

# Cross-ancestry genome-wide association study and systems-level integrative analyses implicate new risk genes and therapeutic targets for depression

Received: 23 January 2024

Accepted: 23 October 2024

Published online: 24 February 2025

 Check for updates

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

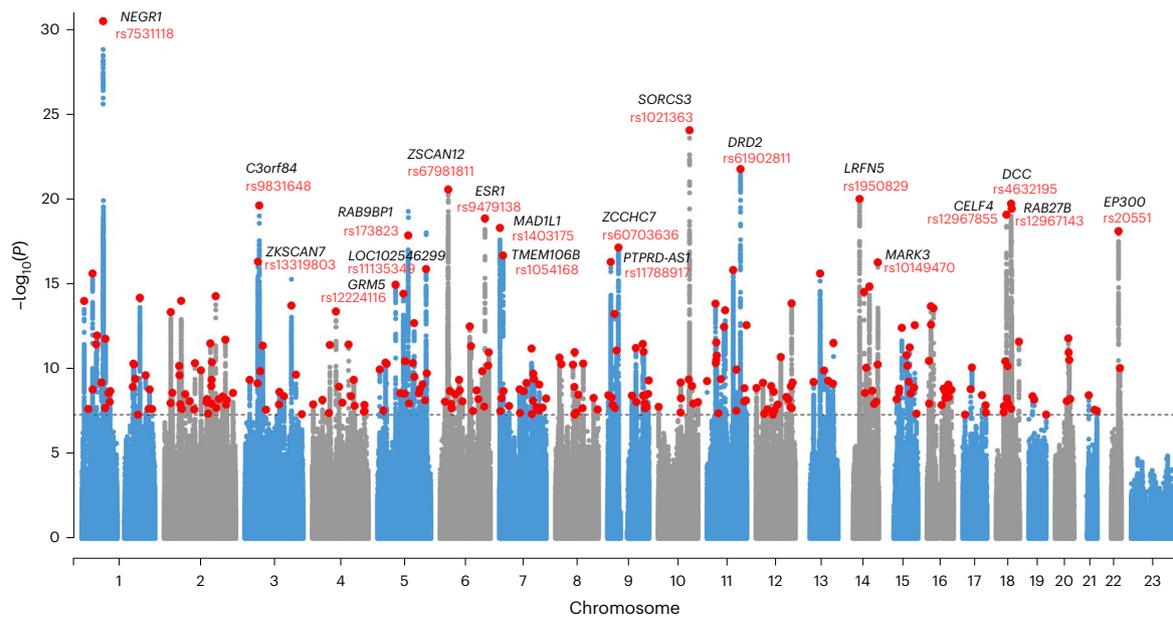
Deciphering the genetic architecture of depression is pivotal for characterizing the associated pathophysiological processes and development of new therapeutics. Here we conducted a cross-ancestry genome-wide meta-analysis on depression (416,437 cases and 1,308,758 controls) and identified 287 risk loci, of which 49 are new. Variant-level fine mapping prioritized potential causal variants and functional genomic analysis identified variants that regulate the binding of transcription factors. We validated that 80% of the identified functional variants are regulatory variants, and expression quantitative trait loci analysis uncovered the potential target genes regulated by the prioritized risk variants. Gene-level analysis, including transcriptome and proteome-wide association studies, colocalization and Mendelian randomization-based analyses, prioritized potential causal genes and drug targets. Gene prioritization analyses highlighted likely causal genes, including *TMEM106B*, *CTNND1*, *AREL1* and so on. Pathway analysis indicated significant enrichment of depression risk genes in synapse-related pathways. Finally, knockdown of *Tmem106b* in mice resulted in depression-like behaviours, supporting the involvement of *Tmem106b* in depression. Our study identified new risk loci, likely causal variants and genes for depression, providing important insights into the genetic architecture of depression and potential therapeutic targets.

Depression is one of the most prevalent mental disorders<sup>1</sup> and a leading cause of disability worldwide<sup>2</sup>. Although recent genome-wide association studies (GWASs) have reported multiple risk loci for depression<sup>3–8</sup>, much of the underlying heritability remains unexplained. In addition, most depression GWASs have been conducted in populations of European ancestry, potentially missing important genetic insights into depression. More importantly, the causal variants and genes remain largely unknown for most reported risk loci, hampering the translation of genetic findings into clinical applications and therapeutics. Hence, the discovery of new genetic risk loci and functional characterization

of the identified risk variants and genes will provide important insights into depression pathophysiology and therapeutic targets.

In this study, we first conducted a large-scale cross-ancestry meta-analysis (416,437 cases and 1,308,758 controls) on depression. Then, based on the results of the meta-analysis, we conducted comprehensive prioritization and integrative analysis to prioritize the potential causal variants and genes. We prioritized potential causal variants and validated the regulatory effect of the identified functional variants. In addition, we also prioritized likely causal genes, including *TMEM106B*, *CTNND1*, *EPHB2* and so on. Finally, we found that *Tmem106b* knockdown

✉ e-mail: [yygyh2000@sina.com](mailto:yygyh2000@sina.com); [litaozjusc@zju.edu.cn](mailto:litaozjusc@zju.edu.cn); [janemengzhang@vip.163.com](mailto:janemengzhang@vip.163.com); [luoxiongjian@seu.edu.cn](mailto:luoxiongjian@seu.edu.cn)



**Fig. 1 | Manhattan plot of the GWAS meta-analysis.** Manhattan plot of the GWAS meta-analysis ( $n = 416,437$  cases and  $1,308,758$  controls). The y axis shows  $-\log_{10}(P)$  of GWAS meta-analysis results from the inverse-variance-weighted fixed-effects model. The dashed line shows the GWS  $P$  threshold ( $P < 5.0 \times 10^{-8}$ ). The red dots highlight all 287 risk loci identified for depression. SNP **rs7531118** near *NEGR1*

shows the most significant association. Only the top 20 risk loci are highlighted, the lead SNPs are shown in red, and the nearest genes to the lead SNPs are shown in black text. The  $P$  values are from fixed-effects model GWAS meta-analysis, and all association analyses were two-sided.

resulted in depression-like behaviours in mice, providing animal model evidence supporting that *TMEM106B* is a depression risk gene. Our study identified new risk loci, likely causal variants and genes associated with depression, providing important insights into the genetic architecture of depression and potential therapeutic targets.

## Results

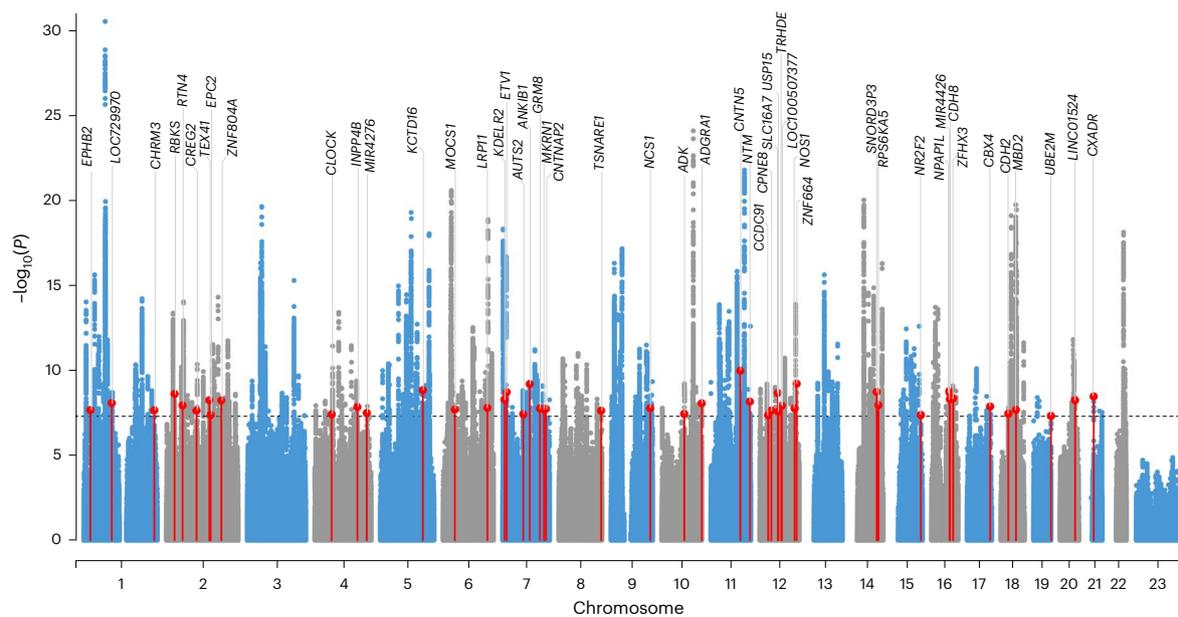
### Cross-ancestry meta-analysis identified 287 depression risk loci

We conducted a meta-analysis by combining genome-wide associations reported by 6 previous studies (Methods, Supplementary Fig. 1 and Supplementary Table 1): Million Veteran Project (MVP) (from the USA, two cohorts were included: cohort of European ancestry, with 83,810 cases and 166,405 controls; and cohort of African ancestry, with 25,843 cases, and 33,757 controls)<sup>3</sup>, FinnGen (<https://www.finnngen.fi/en>) (from Finland, 33,812 cases and 271,380 controls), studies reported by Howard et al. (UK Biobank (UKB) + Psychiatric Genomics Consortium (PGC) + 23andMe) (multiple cohorts, mainly from the USA and Europe, 246,363 cases and 561,190 controls)<sup>5</sup>, the Australian Genetics of Depression Study (AGDS) (from Australia, 13,318 cases and 12,684 controls)<sup>9</sup>, Giannakopoulou et al. (multiple cohorts from East Asia, 12,455 cases and 85,548 controls)<sup>10</sup> and Sakaue et al. (from Japan, BioBank Japan, 836 cases and 177,794 controls)<sup>11</sup>. Fixed-effects meta-analysis (a total of 416,437 cases and 1,308,758 controls) identified 287 independent genomic risk loci ( $P < 5 \times 10^{-8}$ ) (please refer to Methods for risk loci definition) (Fig. 1, Supplementary Fig. 2 and Supplementary Tables 2 and 3), of which 49 were new (genome-wide significant (GWS) loci that have not been reported by each included study and previous depression GWAS<sup>3-5,9-13</sup> were considered as new associations) (Fig. 2 and Supplementary Table 3). The most significant single-nucleotide polymorphism (SNP) **rs7531118** is located approximately 89 kb downstream of the *NEGR1* (1p31.1) (Supplementary Fig. 3a), and the second most significant SNP **rs1021363** is located in the intron 2 of the *SORCS3* (Supplementary Fig. 3b). Linkage disequilibrium score regression (LDSC) showed that polygenicity rather than confounding factors accounts for most of the associations (lambda genomic control 1.54, intercept (standard error (s.e.)) 1.04 (0.01), ratio (s.e.) 0.03 (0.008)). To assess the

variability of allelic effect in cross-ancestry meta-analysis, we further performed a meta-analysis using the random-effects model for variants that showed heterogeneity ( $P_{\text{Cochran's } Q} < 0.05$ ). The results showed that, among the 287 risk loci identified by the fixed-effects model, 272 were also supported by the random-effects model (Supplementary Fig. 4 and Supplementary Table 2), indicating the robustness of the associations. These findings further expand the risk loci of depression.

### Heritability and genetic correlations

Considering the potential effect of linkage disequilibrium (LD) panel mismatch when using cross-ancestry samples to estimate SNP heritability ( $h^2$ )<sup>14</sup>, we used GWAS meta-analysis from populations of European ancestry (377,303 cases and 1,011,659 controls) for heritability estimate. The liability scale  $h^2$  estimate (assuming lifetime risk of 0.15) was 0.073 (estimated by LDSC<sup>15</sup>, s.e. 0.003) and 0.095 (calculated by SumHer<sup>14</sup> from the LDAK software package, with the use of the Baseline-LD (BLD) model from LDAK software, s.e. 0.002). These values are consistent with those previously reported (0.080–0.090 in liability scale)<sup>4,5,9</sup>. We further calculated the genetic correlations between the six included depression studies (MVP, FinnGen, AGDS, 23andMe–UKB–PGC, Howard et al. and Sakaue et al.). Taking into account that the genetic correlation estimates for traits with a low heritability  $Z$  score are generally too noisy to report<sup>11</sup>, we focused on the GWAS datasets with an SNP-based  $Z$  score  $> 4$  (including 23andMe–UKB–PGC, AGDS, FinnGen, MVP–European ancestry (EUR)) and found that the genetic correlations between different depression studies are highly significant (Supplementary Table 5a,b). Given the genetic structure and LD structure differences in different ancestral populations, we further used Popcorn<sup>16</sup>, which can calculate cross-ancestry genetic correlation, to estimate the genetic correlations between the EAS and EUR GWAS datasets (only datasets with an SNP-based  $Z$  score  $> 2$  were included). The results of Popcorn also indicate the high degree of genetic correlations between the included datasets (Supplementary Table 5a,b). Meanwhile, evidence based on LD intercepts (1.04, s.e. 0.010) and attenuation ratio (0.03, s.e. 0.008) also showed negligible inflation or confounding in our meta-analysis results (Supplementary Table 4). Finally, we investigated the genetic



**Fig. 2 | Novel risk loci identified in the cross-ancestry GWAS meta-analysis.** Manhattan plot of the cross-ancestry GWAS meta-analysis, with 49 new risk loci identified in this study highlighted. The lead SNPs are shown in red dots, and the nearest genes to the lead SNPs are shown in italic text. The  $P$  values are from fixed-effects model GWAS meta-analysis, and all association analyses were two-sided.

correlations between depression (that is, the meta-analysis results from populations of European ancestry, including MVP-EUR, 23andMe–UKB–PGC, AGDS and FinnGen datasets) and other brain disorders and intelligence (Supplementary Table 6a,b). The top three disorders that showed the most significant genetic correlations with depression were anxiety disorders, post-traumatic stress disorder and neuroticism (Supplementary Fig. 5 and Supplementary Table 6a,b). Most of the analysed traits showed positive correlations with depression. However, intelligence showed negative correlations.

### Functional genomics identified potential causal variants

Identifying functional (or potential causal) variants is crucial for follow-up mechanistic investigation and functional characterization. To identify the functional variants from the identified risk loci, we conducted a functional genomic analysis, as previously described<sup>17–19</sup>. We identified 64 functional SNPs that affect the binding of transcription factors (TFs) (that is, TF binding-affecting SNPs) (Supplementary Table 7). Among these 64 TF binding-affecting SNPs, 37 affect the binding of CCCTC-binding factor (CTCF), 10 affect the binding of RE-1 silencing transcription factor (REST) and 6 SNPs interfere with more than one TFs simultaneously (Fig. 3 and Supplementary Table 7). About 40% of TF binding-affecting SNPs are located in intronic regions (Fig. 3a), indicating the pivotal role of intronic variants in depression. Notably, 34 TF-disrupting SNPs showed GWS associations with depression, and the CTCF binding-affecting SNP *rs7531118* (located downstream of the *NEGR1*) showed the most significant association (Supplementary Table 7). These results pinpointed the functional (or potential causal) variants from the reported risk loci and suggested that affecting the binding of TFs is a major manner by which these functional SNPs exert their biological effect on depression.

### Regulatory effects of the TF binding-affecting SNPs

To validate the regulatory effect of the identified TF binding-affecting SNPs, we conducted dual-luciferase reporter gene assays for all TF binding-affecting SNPs (Supplementary Table 7). Reporter gene assays revealed that 51 out of 63 SNPs (the vector for one SNP was not successfully constructed owing to its complex genomic sequence) showed a regulatory effect, that is, different alleles of these 51 SNPs affected the

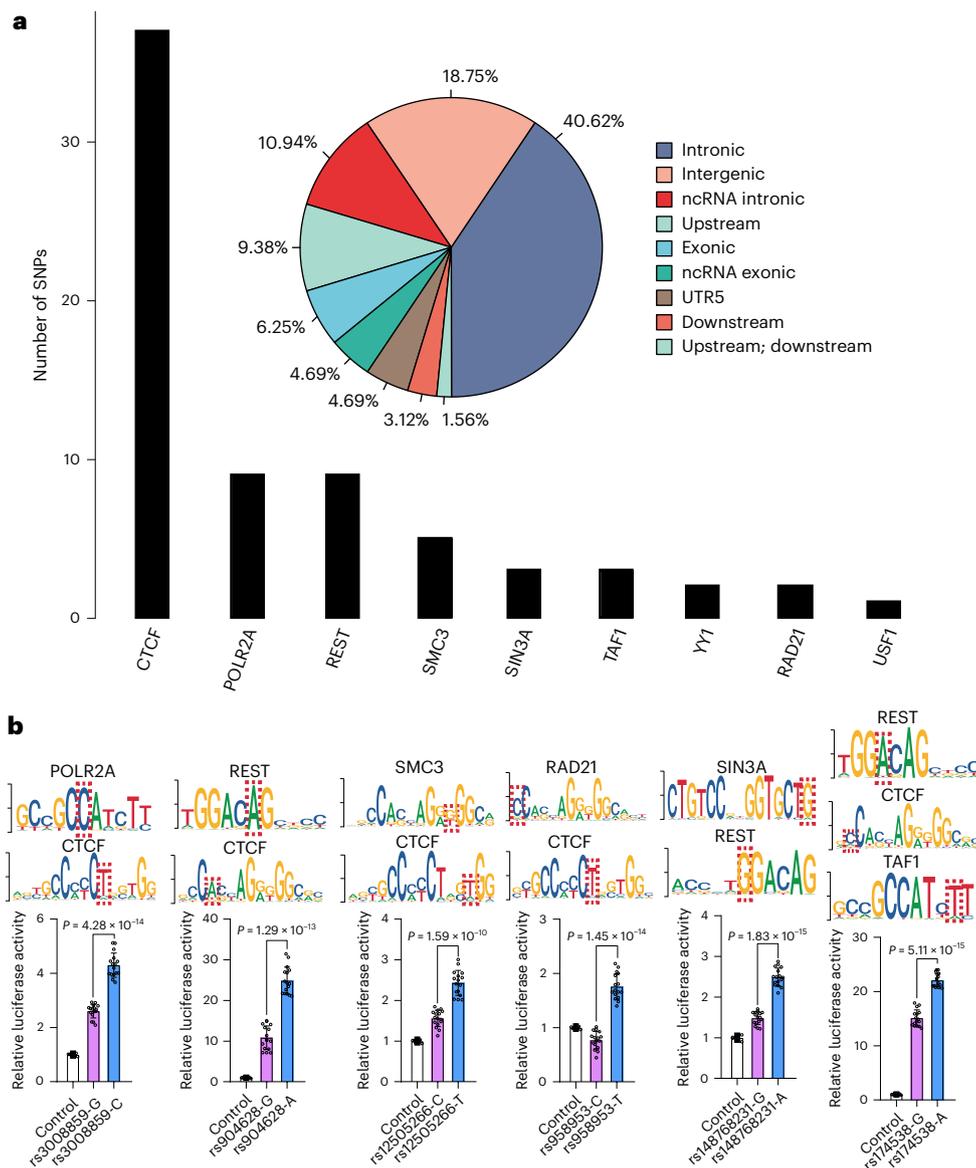
luciferase activity significantly ( $P < 0.05$ ) (Figs. 3b and 4 and Supplementary Fig. 6). These results provided experimental evidence supporting that most of the identified TF binding-affecting SNPs are regulatory variants. Considering that the interaction between TFs and regulatory sequence has a crucial role in expression regulation, these findings also indicate that these functional SNPs may confer risk of depression by regulating gene expression. These results provide important insights into the regulatory mechanisms of depression risk variants.

### Fine mapping prioritized potential causal variants

To identify the potential causal variants from the identified risk loci, we used a sum of the single-effects model fine-mapping approach, MESuSiE<sup>20</sup>, which conducts fine mapping using associations from multi-ancestry GWAS studies. Considering that MESuSiE is developed to identify a set of candidate SNPs shared between two ancestral populations, and the African ancestry cohort included in this study did not identify any GWS risk SNPs ( $\text{GWAS } P < 5 \times 10^{-8}$ ), we conducted fine mapping using GWASs only from the Asian and European ancestries. MESuSiE prioritized 122 high-confidence potential causal variants (MESuSiE posterior inclusion probability (PIP)  $> 0.5$  in either Asian or European cohorts) for 208 risk loci (Supplementary Table 8). Notably, MESuSiE also prioritized ten shared potential causal SNPs in Asian and European ancestry, suggesting that these SNPs may have ancestral shared causal effects in depression (Supplementary Table 8).

### Potential target genes of the risk variants

Most of the identified depression risk variants are located in the non-coding region, implying that the regulation of gene expression is a key potential pathway between risk variants and associated phenotypes. To identify potential target genes of the identified risk variants, we used the BrainMeta v2 (ref. 21) dataset, which contains expression quantitative trait loci (eQTL) data from 2,865 human brain transcriptomes, to identify associations between risk variants and gene expression in the human brain. Lead SNPs, TF binding-affecting SNPs identified by functional genomics, and credible causal SNPs prioritized by MESuSiE were used for eQTL analysis. For lead SNPs, 161 showed significant associations when adjusted by Bonferroni correction ( $P < 3.90 \times 10^{-5}$ ) (Supplementary Table 9). We next examined the



**Fig. 3 | SNPs that affect the binding of TFs.** **a**, Left: the box plot shows the number of depression risk SNPs that affect the binding of individual TF based on matched PWMs from ChIP-seq and PWM database. Right: the distribution of the 64 identified TF binding-affecting SNPs in the human genome. A large proportion of the identified regulatory SNPs were located in intronic regions. **b**, SNPs that affect the binding of two or three TFs. Top: the PWM (that is, binding motif) of the corresponding TFs. Bottom: results of reporter gene assays.

