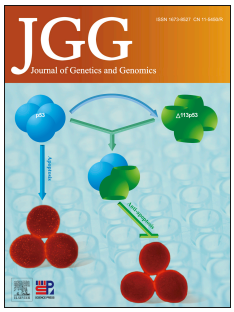


SZDB3.0: an updated comprehensive resource and web tools for schizophrenia research

Changgai Mu, Xinglun Dang, Shan-Shan Dai, Wenqi Lou, Yong Wu, Yong-Gang Yao, Wenqiang Li, Xiong-Jian Luo

PII: S1673-8527(26)00157-8

DOI: https://doi.org/10.1016/j.jgg.2026.04.025

Reference: JGG 1695

To appear in: *Journal of Genetics and Genomics*

Received Date: 19 January 2026

Revised Date: 28 April 2026

Accepted Date: 29 April 2026

Please cite this article as: Mu, C., Dang, X., Dai, S.-S., Lou, W., Wu, Y., Yao, Y.-G., Li, W., Luo, X.-J., SZDB3.0: an updated comprehensive resource and web tools for schizophrenia research, *Journal of Genetics and Genomics*, https://doi.org/10.1016/j.jgg.2026.04.025.

SZDB3.0: an updated comprehensive resource and web tools for schizophrenia research

Schizophrenia (SCZ) is a severe psychiatric disorder with high heritability, affecting about 1% of the global population. Owing to its high mortality, comorbidity, and recurrence rate, SCZ imposes a substantial economic burden on society. Although pharmacological treatments are available for SCZ, many patients do not respond to antipsychotics. Moreover, common side effects of antipsychotic drugs, such as weight gain and metabolic syndrome, are frequently observed. Therefore, there is a pressing need to discover new therapeutic targets and develop new drugs. However, such efforts are hindered by our limited understanding of the pathophysiology underlying SCZ.

To investigate the causes and pathophysiology of SCZ, many omics studies have been conducted at different levels or dimensions. For example, to decipher the genetic structure of SCZ, multiple genome-wide association studies (GWASs) have been conducted in diverse populations, and over 300 SCZ risk loci have been identified (Dang et al., 2025). Through large-scale RNA sequencing, the PsychENCODE consortium identified over four thousand differentially expressed genes (DEGs) in postmortem brain tissues from SCZ cases and healthy controls (Gandal et al., 2018). Transcriptome-wide association study (TWAS) and proteome-wide association study (PWAS) also identified multiple genes and proteins whose genetically regulated expression levels are associated with SCZ (Gandal et al., 2018; Dang et al., 2025).

Although omics studies have provided comprehensive and valuable resources for SCZ research, the omics data generated by each study are usually fragmented and dispersed in different platforms, and data collection and compilation are laborious and time-consuming. To address the pressing needs of the research community, we developed the SZDB, a comprehensive resource for schizophrenia research. SZDB1.0 (http://szdb.org/SZDB/) was first released in 2017 (Wu et al., 2017). In 2020, SZDB was updated to version 2.0 (SZDB2.0: http://szdb.org/) (Wu et al., 2020). Here we release version 3.0 (SZDB3.0) (www.szdb.org.cn).

SZDB3.0 is a comprehensive knowledge base that integrates multi-dimensional datasets, encompassing genomic, transcriptomic, proteomic, and other omics data. While maintaining the user-friendly style of SZDB2.0, the SZDB3.0 optimized classification system, and organized the datasets into seven distinct modules (Fig. 1A), including GWAS, Gene, quantitative trait loci (QTL), Functional SNP, Brainspan, Prioritized and drug target genes, and PPI modules. These modular designs enable researchers to efficiently browse and retrieve specific data types based on their customized needs. Compared to SZDB2.0, SZDB3.0 provides substantially expanded data content, more recent data releases, and enhanced accessibility. The updated and extended datasets are as follows:

(1) GWAS module: genome-wide associations from our recent study (Dang et al., 2025) were incorporated into SZDB3.0. Briefly, GWAS from East Asian (EAS), European (EUR), and cross-ancestry were included in SZDB3.0 (Fig. 1B). Additionally, we also performed comprehensive genomic annotation to identify the putative functional SNPs and target genes.

