive accuracy[30]. The significance thresholds were adjusted for multiple testing ($P_{TWAS} < 1.00 \times 10^{-5}$) to ensure robust associations.

Co-localization. Co-localization analysis examines whether two traits share the same causal variants at a locus[82]. By evaluating association signals across all SNPs in a locus, co-localization detects shared causal variants potentially missed by SMR[83]. We, hence, applied co-localization to assess whether BD risk loci harbored shared causal variants with regulatory effects on gene expression using the COLOC R package (version 5.2.3)[84] implemented in FUSION[30], leveraging PsychENCODE cortex-derived eQTLs[34] as molecular QTLs. Genes meeting a TWAS significance threshold ($P_{TWAS} < 0.05$) were subjected to co-localization, with a posterior probability (PP4) > 0.5 considered evidence of credible co-localization.

Gene-based MAGMA. To complement SNP-based analyses, we performed gene-based association testing using MAGMA version 1.10 (ref. 32). This approach aggregates SNP-level GWAS signals at the gene level while accounting for LD through Brown's method, thereby increasing sensitivity to detect polygenic effects where multiple weakly associated SNPs collectively implicate a gene. We defined gene boundaries with 35-kb upstream and 10-kb downstream windows to capture both coding regions and potential regulatory elements. The conservative significance threshold ($P_{MAGMA} < 2.71 \times 10^{-6}$) controlling for multiple testing was determined.

PoPS gene analysis. We implemented the similarity-based PoPS method[33], which integrates multidimensional evidence, including (1) GWAS association signals, (2) pathway enrichment patterns, (3) spatiotemporal gene expression profiles and (4) protein–protein interaction networks. This approach simultaneously considers both statistical associations and biological coherence and, thereby, addresses the 'missing specificity' problem of GWAS by leveraging the polygenic convergence hypothesis that true risk genes likely share functional characteristics and participate in coherent biological processes. We designated genes ranked in the top 5% of PoPS scores as high-confidence candidates, as this threshold optimally balances discovery power with specificity in simulations of complex trait architectures[33].

Differential expression analyses of credible genes between patients with BD and controls

We examined the mRNA expression alterations of credible genes in the postmortem cortex tissue data of the PsychENCODE dataset, which includes data from 222 patients with BD and 936 healthy controls[34]. Linear regression was conducted for differential expression analysis by adjusting known biological, technical and surrogate variables.

We also leveraged the human postmortem dorsolateral prefrontal cortex (DLPFC) resource from a single-cell multi-omics study by the PsychENCODE Consortium[35]. The cell-type-specific differentially expressed genes between 34 patients with BD and 130 controls of EUR ancestry were downloaded from BrainSCOPE https://brainscope.gersteinlab.org/output-DEG.html. We examined whether the 39 credible genes presented differential expression levels in patients with BD in different types of cells (excitatory neurons: L2/3 IT, L4 IT, L5 IT, L6 IT, L6 IT Car3, L5 ET, L5/6 NP, L6b and L6 CT; inhibitory neurons: SST, SST CHODL, PVALB, Chandelier, LMAP5 LHX6, LAMP5, VIP and PAX6; and non-neuronal cell types: Astro, Oligo, OPC and Micro; detailed information of these cell types was included in the original report[35]).

Functional interpretation of credible genes

MGI knockout mice phenotype. The functional consequences of the 39 BD credible genes in knockout mice were searched in the MGI (https://www.informatics.jax.org/) database. Phenotypes under the Mammalian Phenotype term 'behavior/neurological phenotype' (MP:0005386) were included in our annotation. We categorized the behavioral abnormalities according to the classification of MGI. The 'Cognition', 'Emotion', 'Activity' and 'Social' sections correspond to 'abnormal cognition' (MP:0014114), 'abnormal emotion/affect behavior' (MP:0002572), 'abnormal motor capabilities/coordination/movement' (MP:0002066) and 'abnormal social/conspecific interaction behavior' (MP:0002557).

SynGO. The roles of the 39 BD credible genes in cellular component and biological processes related to synapses were annotated using the SynGO version 1.2 web interface (https://www.syngoportal.org/)[36].