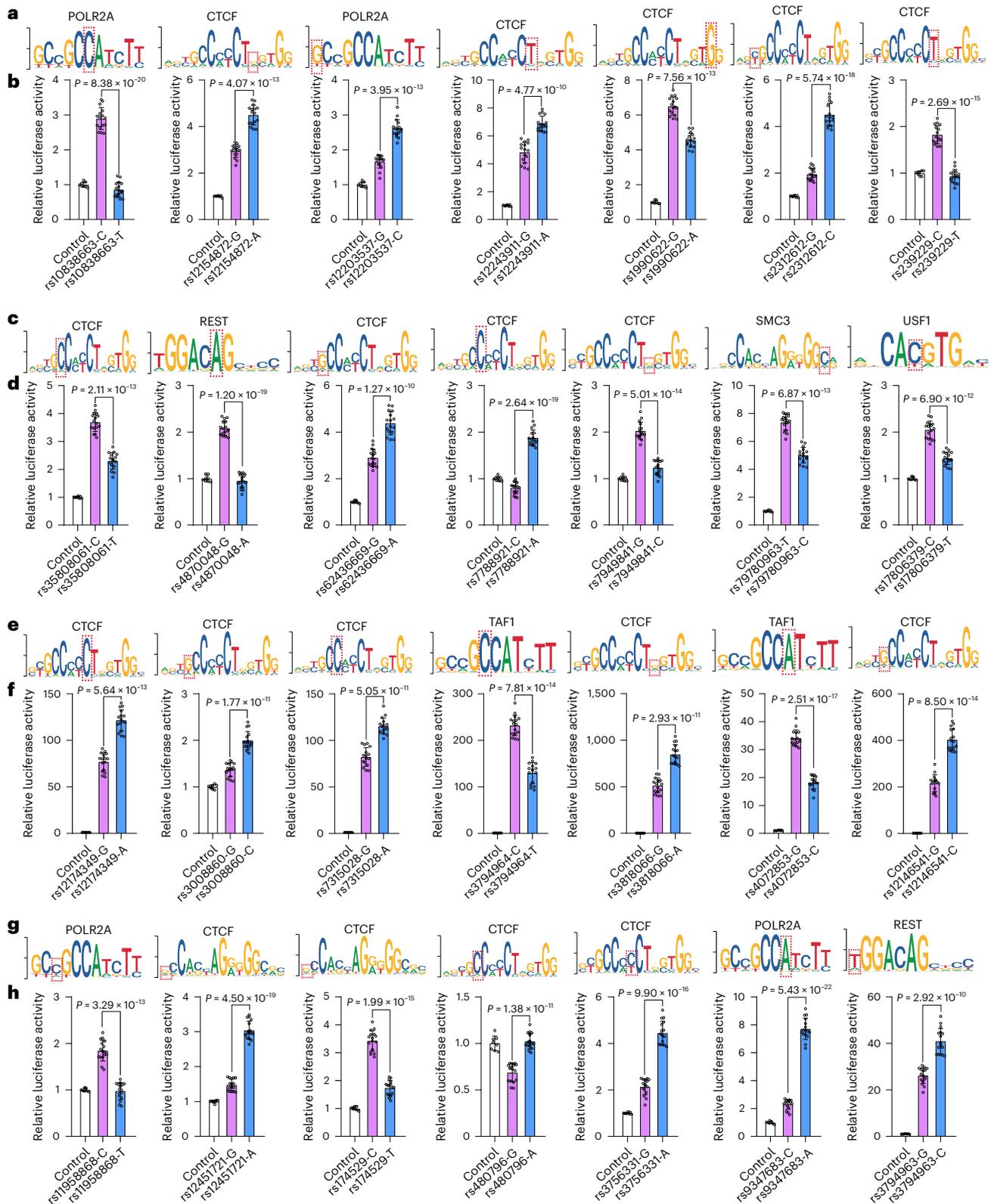
The Renilla internal control was used to normalize the luciferase activity, and the y axis shows the relative luciferase activity.  $n = 8$  for the control group,  $n = 16$  per experimental group for SH-SY5Y cells. Two-tailed Student's *t*-test was used to assess whether the difference reached the significance level ( $P < 0.05$ ). Data represent mean  $\pm$  s.d. The dashed red box highlights the test SNP. Different alleles of all SNPs showed significant differences in luciferase activity, indicating that these SNPs are functional variants. ncRNA, non-coding RNA.

associations between the TF binding-affecting SNPs (Supplementary Table 7) and gene expression. Among the 64 TF binding-affecting SNPs, 50 showed significant associations with gene expression in the human brain (Bonferroni-corrected  $P < 2.40 \times 10^{-5}$ ) (Supplementary Table 10). Finally, we found that 36 SNPs prioritized by MESuSiE showed significant associations with gene expression in the human brain (Supplementary Table 11). Taken together, these results identified the potential target genes regulated by the lead and prioritized functional risk SNPs, suggesting that these functional variants confer the risk of depression through regulating the expression of these target genes.

### TWAS identified depression risk genes

Considering that transcriptome-wide association study (TWAS) first constructs SNP-gene expression weights between SNPs and gene expression using a relatively small reference panel (that is,

TWAS relies on ancestry-specific LD reference panels) and the eQTL data were from populations of European ancestry, we performed a TWAS by integrating depression GWAS results from populations of European ancestry (Supplementary Fig. 7) and brain eQTL data from PsychENCODE (EUR ancestry)<sup>22</sup>. We used FUSION<sup>23</sup> to identify genes whose genetically regulated expression levels are associated with depression in European populations. TWAS identified 179 genes whose genetically regulated expression is associated with depression (Bonferroni-corrected  $P < 3.91 \times 10^{-6}$ ) (Fig. 5a and Supplementary Table 12). Transcriptome-wide significant genes include *RPL31P2*, *ZSCAN12P1*, *AREL1*, *HIST1H4L*, *RP11-73M18.6*, *TMEM106B* and so on. Of note, *TMEM106B* (near the lead SNP rs1054168) is one of the top associations in TWAS. Besides, functional genomics also revealed that the functional SNP rs1990622 located downstream of *TMEM106B* showed a strong association with depression ( $P = 2.30 \times 10^{-16}$ ) (Supplementary



**Fig. 4 | Reporter gene assays validated the regulatory effect of the identified TF binding-affecting SNPs. a–h,** The binding motifs (PWM) of the corresponding TFs (**a, c, e, g**) and the results of reporter gene assays (**b, d, f, h**). DNA fragments (about 600 bp) containing different alleles of the TF binding-affecting SNPs were amplified and cloned into the pGL4.11-basic vector (for promoter activity detection) or pGL3-promoter vector (for enhancer activity detection). For each pair of constructs, there was only one nucleotide difference at the test SNP, so the reporter gene assays will reveal if different alleles of the test SNP

cause differences in luciferase activity. The Renilla internal control was used to normalize the luciferase activity, and the y axis shows the relative luciferase activity.  $n = 8$  for the control group,  $n = 16$  per experimental group for SH-SY5Y cells. Two-tailed Student's *t*-test was used to assess whether the difference reached the significance level ( $P < 0.05$ ), and all the data met the assumptions of the Student's *t*-test, including normality and equal variances. Data represent mean  $\pm$  s.d. The dashed red box highlights the test SNP.

Table 12). Consistent with functional genomics and eQTL analysis, reporter gene assays validated the regulatory effect of **rs1990622**, with the G allele of **rs1990622** conferring significantly higher luciferase activity compared with the A allele ( $P = 7.56 \times 10^{-13}$ ; Fig. 4). These findings not only identified risk genes whose genetically regulated expression is associated with depression but also prioritized **rs1990622** as a functional risk variant that is associated with risk of depression by regulating *TMEM106B* expression.

### PWAS nominated proteins associated with depression

Similar to the TWAS analysis, by integrating depression meta-analysis results from European-ancestry GWAS datasets and two independent human brain protein quantitative trait datasets (also from European populations, the Religious Orders Study/Memory and Aging Project (ROSMAP) and Banner)<sup>24</sup>, we carried out proteome-wide association studies (PWASs) (Supplementary Figs. 8 and 9). To maximize the power of the PWAS analysis, we further performed a meta-analysis to combine the PWAS results from these two datasets using METAL<sup>25</sup> (Fig. 5b). PWAS using protein quantitative trait loci (pQTL) data from the meta-analysis identified 76 proteins whose genetically regulated abundance was associated with depression (Bonferroni-corrected  $P < 2.47 \times 10^{-5}$ ) (Fig. 5b and Supplementary Table 13). Of note, 21 proteome-wide significant proteins also showed significant associations at transcriptome-wide level, and 17 had the same effect direction, including *TMEM106B*, *RAB27B*, *CTNND1*, *GPX1*, *DCC*, *B3GALTL*, *CSE1L*, *LRP4*, *HARS2*, *GMPPB*, *SLC25A12*, *STAU1*, *PPP3CC*, *BTN3A3*, *INPP4B*, *PLEKHB1* and *SLC30A9*, strongly suggesting their genes are promising risk genes for depression. Interestingly, a Mendelian randomization (MR) study conducted by Deng et al.<sup>26</sup> also found that the expression abundances of *RAB27B*, *GMPPB* and *TMEM106B* were associated with the depression risk at both protein and mRNA levels, suggesting a potential causal effect of these genes in depression. Considering the significant genetic correlation between depression and anxiety disorders ( $r_g = 0.94$ ) (Supplementary Fig. 5), we also compared the risk proteins identified in this study and proteins identified in our previous PWAS of anxiety disorders<sup>27</sup>. Three proteins, including *TMEM106B*, *RAB27B* and *CTNND1*, showed significant associations with both depression and anxiety disorders. These data strongly suggested that *TMEM106B*, *CTNND1*, *GMPPB* and *RAB27B* are promising candidates for depression.

### Colocalization analysis

To investigate if the GWAS and QTL signals were driven by the same variants, we conducted colocalization analysis<sup>28</sup>. Colocalization analysis using eQTL (from PsychENCODE) and GWAS signals showed shared variants for 126 genes ( $PP4 > 0.70$ ), including *RPL31P12*, *AREL1*, *TMEM106B*, *RAB27B*, *KLHDC8B*, *ZKSCAN7* and so on (Supplementary Tables 12 and 14). When we restricted colocalization analysis to transcriptome-wide significant genes (Bonferroni-corrected  $P$  threshold  $3.39 \times 10^{-6}$ ), 65 genes showed colocalization signals.

Colocalization of ROSMAP pQTL and GWAS signals identified 34 candidate proteins, including *TMEM106B* ( $PP4$  0.995), *B3GALTL* ( $PP4$  0.987), *CNNM2* ( $PP4$  0.987), *RAB27B* ( $PP4$  0.985) and so on (Supplementary Table 14). Colocalization of Banner pQTL and GWAS signals identified 24 proteins (Supplementary Table 14). Of note, ten genes were supported in the colocalization analysis of all three QTL datasets, including *TMEM106B*, *SLC25A12*, *RAB27B*, *PAMB4* and so on (Supplementary Table 14), suggesting that these genes are candidates for depression.

Given that standard colocalization analysis can infer colocalization between only two traits (such as GWAS and eQTL) at a time, we further utilized the moloc package<sup>29</sup> to perform colocalization analysis for three traits (that is, GWAS, eQTL and pQTL). Of note, six genes were supported in the moloc analysis, including *NCOA3*, *SULF2*, *RNU7-173P*, *SRMP1*, *RNA5SP486* and *RNU7-92P* (Supplementary Table 15), suggesting that these genes are high-confidence candidates for depression.

### Identification of potential therapeutic targets

To identify new drug targets and to seek potential drug repurposing opportunities, we performed MR analysis by using genetic variants associated with gene expression or protein abundance of 1,263 actionable drug proteins (approved or in clinical stage drugs therapeutic targets) curated by Gaziano et al.<sup>30</sup>.

MR analysis using eQTL from BrainMeta v2 dataset and GWAS signals identified nine promising actionable drug targets (Bonferroni-corrected  $P$  threshold  $< 7.50 \times 10^{-5}$ ), including *ESR2*, *P4HTM*, *CD40*, *EPHB2*, *PSMC3*, *GRIK2*, *PRKCD*, *FOLH1* and *CRHR1* (Fig. 6 and Supplementary Table 16). MR analysis using pQTL (ROSMAP dataset) and GWAS associations identified 13 promising candidate proteins, including *DAGLA*, *P4HTM*, *DAGLB*, *EPHB2*, *TAOK3* and so on (Supplementary Table 17). MR analysis using pQTL (Banner dataset) and GWAS associations identified eight proteins (Supplementary Table 18). Of note, *DAGLA*, *P2RX7* and *PSMB4* were supported in the MR analysis of both pQTL datasets. *P4HTM*, *CD40*, *EPHB2* and *GRIK2* showed MR significance at both mRNA and protein levels, suggesting that these genes are potential drug targets for depression. To identify potential therapeutic drugs from MR results, we highlighted potential therapeutic agents for each possible causal gene or protein based on the effect direction of the MR results. For instance, if MR beta  $> 0$ , we listed the inhibitor/blocker targeting the gene or protein, whereas if MR beta  $< 0$ , we listed the agonist/activator/positive modulator drugs (Supplementary Table 19). To mitigate potential biases in cross-population analysis, we also conducted a replication analysis using the EUR-only GWAS meta-analysis results. We observed a high level of reproducibility (the Jaccard index is 0.667 for Banner drug MR, 0.500 for ROSMAP drug MR and 0.700 for BrainMeta drug MR), indicating the reliability of the PWAS results (Supplementary Fig. 10). We further calculated the consistency of drug MR findings across transcriptomic and proteomic datasets. The results showed that the Jaccard index across BrainMeta eQTL and ROSMAP pQTL MR analyses is 0.222, and the Jaccard index between ROSMAP and Banner pQTL MR analyses is 0.167.

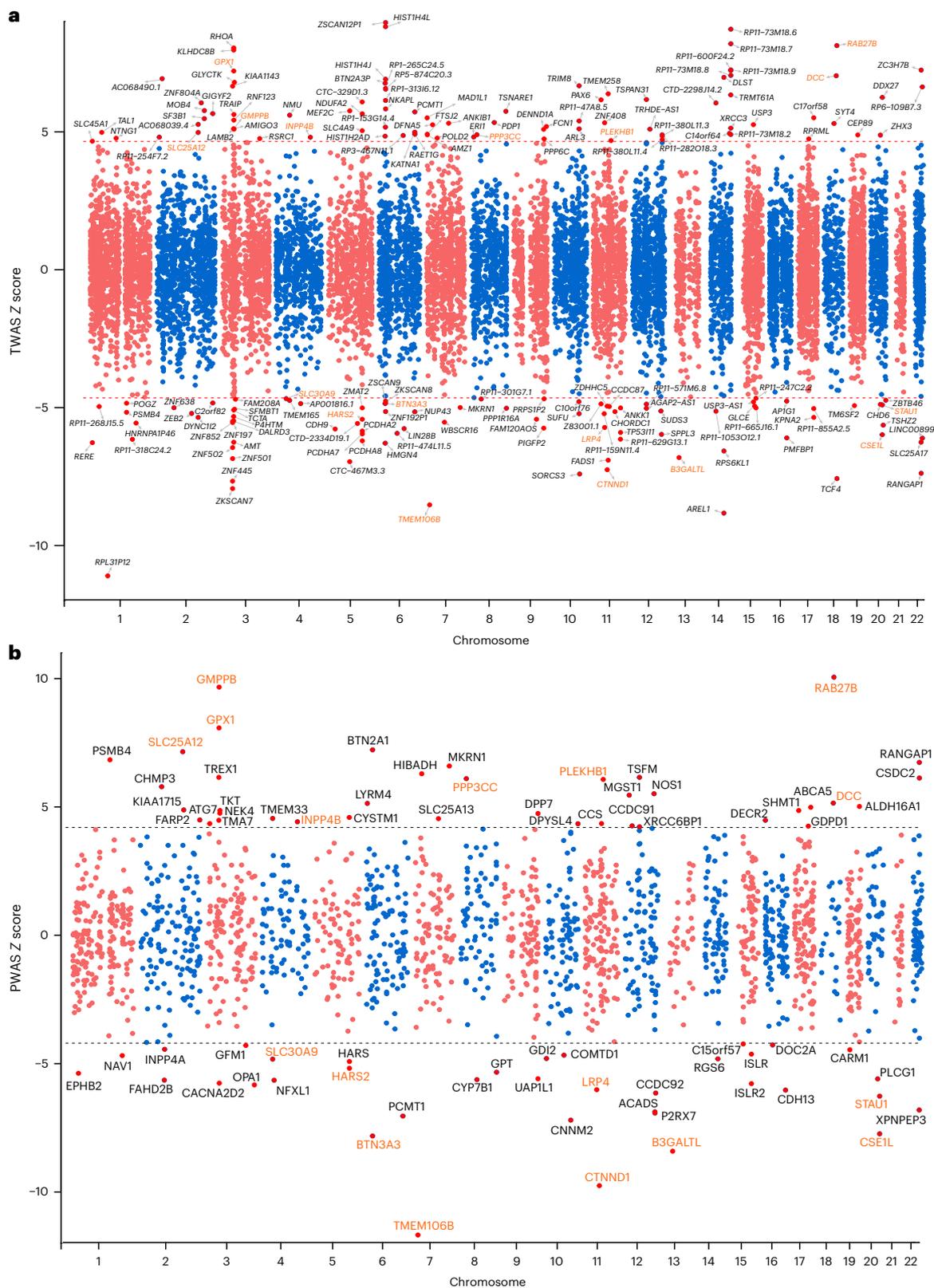
### Heritability enrichment and gene set analysis

Tissue-based heritability enrichment analyses showed that the heritability of depression is mainly enriched in the brain tissues. Of note, depression associations showed the most significant enrichments in the cerebellar hemisphere and frontal cortex (BA9) (Supplementary Figs. 11 and 12). Significant enrichments were also observed in other brain regions (false discovery rate (FDR)  $< 0.05$ ). Cell-type-specific enrichment analysis showed significant enrichments in excitatory and inhibitory neurons, neuroblasts and oligodendrocyte precursor cells (Bonferroni-corrected  $P$  threshold  $< 1.28 \times 10^{-3}$ ). These results identified neurons, neuroblasts and oligodendrocyte precursor cells as possible relevant cell types for depression (Supplementary Fig. 13).

We subsequently used MAGMA to carry out gene set enrichment analysis and identified 12 significant enriched Gene Ontology (GO) terms. Of note, synaptic and postsynaptic membranes showed the most significant enrichment. In addition, the branching morphogenesis of nerves also showed significant enrichment (Supplementary Fig. 14). Taken together, these results identified neurons and synapses as the major cells and components for depression.

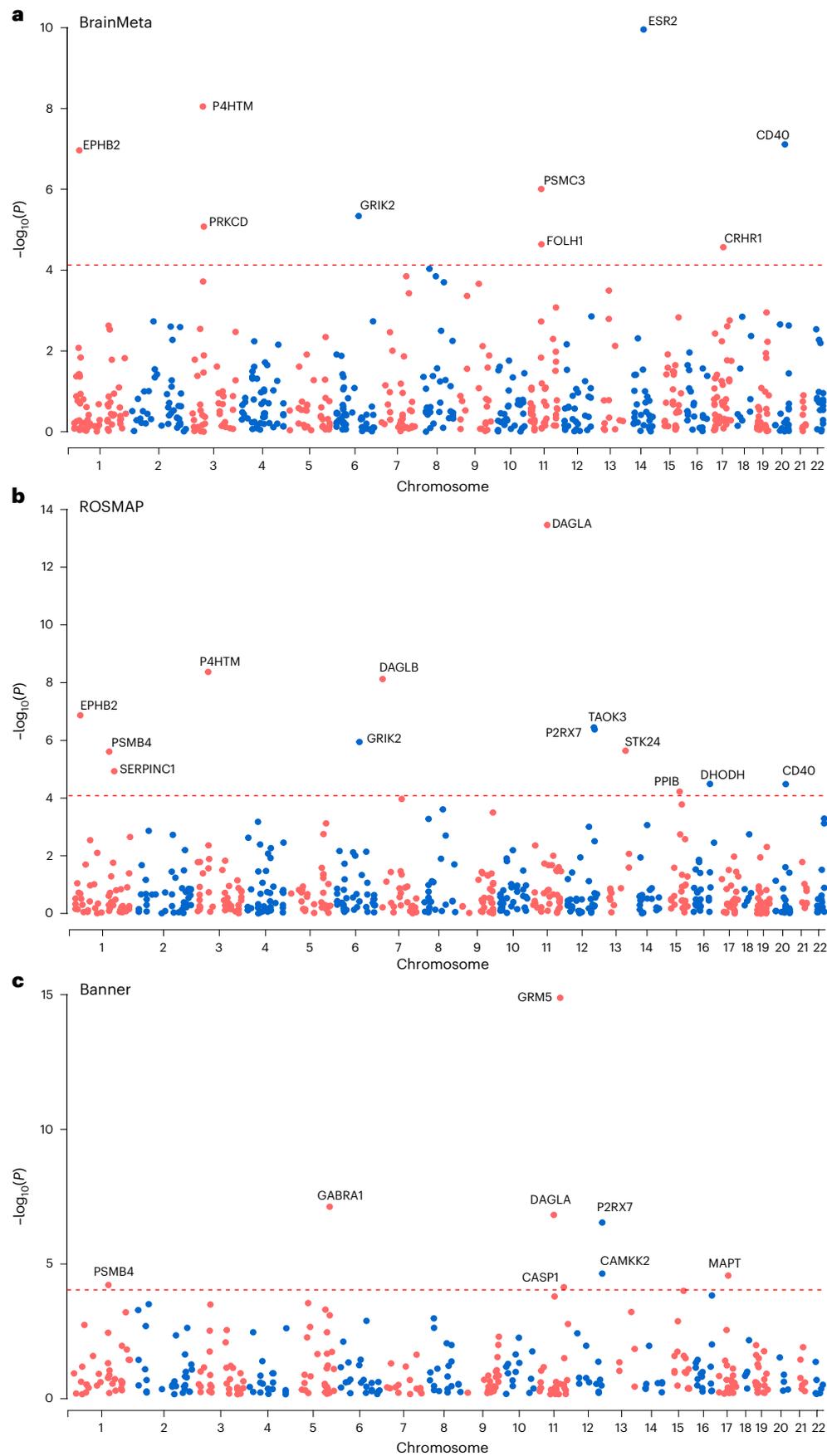
### Prioritization of most likely causal genes

To identify the most likely causal genes, we conducted gene prioritization by combining the evidence from nine different analyses, including genomic location, TWAS, PWAS, colocalization, polygenic priority score (PoPS), functional genomic analysis, eQTL analysis and summary-data-based MR (SMR) (Supplementary Fig. 15 and Supplementary Tables 19–21). Given that we applied stringent Bonferroni correction to each analysis, only genes that survived multiple corrections were considered as reliable candidates. Consequently, the more lines of evidence that support a gene, the higher the probability that

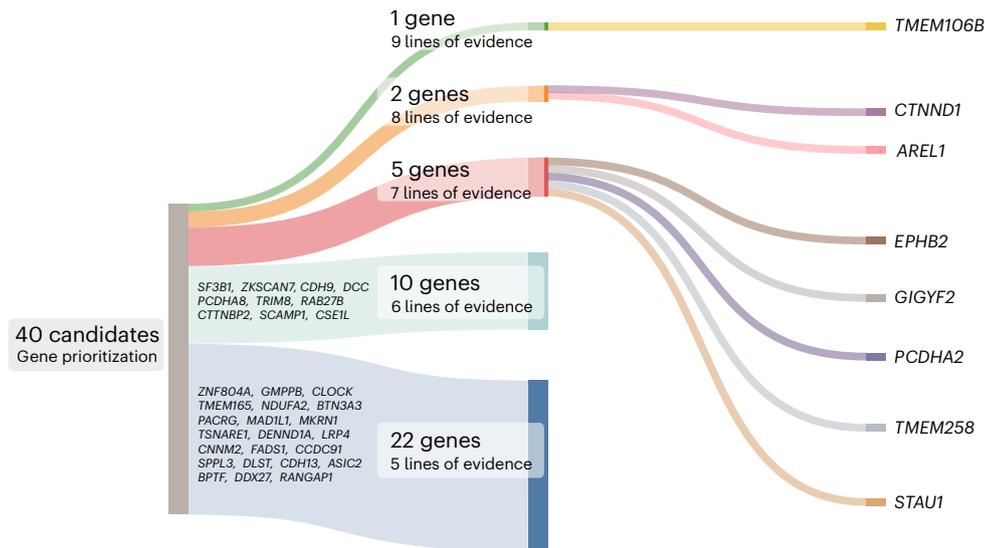


**Fig. 5 | Transcriptome-wide and proteome-wide association results.** Only transcriptome-wide and proteome-wide significant genes (Bonferroni-corrected  $P < 0.05$ ) are shown. **a**, TWAS results. **b**, PWAS results. Z score  $> 0$  indicates that elevated expression or abundance of this gene or protein is associated with the risk of depression, and Z score  $< 0$  indicates that lower expression or abundance

of this gene or protein is associated with the risk of depression. Candidate genes or proteins that reached the Bonferroni-corrected significance threshold in both TWAS and PWAS, and with consistent effect directions in TWAS and PWAS, are highlighted in orange font.