(2) Gene module: in this module, in addition to the DEGs from CommonMind Consortium (CMC) (Fromer et al., 2016), we extended SCZ-associated DEGs from PsychENCODE (543 cases and 899 controls) (Gandal et al., 2018) and Lieber Institute for Brain Development phase2 (LIBD2) (153 cases and 286 controls) (Collado-Torres et al., 2019), as well as SCZ risk genes and proteins identified through TWAS and PWAS based on brain tissue, along with causal genes for SCZ inferred through MR using brain tissue expression QTL (eQTL) data (Wang et al., 2025) (Fig. 1C).

(3) QTL module: in SZDB3.0, with the generation of larger-scale datasets, we have updated both eQTL and methylation QTL (mQTL) data and additionally incorporated large-scale splicing QTL (sQTL) data (Qi et al., 2018; Qi et al., 2022). More importantly, single-cell eQTL (sc-eQTL) data derived from brain tissue, including eQTLs from excitatory neurons, inhibitory neurons, astrocytes, microglia, oligodendrocytes, and oligodendrocyte precursor cells (OPC), were integrated into SZDB3.0 (Emani et al., 2024) (Fig. 1D).

(4) Functional SNP module: although GWASs have identified multiple risk loci for SCZ, identifying functional variants from the abundant SNPs in high linkage disequilibrium (LD) remains a major challenge. Currently, considerable efforts have been made to decipher functional SNPs from the reported risk loci. In a recent study, we employed functional genomic approaches and identified 249 SNPs that affect transcription factor binding (Dai et al., 2026). Additionally, in another study, we conducted fine-mapping and identified hundreds of high-confidence functional genetic variants (Dang et al., 2025). These functional SNPs were updated into SZDB3.0. Finally, two massively parallel reporter assay (MPRA) results were also added in SZDB3.0 (McAfee et al., 2023; Lee et al., 2025) (Fig. 1E).

(5) Other modules: in SZDB3.0, we also constructed three new modules. The Brainspan module contains spatiotemporal gene expression profiles from 12 developmental stages and 3 regions of the human brain (BrainSpan, 2013). The Prioritized and drug target gene module includes 108 high-credibility SCZ risk genes and 3,713 potential small-molecule drugs targeting these genes. We also included a prioritized gene dataset, which was generated by integrating multi-dimensional evidence from SZDB3.0, and identified a total of 1,007 SCZ risk genes supported by at least two lines of evidence. Among these genes, 19 were supported by four or more distinct evidence sources (Fig. 1F). The PPI module contains 9,021,673 non-redundant protein-protein interactions (PPI) obtained from previous publications (Huttlin et al., 2021; Kamburov et al., 2021; Kim et al., 2022).

The SZDB3.0 will facilitate the mechanistic dissection of SCZ. First, we updated the GWAS of EUR ancestry and expanded the datasets to include EAS and cross-ancestry populations (Dang et al., 2025). This expansion not only substantially enhanced the statistical power for identifying SCZ risk loci but also enabled the detection of ancestry-specific risk loci. Second, we expanded multiple SCZ risk genes into SZDB3.0, including DEGs from LIBD2 (Collado-Torres et al., 2019) and PsychENCODE (Gandal et al., 2018), as well as risk genes identified through TWAS, PWAS, and MR analyses (Dang et al., 2025; Wang et al., 2025). These risk genes offer crucial starting points for elucidating the pathogenic mechanisms of SCZ. Third, SZDB3.0 not only includes significantly

updated bulk-tissue eQTL, sQTL (Qi et al., 2022), and mQTL (Qi et al., 2018), but also extends sc-eQTLs data to include multiple brain cell types (Emani et al., 2024). These QTL resources provide deep insights into the regulatory effect of risk variants on gene expression and highlight the cell-type-dependent regulatory effect of risk variants. Fourth, we incorporated functional SNPs identified through multiple approaches into SZDB3.0. These functional SNPs offer important genetic evidence for understanding the pathogenic mechanisms underlying SCZ. Finally, high-confidence SCZ risk genes that were supported by multiple lines of evidence and the potential drugs targeting these genes were also included in SZDB3.0. These results offer valuable resources for investigating the pathophysiology of SCZ and the development of antipsychotic drugs. Overall, SZDB3.0 not only provides a valuable resource for SCZ research but also facilitates the translation of omics findings into disease biology and potential therapeutic targets.