Drug target analyses. We examined the drug–gene interactions for the 39 BD credible genes in DGIdb 5.0 (https://dgidb.org/api)[37], which includes data mined from DrugBank, PharmGKB, ChEMBL, Drug Target Commons and others. The drug target enrichment for the 39 BD credible genes was analyzed through a hypergeometric test with the phyper() function in R (1,103 drugs with at least 10 target genes were included). The druggability was evaluated using the Open Targets web interface (https://platform.opentargets.org/). The druggable genes were predicted on the basis of the 'small molecule' modality in tractability assessment with any one of the labels ('Structure with Ligand', 'High-Quality Ligand', 'High-Quality Pocket', 'Med-Quality Pocket', 'Druggable Family' and 'Phase 1 Clinical') labeled 'TRUE'.

Reporting summary

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

Data availability

The genome-wide summary statistics of BD PGC4 GWAS are available via figshare at https://doi.org/10.6084/m9.figshare.27216117.v2 (ref. 85). The SNP expression weights of PsychENCODE used in this study are available at http://resource.psychencode.org/. The PsychENCODE *cis*-eQTL data and the BrainMeta version 2 *cis*-eQTL data are available via the BrainMeta portal at https://yanglab.westlake.edu.cn/software/smr/#DataResource. The cell-type-specific gene expression matrices from two BICCN studies[29,30] for enrichment analysis were downloaded from the CELLxGENE database at https://cellxgene.cziscience.com/collections/bacccb91-066d-4453-b70e-59de0b4598cd and https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443 separately. The summary statistics of BD cross-ancestry meta-analysis are publicly available on the Scientific Data Center of the Kunming Institute of Zoology (https://datacenter.kiz.ac.cn/Home/DataContent?data_gd=4c22800a-d798-169c-7184-430976572334). According to the related policy of the Ministry of Science and Technology of the People's Republic of China, depositing genetic data (including the summary statistics) of Chinese populations in a third-party website (without application) is not allowed before approval; alternatively, GWAS summary statistics in Chinese populations can be requested from the China National Genomics Data Center (https://ngdc.cncb.ac.cn/gvm/), with data accession number GVP000051.

Code availability

No custom code was developed for this study. All software and tools used for the analyses presented are publicly available and referenced within the respective sections in Methods of the article.

References

57. Anderson, C. A. et al. Data quality control in genetic case–control association studies. *Nat. Protoc.* **5**, 1564–1573 (2010).
58. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
59. Delaneau, O., Zagury, J. F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
60. Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F. & Marchini, J. Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
61. Rubinacci, S., Delaneau, O. & Marchini, J. Genotype imputation using the Positional Burrows Wheeler Transform. *PLoS Genet.* **16**, e1009049 (2020).
62. Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
63. Bahrami, S. et al. Genetic loci shared between major depression and intelligence with mixed directions of effect. *Nat. Hum. Behav.* **5**, 795–801 (2021).
64. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
65. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
66. Moore, J. E. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
67. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
68. Vorontsov, I. E. et al. Genome-wide map of human and mouse transcription factor binding sites aggregated from ChIP-Seq data. *BMC Res. Notes* **11**, 756 (2018).
69. Girskis, K. M. et al. Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* **109**, 3239–3251 (2021).
70. Hubisz, M. J. & Pollard, K. S. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Curr. Opin. Genet. Dev.* **29**, 15–21 (2014).
71. Gittelman, R. M. et al. Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res.* **25**, 1245–1255 (2015).
72. Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).
73. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet* **50**, 825–833 (2018).
74. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* **37**, 4593–4596 (2021).
75. Skene, N. G. & Grant, S. G. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.* **10**, 16 (2016).
76. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
77. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
78. Mai, J., Lu, M., Gao, Q., Zeng, J. & Xiao, J. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun. Biol.* **6**, 899 (2023).
79. Stacey, D. et al. ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Res.* **47**, e3 (2019).
80. Hemerich, D. et al. An integrative framework to prioritize genes in more than 500 loci associated with body mass index. *Am. J. Hum. Genet.* **111**, 1035–1046 (2024).
81. Wainberg, M. et al. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* **51**, 592–599 (2019).
82. Zuber, V. et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am. J. Hum. Genet.* **109**, 767–782 (2022).
83. Hukku, A., Sampson, M. G., Luca, F., Pique-Regi, R. & Wen, X. Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. *Am. J. Hum. Genet.* **109**, 825–837 (2022).
84. Wallace, C. Statistical testing of shared genetic control for potentially related traits. *Genet. Epidemiol.* **37**, 802–813 (2013).
85. Sullivan, P. bip2024. *figshare* https://doi.org/10.6084/m9.figshare.27216117.v2 (2024).