**Fig. 6 | Drug target gene MR analysis.** Only MR significant genes (Bonferroni-corrected  $P < 0.05$ ) are shown. **a**, BrainMeta eQTL MR results. **b,c**, ROSMAP (**b**) and Banner (**c**) pQTL MR results. The  $P$  values are from MR analyses, and all association analyses were two-sided.



**Fig. 7 | Gene prioritization results.** Genes supported by at least five lines of evidence are considered as high-confidence candidate genes. *TMEM106B* was supported by all of the analyses. *CTNND1* and *AREL1* were supported by eight lines of evidence. *EPHB2*, *GIGYF2*, *PCDHA2*, *STAU1* and *TMEM258* were supported by seven lines of evidence.

this gene is causal. A total of 40 genes located in 32 depression risk loci were supported by at least five lines of evidence, of which 8 genes were supported by at least 7 lines of evidence (Fig. 7 and Supplementary Table 22). These genes include *TMEM106B*, *AREL1*, *CTNND1*, *EPHB2*, *GIGYF2*, *PCDHA2*, *STAU1* and *TMEM258*. *TMEM106B* was supported by all analyses and ranked the highest among these genes, suggesting the causality of this gene. In addition, *CTNND1* and *AREL1* were supported by eight lines of evidence. Of note, *EPHB2* was also prioritized as a potential therapeutic target for depression in drug MR analysis. These results prioritized the most likely causal genes for depression and provided important candidate genes for further functional characterization and animal studies.

#### Enrichment of depression risk gene in synaptic process

Our gene prioritization analysis identified 40 high-confidence depression risk genes (supported by at least five lines of evidence; Fig. 7 and Supplementary Table 22). We used SynGO annotations<sup>31</sup> to assess the role of these genes in synapses. Genes encoding postsynaptic specialization (*CTNND1*, *ASIC2* and *DCC*) and presynapse ontology terms (*CDH9*, *CTNND1*, *EPHB2*, *RBA27B* and *SCAMP1*) showed significant enrichment (Supplementary Fig. 16 and Supplementary Table 23). Nine genes were mapped to SynGO biological processes annotations, involving synaptic signalling, synapse organization and postsynapse and presynapse process (Supplementary Fig. 17 and Supplementary Table 23). These results indicate significant enrichment of depression risk genes in synaptic process.

#### *Tmem106b* knockdown resulted in depression-like behaviours

Animal models are pivotal to validate if risk genes identified by human genetic studies are involved in disease pathogenesis. Our above analyses indicated that *TMEM106B* ranked as the top risk gene for depression (Fig. 7). To further investigate if *TMEM106B* is a bona fide depression risk gene, we knocked down *Tmem106b* (the mouse orthologue of human *TMEM106B*, Ensembl ID: ENSMUSG0000029571) in the mouse hippocampus, a brain region that has been previously reported to have a critical role in depression on the basis of mouse models and human studies<sup>32–35</sup> and where *Tmem106b* is highly expressed<sup>36</sup> (Supplementary Fig. 18), and performed serial behavioural experiments (Fig. 8a–d). The open-field test showed that *Tmem106b* knockdown resulted in anxiety-like behaviours (Fig. 8e). Light–dark transition assays showed that *Tmem106b*-knockdown mice preferred to stay in the dark box and

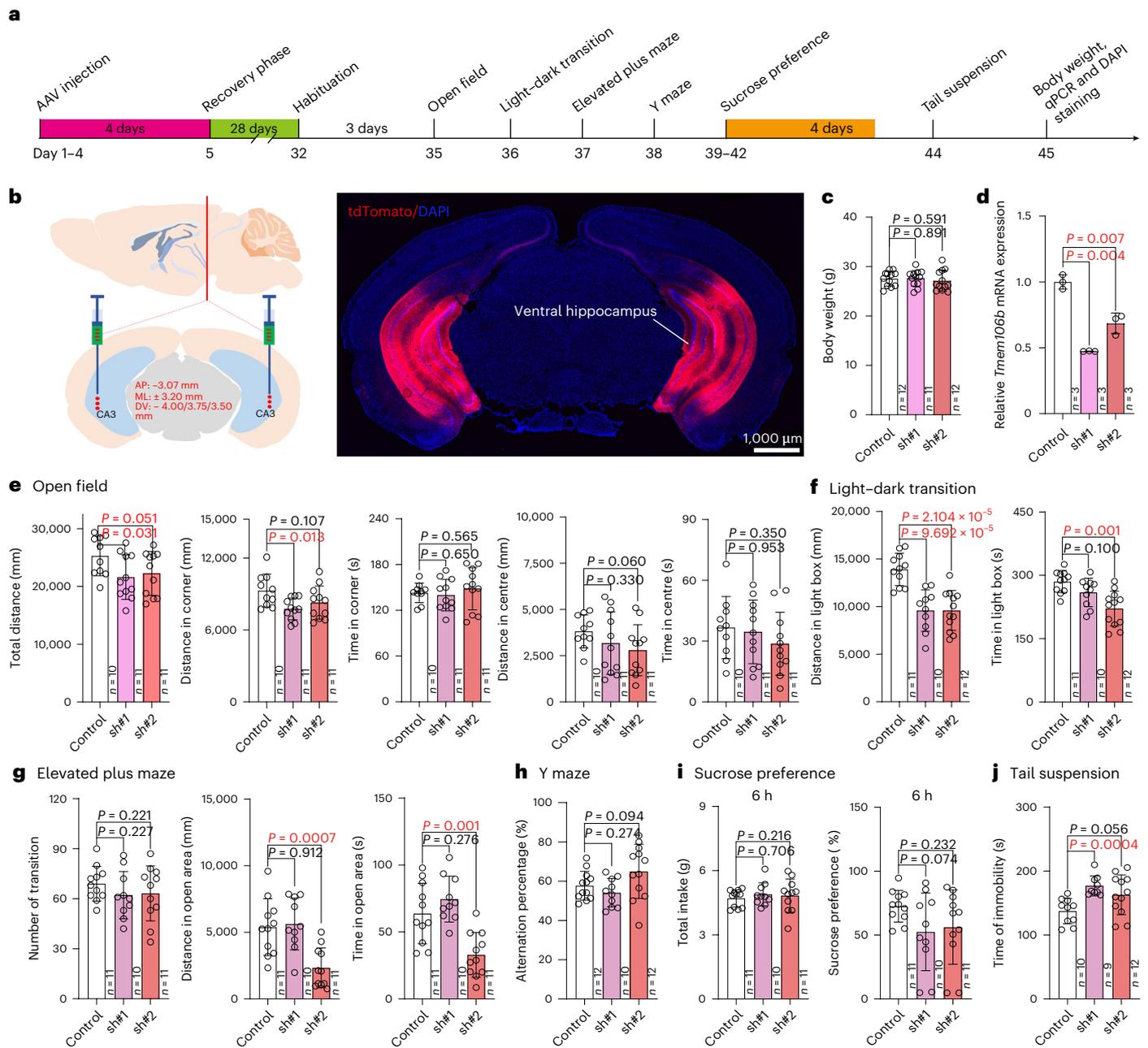
moved less distance in the light box compared with controls (Fig. 8f). The elevated plus maze test showed that the *Tmem106b*-knockdown mice preferred to stay in the close arm (not in the open arm) compared with controls (Fig. 8g). These results indicated that *Tmem106b* knockdown resulted in anxiety-like behaviours. We also evaluated the spatial working memory using the Y maze test and found that the spatial working memory of *Tmem106b*-knockdown mice did not show a significant difference from controls (Fig. 8h). We next assessed depression-like behaviours using the sucrose preference and tail suspension tests. We found that the sucrose preference of *Tmem106b*-knockdown mice was not significantly decreased compared with controls (Fig. 8i). However, the immobility time of *Tmem106b*-knockdown mice was significantly increased compared with controls (Fig. 8j). These results indicated that *Tmem106b* knockdown resulted in depression-like behaviours. Taken together, these behavioural results indicated that knockdown of *Tmem106b* in the ventral hippocampus can lead to anxiety- and depression-like behaviours in mice. The recapitulation of depression-like behaviours in *Tmem106b*-knockdown mice provides robust evidence that supports our genetic findings, that is, *TMEM106B* is a risk gene for depression.

#### Discussion

In this study, we carried out a large-scale depression GWAS meta-analysis. By including 416,437 depression cases and 1,308,758 healthy controls, we identified 287 risk loci, of which 49 are new. In addition to reporting multiple novel risk loci for depression, we also performed a systems-level in-depth analysis to prioritize the likely causal variants and genes. We generated the landscape of potentially causal genes for depression, which provides important opportunities to elucidate pathophysiology and develop new therapeutic targets.

One of the strengths of this study is the use of multi-ancestry cohorts. Considering that most genetic studies of depression have been reported in European populations, it is of great importance to include other populations. In this study, with the use of European, Asian and African ancestries, we identified 49 new risk loci for depression. Our cross-ancestry GWAS not only expands the risk loci of depression but also provides important insights into the genetic architecture of depression.

Compared with two recent large-scale depression GWASs<sup>12,13</sup>, our study has the following strengths. First, our study has the largest sample size of cases ( $n = 416,437$ ). Accordingly, we identified more



**Fig. 8 | *Tmem106b* downregulation in ventral hippocampus causes anxious and depressive behaviours in mice.**

**a**, A schematic diagram of stereotaxic injection and behavioural tests. **b**, Left: a schematic illustration of the stereotaxic injection location in the mouse ventral hippocampus. AP, anteroposterior; ML, mediolateral; DV, dorsoventral. Right: a representative immunofluorescence image showing the expression of tdTomato protein (red), and nuclei are counterstained with DAPI (blue). **c**, The body weight did not show a significant difference in *Tmem106b*-knockdown and control mice (sh#1 versus control: degrees of freedom (d.f.) 21.92,  $P = 0.891$ ,  $t$ -test statistic ( $t$ ) = 0.14, 95% confidence interval (95% CI) -1.23 to 1.40; sh#2 versus control: d.f. 20.01,  $P = 0.591$ ,  $t = 0.547$ , 95% CI -1.95 to 1.14). **d**, qPCR showed that *Tmem106b* expression in the ventral hippocampus was significantly downregulated in *Tmem106b*-knockdown mice compared with controls.  $n = 3$  (three independent biological replicates) for each group and three technical replicates were performed for each biological replicate. Two pairs of complementary shRNA targeting *Tmem106b* were designed, and the first pair of shRNA (sh#1) showed higher knockdown efficiency (sh#1 versus control: d.f. 2.006,  $P = 0.004$ ,  $t = 16.73$ , 95% CI -0.66 to 0.39; sh#2 versus control: d.f. 3.57,  $P = 0.007$ ,  $t = 5.717$ , 95% CI -0.48 to -0.15). All the data met the assumptions of the Welch's  $t$ -test. **e**, Results of the open-field test. Total distance moved, distance moved in corner, time spent in corner,

distance moved in centre and time spent in centre area were analysed. The results showed that total moved distance was significantly decreased in both *Tmem106b*-knockdown groups (shRNA#1 and #2) compared with control group. Distance moved in corner was significantly decreased in *Tmem106b*-knockdown (shRNA#1) mice compared with controls. **f, g**, Light-dark transition (**f**) and elevated plus maze tests (**g**) indicated that *Tmem106b* knockdown resulted in anxiety-like behaviours. The *Tmem106b*-knockdown mice (shRNA#1 and #2) moved less distance in the light box compared with controls. Besides, the *Tmem106b*-knockdown mice (shRNA#2) preferred to stay in the dark box rather than in the light box compared with controls. The elevated plus maze test showed that *Tmem106b*-knockdown mice preferred to stay in close arm rather than in the open arm compared with controls. **h**, The Y maze test showed that the spatial working memory of *Tmem106b*-knockdown and control mice did not show significant differences. **i**, The sucrose preference test showed that there was no significant difference in total intake (water and sucrose) in *Tmem106b*-knockdown and control mice. **j**, The tail suspension test showed that the immobility time was significantly increased in *Tmem106b*-knockdown mice compared with controls. The two-tailed Welch's  $t$ -test was used for statistical testing. The significance threshold was set at  $P < 0.05$ . Data represent mean  $\pm$  s.d. The sample size used in each test is indicated on the right side of each bar graph.

risk loci. In fact, we identified a total of 49 new risk loci in this study, and these newly identified risk loci were not reported by previous GWASs, including studies by Als et al.<sup>12</sup> and Meng et al.<sup>13</sup> Second, we performed serial in-depth analyses to prioritize potential causal variants and genes, and we conducted reporter gene assays to verify the regulatory effect of the identified TF-binding risk variants. Third, our drug–gene MR analysis identified potential therapeutic targets for depression. Finally, we carried out animal experiments to validate if risk genes identified in this study could result in depression-like behaviours in mice.

The lead SNP **rs7531118** showed the most significant association with depression ( $P = 3.20 \times 10^{-31}$ ) in our meta-analysis. Interestingly, this SNP is located near the *NEGR1*, a gene that has been frequently reported to be associated with depression<sup>3,18,26,37,38</sup>. *NEGR1* encodes neuronal growth regulator 1, a cell adhesion molecule of the immunoglobulin LON family. *NEGR1* has been reported to regulate neuronal migration and spine density during mouse cortical development<sup>39</sup>. In addition, Noh et al. found that *NEGR1* has a critical role in hippocampal neurogenesis, and loss of *NEGR1* resulted in anxiety- and depression-like behaviours in mice<sup>38</sup>. Interestingly, Carboni et al. showed that antidepressant treatment altered *Negr1* expression significantly<sup>40</sup>, further supporting the role of *NEGR1* in depression. These genetic and animal findings implicate that *NEGR1* may have a pivotal role in the pathophysiology of depression. In addition to replicating known GWAS associations, this study identified 49 novel risk loci, providing further insights into the genetic architecture of depression. For example, *CNTNAP2*, a depression risk gene previously identified in the Han Chinese population<sup>41</sup>, was validated in our cross-ancestry analysis (Supplementary Table 3). In addition, *EPHB2*, a tyrosine kinase associated with *N*-methyl-D-aspartate receptors and depression-like behaviour in prior cellular and animal studies<sup>42,43</sup>, was supported by GWAS signals in our study, suggesting its potential role in depression. Follow-up functional and mechanistic studies on these newly identified risk loci will help to elucidate the pathogenesis of depression.

In addition to identifying novel risk loci, we also identified functional (or potentially causal) variants from GWS loci using functional genomics analysis. We identified 64 TF binding-affecting SNPs from 287 risk loci, and our reporter gene assays showed that 80% (51 out of 63 tested SNPs) of the identified functional SNPs have a regulatory effect, that is, different alleles of these functional SNPs conferred significant luciferase activity differences. In fact, this is one of the most comprehensive and systematic functional genomic studies to identify and elucidate the regulatory effect of depression risk variants. Considering the fact that identifying causal variants from risk loci is a major challenge in genetic study, the prioritization of functional variants provides an important starting point and opportunity to elucidate the regulatory mechanisms of depression risk variants. Our findings indicate that these functional SNPs may exert their biological effect through altering their binding affinity to TFs, and we dissected the regulatory mechanisms of the functional variants at the single-nucleotide level. Notably, the CTCF binding-affecting SNP **rs7531118** (which is located near the *NEGR1*) also showed the most significant association with depression among the 64 TF binding-affecting risk SNPs, suggesting that this functional SNP may confer depression risk by modulating *NEGR1* expression. Furthermore, *TMEM106B*, which is located 794 base pairs (bp) upstream of the **rs1990622** (affects binding of CTCF), was supported by all lines of evidence in our gene prioritization analyses, strongly suggesting that the functional variant **rs1990622** may confer depression risk by modulating *TMEM106B* expression.

Gene set enrichment analysis revealed that depression associations were significantly enriched in synapse-related cellular components and biological processes (Supplementary Fig. 14), suggesting the pivotal role of depression risk genes in synapse-related functions. Notably, some previous studies have also suggested that the synapse may have

an important role in depression<sup>8,44–48</sup>. Our results further support that synapse dysfunction may play a key role in the aetiology of depression.

Our drug–gene MR analysis implicates potential therapeutic targets for depression, including *DAGLA*, *P2RX7*, *PSMB4*, *P4HTM*, *CD40*, *EPHB2* and *GRIK2* (supported by at least two MR analyses; Fig. 6 and Supplementary Tables 16–18). Among these genes, *EPHB2* was supported by seven lines of evidence in our gene prioritization analysis (Fig. 7), *DAGLA* reached the genome-wide significance level (Supplementary Table 3), and *P2RX7* and *PSMB4* were supported by both PWAS and colocalization analysis (Figs. 5 and 7). These findings indicated that these genes may represent promising therapeutic targets for depression. Drug–gene MR and previous findings from cellular and animal experiments suggest that downregulation of *EPHB2* expression is a risk factor for depression. However, currently approved drugs targeting *EPHB2* are primarily inhibitors (Supplementary Table 19). Based on our drug–gene MR result, agonists or activators targeting this gene may have therapeutic effect on depression. In addition, another potential therapeutic target is *DAGLA*, a gene that encodes diacylglycerol lipase  $\alpha$  (*DAGLA*) and is involved in the biosynthesis of the endocannabinoid, 2-arachidonoyl-glycerol<sup>49–51</sup>. *DAGLA* knockout led to depression-related behaviours<sup>52</sup>. Furthermore, clinical studies have also shown that antagonists of *DAGLA* can increase the incidence and severity of depression<sup>53,54</sup>. Behavioural changes induced by the loss of *DAGLA* include a range of classic depressive phenotypes. Interestingly, among the approved drugs, adenosine triphosphate and caffeine are both activators of *DAGLA* (Supplementary Table 19). Population cohort data indicate a negative correlation between caffeine intake and depressive symptoms<sup>55</sup>, and MR studies suggest that higher long-term plasma caffeine levels may reduce the risk of depression<sup>56</sup>. Based on these findings, additional prospective studies and animal experiments are warranted to further determine the role of adenosine triphosphate and caffeine in depression. *P2RX7* encodes ATP-channel, purinergic receptor P2X, ligand-gated ion channel 7, *P2RX7*<sup>57</sup>. *P2RX7* is primarily located in immune cells and glial cells of the central nervous system<sup>58</sup>. The role of dysregulation of the neuroimmune system, particularly the inflammatory response, in the development of symptoms of depression has been well described<sup>59,60</sup>. Of note, a previous study also nominated *P2RX7* as a potential therapeutic target for depression, suggesting that *P2RX7* receptor antagonists may represent potential drugs for depression<sup>61–63</sup>. *PSMB4* encodes proteasome 20S subunit beta 4. The 20S proteasome plays an important role in major histocompatibility complex class I protein antigen peptide presentation<sup>64</sup>, which has been reported to be associated with susceptibility of depression in the previous genetic association studies<sup>65</sup>. Interestingly, *PSMB4* and *P2RX7* were also identified as actionable drug targets for depression in our previous MR study<sup>66</sup>. These lines of evidence suggest that *EPHB2*, *DAGLA*, *P2RX7* and *PSMB4* are promising drug targets for depression. Further animal and clinical studies on these candidates will help to develop new therapeutic drugs for depression.

In addition to identifying risk loci and performing comprehensive integrative analysis, we also provided animal model evidence supporting that *TMEM106B* is a risk gene for depression. Recapitulation of disease phenotypes is one of the most important and useful ways to validate if risk genes identified by genetic studies are involved in disease. We found that *Tmem106b* knockdown resulted in depression-like behaviours, indicating that this gene has a critical role in depression. These animal model data provide robust evidence that supports our GWAS findings.

There are several limitations of this study. First, we conducted animal studies only on the top-prioritized gene, *TMEM106B*. Whether the knockdown of other risk genes also leads to depression-like behaviours remains unclear, further cellular and animal experiments are needed to elucidate their involvement in depression. Second, due to the limited throughput, we performed reporter gene assays only

for variants supported by functional genomics analyses. Functional characterization of the fine-mapped causal variants using methods such as massively parallel reporter assays would provide a more comprehensive understanding of the functionality of the causal variants. Finally, although populations of European, Asian and African ancestry were included in this study, the sample size of Asian and African populations was much smaller than that of the European population, including more cohorts or individuals of non-European ancestry will provide important insights into the genetic architecture of depression.

In summary, we identified 49 novel risk loci and prioritized likely causal variants and genes for depression. More importantly, we validated the regulatory effect of most functional variants with reporter gene assays. Our study provides important insights into the genetic architecture of depression. Further functional characterization and mechanistic study of the identified risk genes will help to elucidate the neurobiology of depression and provide new therapeutics.