The core value of this update is establishing a cross-omics bridge that connects genetic associations to biological function. While traditional genetic studies can identify disease-associated loci, the underlying causal genes and mechanisms often remain elusive. SZDB3.0 substantially addresses this bottleneck. First, the integration of genomic and QTL data enables researchers to directly link SCZ risk variants to gene expression regulation. Bulk tissue and sc-eQTL data offer complementary advantages, bulk tissue eQTLs provide a global perspective but cannot resolve cellular heterogeneity, whereas sc-eQTLs allow the detection of context-specific regulatory effects present only in particular cell types (e.g., excitatory neurons or microglia). This is particularly critical for SCZ, a disorder with complex cell-type-specific mechanisms in the brain (Emani et al., 2024). Second, our functional SNP module incorporates functional SNPs identified through functional genomics, fine-mapping, and MPRA, offering critical insights for further exploration of the pathogenic mechanisms of these variants. Third, our gene module serves as a critical resource for identifying promising candidate SCZ risk genes and characterizing the expression patterns of these genes in SCZ patients. Lastly, data from Brainspan, PPI, and drug target gene modules offer important insights into elucidating

the roles of SCZ target genes in disease pathogenesis and facilitating the development of potential therapeutic drugs for this disorder.

The following example illustrates how to elucidate the mechanisms of risk variants in disease pathogenesis using SZDB3.0. A previous study has suggested that rs6996860 is a risk variant for SCZ, with *DDHD2* as its putative target gene (Trubetskoy et al., 2022). By leveraging the GWAS findings included in SZDB3.0, we found that this SNP is significantly associated with SCZ in both European ($P = 5.14\times10^{-7}$) and cross-ancestry ($P = 1.91\times10^{-13}$) GWASs. The functional score of rs6996860 from the RegulomeDB database (Dong et al., 2023) is 0.55, suggesting the SNP is a potential functional variant. Furthermore, functional genomics data reveal that this SNP influences the binding of the transcription factor POU2F1, and MPRA assays demonstrate that it acts as a functional variant regulating gene expression in human neural progenitors. Bulk-tissue eQTL confirms the association between rs6996860 and *DDHD2* expression ($P = 1.49\times10^{-32}$), and sc-eQTLs further indicate significant associations between rs6996860 and the expression of *DDHD2* in excitatory neurons ($P = 2.99\times10^{-5}$), inhibitory neurons ($P = 8.53\times10^{-5}$), and oligodendrocyte ($P = 1.43\times10^{-5}$). These lines of evidence collectively suggest that rs6996860 is a functional SNP that modulates *DDHD2* expression in specific cell types. Furthermore, TWAS and MR also support *DDHD2* as a SCZ risk gene. Consistently, differential expression analyses from CMC (Fromer et al., 2016) and PsychENCODE (Gandal et al., 2018) datasets demonstrate *DDHD2* is downregulated in SCZ patients (Fig. 1G and 1H). Additionally, this gene is highly expressed in the human brain across developmental stages (Fig. 1I). Therefore, these results indicate that rs6996860 may increase the risk of SCZ by downregulating the expression of *DDHD2* in specific cell types. Collectively, this case demonstrates that SZDB3.0 enables users to gain mechanistic insights into the functional roles of genetic variants in SCZ pathogenesis.

It should be noted that most data included in SZDB3.0 were derived from European populations (Table S1). Therefore, findings based on these European cohorts may not be generalizable to other populations. Besides, for most datasets, we only included the most

representative and high-powered studies into SZDB3.0, this may limit the breadth of its data sources. In the future, we will incorporate more data to progressively expand its data coverage.

In summary, the current update to SZDB3.0 establishes a more comprehensive and multidimensional multi-omics database for SCZ research. This comprehensive knowledge base will accelerate the translation of genetic discoveries into biological mechanisms, laying a solid foundation for elucidating the complex etiology of SCZ and developing novel therapeutic strategies.

Data availability

All data presented in this study were obtained from previously published studies and have been deposited in SZDB3.0 (http://www.szdb.org.cn/). The sources of these data are cited in the supplementary materials and methods.

Conflict of interest

The authors declare no competing interests.

Acknowledgements

This study was equally supported by the Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project (2025ZD0216000) and the Youth Science Fund Project (Category A) of the National Natural Science Foundation of China (Grant No. 82525026 to X.-J.L.). This study was also supported by the National Natural Science Foundation of China (No. 82401767, No. 82460289, No. 82301690, No. 82501799, and No. 82501798), Fund from Yunnan Province (202305AH340006), the startup funds from Southeast University (RF1028623032), and the SEU Innovation Capability Enhancement Plan for Doctoral Students (CXJH_SEU 25233).

Supplementary data

Supplementary data to this article can be found online at xxx.