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (82225016 to Ming Li and 82222024 and U24A20699 to X.X.); the National Key Research and Development Program of China (2023YFA1800500 to Ming Li, 2024YFA1803200 to X.X., 2023YFE0119400 to W.Y. and 2023YFC2605400 to S.X.); the National Natural Science Foundation of China (82230044 to Y.G., U21A20364 to Z.L., 82330042 to W.Y., 81930033 to Y.F., 32030020 to S.X., 32470649 to Y.L. and 323B2013 to X.T.); the Yunnan Fundamental Research Projects (202201AS070048 and 202401AS070080 to Ming Li, 202401AS070084 and 202501AV070009 to X.X. and 202305AH340006 to Y.-G.Y.); the Municipal Key R&D Program of Ningbo (2022Z127 to C.W.); the Jiangsu Provincial Key Research and Development Program (BE2020661 to L.H.); the Spring City Plan: the High-level Talent Promotion and Training Project of Kunming (2022SCP001 to Ming Li); the Shanghai Mental Health Centre Clinical Research Center Major Project (CRC2018ZD02 to Y.F.); the Shanghai Science and Technology Commission Program (23JS1410100 to S.X.); the Chinese Academy of Sciences (CAS) 'Light of West China' Program (xbzg-zdsys-202312 to X.X. and xbzg-zdsys-202404 to Ming Li); the Japan Agency for Medical Research and Development (21wm0425008 to N.I.); and the Open Program of Yunnan Key Laboratory of Animal Models and Human Disease Mechanisms (AMHD-2024-2, AMHD-2024-6 and AMHD-2024-8 to Ming Li). X.X. was also supported by the CAS 'Light of West China' Program, the CAS Youth Innovation Promotion Association and the Yunnan Revitalization Talent Support Program Young Talent Project. Ming Li was also supported by the Yunnan Revitalization Talent Support Program Yunling Scholar Project. PsychENCODE Consortium data were generated as part of the PsychENCODE Consortium. See https://doi.org/10.7303/syn26365932 for a complete list of grants and principal investigators. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Ming Li and X.X. conceived and designed the study. C.-Y.Z. performed most statistical analyses on the GWAS data, with the help of Miao Li and Y.W. P.S., L.H., Y.G., J.-Z.Y., N.Z., X.F., L.G., J.Y., H.-Y.J., Y.-Q.C., S.M., Q.G., Y.S., Y.L., N.Q., X.-Y.Y., L.W., Y.Y., C.W., L.L., D.Z., X.L., X.C., C.Z., J.C., X.S., J.T., J.C., W.F., Wei Tang, Wenxin Tang, W.L., X.T., X.Z., Y.L., C.W., Z.L., S.X., W.Y., Y.F. and F.Z. contributed to sample collections. The GeseDNA Research Team helped with sample collections. N.I., M.I. and T.S. contributed to the data in the Japanese sample. Y.-G.Y. and H.-C.S. helped with all aspects of study design and results interpretation. Ming

Li, X.X. and C.-Y.Z. drafted the first version of the paper. All authors revised the paper critically and approved the final version.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41593-025-02147-2.

Correspondence and requests for materials should be addressed to Yiru Fang, Feng Zhu, Xiao Xiao or Ming Li.