## Methods

### Genome-wide summary statistics used for meta-analysis

Four of the six included GWAS datasets were publicly available, and detailed information about data download are described in the 'Data availability' section. Summary statistics of the AGDS were from B.L.M. The genome-wide summary statistics of 23andMe were obtained under a data transfer agreement<sup>67</sup>. We performed an inverse-variance-based fixed-effects meta-analysis by combining genome-wide associations from six large-scale multiancestry GWASs (Supplementary Table 1), including MVP (83,810 cases and 166,405 controls of European ancestry, 25,843 cases and 33,757 controls of African ancestry)<sup>3</sup>, study from Howard et al. (23andMe–UKB–PGC, 246,363 cases, and 561,190 controls)<sup>5</sup>, AGDS (13,318 cases and 12,684 controls)<sup>9</sup>, BioBank Japan (836 cases and 177,794 controls)<sup>11</sup>, East Asian ancestry (12,455 cases and 85,548 controls)<sup>10</sup> and FinnGen (33,812 cases and 271,380 controls) (<https://www.finnngen.fi/en>)<sup>68</sup>. The total sample size of the multi-ancestry meta-analysis was 1,725,195 ( $N_{\text{cases}} = 416,437$  and  $N_{\text{controls}} = 1,308,758$ ). Of the total effective sample size (calculated according to the formula  $N_{\text{eff}} = 4/(1/N_{\text{cases}} + 1/N_{\text{controls}})$  (ref. 25)) included in the cross-ancestry meta-analysis, 90.9% were of European ancestry, 5.1% were of African ancestry and 4.0% were of East Asian ancestry. In brief, for MVP, cases were defined using an International Classification of Diseases (ICD) code-based algorithm to determine depression case status. Cases with at least one MDD inpatient diagnosis code or two MDD outpatient diagnosis codes were included in GWAS<sup>3,69</sup>. For the 23andMe–UKB–PGC dataset, a broad definition of depression was used for UK Biobank and more detailed phenotypic information can be found in the original paper<sup>8</sup>. For 23andMe data, samples were classified on the basis of responses to web surveys, and those individuals who self-reported receiving a clinical diagnosis or treatment of depression were included as cases. For PGC samples, the clinically derived phenotype of depression was used and detailed information has been described previously<sup>4</sup>. The sample of the AGDS dataset is from the AGDS, and participants met the DSM-5 MDD standard (lifelong MDD) at some time in their lives but were not diagnosed as having schizophrenia, bipolar disorder or attention-deficit/hyperactivity disorder. More detailed information can be found in the original publication<sup>9,70</sup>. For the BioBank Japan dataset, Sakaue et al. combined the past medical history and text mining of electronic medical records to conduct 220 deep-phenotype GWASs (including diseases, biomarkers and drug use) in the Japanese Biological Bank. The dataset of depression was downloaded from <https://phweb.jp/download/Depression>. For East Asian ancestry data, the samples from China Kadoorie Biobank, CONVERGE and Taiwan MDD research and the East Asian DNA samples of 23andMe companies in the USA and the UK were included. The details of each cohort can be found in ref. 10. For FinnGen, cases were assessed using ICD criteria. Cases that met the unipolar depression criteria (FinnGen code: F5\_DEPRESSIO, ICD-10 code: F32, F33) were included.

### Data processing and meta-analysis

As several analytic methods, including logistic regression, linear regression and logistic mixed models, were used in the original studies, we processed each dataset separately to obtain a unified data format for meta-analysis. For results obtained using logistic regression, effect size ( $\beta$ ) was converted into odds ratio (OR) using  $\text{OR} = \exp(\beta)$ . We further processed each dataset to ensure the OR was based on the same effect allele. Besides, the standard error of  $\ln(\text{OR})$  and the corresponding  $P$  value were also calculated for all datasets.

To maximize discovery power<sup>71</sup>, we performed the meta-analysis using the inverse-variance-based fixed-effects meta-analysis approach implemented in PLINK (v1.90)<sup>72</sup>, which conducts meta-analyses based on the standard-error-weighted method. The combined effect size of the meta-analysis is the weighted average of individual study effect sizes (that is, OR), which was calculated on the basis of the logarithm of the effect size (OR) of each study and the inverse of the within-study standard errors of the effect size. To assess the variability of the allelic effect in the cross-ancestry meta-analysis, we first used Cochran's  $Q$  test implemented in the PLINK v1.90 (ref. 72) to evaluate the heterogeneity of each variant in the meta-analysis. For variants that showed significant heterogeneity (that is  $P_{\text{Cochran's } Q} < 0.05$ ), we used the random-effects (DerSimonian and Laird) method to estimate the OR and the  $P$  value. For fixed-effects and random-effects model GWAS meta-analysis, FUMA v1.3.7 was used to define the risk loci<sup>73</sup>, with the default parameters. In brief, FUMA first defines independent lead SNPs ( $P < 5 \times 10^{-8}$ ,  $r^2 < 0.1$ ), then defines risk loci by merging physically close or overlapping SNPs; lead SNPs that are closer than 250 kb were merged into one genomic risk locus. LD information was calculated using the European reference panel from the 1000 Genomes phase 3.

### Heritability estimate and genetic correlation analyses

LDSC (<https://github.com/bulik/ldsc>)<sup>15</sup> was used to estimate the SNP-based heritability of depression GWAS meta-analysis from populations of European ancestry (including MVP-EUR, 23andMe–UKB–PGC, AGDS and FinnGen datasets) and pairwise genetic correlations between the six depression GWASs (MVP, 23andMe–UKB–PGC, AGDS, BioBank Japan, East Asian ancestry and FinnGen). To control the genomic inflation, we also calculate the LDSC regression intercepts and attenuation ratio for the six contributing cohorts and the final meta-analysis results (Supplementary Table 5a,b). LDSC quantifies the genetic variability (LD score) for each SNP marker by the degree of LD ( $r^2$ ) between SNPs. Then, linear regression was used to fit the relationship between the LD score and the chi-square statistic to determine whether there were confounding factors in the GWAS results. We further used the `--sum-hers` flag and BLD-LDAK model<sup>14</sup> from the Genome-wide Complex Trait Analysis software to perform the liability scale SNP-based heritability estimate.

LDSC can analyse the genetic correlations between traits by replacing the chi-square statistic with the product of the  $Z$  scores from two traits<sup>74</sup>. In this study, we reformat the GWAS summary statistics using the standard procedures (described in detail at <https://github.com/bulik/ldsc/wiki/Heritability-and-Genetic-Correlation>) and use the `--r_g` command and the European reference panel from the 1000 Genomes project to calculate the genetic correlations of the pairwise European ancestry depression GWAS datasets and genetic correlation between depression and other traits (Supplementary Table 6a,b). To obtain meaningful and robust results<sup>74</sup>, only European GWAS datasets with an SNP-based heritability  $Z$  score  $>4$  were included in LDSC genetic correlation analyses. To calculate the genetic correlations between different ancestral populations, we further performed the genetic correlation analysis for GWAS datasets with a heritability  $Z$  score  $>2$  using a cross-ancestry method, Popcorn<sup>16</sup>, which uses the approximate weighted likelihood model genetic correlation method and considers the LD panel from different ancestry. Detailed information on the Popcorn method can be found in the original publication<sup>16</sup>.

### Identifying regulatory variants with functional genomics

We used the functional genomics approach<sup>17,18</sup> to identify the functional (or potential causal) variants from the identified risk loci. In brief, chromatin immunoprecipitation sequencing (ChIP-seq) assays that used brain tissues or neuroblastoma cell lines (a total of 30 TFs) were downloaded and processed<sup>17,18</sup>. On the basis of these ChIP-seq data, we derived the DNA binding motifs of these TFs. We further used FIMO (find individual motif occurrences)<sup>75</sup> to compare the derived binding motifs with the publicly available position weight matrix (PWM)<sup>19</sup>, and the best-matched motifs were used for further analysis. We extracted the SNPs that were in LD ( $r^2 > 0.8$ ) with the lead SNPs using the 1000 Genome phase 3 data of the European population from the SNIIPA website (<https://snipa.helmholtz-muenchen.de/snipa3/>)<sup>76</sup>. Finally, we investigated if the lead SNPs or SNPs in LD with the lead SNPs affect the binding of TFs.

### Identifying potential causal variants using fine-mapping

To identify the potential causal variants, we carried out statistical fine-mapping using MESuSiE<sup>20</sup>. MESuSiE extends the sum of single effect model (SuSiE<sup>77</sup>) to accommodate multiple ancestries. It utilizes GWAS summary statistics from diverse ancestral populations as input, and it considers the LD structure within local genomic regions across different ancestries and incorporates models for shared and ancestry-specific causal signals, for fine-mapping. This framework accommodates both the similarity in causal effect sizes and the heterogeneity across ancestries<sup>20</sup>. To perform the MESuSiE fine-mapping analysis, we first extracted the SNPs in LD ( $r^2 \geq 0.8$ ) with the lead SNPs (from the cross-ancestry GWAS meta-analysis results) using Asian and European genotype data from the 1000 Genomes projects (phase 3)<sup>78</sup> as candidate SNPs sets. Subsequently, we conducted ancestral-specific GWAS meta-analysis for the Asian ancestry (including BioBank Japan<sup>11</sup> and East Asian ancestry<sup>10</sup> cohorts) and European ancestry (including MVP-EUR<sup>3</sup>, 23andMe-UKB-PGC<sup>5</sup>, AGDS<sup>9</sup> and FinnGen<sup>68</sup> cohorts). The results of ancestral-specific meta-analyses were then used as input for MESuSiE. After extracting locus LD matrices from corresponding reference panels, we performed MESuSiE analysis using the `meSuSiE_core()` function with default settings: up to ten putative causal variants per locus ( $L = 10$ ), and the Bayes factor value was set to 3, implying PIPs of 3/7, 3/7 and 1/7 for the hypothesis of being uniquely causal in ancestry 1, ancestry 2 and in both ancestries, respectively (the summation of all hypotheses is 1). In total, we fine-mapped 208 GWS risk loci and considered SNPs with a PIP > 0.5 in either ancestries as potential causal SNPs (Supplementary Table 8).

### eQTL analysis

To identify genes whose expression was associated with the risk variants, we examined the associations between the identified risk variants (including lead SNPs, functional SNPs identified by functional genomics, and causal variants prioritized by fine-mapping approaches) and gene expression in the human brain. Brain eQTL data from the BrainMeta v2 dataset<sup>21</sup> ( $n = 2,865$ ) were used in this study. The BrainMeta v2 dataset contains eQTL results for brain cortical tissue from seven cohorts, including BrainGVEX, the Lieber Institute for Brain Development, the CommonMind Consortium, the CommonMind Consortium's National Institute of Mental Health Human Brain Collection Core, Mount Sinai Brain Bank (including four cortex regions: BM10, BM22, BM36 and BM44), Mayo Clinic and ROSMAP; detailed information can be found in the original publications<sup>21,79–81</sup>.

### Transcriptome-wide association study

To identify genes whose genetically regulated expression changes are associated with depression, we performed a TWAS through the FUSION package<sup>23</sup> by integrating the depression GWAS summary statistics from European ancestry and brain eQTL data ( $n = 1,321$ ) from PsychENCODE<sup>22</sup> with default settings. The EUR reference panel from

the 1000 Genomes project was used for LD calculation. TWAS analysis utilizes several linear models (including best linear unbiased predictor (BLUP), Bayesian sparse linear model (BSLMM), LASSO regression (LASSO), elastic net and top SNPs) to calculate SNP-expression weights representing the association between SNPs and gene expression in a reference set. The SNP-expression weights of PsychENCODE used in this study were downloaded from <http://resource.psychencode.org/>. Detailed information on the calculation of SNP-expression weights is available on the PsychENCODE website and related publications<sup>22</sup>. Finally, the Z-score result calculated by TWAS was used to evaluate the association between genes and depression, and the absolute value of the Z-score reflected the strength of the association between risk genes and diseases. Significance thresholds for TWAS results were adjusted using the Bonferroni correction (corrected  $P$  threshold  $2.86 \times 10^{-5}$ ). Genes that passed the Bonferroni correction were considered statistically significant TWAS results.

### Proteome-wide association study

To identify proteins whose genetically regulated abundance is associated with depression, we conducted PWAS by integrating genome-wide meta-analysis results (considering that the EUR reference panel from the 1000 Genomes project was used for LD calculation and the pQTL data were also from populations of European ancestry, only depression GWAS meta-analysis from European ancestry were used) and two independent protein quantitative trait loci (pQTL) datasets from the dorsolateral prefrontal cortex (ROSMAP  $n = 376$ , Banner  $n = 152$ )<sup>24</sup>. In brief, Wingo et al. conducted a proteomic study and used FUSION to estimate protein weights in discovery and confirmation datasets separately. We downloaded the processed protein weight files from <https://www.synapse.org/> (Synapse ID: syn9884314 and syn23245237) and performed PWAS analysis as described for TWAS. To maximize the power of the PWAS analysis, we further performed a meta-analysis to combine PWAS results from Banner and ROSMAP datasets using METAL<sup>25</sup>. For proteins included in both Banner and ROSMAP PWAS, the sample size was used as weight, and PWAS Z scores and  $P$  values were used for METAL meta-analysis. The proteins that passed the Bonferroni-corrected significance threshold ( $0.05/2023 = 2.47 \times 10^{-5}$ ) were considered PWAS significant.

### Colocalization analysis

To explore whether the eQTL (from PsychENCODE<sup>22</sup>), pQTL (from the Banner and ROSMAP<sup>24</sup>) and GWAS signals were colocalized, we performed colocalization analysis with Coloc package<sup>28</sup> (implemented in FUSION<sup>23</sup>). Only genes with corrected TWAS or PWAS  $P < 0.05$  were included for colocalization analysis. To explore whether the eQTL, pQTL and GWAS signals were simultaneously colocalized, we further performed colocalization analysis using the moloc package<sup>29</sup>. We used the 287 risk loci identified from the cross-ancestry GWAS meta-analysis as target regions for colocalization analysis, setting prior probabilities for the association of a SNP with one trait ( $p_1 = 1 \times 10^{-4}$ ), two traits ( $p_2 = 1 \times 10^{-6}$ ) or three traits ( $p_3 = 1 \times 10^{-7}$ ). Colocalization signals were defined as those with a posterior probability (PPA.abc, assuming one SNP is associated with all three traits) greater than 0.85.

### Drug-gene MR analysis

For the drug-gene MR analysis, the TwoSampleMR R package was used to perform two-sample MR analysis (v0.5.6, <https://mrcieu.github.io/TwoSampleMR/>)<sup>82</sup>. We used 1,263 actionable druggable genes selected by Gaziano et al. as potential candidates in this study<sup>30</sup>. To identify drug repurposing opportunities, Gaziano et al. used data from ChEMBL (version 26) to identify 1,263 human proteins as actionable (that is, the approved or clinical-stage drugs therapeutic targets) drug targets<sup>30,83</sup>. More detailed information on these actionable drug targets can be found in the study by Gaziano et al.<sup>30</sup>. Associations between genetic variants and gene or protein expression abundance (that is, eQTL and

pQTL) were used as exposure instruments, and the meta-analysis result was used as outcome data. The brain eQTL data were from the Brain-Meta v2 dataset<sup>21</sup>, and the brain pQTL data were from the Banner and ROSMAP datasets<sup>24</sup>. Considering the sample size of the pQTL analysis was small, we used a relatively relaxed  $P$  threshold (that is, 0.05) for pQTL datasets, as described in previous studies<sup>84–86</sup>, and used a more stringent  $P$  value threshold (that is,  $5 \times 10^{-8}$ ) for the eQTL dataset. Subsequently, actionable druggable genes from pQTL and eQTL results were obtained as QTL datasets for the MR analysis. Independent SNPs were defined if the LD value ( $r^2$ ) was less than 0.001. Then, the exposure and outcome data were harmonized using the `harmonise_data()` function to ensure the same effect allele in exposure and outcome data. For exposures with only one instrumental variable (IV), the Wald ratio method was used, and for exposures with two or more IVs, the inverse-variance-weighted method was utilized.

### SMR

We used the SMR method<sup>87</sup> (which integrates GWAS and gene eQTL data) to identify causal genes associated with depression. The cis-eQTL data used as IVs for gene expression were from the PsychENCODE project ( $n = 1,387$ )<sup>80</sup> and were downloaded from the SMR website ([https://yanglab.westlake.edu.cn/data/SMR/PsychENCODE\\_cis\\_eqtl\\_HCP100\\_summary.tar.gz](https://yanglab.westlake.edu.cn/data/SMR/PsychENCODE_cis_eqtl_HCP100_summary.tar.gz)). We ran SMR analysis with default parameters (`--peqt-smr 5.0e-8, --maf 0.01, --peqt-heidi 1.57e-3, --heidi-mtd 1`). Heterogeneity tests were conducted with the HEIDI test. We excluded results with significant heterogeneity ( $P_{\text{HEIDI}} > 0.01$ ) from the SMR results and used the Bonferroni-corrected significance threshold (corrected  $P$  threshold  $4.88 \times 10^{-6}$ ) to identify potential causal genes that reach SMR significance.

### Tissue and cell type heritability enrichment analysis

To explore whether the genome-wide associations of depression were enriched in specific tissues or cell types, we performed heritability enrichment analysis using MAGMA (v1.1.0)<sup>88</sup>. For the tissue enrichment analysis, we used gene expression data (from the Genotype-Tissue Expression (GTEx) consortium, GTEx\_Analysis\_2017-06-05\_v8\_RNASeQCv1.1.9) of different human tissues to identify genes that are expressed in specific tissues. Based on the GWAS  $P$  value, MAGMA uses a multiple linear principal component regression model to quantify the degree of association between genes and depression. MAGMA then tested whether these genes were enriched among genes specifically expressed in specific tissues. Cell type enrichment analysis was also performed using MAGMA. The single-cell RNA sequencing data that we used for cell type enrichment analysis were downloaded from ref. 89 ([https://github.com/jbryois/scRNA\\_disease/tree/master/Code\\_Paper/Code\\_Zeisel](https://github.com/jbryois/scRNA_disease/tree/master/Code_Paper/Code_Zeisel)). A detailed description of this dataset can be found in previous publications<sup>90,91</sup>.  $P$  values for MAGMA enrichment analyses were corrected for the FDR using the `p.adjust(, method = 'fdr')` function from R4.1.1.

### Gene set enrichment analysis

Gene set analysis was performed using MAGMA software. First, MAGMA maps SNPs to genes by using GWAS summary statistics and gene annotation files. Then, MAGMA calculates the association between specific genes and depression. Finally, MAGMA performs gene set enrichment analysis by testing whether the target gene set is more associated with depression than other genes that are not included in the gene set.  $P$  values for competing gene sets were used to determine the level of significance. We downloaded all GO terms (including cellular components, biological processes and molecular functions) and Kyoto Encyclopedia of Genes and Genomes pathway gene sets from the MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#C5>, v2022.1.Hs). The final  $P$  value of the MAGMA gene set enrichment analysis was corrected for the FDR using the `p.adjust(, method = 'fdr')` function from R4.1.1.

### PoPS prioritization of candidate causal genes

We used the PoPS method to prioritize the genes near all SNPs in high LD with the lead SNP ( $r^2 > 0.8$ ). PoPS is a gene prioritization method<sup>92</sup> that uses the whole-genome signals in the GWAS summary statistical data and combines a large number of bulk and single-cell expression datasets, biological pathways and predicted protein–protein interaction data to prioritize candidate causal genes. We used the gene list and feature files downloaded from the PoPS GitHub page (<https://github.com/FinucaneLab/pops>) to calculate the PoPS score of all candidate genes. For each locus, only the gene with the highest PoPS score will be marked as the final candidate gene.

### Integrating lines of evidence for gene prioritization

To prioritize the most likely causal genes for each risk locus, we combined evidence from nine different analyses, including gene location, PoPS, colocalization, TWAS, PWAS, functional genomics, eQTL and SMR. Considering that we applied stringent Bonferroni correction to each analysis, we assigned one point to each line of evidence, and the total score of each gene was the sum of all lines of evidence. Higher scores indicate a higher probability of being causal.

### Enrichment of depression risk gene in synaptic process

We tested the enrichment of 40 prioritized depression risk genes in the synaptic-related process (biological processes and cellular components) using SynGO (SynGO release: 20210225)<sup>31</sup>. The background gene set was used as the default 'brain expression', the evidence filters were set to 'medium stringency' and the minimum number of genes for each term in the GSEA analysis was set to 3. For more detailed content, see the SynGO website description (<https://syngoportal.org/help>).

### Reporter gene assays

Reporter gene assays were used to examine the transcriptional activity of DNA fragments<sup>17,93,94</sup>. The DNA sequences (approximately 600 bp) containing the candidate functional SNPs were inserted into the pGL4.11 vector, which is a basic vector without a promoter. Thus, this vector can be used to test the regulatory activity of SNPs located in promoter regions (referred as promoter assays). For SNPs that are not located in promoter region, DNA sequences containing the test SNPs were inserted into pGL3-Promoter vectors to detect enhancer activity (enhancer assays). All the inserted DNA sequences were verified by Sanger sequencing and DNA sequences containing SNPs were synthesized by Sangon Biotech.

The human neuroblastoma cells (SH-SY5Y), which were originally obtained from the Kunming Cell Bank, Kunming Institute of Zoology, were used for reporter gene assays. PCR was conducted periodically to detect mycoplasma during cell culture to ensure no mycoplasma contamination. High-glucose Dulbecco's modified Eagle medium supplemented with 10% foetal bovine serum (Gibco, 10091148), 1% penicillin and streptomycin ( $100 \text{ U ml}^{-1}$ ), 10 mM sodium pyruvate solution (Gibco, 11360070) and 1× Minimum Essential Medium non-essential amino acid solution (Gibco, 11140050) was used for cell culture. All cells were cultured in a cell culture incubator with 5%  $\text{CO}_2$  at 37 °C.

For cell transfection and reporter gene detection, the constructed vectors (150 ng) and the internal control vector pRL-TK (Promega, E2241) (50 ng) were cotransfected into the SH-SY5Y cells using Lipofectamine 3000 (Invitrogen, L3000-015). SH-SY5Y cells were plated into 96-well plates containing 100  $\mu\text{l}$  medium at densities of  $6.0 \times 10^4$  cells per well. Forty-eight hours after transfection, a dual-luciferase reporter gene assay system (Promega, E1960) was used to measure the luciferase activity. Differences were calculated with two-tailed Student's  $t$ -test, and the significance threshold was set at  $P < 0.05$ .