References

BrainSpan. 2013. BrainSpan: Atlas of the developing human brain. http://www.brainspan.org (accessed April 28 2015).

Collado-Torres, L., Burke, E. E., Peterson, A., Shin, J., Straub, R. E., Rajpurohit, A., Semick, S. A., Ulrich, W. S., Price, A. J., Valencia, C., et al., 2019. Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. Neuron 103, 203–216.

Dai, S., Dang, X., Gong, D., Li, D., Mu, C., Teng, Z., Luo, X.J., 2026. Molecular regulatory mechanisms of schizophrenia-associated functional non-coding variants. Mol. Psychiatry. doi: 10.1038/s41380-026-03566-8.

Dang, X., Teng, Z., Yang, Y., Li, W., Liu, J., Hui, L., Zhou, D., Gong, D., Dai, S.-S., Li, Y., et al., 2025. Gene-level analysis reveals the genetic aetiology and therapeutic targets of schizophrenia. Nat. Hum. Behav. 9, 609–624.

Dong, S., Zhao, N., Spragins, E., Kagda, M. S., Li, M., Assis, P., Jolanki, O., Luo, Y., Cherry, J. M., Boyle, A. P., et al., 2023. Annotating and prioritizing human non-coding variants with RegulomeDB v.2. Nat. Genet. 55, 724–726.

Emani, P. S., Liu, J. J., Clarke, D., Jensen, M., Warrell, J., Gupta, C., Meng, R., Lee, C. Y., Xu, S., Dursun, C., et al., 2024. Single-cell genomics and regulatory networks for 388 human brains. Science 384, eadi5199.

Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., Ruderfer, D. M., Oh, E. C., Topol, A., Shah, H. R., et al., 2016. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. 19, 1442–1453.

Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al., 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362, eaat8127.

Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., Baltier, K., Gebreab, F., Gygi, M. P., Thornock, A., Zarraga, G., Tam, S., et al., 2021. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184, 3022–3040.e3028.

Kamburov, A., & Herwig, R., 2021. ConsensusPathDB 2022: molecular interactions update as a resource for network biology. Nucleic Acids Res. 50, D587–D595.

Kim, C. Y., Baek, S., Cha, J., Yang, S., Kim, E., Marcotte, E. M., Hart, T., & Lee, I., 2022. HumanNet v3: an improved database of human gene networks for disease research. Nucleic Acids Res. 50, D632–D639.

Lee, S., McAfee, J. C., Lee, J., Gomez, A., Ledford, A. T., Clarke, D., Min, H., Gerstein, M. B., Boyle, A. P., Sullivan, P. F., et al., 2025. Massively parallel reporter assay investigates shared genetic variants of eight psychiatric disorders. Cell 188, 1409–1424.

McAfee, J. C., Lee, S., Lee, J., Bell, J. L., Krupa, O., Davis, J., Insigne, K., Bond, M. L., Zhao, N., Boyle, A. P., et al., 2023. Systematic investigation of allelic regulatory activity of schizophrenia-associated common variants. Cell Genom. 3, 100404.

Qi, T., Wu, Y., Fang, H., Zhang, F., Liu, S., Zeng, J., & Yang, J., 2022. Genetic control of RNA splicing and its distinct role in complex trait variation. Nat. Genet. 54, 1355–1363.

Qi, T., Wu, Y., Zeng, J., Zhang, F., Xue, A., Jiang, L., Zhu, Z., Kemper, K., Yengo, L., Zheng, Z., et al., 2018. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat. Commun. 9, 2282.

Trubetskoy, V., Pardinas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C. Y., Dennison, C. A., Hall, L. S., et al., 2022. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604, 502–508.

Wang, D., Li, D., Dang, X., Mu, C., Liu, C., Zeng, Y., Yuan, Y., Teng, Z., Li, Y., & Luo, X. J., 2025. Mendelian randomization reveals causalities between DNA methylation and schizophrenia. Biol. Psychiatry 98, 712–723.

Wu, Y., Li, X., Liu, J., Luo, X.-J., & Yao, Y.-G., 2020. SZDB2.0: an updated comprehensive resource for schizophrenia research. Hum. Genet. 139, 1285–1297.

Wu, Y., Yao, Y. G., & Luo, X. J., 2017. SZDB: a database for schizophrenia genetic research. Schizophr. Bull. 43, 459–471.