Peer review information *Nature Neuroscience* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Yiru Fang, Feng Zhu, Xiao Xiao, Ming Li

Last updated by author(s): Oct 16, 2025

Reporting Summary

Nature Portfolio 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 Portfolio 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	No softwares used for data collection.
Data analysis	Pre-GWAS QC, imputation, association analysis and meta-analysis: PLINK v1.9 (https://www.cog-genomics.org/plink/) was used for GWAS and meta analysis; EIGENSTRAT v7.1.2 (https://github.com/DReichLab/EIG/tree/master/EIGENSTRAT) was used for principle component analysis; SHAPEIT4 (https://github.com/odelaneau/shapeit4) and IMPUTE5 (https://jmarchini.org/software/#impute-5) were used for genotype imputation. SNP2HLA v1.0.3 (https://software.broadinstitute.org/mpg/snp2hla/) was used for imputation of the MHC region in Han Chinese samples. Genetic comparison across traits and populations: LDSC v1.0.1 (https://github.com/bulik/ldsc) and MiXeR v1.3 (https://github.com/precimed/mixer) were used for cross-trait genetic relationship estimation; Popcorn v1.0 (https://github.com/brielin/Popcorn) and PRS-CS v1.1.0 (https://github.com/getian107/PRScs) were used for trans-ancestry genetic correlation analyses. Post-GWAS analysis: LOLA R package, MAGMA v1.10 (https://ctg.cncr.nl/software/magma), MAGMA_Celltyping R package v2.0.11 (https://github.com/neurogenomics/MAGMA_Celltyping) and EWCE R package v1.6.0 (https://github.com/NathanSkene/EWCE) were used for enrichment analyses; FUMA v1.6.4 (https://fuma.ctglab.nl/) was used for SNP-to-GENE annotation;

SMR v1.0.3 (https://yanglab.westlake.edu.cn/software/smr/#Overview), MAGMA v1.10 (https://ctg.cncr.nl/software/magma), Polyfun (https://github.com/omerwe/polyfun), PoPS (https://github.com/FinucaneLab/pops), FUSION (https://github.com/gusevlab/fusion_twas) and COLOC v5.2.3 R package were used for credible gene prioritization;
SynGO tools v1.2 (https://www.syngoportal.org/), OpenTargets (https://platform.opentargets.org/api) and DGIdb 5.0 API (https://dgidb.org/api) were used for credible gene annotations;
R v4.0.3 was used for statistical tests.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The genome-wide summary statistics of PGC4 GWAS are available via Figshare at https://figshare.com/articles/dataset/bip2024/27216117. The SNP expression weights of PsychENCODE used in this study are available at http://resource.psychencode.org/. The PsychENCODE cis-eQTL data and the BrainMeta V2 cis-eQTL data are available via the BrainMeta portal at https://yanglab.westlake.edu.cn/software/smr/#DataResource. The cell-type specific gene expression matrices from two BICCN studies29,30 for enrichment analysis were downloaded from CELLxGENE database at https://cellxgene.cziscience.com/collections/bacccb91-066d-4453-b70e-59de0b4598cd and https://cellxgene.cziscience.com/collections/283d65eb-dd53-496d-adb7-7570c7caa443 separately. The summary statistics of BD cross-ancestry meta-analysis is publicly available on the Scientific Data Center of Kunming Institute of Zoology (https://datacenter.kiz.ac.cn/Home/DataContent?data_gd=4c22800a-d798-169c-7184-430976572334). However, according to the related policy, depositing genetic data (including the summary statistics) of Han Chinese only in a third-party website (without application) is not allowed before approval. Alternatively, GWAS summary statistics in Chinese population can be requested from the China National Genomics Data Center (https://ngdc.cncb.ac.cn/gvm/), with the data accession number GVP000051, and the GWAS summary statistics can also be available upon reasonable request from the corresponding author.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Findings in our study were not sex-specific and we included both males (52.8%) and females (47.2%) in our cohort. Self-reporting gender was recorded upon recruitment of both cases and controls and the biological sex was also determined by homozygosity rate across all X-chromosome SNPs. The biological sex info was included in the subsequent logistic regression model as a covariate.
Reporting on race, ethnicity, or other socially relevant groupings	In our study, the primary socially constructed categorization variable used is ethnicity, specifically focusing on the Han Chinese population. The focus on the Han Chinese population is driven by the significant lack of BD GWAS in Chinese population despite the large number of affected individuals. The samples were collected from public hospitals across various provinces in China. We ensured that our sampling did not target any specific socioeconomic group or demographic segment. Since public hospitals serve a wide demographic, our sample is representative of the general population and does not exhibit bias towards any particular socioeconomic status.
Population characteristics	We collected 5,164 cases and 13,460 controls of Han Chinese origin.Genomic DNA was extracted from the peripheral blood leukocytes or saliva of the participants. Individuals were genotyped using the Illumina Genome-Wide Asian Screening Array (ASA) Chip or the Illumina Infinium Global Screening Array (GSA) Chip. There are 45.43% males + 54.56% females in 5,164 BD patients and 55.6% males + 44.4% females in 13,460 healthy controls. The average age of cases and controls in our Han Chinese cohort were 33.12±11.28 and 27.70±6.3 years, respectively. The population characteristics details for the included public GWAS studies have been described in the original GWAS publications and no changes due to additional inclusion or exclusion criteria have been imposed.
Recruitment	All BD patients signed up for the study at their local mental health centers and psychiatric departments of general hospitals in multiple provinces (Yunnan, Jiangsu, Shanghai, Shandong, Henan, Zhejiang, Hubei, Hunan etc.) of Chinese mainland. Patients were diagnosed with BD by one chief psychiatrist and one attending psychiatrist. Each BD patient was independently interviewed by both clinicians, and those diagnosed with the same Axis I disorder by both psychiatrists were recruited. All diagnoses were confirmed by a research psychiatrist through an Extensive Clinical Interview and a Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Version (SCID-P). Cases were excluded if they were diagnosed with other psychiatric disorders or neurological disorders, being pregnant, or breast-feeding at the time of study. The control subjects were recruited from Chinese mainland, and were not specially screened for psychiatric disorders by psychiatrists but had no current self-reported serious illnesses or disabilities, or any personal or family histories (including first-, second- and third-degree relatives) of psychiatric illnesses.
Ethics oversight	The study protocol of BD GWAS was approved by the Institutional Review Board of Kunming Institute of Zoology, Chinese Academy of Sciences, and ethics committees of all participating hospitals and universities.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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