### Behavioural assays

**Construction of *Tmem106b* knockdown plasmid.** To knock down *Tmem106b* expression, we designed two small hairpin RNAs

(shRNAs) (shRNA#1:5'-GCGTTACATCGACAGACAA-3'; shRNA#2: 5'-GGAATTTACTGGAAGAGAT-3') targeting mouse *Tmem106b* exon-3. Non-specific shRNA sequence (5'-GATTTGCTGTTCGCCCAAG-3') was used as negative control. The DNA sequences (containing U6 promoter and shRNAs) were synthesized and inserted into pAAV-CAG-tdTomato (Plasmid #59462, Addgene) at NdeI restriction site, and the recombinant vectors (pAAV-U6-shRNA#c/1/2-CAG-tdTomato) were purchased from TSINGKE Biological Technology.

### AAV production and purification

HEK293T cells with 80–90% confluence (cultured in 15 cm cell-culture dishes) were used to produce adeno-associated viruses (AAVs). Ten micrograms of recombinant vectors and AAV packaging plasmids (10 µg AAV-DJ and 20 µg pHHelper) were cotransfected into HEK293T cells using the polyethyleneimine method. Seventy-two hours after transfection, cells containing target AAVs were collected and resuspended with cell lysis buffer (0.15 M NaCl and 20 mM Tris-HCl) and subjected to four freeze-thaw cycles (between liquid nitrogen and 37 °C water) to release AAVs. Then, 0.5% sodium deoxycholate and 50 U ml<sup>-1</sup> benzonase (Sigma, E1014) were added into the above cell lysis buffers to digest non-viral DNA at 37 °C for 30 min, then centrifugated by 5,588g for 90 min at 4 °C. The supernatants containing target AAVs were then subjected to gradient centrifugation using 15%, 25%, 40%, 54% iodixanol (Sigma-Aldrich, D1556) under 255,300g for 3 h at 18 °C. The target AAVs were centrifugated into 40% iodixanol, and then the 40% iodixanol was collected for further purification. The iodixanol was removed by using Amicon ultrafiltration tube (Millipore, UFC910096) and 50 ml precooled 1× HBSS. Finally, 100–300 µl HBSS residues were used to suspend the purified target AAVs, and the collected AAVs were stored at -80 °C. The virus titre was determined following the protocol of AAV titration by quantitative PCR (qPCR) using SYBR Green technology (<https://www.addgene.org/protocols/aav-titration-qpcr-using-sybr-green-technology/>)<sup>95</sup>.

### Brain stereotactic injection

Six-to-eight-week-old C57BL/6J male mice were purchased from Institute of Model Animals, Nanjing University (Nanjing, China) and were raised in 12 h light/dark cycle specific pathogen-free conditions with 23 ± 2 °C, 50–60% humidity. Mice were allowed to intake food and water freely. All animal experiments were approved by the Animal Ethic Committee of Kunming Institute of Zoology (IACUC-RE-2023-02-002).

Mice were adapted to new breeding environment for 1–2 weeks before brain stereotactic injection. Mice were anaesthetized with isoflurane gas, and then the heads of mice were fixed on the brain stereotactic apparatus horizontally (RWD Life Science). Detailed information about viral injection is provided in the Supplementary Information.

### Real-time qPCR

Total RNA was extracted from fresh hippocampal tissues with TRIzol reagent (Invitrogen, 10296028). One microgram of total RNA was used as template to synthesize cDNAs using PrimeScript RT reagent Kit with gDNA Eraser (Takara, RR047A). *Actb* and *Tmem106b* expression was quantified using TB Green Premix Ex Taq II (Tli RNaseH Plus) (Takara, RR820A) and CFX96 Touch Real-Time PCR Detection System (Bio-Rad). *Actb* was set as an internal gene to calculate relative gene expression using the 2<sup>-ΔΔC<sub>t</sub></sup> method<sup>96</sup>. Primers for *Actb* were 5'-GGCTGTATTCCCCTCCATCG-3' (forward) and 5'-CCAGTTGGTAACAATGCCATGT-3' (reverse), and primers for *Tmem106b* were 5'-AACATTGGCCCACTTGATATGAA-3' (forward) and 5'-GAGTGTCCTAAAGTATGCTGTTGT-3' (reverse).

### Behavioural tests

Six behavioural tests (open field, light-dark transition, elevated plus maze, Y maze, sucrose preference and tail suspension) were performed to evaluate the depressive and anxious behaviours. All behavioural tests

were conducted in an independent and undisturbed room between 10:00 and 18:00. An approximately 24 h interval was applied to minimize the stress effects. SuperMaze and SuperTst software (Shanghai XinRuan Information Technology Company) were used to record the test mice behaviours. For the results of the mouse experiment, we used the median absolute deviation to remove outliers (defined as values deviation from median ± 3 median absolute deviation) and then conducted two-tailed Welch's *t*-test for all the behavioural tests in the GraphPad Prism software (version 8.0.1). Further details on the behavioural experiments can be found in the Supplementary Information.

### Reporting summary

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

### Data availability

Genome-wide summary statistics of MVP were obtained from dbGaP via application (accession no. [phs001672.v1.p.1](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE245219)), summary statistics of FinnGen (publicly available) were downloaded from FinnGen website (<https://www.finnngen.fi/en>), summary statistics of Howard et al. (UKB + PGC) were publicly available and downloaded from <https://doi.org/10.7488/ds/2458>, summary statistics of the AGDS were from B.L.M., summary statistics of Giannakopoulou et al. were downloaded from PGC website (<https://pgc.unc.edu/>) and summary statistics of Sakaue et al. were publicly available and downloaded from BioBank Japan (BBJ) (<https://pheweb.jp/downloads>). The genome-wide summary statistics of 23andMe were obtained under a data transfer agreement. The genome-wide summary statistics (not including 23andMe and AGDS) will be made publicly available (at <https://doi.org/10.6084/m9.figshare.24521968.v1> (ref. 97)) once the article has been published. The full GWAS summary statistics for the 23andMe discovery dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit <https://research.23andme.com/collaborate/#dataset-access/> for more information and to apply to access the data. GWAS summary statistics for the AGDS dataset will be available upon reasonable request (please contact B.L.M. at [Brittany.mitchell@qimrberghofer](mailto:Brittany.mitchell@qimrberghofer)). The SNP-expression weights of PsychENCODE used in this study were downloaded from <http://resource.psychencode.org/>. The processed protein weight files were downloaded from <https://www.synapse.org/> (Synapse ID: syn9884314 and syn23245237). The PsychENCODE cis-eQTL data were downloaded from the SMR website ([https://yanglab.westlake.edu.cn/data/SMR/PsychENCODE\\_cis\\_eqtl\\_HCP100\\_summary.tar.gz](https://yanglab.westlake.edu.cn/data/SMR/PsychENCODE_cis_eqtl_HCP100_summary.tar.gz)). The gene expression data used for MAGMA were from the GTEx consortium, GTEx\_Analysis\_2017-06-05\_v8\_RNASeQCv1.1.9. All GO terms (including cellular components, biological processes and molecular functions) and Kyoto Encyclopedia of Genes and Genomes pathway gene sets were downloaded from the MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#C5>, v2022.1.Hs). The gene list and feature files were downloaded from the PoPS GitHub page (<https://github.com/FinucaneLab/pops>) to calculate the PoPS score of all candidate genes. Source data are provided with this paper.

### Code availability

GWAS meta-analysis was used an inverse-variance-based fixed-effects meta-analysis implemented in PLINK (v1.90, <https://www.cog-genomics.org/plink/>). FUMA v1.3.7 (<https://fuma.ctglab.nl/home>) was used to define the risk loci, with the default parameters. LDSC (<https://github.com/bulik/ldsc>) was used to estimate the SNP-based heritability and pairwise genetic correlations between the GWASs. FIMO was used to compare the derived binding motifs with the publicly available PWM and search the best-matched motifs. MESuSIE (<https://github.com/borangao/meSuSIE>) was used for statistical fine mapping. The TwoSampleMR R package was used to perform two-sample MR

analysis (v0.5.6, <https://mrcieu.github.io/TwoSampleMR/>). Other custom codes is available via Zenodo at <https://doi.org/10.5281/zenodo.13856052> (ref. 98).

## References

- Malhi, G. S. & Mann, J. J. Depression. *Lancet* **392**, 2299–2312 (2018).
- Friedrich, M. J. Depression is the leading cause of disability around the world. *JAMA* **317**, 1517 (2017).
- Levey, D. F. et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat. Neurosci.* **24**, 954–963 (2021).
- Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
- Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
- Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
- CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
- Howard, D. M. et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* **9**, 1470 (2018).
- Mitchell, B. L. et al. The Australian Genetics of Depression Study: new risk loci and dissecting heterogeneity between subtypes. *Biol. Psychiatry* **92**, 227–235 (2021).
- Giannakopoulou, O. et al. The Genetic Architecture of Depression in individuals of East Asian ancestry: a genome-wide association study. *JAMA Psychiatry* **78**, 1258–1269 (2021).
- Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
- Als, T. D. et al. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).
- Meng, X. et al. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nat. Genet.* <https://doi.org/10.1038/s41588-023-01596-4> (2024).
- Speed, D., Holmes, J. & Balding, D. J. Evaluating and improving heritability models using summary statistics. *Nat. Genet.* **52**, 458–462 (2020).
- Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 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).
- Huo, Y., Li, S., Liu, J., Li, X. & Luo, X. J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat. Commun.* **10**, 670 (2019).
- Li, S. et al. Regulatory mechanisms of major depressive disorder risk variants. *Mol. Psychiatry* **25**, 1926–1945 (2020).
- Whittington, T. et al. Gene regulatory mechanisms underpinning prostate cancer susceptibility. *Nat. Genet.* **48**, 387–397 (2016).
- Gao, B. & Zhou, X. MESuSIE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. *Nat. Genet.* **56**, 170–179 (2024).
- Qi, T. et al. Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.* **54**, 1355–1363 (2022).
- Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* <https://doi.org/10.1126/science.aat8127> (2019).
- Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
- Wingo, A. P. et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer’s disease pathogenesis. *Nat. Genet.* **53**, 143–146 (2021).
- Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Deng, Y. T. et al. Identifying causal genes for depression via integration of the proteome and transcriptome from brain and blood. *Mol. Psychiatry* **27**, 2849–2857 (2022).
- Li, W. et al. Genome-wide meta-analysis, functional genomics and integrative analyses implicate new risk genes and therapeutic targets for anxiety disorders. *Nat. Hum. Behav.* **8**, 361–379 (2024).
- Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
- Giambartolomei, C. et al. A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018).
- Gaziano, L. et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat. Med.* **27**, 668–676 (2021).
- Koopmans, F. et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234 e4 (2019).
- Tartt, A. N., Mariani, M. B., Hen, R., Mann, J. J. & Boldrini, M. Dysregulation of adult hippocampal neuroplasticity in major depression: pathogenesis and therapeutic implications. *Mol. Psychiatry* **27**, 2689–2699 (2022).
- Krishnan, V. & Nestler, E. J. The molecular neurobiology of depression. *Nature* **455**, 894–902 (2008).
- Abe-Higuchi, N. et al. Hippocampal sirtuin 1 signaling mediates depression-like behavior. *Biol. Psychiatry* **80**, 815–826 (2016).
- Cui, L. et al. Major depressive disorder: hypothesis, mechanism, prevention and treatment. *Signal Transduct. Target. Ther.* **9**, 30 (2024).
- Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
- Dall’Aglio, L., Lewis, C. M. & Pain, O. Delineating the genetic component of gene expression in major depression. *Biol. Psychiatry* **89**, 627–636 (2021).
- Noh, K. et al. Negr1 controls adult hippocampal neurogenesis and affective behaviors. *Mol. Psychiatry* **24**, 1189–1205 (2019).
- Szczurkowska, J. et al. NEGR1 and FGFR2 cooperatively regulate cortical development and core behaviours related to autism disorders in mice. *Brain* **141**, 2772–2794 (2018).
- Carboni, L. et al. Depression-associated gene Negr1-Fgfr2 pathway is altered by antidepressant treatment. *Cells* **9**, 1818 (2020).
- Ji, W. et al. CNTNAP2 is significantly associated with schizophrenia and major depression in the Han Chinese population. *Psychiatry Res.* **207**, 225–228 (2013).
- Zhang, R. X. et al. EphB2 in the medial prefrontal cortex regulates vulnerability to stress. *Neuropsychopharmacology* **41**, 2541–2556 (2016).
- Grunwald, I. C. et al. Kinase-independent requirement of EphB2 receptors in hippocampal synaptic plasticity. *Neuron* **32**, 1027–1040 (2001).
- Calabrese, F., Riva, M. A. & Molteni, R. Synaptic alterations associated with depression and schizophrenia: potential as a therapeutic target. *Expert Opin. Ther. Targets* **20**, 1195–1207 (2016).

45. Holmes, S. E. et al. Lower synaptic density is associated with depression severity and network alterations. *Nat. Commun.* **10**, 1529 (2019).
46. Price, R. B. & Duman, R. Neuroplasticity in cognitive and psychological mechanisms of depression: an integrative model. *Mol. Psychiatry* **25**, 530–543 (2020).
47. Duman, R. S., Aghajanian, G. K., Sanacora, G. & Krystal, J. H. Synaptic plasticity and depression: new insights from stress and rapid-acting antidepressants. *Nat. Med.* **22**, 238–249 (2016).
48. Duman, R. S. & Aghajanian, G. K. Synaptic dysfunction in depression: potential therapeutic targets. *Science* **338**, 68–72 (2012).
49. Shonesy, B. C. et al. CaMKII regulates diacylglycerol lipase- $\alpha$  and striatal endocannabinoid signaling. *Nat. Neurosci.* **16**, 456–463 (2013).
50. Ogasawara, D. et al. Rapid and profound rewiring of brain lipid signaling networks by acute diacylglycerol lipase inhibition. *Proc. Natl Acad. Sci. USA* **113**, 26–33 (2016).
51. Bisogno, T. et al. Cloning of the first sn1-DAG lipases points to the spatial and temporal regulation of endocannabinoid signaling in the brain. *J. Cell Biol.* **163**, 463–468 (2003).
52. Jenniches, I. et al. Anxiety, stress, and fear response in mice with reduced endocannabinoid levels. *Biol. Psychiatry* **79**, 858–868 (2016).
53. Liu, X. et al. Plasma lipidomics reveals potential lipid markers of major depressive disorder. *Anal. Bioanal. Chem.* **408**, 6497–6507 (2016).
54. Hill, M. N., Miller, G. E., Carrier, E. J., Gorzalka, B. B. & Hillard, C. J. Circulating endocannabinoids and *N*-acyl ethanolamines are differentially regulated in major depression and following exposure to social stress. *Psychoneuroendocrinology* **34**, 1257–1262 (2009).
55. Iranpour, S. & Sabour, S. Inverse association between caffeine intake and depressive symptoms in US adults: data from National Health and Nutrition Examination Survey (NHANES) 2005–2006. *Psychiatry Res.* **271**, 732–739 (2019).
56. Woolf, B. et al. Appraising the causal relationship between plasma caffeine levels and neuropsychiatric disorders through Mendelian randomization. *BMC Med.* **21**, 296 (2023).
57. Harvey, M., Belleau, P. & Barden, N. Gene interactions in depression: pathways out of darkness. *Trends Genet.* **23**, 547–556 (2007).
58. Vereczkei, A. et al. Association of purinergic receptor P2RX7 gene polymorphisms with depression symptoms. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **92**, 207–216 (2019).
59. Hodes, G. E., Kana, V., Menard, C., Merad, M. & Russo, S. J. Neuroimmune mechanisms of depression. *Nat. Neurosci.* **18**, 1386–1393 (2015).
60. Wohleb, E. S., Franklin, T., Iwata, M. & Duman, R. S. Integrating neuroimmune systems in the neurobiology of depression. *Nat. Rev. Neurosci.* **17**, 497–511 (2016).
61. Boucher, A. A. et al. Resilience and reduced c-Fos expression in P2X7 receptor knockout mice exposed to repeated forced swim test. *Neuroscience* **189**, 170–177 (2011).
62. Basso, A. M. et al. Behavioral profile of P2X7 receptor knockout mice in animal models of depression and anxiety: relevance for neuropsychiatric disorders. *Behav. Brain Res.* **198**, 83–90 (2009).
63. Deussing, J. M. & Arzt, E. P2X7 receptor: a potential therapeutic target for depression? *Trends Mol. Med.* **24**, 736–747 (2018).
64. Groettrup, M. et al. A role for the proteasome regulator PA28 $\alpha$  in antigen presentation. *Nature* **381**, 166–168 (1996).
65. Wong, M. L., Dong, C., Maestre-Mesa, J. & Licinio, J. Polymorphisms in inflammation-related genes are associated with susceptibility to major depression and antidepressant response. *Mol. Psychiatry* **13**, 800–812 (2008).
66. Liu, J. et al. Genome-wide Mendelian randomization identifies actionable novel drug targets for psychiatric disorders. *Neuropsychopharmacology* **48**, 270–280 (2023).
67. Li, X. et al. Common variants on 6q16.2, 12q24.31 and 16p13.3 are associated with major depressive disorder. *Neuropsychopharmacology* **43**, 2146–2153 (2018).
68. FinnGen (documentation of R5 release). *FinnGen* <https://finngen.gitbook.io/documentation/> (2020).
69. Gelernter, J. et al. Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. *Nat. Neurosci.* **22**, 1394–1401 (2019).
70. Byrne, E. M. et al. Cohort profile: the Australian Genetics of Depression Study. *BMJ Open* **10**, e032580 (2020).
71. Evangelou, E. & Ioannidis, J. P. A. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14**, 379–389 (2013).
72. 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).
73. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
74. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
75. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
76. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNIpA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334–1336 (2015).
77. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B* **82**, 1273–1300 (2020).
78. Abecasis, G. R. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
79. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
80. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
81. Vialle, R. A., de Paiva Lopes, K., Bennett, D. A., Crary, J. F. & Raj, T. Integrating whole-genome sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the human brain. *Nat. Neurosci.* **25**, 504–514 (2022).
82. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenotype. *eLife* **7**, e34408 (2018).
83. Mendez, D. et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2019).
84. Yang, Z., Yang, J., Liu, D. & Yu, W. Mendelian randomization analysis identified genes pleiotropically associated with central corneal thickness. *BMC Genomics* **22**, 517 (2021).
85. Wang, F. et al. Mendelian randomization analysis identified genes potentially pleiotropically associated with periodontitis. *Saudi J. Biol. Sci.* **28**, 4089–4095 (2021).
86. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
87. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
88. 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).

89. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
90. Liu, J. et al. Genome-wide association study followed by trans-ancestry meta-analysis identify 17 new risk loci for schizophrenia. *BMC Med.* **19**, 177 (2021).
91. Bryois, J. et al. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson's disease. *Nat. Genet.* **52**, 482–493 (2020).
92. Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* <https://doi.org/10.1038/s41588-023-01443-6> (2020).
93. Li, Y. et al. Regulatory variant rs2535629 in ITIH3 intron confers schizophrenia risk by regulating CTCF binding and SFMBT1 expression. *Adv. Sci.* **9**, e2104786 (2022).
94. Wang, J. et al. Functional variant rs2270363 on 16p13.3 confers schizophrenia risk by regulating NMRAL1. *Brain* **145**, 2569–2585 (2022).
95. Aurnhammer, C. et al. Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. *Hum. Gene Ther. Methods* **23**, 18–28 (2011).
96. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* **25**, 402–408 (2001).
97. Dang, X. Genome-wide summary statistics from a cross-ancestry meta-analysis of depression. *figshare* <https://doi.org/10.6084/m9.figshare.24521968.v1> (2025).
98. Dang, X. & Luo, X. Codes for 'Cross-ancestry genome-wide association study and systems-level integrative analyses implicate new risk genes and therapeutic targets for depression'. *Zenodo* <https://doi.org/10.5281/zenodo.13856052> (2024).

## Acknowledgements

This study was equally supported by the startup funds from Southeast University (RF1028623032 to X.-J.L.) and the Special Project for Social Development of Yunnan Province (nos. 202203AC100007 and 202305AH340006). This study was also supported by the National Natural Science Foundation of China (nos. U2102205, 82271570, 82260276, 82301690, 82130042, 82471552, 82230046 and 81920108018), the Talent and Fundamental Research Project of Yunnan Province (nos. 202301AY070001-299, 202105AC160004 and 202301AT070037), the China Science and Technology Innovation

2030 – Major Project (nos. 2022ZD0211701 and 2021ZD0200700), and the SEU Innovation Capability Enhancement Plan for Doctoral Students (CXJH\_SEU 25233). We thank Q. Li and Z. Ding for their technical assistance.

## Author contributions

X.-J.L. conceived, designed and supervised the whole study. X.D. performed all the analyses, including meta-analysis, TWAS, PWAS, colocalization, fine mapping, SMR and so on. Y.L., J.W. and S.L. conducted the behavioural tests. R.C. performed reporter gene assays. X.-J.L., Z.T., Y.Z., Y. Yue, B.L.M., Y.-G.Y., M.L., Z.L., Y. Yuan, T.L. and Z.Z. contributed to this work in study design, data interpretation, paper writing and revision. X.D. drafted the paper. X.-J.L. oversaw the project and finalized 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/s41562-024-02073-6>.

**Correspondence and requests for materials** should be addressed to Yonggui Yuan, Tao Li, Zhijun Zhang or Xiong-Jian Luo.