Changgai Mu

Department of Psychosomatics and Psychiatry, Zhongda Hospital, State Key Laboratory of Digital Medical Engineering, School of Medicine, Advanced Institute for Life and Health, Jiangsu Provincial Key Laboratory of Brain Science and Medicine, Southeast University, Nanjing, Jiangsu 210096, China

Medical Research Center, Nantong First People's Hospital, Southeast University, Nanjing, Jiangsu 210096, China

Xinglun Dang, Shan-Shan Dai, Wenqi Lou

Department of Psychosomatics and Psychiatry, Zhongda Hospital, State Key Laboratory of Digital Medical Engineering, School of Medicine, Advanced Institute for Life and Health, Jiangsu Provincial Key Laboratory of Brain Science and Medicine, Southeast University, Nanjing, Jiangsu 210096, China

Yong Wu

State Key Laboratory of Genetic Evolution & Animal Models, Yunnan Key Laboratory of Animal Models and Human Disease Mechanisms, and KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, 650204, China

Yong-Gang Yao

State Key Laboratory of Genetic Evolution & Animal Models, Yunnan Key Laboratory of Animal Models and Human Disease Mechanisms, and KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, 650204, China

National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), National Resource Center for Non-Human Primates, Yunnan Engineering Center on Brain Disease Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, 650107, China

Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming,

Yunnan 650204, China

Wenqiang Li

Henan Mental Hospital, the Second Affiliated Hospital of Xinxiang Medical University, Xinxiang, Henan 453002, China

Xiong-Jian Luo*

Department of Psychosomatics and Psychiatry, Zhongda Hospital, State Key Laboratory of Digital Medical Engineering, School of Medicine, Advanced Institute for Life and Health, Jiangsu Provincial Key Laboratory of Brain Science and Medicine, Southeast University, Nanjing, Jiangsu 210096, China

Medical Research Center, Nantong First People's Hospital, Southeast University, Nanjing, Jiangsu 210096, China

*Corresponding author.