Life sciences study design

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

Sample size	For Han Chinese BD GWAS, sample size (Ncase = 5,164, Ncontrol = 13,460) was not predetermined but reflected our best efforts to collect all available BD patients and healthy controls in Han Chinese population, which is also the largest-to-date GWAS cohort in this population. For the EAS GWAS, we integrated the largest EAS BD GWAS from PGC4 (Ncase = 4,479, Ncontrol = 75,725) reaching a doubled sample size including 9,643 BD cases and 89,185 controls. For the trans-ancestry meta-analysis, we integrated data from the largest EUR BD GWAS (Ncase = 59,287, Ncontrol = 781,022) reaching a final sample size including 68,930 BD cases and 870,207 controls. For public datasets (eQTL, scRNA-seq) used in the post-GWAS analyses, we used the largest and most credible publicly available datasets appropriate for our study.
Data exclusions	In Han Chinese BD GWAS, the exclusion thresholds for subjects were: (1) inconsistent sex information estimated by the homozygosity rate across all X-chromosome SNPs; (2) a genotyping missing rate exceeding 3% or heterozygosity rates over 6 s.d.; (3) duplicates and relatives (IBD, identity by decent) > 0.1875; (4) On the basis of visual inspection of plots of principal component 1 (PC1) and principal component 2 (PC2) for each dataset, we excluded obvious outliers to obtain more homogeneous datasets based on the principal component analysis (PCA). SNPs with a call rate < 95%, a minor allele frequency (MAF) < 1%, deviation from Hardy-Weinberg equilibrium (HWE) P < 1e–5 in all samples, or with significantly different missing genotype rates between cases and controls were excluded from further analyses.
Replication	We included all samples in the primary analysis to maximize statistical power. No replication samples were available.
Randomization	No randomization was conducted. In this large-scale genetics study, randomness is achieved by the nature of how alleles are distributed in the population.
Blinding	No blinding was carried out. The nature of the study is in of itself blind to study recruiters. No one would beforehand know the genotype of recruited participants.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*