**Peer review information** *Nature Human Behaviour* thanks Jonathan Coleman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

Yifan Li<sup>1,10</sup>, Xinglun Dang<sup>1,10</sup>, Rui Chen<sup>2,10</sup>, Zhaowei Teng<sup>3,10</sup>, Junyang Wang<sup>4,10</sup>, Shiwu Li<sup>2,10</sup>, Yingying Yue<sup>1</sup>, Brittany L. Mitchell<sup>5</sup>, Yong Zeng<sup>3</sup>, Yong-Gang Yao<sup>2</sup>, Ming Li<sup>2</sup>, Zhongchun Liu<sup>6,7</sup>, Yonggui Yuan<sup>1</sup>✉, Tao Li<sup>8</sup>✉, Zhijun Zhang<sup>1,9</sup>✉ & Xiong-Jian Luo<sup>1</sup>✉

<sup>1</sup>Department of Psychiatry and Psychosomatics, Zhongda Hospital, School of Medicine, Advanced Institute for Life and Health, Jiangsu Provincial Key Laboratory of Brain Science and Medicine, Southeast University, Nanjing, China. <sup>2</sup>Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. <sup>3</sup>Key Laboratory of Neurological and Psychiatric Disease Research of Yunnan Province, The Second Affiliated Hospital of Kunming Medical University, Yunnan Provincial Department of Education Gut Microbiota Transplantation Engineering Research Center, Kunming, China. <sup>4</sup>Department of Human Anatomy, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou, China. <sup>5</sup>Mental Health and Neuroscience Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. <sup>6</sup>Department of Psychiatry, Renmin Hospital of Wuhan University, Wuhan, China. <sup>7</sup>Taikang Center for Life and Medical Sciences, Wuhan University, Wuhan, China. <sup>8</sup>Affiliated Mental Health Center, Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou, China. <sup>9</sup>Department of Mental Health and Public Health, Faculty of Life and Health Sciences, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. <sup>10</sup>These authors contributed equally: Yifan Li, Xinglun Dang, Rui Chen, Zhaowei Teng, Junyang Wang, Shiwu Li. ✉e-mail: [yygyth2000@sina.com](mailto:yygyth2000@sina.com); [litaozjusc@zju.edu.cn](mailto:litaozjusc@zju.edu.cn); [janemengzhang@vip.163.com](mailto:janemengzhang@vip.163.com); [luoxiongjian@seu.edu.cn](mailto:luoxiongjian@seu.edu.cn)

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure 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

The luciferase activities of reporter gene assays were collected by the Luminoskan ascent microplate luminometer (ThermoFisher) and the ascent software (Version 1.0). RT-qPCR were collected by a 7500 real-time fluorescent quantitative PCR system (ThermoFisher) and the QuantStudio real-time PCR software.

The SNP-expression weights of PsychENCODE used in this study were downloaded from <http://resource.psychencode.org/>.

The processed protein weight files were download from <https://www.synapse.org/> (Synapse ID: syn9884314 and syn23245237)

The PsychENCODE cis-eQTL data was downloaded from the SMR website ([https://yanglab.westlake.edu.cn/data/SMR/PsychENCODE\\_cis\\_eqt1\\_HCP100\\_summary.tar.gz](https://yanglab.westlake.edu.cn/data/SMR/PsychENCODE_cis_eqt1_HCP100_summary.tar.gz))

The gene expression data used for MAGMA was from the Genotype-Tissue Expression (GTEx) consortium, GTEx\_Analysis\_2017-06-05\_v8\_RNASeQCv1.1.9

All GO terms (including cellular components, biological processes, and molecular functions) and KEGG pathway gene sets was downloaded from the MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#C5>).

The gene list and feature files was downloaded from the PoPS GitHub page (<https://github.com/FinucaneLab/pops>) to calculate the PoPS score of all candidate genes.

Data analysis

Experimental data analysis was done with Prism (version 8.0.1, GraphPad Software). Quantitative data were expressed as mean  $\pm$  standard deviation (S.D.). Two group comparisons were analyzed by the Student's t-test (two-tailed) and Welch's t-test (two-tailed) otherwise by ANOVA test. Statistical significance threshold was set at  $P < 0.05$ .

GWAS meta-analysis was performed using the inverse variance-based fixed-effects meta-analysis approach implemented in PLINK (v1.90). FUMA v1.3.7 was used to define the risk loci, with the default parameters.

Linkage disequilibrium Score regression (LDSC, <https://github.com/bulik/ldsc>) was used to estimate the SNP-based heritability and pairwise

genetic correlations between the GWASs.

FIMO was used to compare the derived binding motifs with the publicly available position weights matrix (PWM), and the best-matched motifs were used.

MESuSiE (<https://github.com/borangao/meSuSie>) was used for statistical fine-mapping.

TwoSampleMR R package was used to perform two-sample MR analysis (v0.5.6, <https://mrcieu.github.io/TwoSampleMR/>)

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](#)

Genome-wide summary statistics of MVP were obtained from dbGaP via application (accession no: phs001672.v1.p.1), summary statistics of FinnGen (public available) were downloaded from FinnGen website (<https://www.finnngen.fi/en>), summary statistics of Howard et al. (UKB+PGC) were public available and downloaded from <https://doi.org/10.7488/ds/2458>, summary statistics of the AGDS were from Dr. Mitchell (a co-author of this study), summary statistics of Giannakopoulou et al. were downloaded from PGC website (<https://pgc.unc.edu/>), and summary statistics of Sakaue et al. were public available and downloaded from BioBank Japan (BBJ) (<https://phweb.jp/downloads>). The genome-wide summary statistics of 23andMe were obtained under a data transfer agreement. The genome-wide summary statistics (not including 23andme and AGDS) will be made public available (at <https://doi.org/10.6084/m9.figshare.24521968.v1>) once the paper has been published. The full GWAS summary statistics for the 23andMe discovery dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit <https://research.23andme.com/collaborate/#dataset-access/> for more information and to apply to access the data.

## 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

No human subjects were recruited in this study.

Reporting on race, ethnicity, or other socially relevant groupings

No human subjects were recruited in this study.

Population characteristics

No human subjects were recruited in this study.

Recruitment

No human subjects were recruited in this study.

Ethics oversight

No human subjects were recruited in this study.

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](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size

Sample sizes required for this study were estimated based on pilot studies, no statistical tests were performed to predetermine sample sizes. We knocked-down Tmem106b expression by injecting adeno-associated virus (AAV) expressing shRNAs (targeting Tmem106b) into the ventral hippocampus of the 10 week old mice. We conducted serial behavioral experiments after 5 weeks.

Data exclusions

No data were excluded from the analyses.

Replication

Biological replicates (different mice) were made for every experiment, and used as validations. Replications were successful and we report differences between Control and Tmem106b-shRNAs mice which are confirmed by all samples.

Randomization

For knockdown of Tmem106b in ventral hippocampus, mice were randomly allocated for Tmem106b knockdown group and control group.

Blinding

All Behavioral tests (open-field test, Y-maze, elevated-plus maze, tail suspension test, Sucrose preference test, and light-dark transition test) were performed in a completely blinded manner.

# 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	Involvement	Material/System
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Plants

## Methods

n/a	Involvement	Method
<input checked="" type="checkbox"/>	<input type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/>	MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	SH-SY5Y (human neuroblastoma cell line) cells were purchased from the Kunming Cell Bank at Kunming Institute of Zoology, Chinese Academy of Sciences.
Authentication	Cells were validated by short tandem repeat analysis.
Mycoplasma contamination	Cells were periodically tested for mycoplasma and all cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used.

## Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	The SPF grade C57BL/6J mice, at ages: 8 weeks old.
Wild animals	The study did not include any wild animals.
Reporting on sex	Behavioral assays in the study include only male mice. Because female mice have a physiological cycle, which has an impact on behavior.
Field-collected samples	Field-collected samples were not studied in the study.
Ethics oversight	All animal experiments were approved by the Animal Ethic Committee of Kunming Institute of Zoology (IACUC-RE-2023-02-002).

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

## Plants

Seed stocks	The study did not include any plants.
Novel plant genotypes	The study did not include any plants.
Authentication	The study did not include any plants.

# **Cross-ancestry genome-wide association study and systems-level integrative analyses implicate new risk genes and therapeutic targets for depression**

---

In the format provided by the authors and unedited

# Content

<b>Supplementary methods</b> .....	3
<i>Behavioral tests</i> .....	3
<i>Open field test</i> .....	3
<i>Light-dark transition test</i> .....	3
<i>Elevated plus maze test</i> .....	3
<i>Y maze test</i> .....	4
<i>Sucrose preference test</i> .....	4
<i>Tail suspension test</i> .....	5
<b>Supplementary figures</b> .....	6
Supplementary Figure 1. Overview of the analysis included in this study.....	6
Supplementary Figure 2. Quantile-quantile plot illustrating the GWAS Meta-Analysis for major depressive disorder.....	7
Supplementary Figure 3. Locuszoom plots for the top six genome-wide significant association loci. ....	8
Supplementary Figure 4. Manhattan plot of the cross-ancestry random-effect model GWAS meta-analysis.....	9
Supplementary Figure 5. Genetic correlations between major depressive disorder (EUR ancestry meta-analysis results) and other brain disorders and intelligence. ....	10
Supplementary Figure 6. Reporter gene assays validated the regulatory effect of the identified TF binding-affecting SNPs. ....	11
Supplementary Figure 7. Manhattan plot of the EUR-only GWAS meta-analysis. ....	12
Supplementary Figure 8. PWAS analysis result using Banner cis-pQTL data and EUR population MDD GWAS meta-analysis dataset.....	12
Supplementary Figure 9. PWAS analysis result using ROSMAP cis-pQTL data and EUR population MDD GWAS meta-analysis dataset.....	13
Supplementary Figure 10. Drug target gene mendelian randomization analysis using EUR population MDD GWAS meta-analysis dataset as outcome.....	14
Supplementary Figure 11. Heritability enrichment analysis in 30 general tissue types. ....	15
Supplementary Figure 12. Heritability enrichment analysis in GTEx V8 detailed tissue types. ....	16
Supplementary Figure 13. Heritability enrichment analysis in different cell type types....	17

Supplementary Figure 14. Heritability enrichment analysis of Gene Ontology (GO) and KEGG terms.....	18
Supplementary Figure 15. SMR analysis result using PsychENCODE cis-eQTL data. ....	18
Supplementary Figure 16. MDD risk genes in SynGO cellular components annotations. ..	19
Supplementary Figure 17. MDD risk genes in SynGO biological processes annotations....	20
Supplementary Figure 18. The mRNA expression of Tmem106b in adult mouse brain. ....	20

## **Supplementary methods**

### ***Behavioral tests***

#### ***Open field test***

Open field test is a method to evaluate movement and anxious behaviors of animals in novel environments<sup>1</sup>. Briefly, the open field arena (Length: 40 cm; Width: 40 cm; Height: 40 cm) was divided into sixteen squares using SuperMaze software, the four squares at corner were defined as corner areas, and the four squares at center were defined as central areas. Each test mouse was placed in the same corner and allowed to freely explore novel environment for 5 min. Their movement tracks were recorded by video camera and SuperMaze software with the same parameters. The apparatus was cleaned with 10% ethanol before each trial. A total of 32 mice were tested successfully (control: n = 10; *Tmem106b*-shRNA#1: n = 11; *Tmem106b*-shRNA#2: n = 11). The time and distance of central and corner areas are indicators to reflect anxious levels.

#### ***Light-dark transition test***

The light-dark transition test uses the conflict between innate aversion to light areas and exploratory behavior of rodents to study anxious behaviors of mice<sup>1</sup>. Briefly, the apparatus consists of two same boxes (length: 20 cm; width: 15 cm; height: 25 cm), a light and dark box. The test mouse was placed in the center of the light box with head back to the hole, and allowed to freely shuttle between light and dark box for 10 mins. Their movement tracks in light box were recorded by video camera and SuperMaze software with the same parameters, and transition times were manually recorded. The apparatus was cleaned with 10% ethanol before each trial. A total of 33 mice were tested successfully (control: n = 11; *Tmem106b*-shRNA#1: n = 10; *Tmem106b*-shRNA#2: n = 12). Time and distance spent in light box and the transition times evaluated their anxiety levels.

#### ***Elevated plus maze test***

Elevated plus maze investigates the anxiety state of animals based on the contradictory behaviors

of exploring the new environment and the fear of hanging in open arms<sup>1</sup>. Briefly, the elevated plus maze apparatus is above the ground 75 cm and consists of four arms (two close arms (length: 35 cm; width: 5 cm; height: 15 cm) and two open arms (length: 35 cm; width: 5 cm) and a cross area (length: 5; width: 5 cm). The test mouse was placed in cross area facing the same open arms. Their movement tracks in elevated plus maze apparatus were recorded by video camera and SuperMaze software with the same parameters. The apparatus was cleaned with 10% ethanol before each trial. A total of 32 mice were tested successfully (control: n = 11; *Tmem106b*-shRNA#1: n = 10; *Tmem106b*-shRNA#2: n = 11). Time and distance spent in the open arms were recorded to reflect anxiety levels.

### ***Y maze test***

Y-maze investigates spatial working memory in mice based on spontaneous tendency to explore new environments<sup>1</sup>. Briefly, Y-shaped geometric arena with three arms (arm length, width, height: 35 cm, 5 cm, 15 cm) were defined as A, B, and C arms, and the angle between arm is 120°. The test mouse was placed in A arm and permitted to freely explore Y maze apparatus for 8 mins. Their movement tracks in Y maze apparatus were recorded by video camera and SuperMaze software with the same parameters, and manually recorded the series of arm entries. The spontaneous alternation event was defined as three consecutive arm entries that are different. The apparatus was cleaned with 10% ethanol before each trial. A total of 34 mice were tested successfully (control: n = 12; *Tmem106b*-shRNA#1: n = 10; *Tmem106b*-shRNA#2: n = 12). Spontaneous alternation percent in Y-maze was evaluated their spatial working memory levels.

### ***Sucrose preference test***

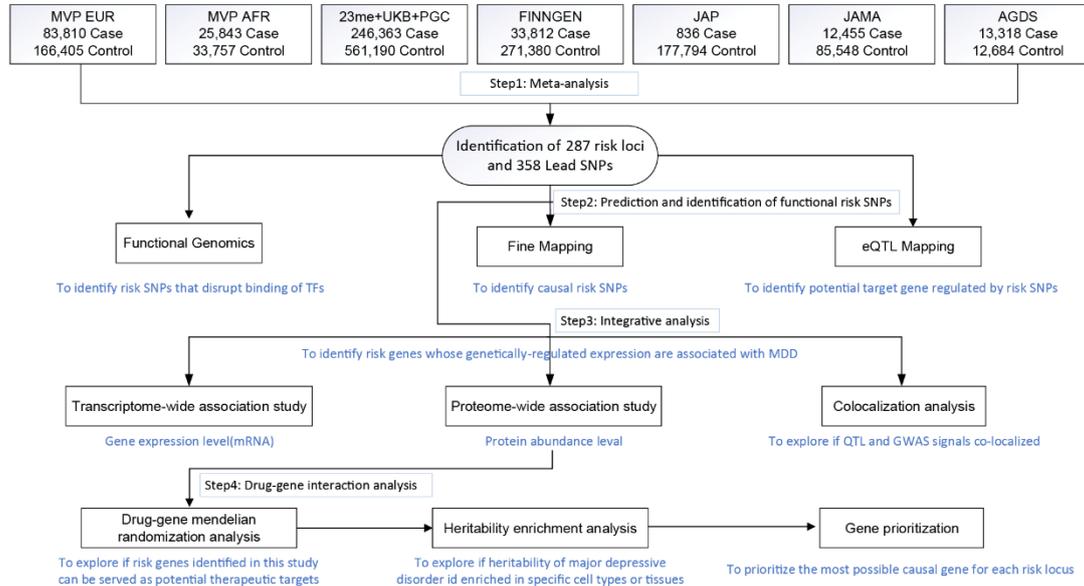
Sucrose preference test utilizes rodents' preference for sweetness to evaluate the anhedonia and depression of mice<sup>2</sup>. Briefly, the test mouse was first fed in individual cage and adapted to two bottles (one containing water, the other containing 1% sucrose solution) for 48 hours, and the position of two bottles was changed at 24 hours to avoid position and bottle preference. Secondly,

the two bottles were taken away for 24 hours after adaptation period. Finally, the mice were permitted to freely access water and sucrose solution for 6 hours and switched the positions of two bottles at 3 hours. The water and sucrose solution in each cage were measured before and after 6 hours. Sucrose preference (%) = sucrose consumption/ (sucrose consumption + water consumption) × 100%. A total of 32 mice were tested successfully (control: n = 11; *Tmem106b*-shRNA#1: n = 10; *Tmem106b*-shRNA#2: n = 11). Sucrose preference was evaluated their anhedonia and depressive levels.

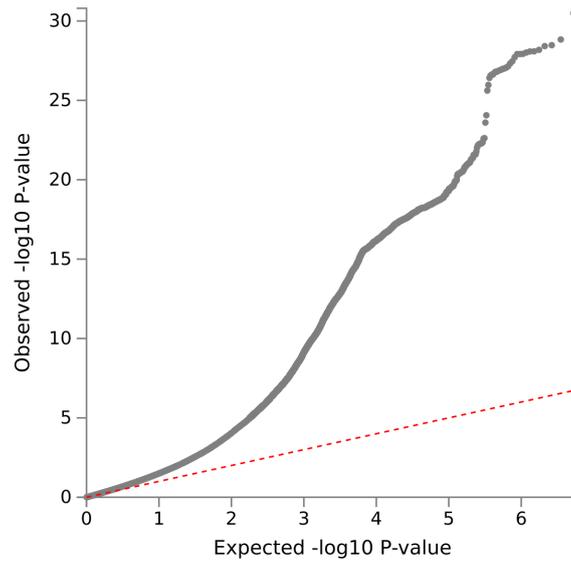
### ***Tail suspension test***

Tail suspension test was used to detect desire to survive of mouse in a desperate situation, the mouse frequently has no desire to survive in depressive state<sup>1</sup>. Briefly, mouse tail (about 1 cm) was fixed on the suspension hook in chamber box (length: 20 cm; width: 20 cm; height: 32 cm) using adhesive tape, and let the mouse body freely hung down in the air at a fixed height for 6 mins. Their activity (swing, struggle, and climb) were recorded by video camera and SuperTst software with the same parameters. A total of 31 mice were tested successfully (control: n = 10; *Tmem106b*-shRNA#1: n = 9; *Tmem106b*-shRNA#2: n = 12). Immobility state (swing) in the last 4 mins was used to evaluate their depressive levels.

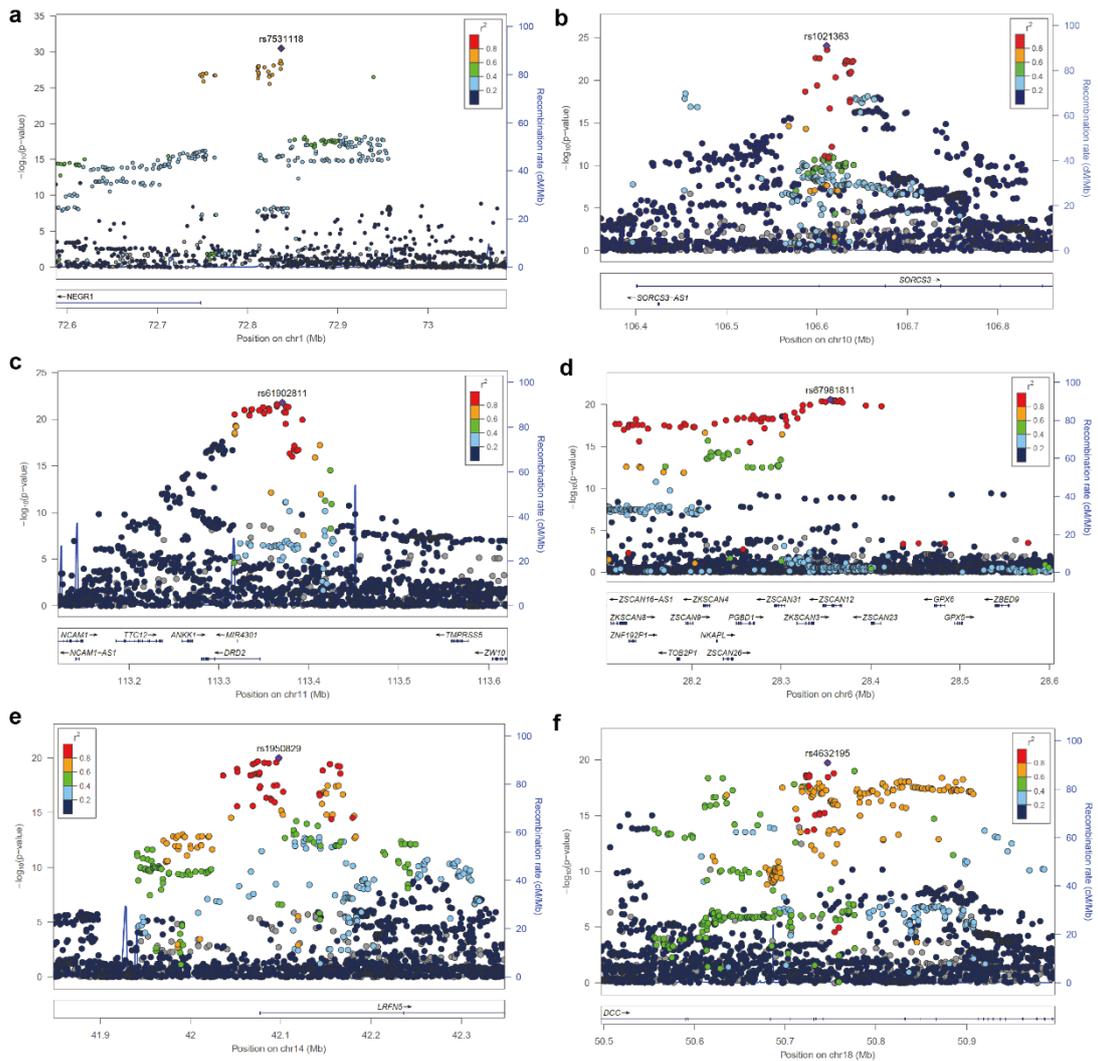
## Supplementary figures



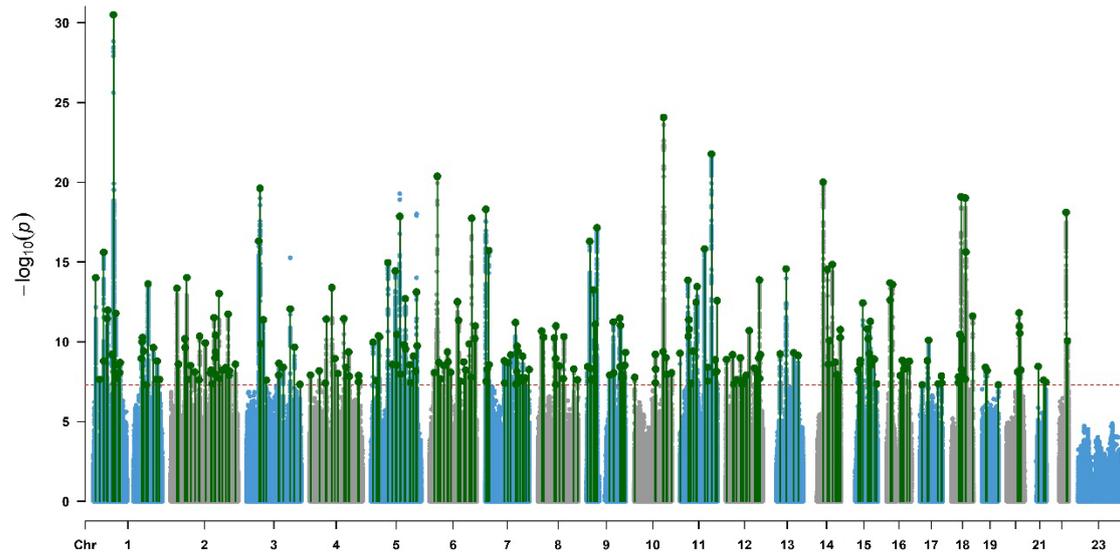
**Supplementary Figure 1. Overview of the analysis included in this study.** A meta-analysis was first conducted to identify novel risk loci for major depressive disorder. Based on the results of the meta-analysis, several analyses were then performed. SNP-based analyses, including fine-mapping, functional genomics, and expression quantitative trait loci (eQTL), were conducted to prioritize the functional (or potential causal) risk variants. Fine-mapping was conducted to prioritize the potential causal variants, functional genomics was performed to identify the SNPs that disrupt the binding of transcription factors (TFs), and eQTL was used to examine if the risk variants were associated with gene expression in the human brain. We further carried out gene-based analyses to prioritize the potential risk genes. Transcriptome-wide association study (TWAS) and proteome-wide association study (PWAS) were conducted to identify genes whose genetically-regulated expression is associated with major depressive disorder. Co-localization analysis was performed to explore if the QTL and GWAS signals were driven by the same variants. We conducted a drug-gene Mendelian randomization analysis to identify potential therapeutic targets, and we investigated whether the heritability of major depressive disorder is enriched in specific tissues or cell types. Based on lines of evidence from the above analyses, we performed a gene prioritization analysis to prioritize the most possible risk genes for major depressive disorders. Finally, we explored the potential role of *Tmem106b* (the rat homolog of the human *TMEM106B* gene) in mouse behavior.