E-mail: luoxiongjian@seu.edu.cn

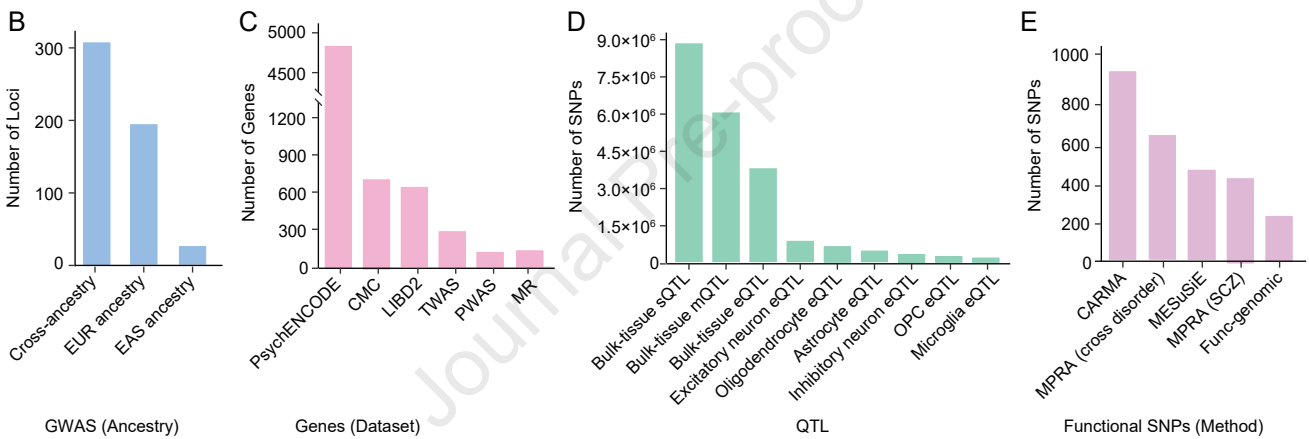
Figure legends

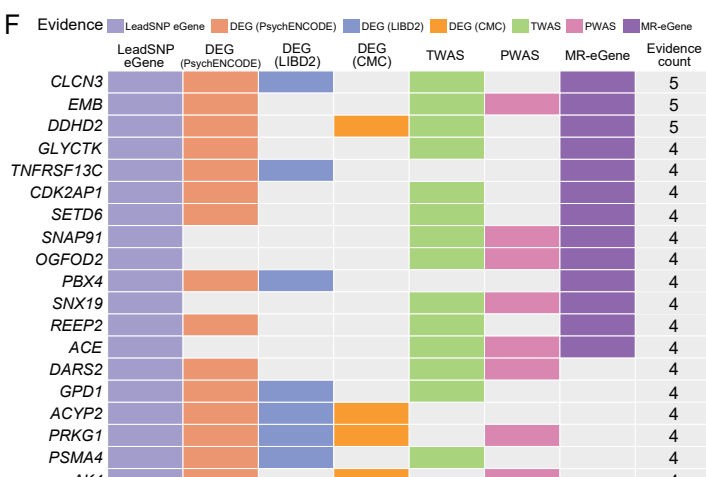
Fig. 1. Overview of SZDB3.0 (www.szdb.org.cn). **A**: Database structure of SZDB3.0, comprising four core modules. **B**: The SCZ risk loci identified in different ancestries. **C**: SCZ risk genes identified using different datasets and methods. **D:** Significant QTLs identified in different datasets, including bulk-tissue sQTL, eQTL, mQTL, and single-cell eQTL. **E**: Functional SNPs identified using different methods. **F**: Prioritized risk genes. The SCZ candidate genes supported by at least four lines of evidence are displayed. **G** and **H**: *DDHD2* is significantly downregulated in SCZ patients. **I**: The dynamics of *DDHD2* expression at different developmental time points in the dorsolateral prefrontal cortex. LeadSNP eGene, genes whose expression levels are associated with the lead SNP; DEG (PsychENCODE), DEG identified using samples from PsychENCODE; DEG (LIBD2), DEG identified using samples from LIBD2; DEG (CMC), DEG identified using samples from CMC; TWAS, SCZ risk genes identified by TWAS; PWAS, SCZ risk genes identified by PWAS; MR-eGene, SCZ causal genes identified by MR. T1: 8–9 post-conceptional weeks (PCW); T2: 12–13 PCW; T3: 16–17 PCW; T4: 19–21 PCW; T5: 24–26 PCW; T6: 4–6 months postnatally (mos); T7: 10–12 mos; T8: 2–4 years postnatally (yrs); T9: 8–11 yrs; T10: 13–15 yrs; T11: 18–23 yrs; T12: 30–40 yrs. SCZ, schizophrenia; DEG, differentially expressed gene; LIBD2, Lieber Institute for Brain Development phase2; CMC, CommonMind Consortium; TWAS, transcriptome-wide association study; PWAS, proteome-wide association study.

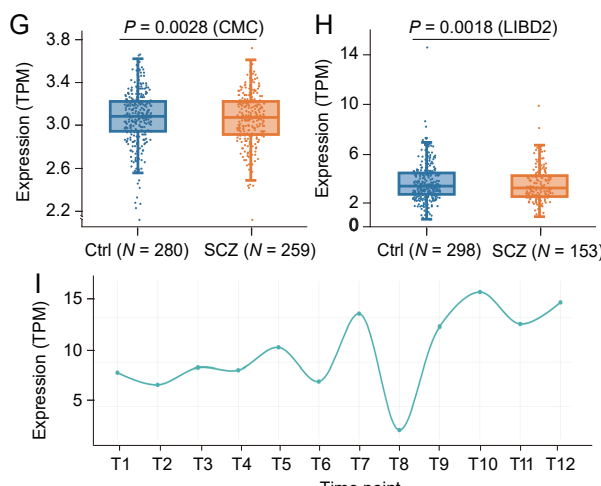
A

SZDB v3.0

GWAS	Genes	QTL	Functional SNPs	Others
1. European ancestry	1. DEGs from PsychENCODE	1. Brain bulk-tissue mQTL	1. SNPs affecting TF binding site	1. Prioritized and drug target genes
2. East Asian ancestry	2. DEGs from LIBD2	2. Brain bulk-tissue eQTL	2. SNPs identified by CARMA Fine-mapping	2. Brainspan
3. Cross-ancestry	3. DEGs from CMC	3. Brain bulk-tissue sQTL	3. SNPs identified by MESuSiE Fine-mapping	3. Protein-protein interactions
	4. Genes identified by TWAS	4. Astrocyte eQTL	4. SNPs identified by MPRA	
	5. Genes identified by PWAS	5. Microglia eQTL		
	6. Genes identified by MR	6. Excitatory neuron eQTL		
		7. Inhibitory neuron eQTL		
		8. Oligodendrocyte eQTL		
		9. OPC eQTL		