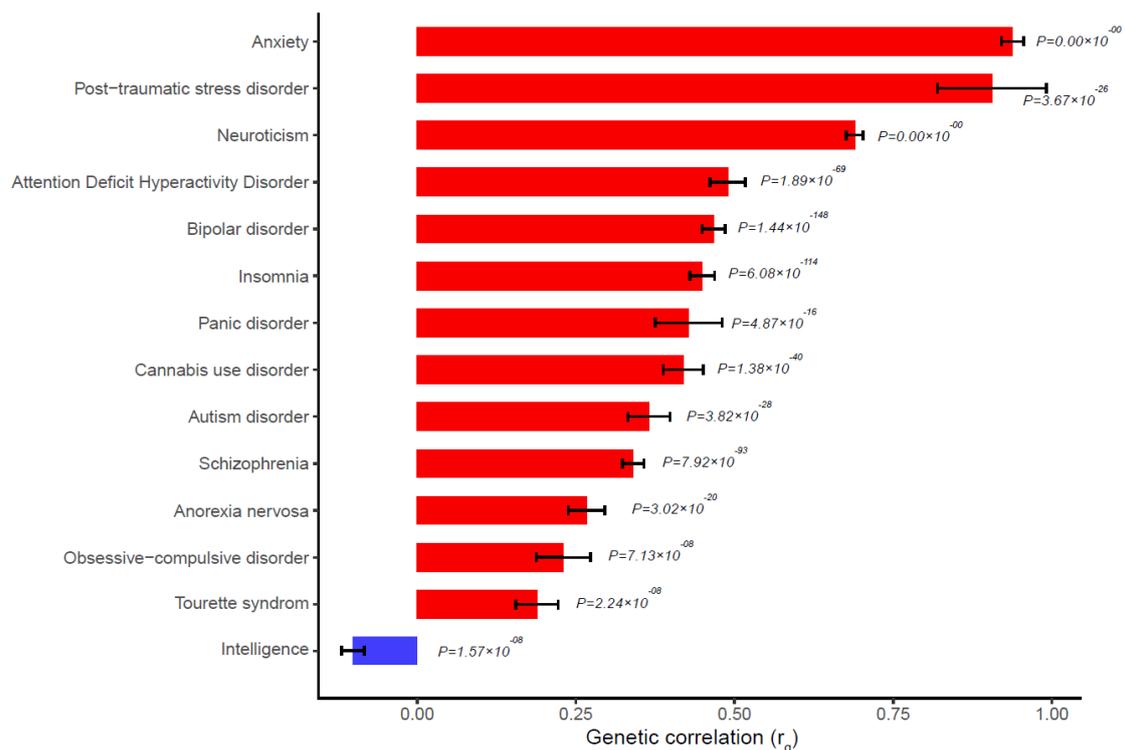
**Supplementary Figure 2. Quantile-quantile plot illustrating the GWAS Meta-Analysis for major depressive disorder.** The observed  $P$ -values are from cross-ancestry major depressive disorder GWAS meta-analyses, and all association analyses were two-sided. The red dashed line represents the predicted  $P$ -value line.



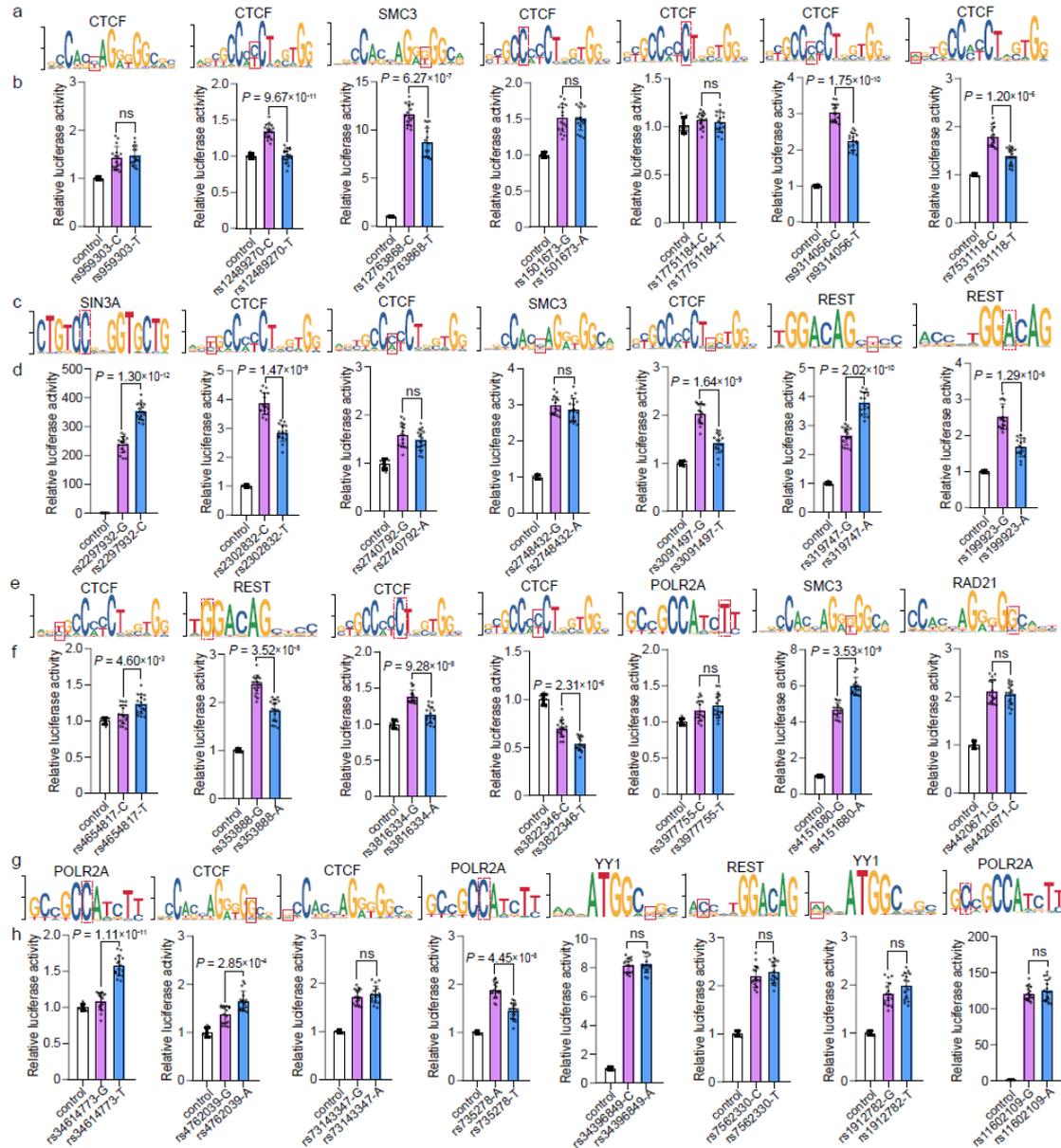
**Supplementary Figure 3. Locuszoom plots for the top six genome-wide significant association loci.** Among the top associations, rs7531118 is located in downstream of *NEGR1*, rs1021363 is located in intronic of *SORCS3*, rs61902811 is located in downstream of *DRD2*, rs67981811 is located in UTR3 of *ZSCAN12*, rs1950829 is located in intronic region of *LRFN5*, and rs4632195 is located in intronic of *DCC*. The reference LD panel for calculating LD between SNPs was based on the 1000 Genomes EUR reference panel. The *P*-values are from fixed-effects model GWAS meta-analyses, and all association analyses were two-sided.



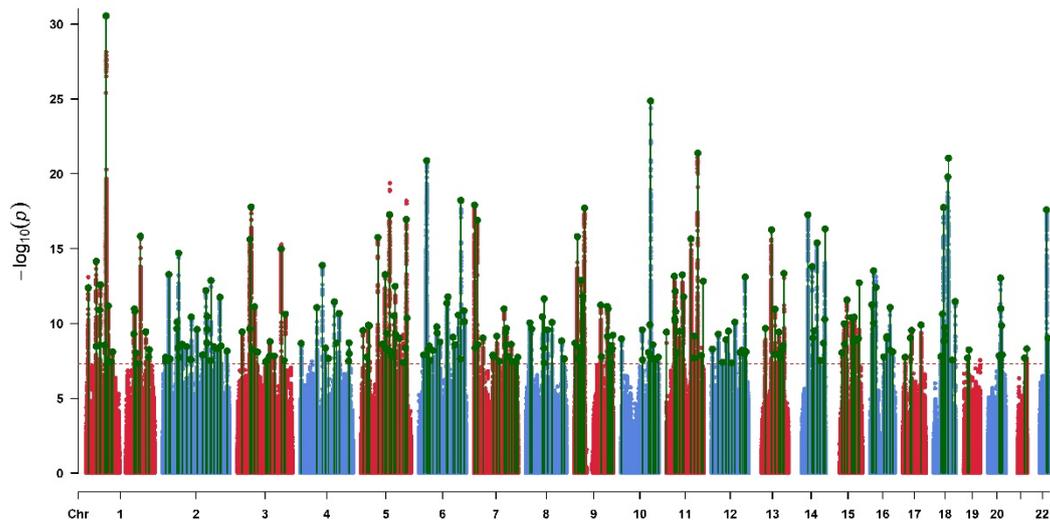
**Supplementary Figure 4. Manhattan plot of the cross-ancestry random-effect model GWAS meta-analysis.** Shows the associations of the random-effect model cross-ancestry GWAS meta-analysis (including MVP, 23andme-UKB-PGC, AGDS, study from Giannakopoulou et al., Biobank Japan, and FINNEG datasets, a total of 416,437 cases and 1,308,758 controls). The red line shows the genome-wide significant  $P$  threshold ( $P < 5.0 \times 10^{-8}$ ). A total of 272 genomic risk loci were identified and marked in green. The  $P$ -values are from the random effect model GWAS meta-analysis, and all association analyses were two-sided.



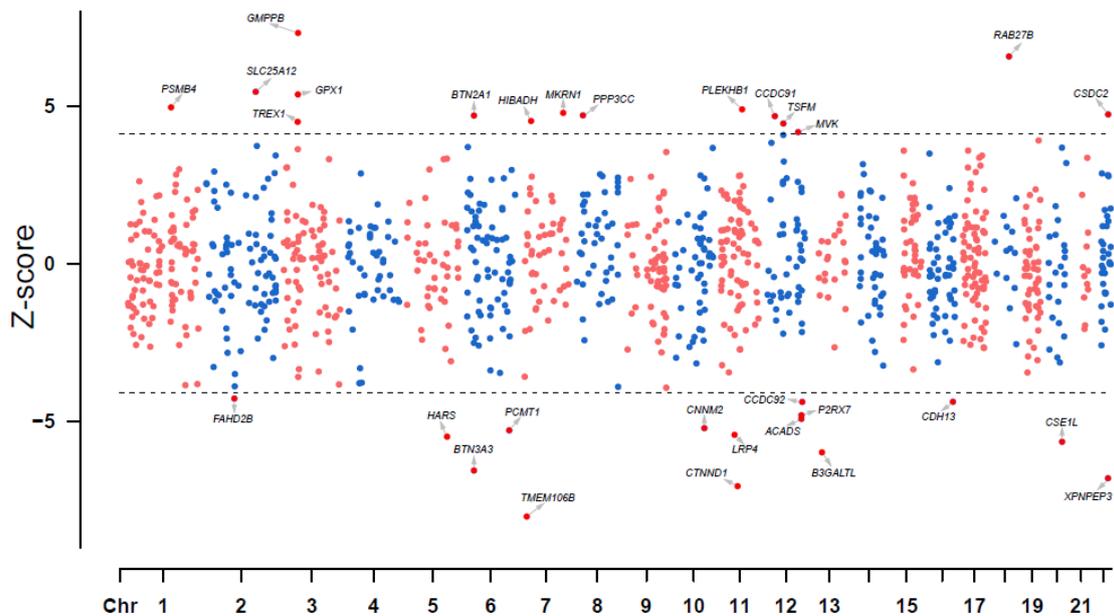
**Supplementary Figure 5. Genetic correlations between major depressive disorder (EUR ancestry meta-analysis results) and other brain disorders and intelligence.** Considering that the LDSC method is not suitable for admixed populations, we restricted LDSC genetic correlation analysis to the meta-analysis results from European ancestry (including MVP-EUR, 23andme-UKB-PGC, AGDS, and FINNEGNN datasets, a total of 377,303 cases and 1,011,659 controls). The horizontal axis indicates genetic correlation ( $r_g$ ), and the vertical axis indicates phenotypes. The bar chart shows the value of the genetic correlation between each phenotype and major depressive disorder, error bars represent standard errors, the  $P$ -value represents the  $P$ -value of genetic correlation calculated by LDSC, and all association analyses were two-sided. Anxiety disorders show the most significant genetic correlation with depression.



**Supplementary Figure 6. Reporter gene assays validated the regulatory effect of the identified TF binding-affecting SNPs. Upper panel, the position weight matrix (i.e., binding motif) of the corresponding TFs. Lower panel, results of reporter gene assays.** Reporter gene assays validated the regulatory effect of the identified TF binding-disrupting SNPs. N = 8 for the control group (8 independent technical replicates), n = 16 per experimental group for SH-SY5Y cells (16 independent technical replicates). Two-tailed *Student's t-test* was used for statistical analyses. Data represent mean  $\pm$  SD. The red dashed box highlights the test SNP. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

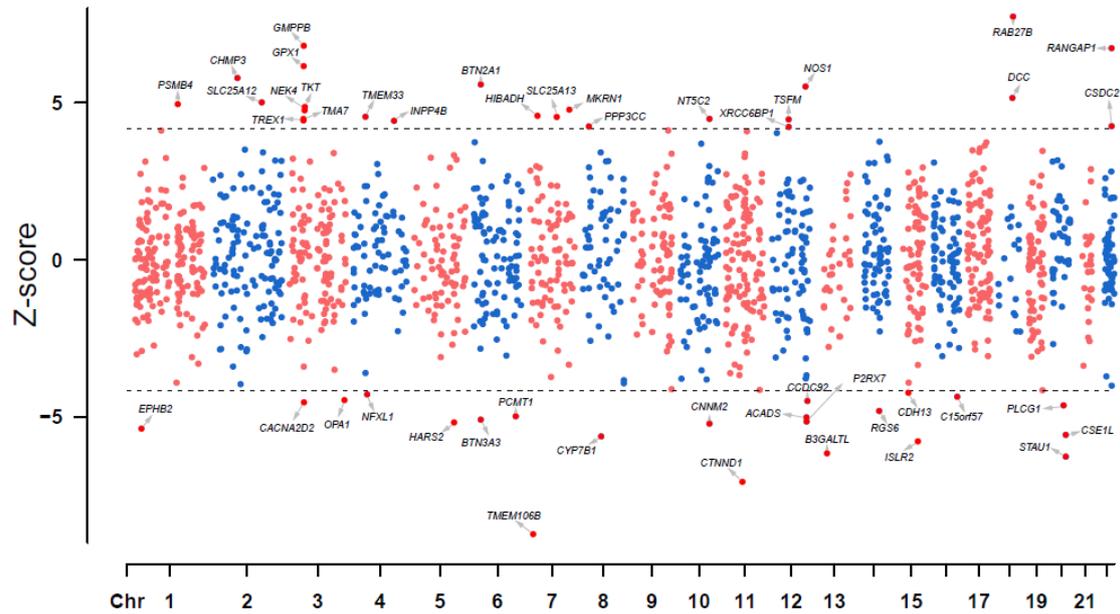


**Supplementary Figure 7. Manhattan plot of the EUR-only GWAS meta-analysis.** Associations of the EUR-only GWAS meta-analysis (including MVP-EUR, 23andme-UKB-PGC, AGDS, and FINNEGNET datasets, total 377,303 cases and 1,011,659 controls). The red line shows the genome-wide significant  $P$  threshold ( $P < 5.0 \times 10^{-8}$ ). A total of 267 genomic risk loci were identified and marked in green. The  $P$ -values are from fixed effect model GWAS meta-analysis, and all association analyses were two-sided.

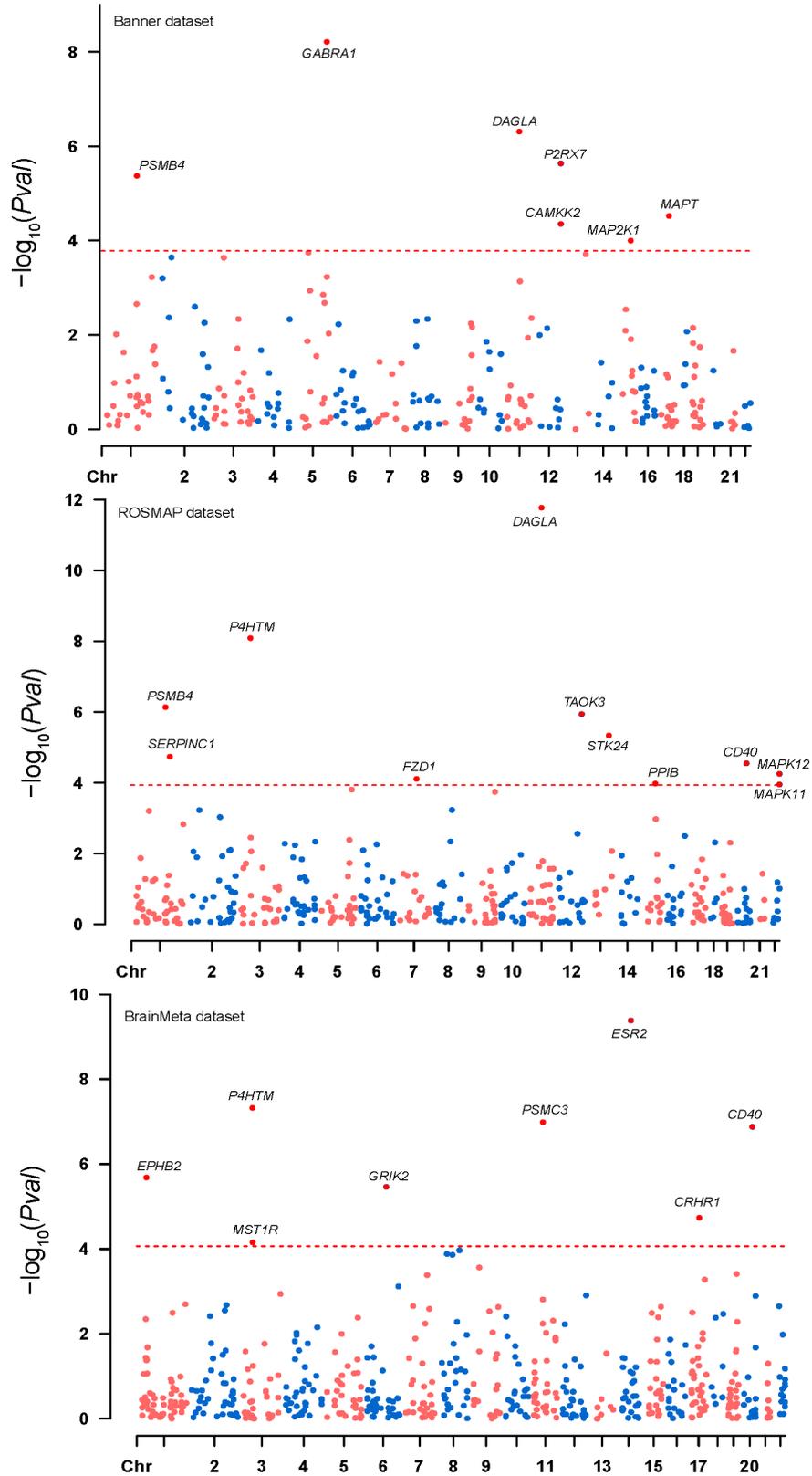


**Supplementary Figure 8. PWAS analysis result using Banner cis-pQTL data and EUR population MDD GWAS meta-analysis dataset.** The black dotted lines show the Bonferroni-corrected

significance Z-score threshold. Only PWAS-significant proteins are shown.

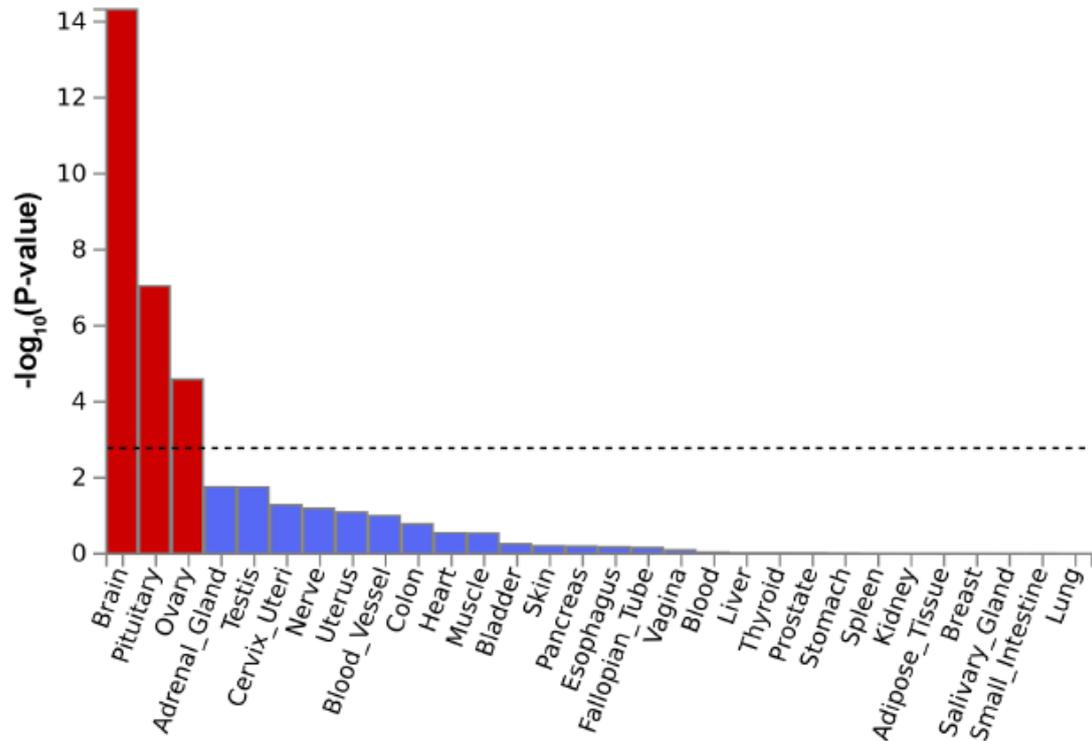


**Supplementary Figure 9. PWAS analysis result using ROSMAP cis-pQTL data and EUR population MDD GWAS meta-analysis dataset.** The black dotted lines show the Bonferroni-corrected significance Z score threshold. Only PWAS-significant proteins are shown.

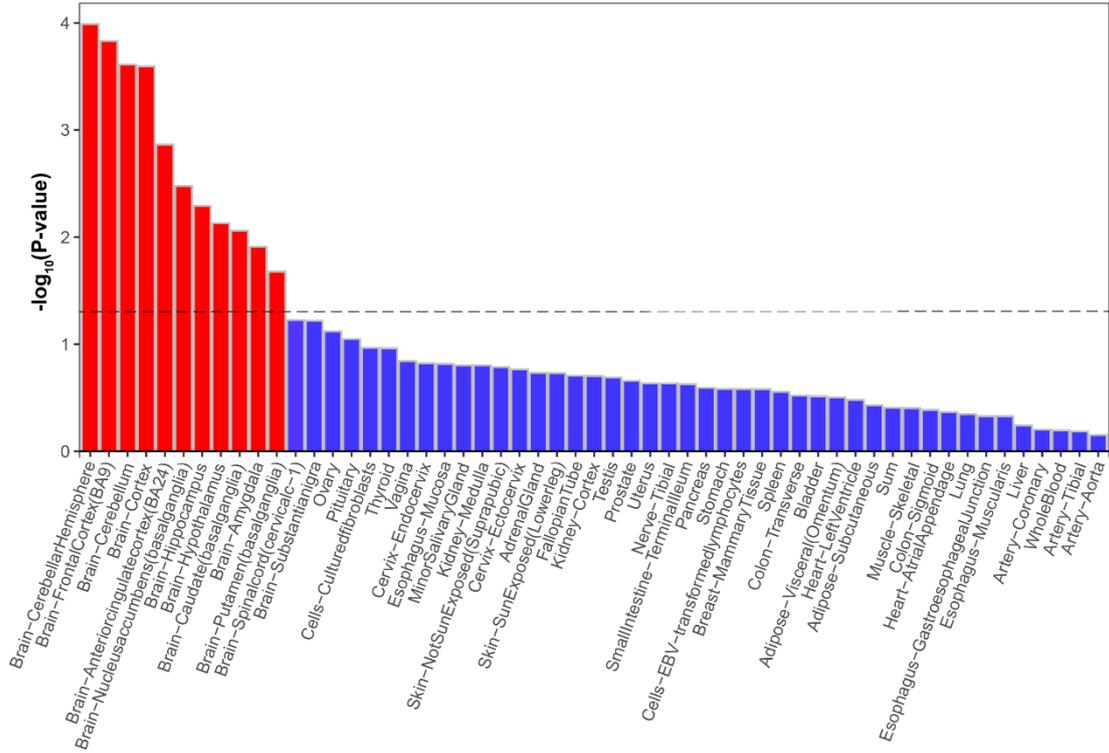


Supplementary Figure 10. Drug target gene mendelian randomization analysis using EUR population MDD GWAS meta-analysis dataset as outcome. Only MR-significant genes (Bonferroni corrected  $P < 0.05$ ) are shown. (a,b) ROSMAP and Banner pQTL MR results. (c) BrainMeta eQTL MR

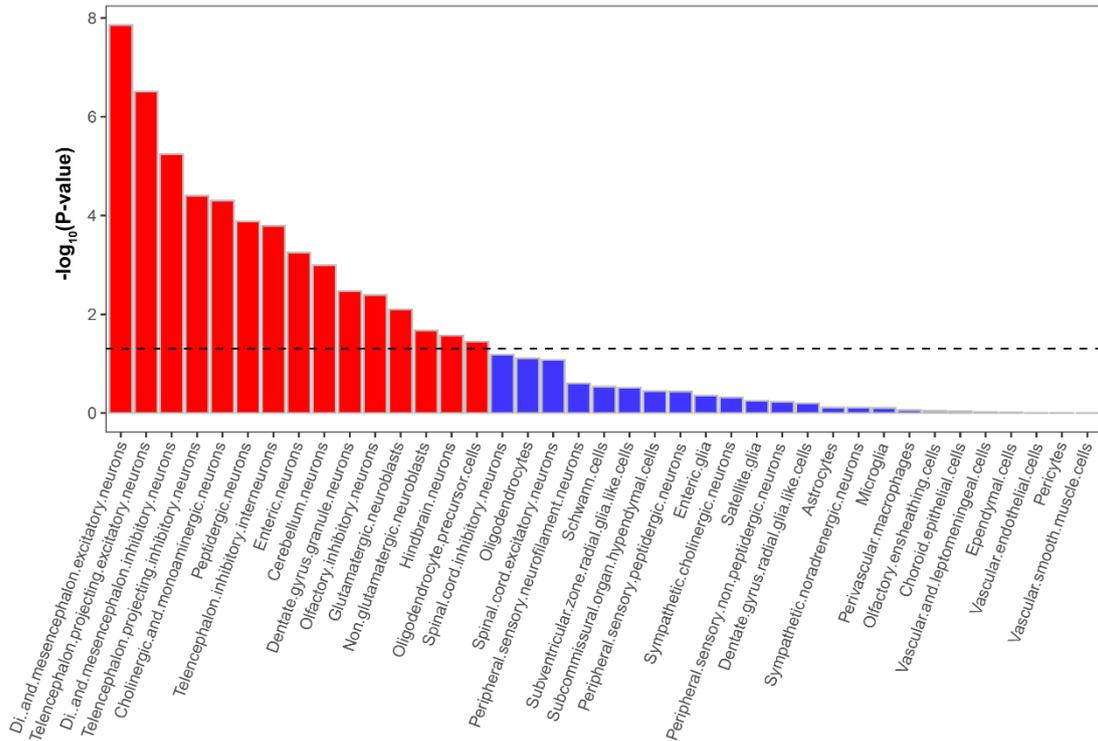
results. The  $P$ -values are from Mendelian randomization analyses, and all association analyses were two-sided.



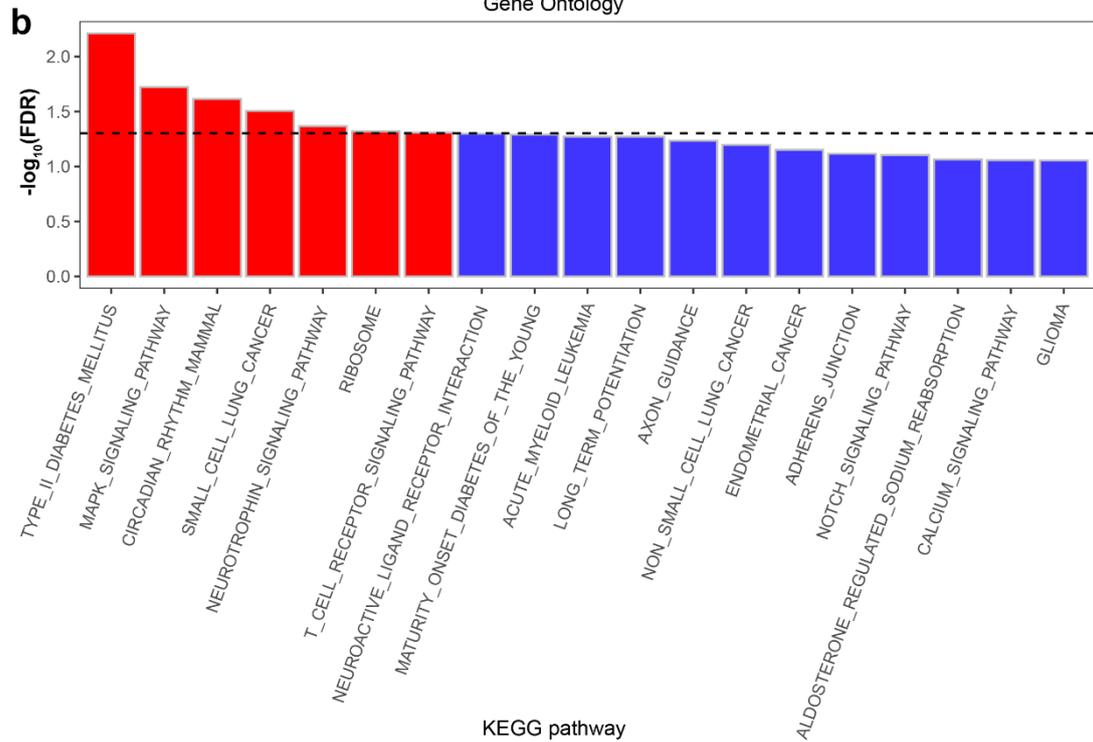
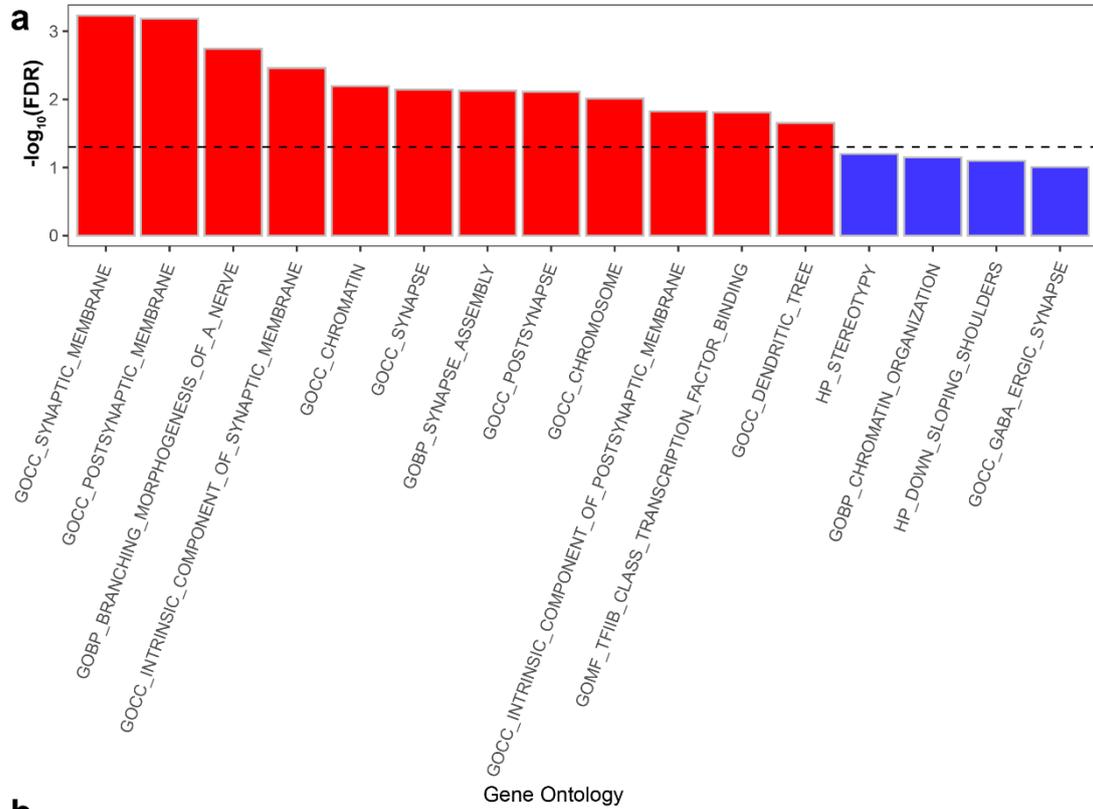
**Supplementary Figure 11. Heritability enrichment analysis in 30 general tissue types.**  $P$ -values are from the tissue-based MDD heritability enrichment analysis, the dashed lines indicate Bonferroni corrected significance threshold ( $-\log_{10}P = 3.04$ ) and all analyses were two-sided.



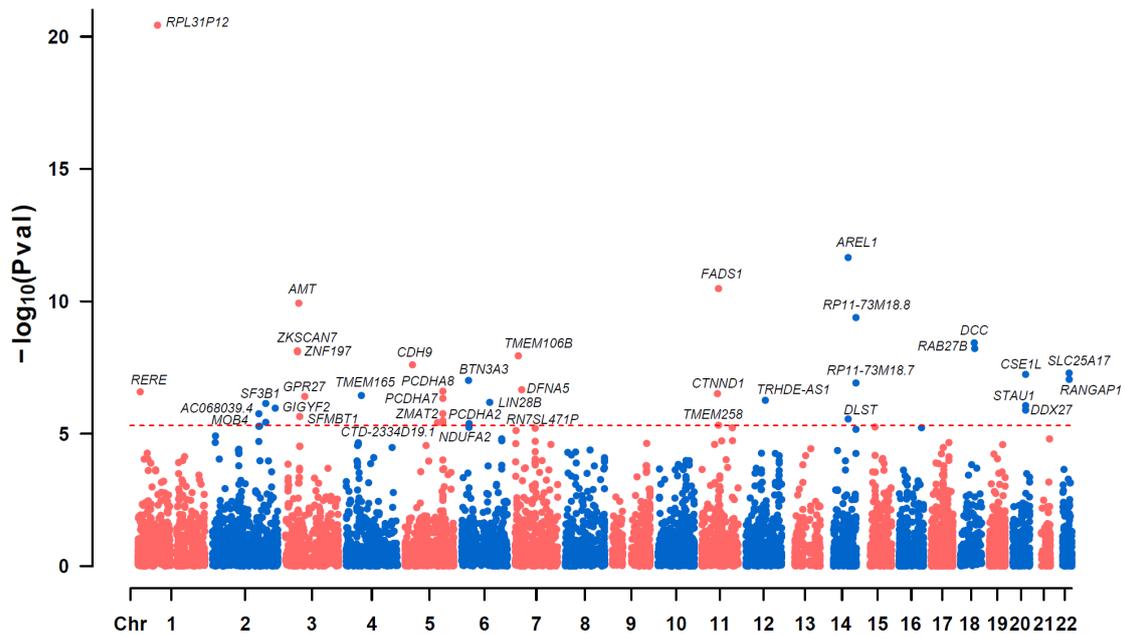
**Supplementary Figure 12. Heritability enrichment analysis in GTEx V8 detailed tissue types.**  $P$ -values are from the tissue-based MDD heritability enrichment analysis, the dashed lines indicate nominal significance thresholds ( $-\log_{10}P = 1.30$ ) and all analyses were two-sided.



**Supplementary Figure 13. Heritability enrichment analysis in different cell type types.** *P*-values are from the cell type based MDD heritability enrichment analysis, the dashed lines indicate nominal significance thresholds ( $-\log_{10}P = 1.30$ ) and all analyses were two-sided.

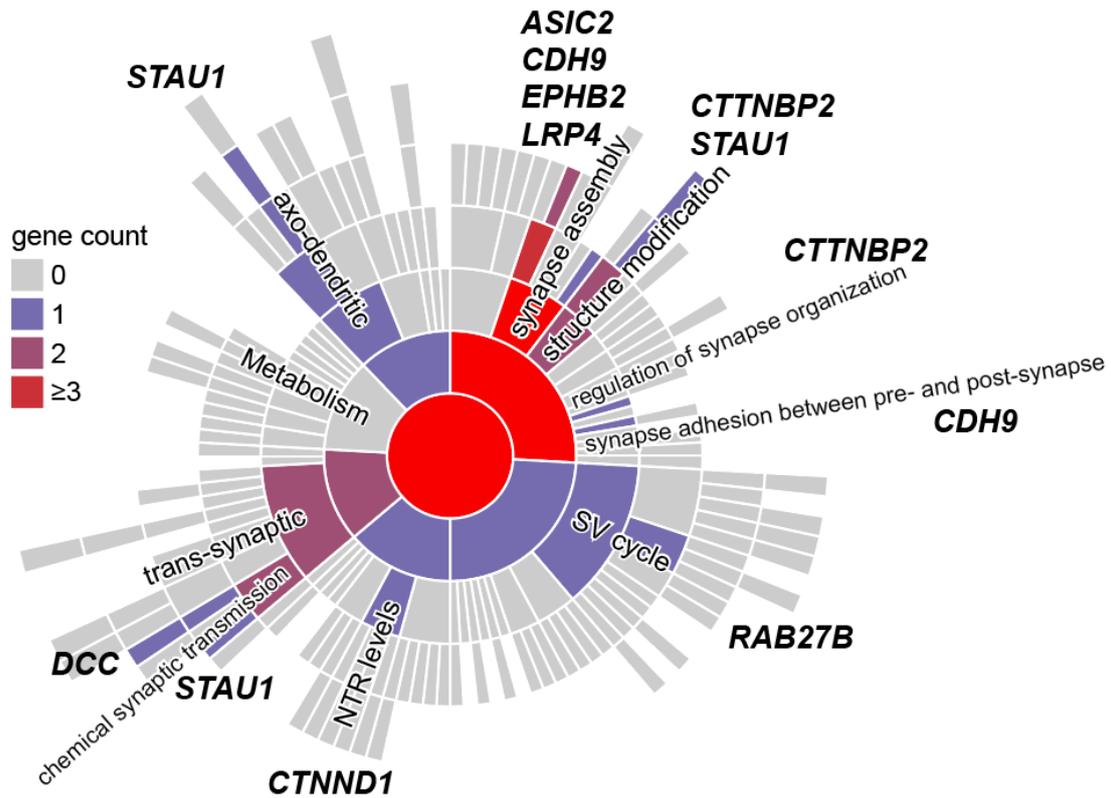


**Supplementary Figure 14. Heritability enrichment analysis of Gene Ontology (GO) and KEGG terms.** *P*-values are from the gene set enrichment analysis, the dashed lines indicate nominal significance thresholds ( $-\log_{10}P = 1.30$ ) and all analyses were two-sided.

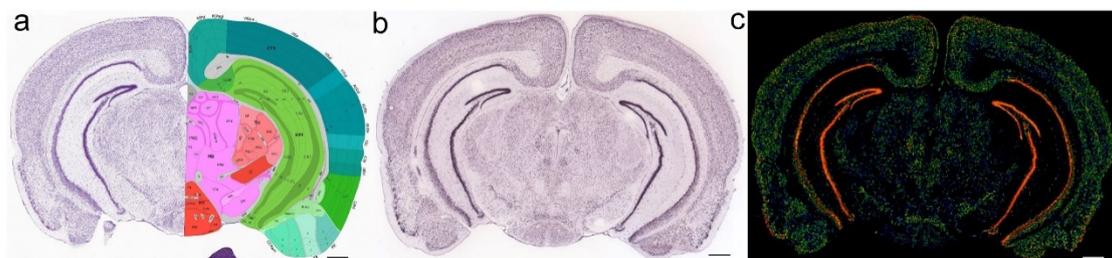


**Supplementary Figure 15. SMR analysis result using PsychENCODE cis-eQTL data.** The red line shows the Bonferroni-corrected significance threshold ( $0.05/10243, 4.88 \times 10^{-6}$ ). Only MR-significant genes are shown. *P*-values are from the SMR analysis, the dashed lines indicate nominal significance thresholds ( $P = 4.88 \times 10^{-6}$ ) and all analyses were two-sided.





**Supplementary Figure 17. MDD risk genes in SynGO biological processes annotations.** The colors show the number of MDD risk genes annotated in SynGO. SV: synaptic vesicle; NTR: neurotransmitter receptor.



**Supplementary Figure 18. The mRNA expression of *Tmem106b* in adult mouse brain.** (a) Nissl (left) and anatomical annotations (right) from the Allen Mouse Brain Atlas and Allen Reference Atlas – Mouse Brain, at the same slice position as **b** and **c** (scale bars represent 500  $\mu$ m). (b) In situ hybridization (ISH) staining of *Tmem106b* mRNA in adult mouse brain (scale bars represent 500  $\mu$ m). (c) Visualization image of *Tmem106b* expression level in **b**, blue indicates low expression and red indicates high expression (scale bars represent 500  $\mu$ m). Allen Mouse Brain Atlas, <https://mouse.brain->

[map.org/gene/show/47741](http://map.org/gene/show/47741).

## Reference

1. Li, S. *et al.* The schizophrenia-associated missense variant rs13107325 regulates dendritic spine density. *Translational Psychiatry* **12**, 361 (2022).
2. Luo, A. *et al.* The soluble epoxide hydrolase inhibitor TPPU improves comorbidity of chronic pain and depression via the AHR and TSPO signaling. *Journal of Translational Medicine* **21**, 71 (2023).