Supplementary Materials and Methods

GWAS datasets and SNP module

Genome-wide association studies (GWASs) summary statistics from our recent study (Dang et al., 2025) were incorporated into SZDB3.0. Briefly, we performed a large-scale GWAS meta-analysis by combining association data from populations of East Asian (EAS), European (EUR), African American (AA) and Latino (LAT). In the EAS cohort, a total of 29,519 cases and 44,392 controls were included, and 23 SCZ risk loci were identified (ten of which were EAS-specific risk loci) (Lam et al., 2019; Liu et al., 2021; Trubetskoy et al., 2022). For the EUR population, summary statistics from 59,901 cases and 441,418 controls were included (Trubetskoy et al., 2022; Kurki et al., 2023). Additionally, a cross-ancestry meta-analysis incorporating EAS, EUR, AA, and LAT populations was conducted, and 308 risk loci were identified. The details about the study can be found in the original publication (Dang et al., 2025).

We performed comprehensive functional annotation for SNPs identified by GWAS. First, we utilized ANNOVAR software to annotate SNPs by using the MANE SELECT transcript (MANE.GRCh38.v1.5.ensembl_genomic.gtf), including predicting their potential target genes and mapping them to specific genomic regions (e.g., intron, 5' UTR, 3' UTR) (Wang et al., 2010). Second, we used the probabilistic score from the RegulomeDB database (Dong et al., 2023) to annotate each SNP. The probabilistic score was assigned based on a comprehensive integration of lines of evidence, such as chromatin accessibility, transcription factor binding, histone modifications, and expression quantitative trait loci (eQTL). Consequently, a higher score indicates a higher functional probability of the SNP (Dong et al., 2023).

Gene module

Differentially expressed genes: Gene expression dysregulation has a critical role in the pathogenesis of SCZ. Multiple transcriptomic studies have been conducted to investigate the differences in gene expression between patients with SCZ and healthy controls. In SZDB2.0, differential gene expression data from the CommonMind Consortium (CMC, 258 cases and 279 controls) (Fromer et al., 2016) were included. In SZDB3.0, two large-scale transcriptomic datasets

were incorporated, including datasets from PsychENCODE Consortium (Gandal et al., 2018) (543 cases and 899 controls) and Lieber Institute for Brain Development phase2 (LIBD2, 153 cases and 286 controls) (Collado-Torres et al., 2019). Briefly, PsychENCODE collected the largest RNA-seq data from postmortem human brains of SCZ patients and non-psychiatric controls (Gandal et al., 2018). For the postmortem human brains, PsychENCODE conducted strict quality control on all samples (samples with an ambiguous diagnosis, age, sex, as well as individuals with low sequencing quality were discarded, and obtained 1,442 high-quality RNA-seq samples (543 SCZ patients and 899 non-psychiatric controls). After read mapping and quantification, differential gene expression analysis was conducted, and 4,821 differentially expressed genes (DEGs) (FDR < 0.05) were identified (Gandal et al., 2018). LIBD2 collected 460 adult postmortem human brain tissues and performed RNA-seq in hippocampi and dorsolateral prefrontal cortices (DLPFCs). Differential expression analysis was conducted with adjustment for age, ethnicity and technical covariates. More details about the LIBD2 dataset can be found in the original study (Collado-Torres et al., 2019).

Genes identified by TWAS and PWAS*:* Transcriptome-wide association study (TWAS) and proteome-wide association study (PWAS) are powerful approaches in identifying risk genes or proteins whose genetically regulated expression levels are associated with SCZ (Zhu et al., 2016; Brandes et al., 2020). In SZDB3.0, we incorporated TWAS and PWAS datasets from our recent study (Dang et al., 2025). For TWAS, the gene expression weight matrix were obtained from PsychENCODE (N = 1,321), and a total of 277 risk genes were identified (Gandal et al., 2018). For PWAS, human brain protein expression weight matrix from Banner Brain Database (Banner, N = 155) (Wingo et al., 2021) and Religious Orders Study and Memory and Aging Project (ROS/MAP, N = 376) (Bennett et al., 2018) were utilized. A total of 58 and 52 significant risk proteins were identified in Banner (Wingo et al., 2021) and ROS/MAP(Bennett et al., 2018) datasets, respectively.

Genes identified by MR: Mendelian randomization (MR) is a powerful statistical approach which leverages genetic variants as instrumental variables (IVs) to estimate the causal effect of exposure on outcome (Sanderson et al., 2022). Wang *et al*. (Wang et al., 2025) performed MR analyses using eQTL from BrainMeta V2 (N = 2,865) (Qi et al., 2022) and SCZ GWAS from PGC3 (Trubetskoy et al., 2022), and identified 123 potential causal risk genes. We collected and incorporated these SCZ causal genes into SZDB3.0 (Wang et al., 2025).

QTL module

GWASs have identified multiple SCZ risk variants. However, the mechanisms by which these variants confer SCZ risk remain largely unknown. Most risk variants identified by GWASs are localized in non-coding regions, suggesting that most risk variants confer disease susceptibility by modulating gene expression. eQTL mapping enables the identification of genetic variants associated with gene expression levels. Methylation QTL (mQTL) analysis can pinpoint variants correlated with DNA methylation levels. Splicing QTL (sQTL) analysis allows for the detection of variants linked to alternative splicing of RNA. We incorporated three QTL datasets into SZDB3.0.

eQTL and sQTL datasets: The eQTL and sQTL datasets from Qi et al. (Qi et al., 2022) were included in SZDB3.0. These two datasets represent the largest brain eQTL and sQTL to date. Briefly, Qi et al. (Qi et al., 2022) collected ten brain transcriptomic datasets from seven cohorts and obtained a total of 2,865 samples from 2,443 unrelated individuals of European ancestry. In total, they identified 1,342,073 unique cis-sQTL with $P < 5\times10^{-8}$ for 9,305 genes, 1,962,048 unique cis-eQTL with $P < 5\times10^{-8}$ for 16,704 genes.

mQTL dataset: DNA methylation is an important epigenetic mechanism for gene expression regulation. Numerous studies have investigated the associations between genetic variants and DNA methylation. To enhance statistical power, Qi et al. (Qi et al., 2018) performed a meta-analysis (N = 1,160) by combining three mQTL studies that were from Hannon et al. (Hannon et al., 2016), Jaffe et al. (Jaffe et al., 2016) and ROS/MAP (Ng et al., 2017). In total, they identified 1,868,200 unique cis-mQTL with $P < 5\times10^{-8}$ for 97,263 DNA methylation probes. mQTL from Qi *et al*. (Qi et al., 2018) were incorporated into SZDB3.0.

Single-cell eQTL datasets: Tissues and organs are composed of highly heterogeneous cell populations, and the brain exhibits the highest diversity of cell types. Conventional bulk-tissue eQTL studies have provided important insights into the relationships between genetic variants and gene expression. However, given gene expression and genetic regulation are highly dependent on cell types, it is important to investigate the correlations between genetic variants and gene expression in specific cell types. Recently, PsychENCODE released a single-cell eQTL dataset from 388 brain samples obtained from 182 healthy controls and 206 patients with mental disorders,

covering excitatory neurons, inhibitory neurons, astrocytes, microglia, oligodendrocytes and oligodendrocyte precursor cells (Emani et al., 2024). These single-cell eQTL data from PsychENCODE were incorporated into SZDB3.0.

Functional SNP module

Functional SNP identified by functional genomics: To prioritize functional (or potential causal) variants from GWAS-identified risk loci, we conducted a functional genomics study to identify risk variants that affect the binding affinity of transcription factors. Through integrating binding motifs of transcription factors from ENCODE (Moore et al., 2020) and SCZ GWAS summary statistics, we identified 249 SNPs that affect the binding of transcription factors (Dai et al., 2026).

Functional SNP identified by MPRA: Massively parallel reporter assay (MPRA) is a high-throughput experimental methodology which identifies expression modulating variants (emVar). Owing to the scalability and direct functional readout, MPRA has become a useful tool to functionally characterize non-coding risk variants identified by GWASs (Townsley et al., 2020). McAfee et al. (McAfee et al., 2023) and Lee et al. (Lee et al., 2025) identified 439 and 631 SCZ-associated emVar, respectively, and these MPRA results were incorporated into SZDB3.0.

Functional SNP identified by fine-mapping: Fine-mapped variants from our recent research (Dang et al., 2025) were incorporated into SZDB3.0. Briefly, by using CARMA fine-mapping framework (Yang et al., 2023), we identified 957 potential causal SNPs with posterior probability (PIP) > 0.7. We also performed fine-mapping using MESuSiE (Gao et al., 2024), and identified 490 and 1,209 causal SNPs with PIP > 0.7 in European and cross-ancestry GWAS, respectively (Dang et al., 2025).

Other modules

***Protein-protein interaction (PPI)*:** Protein interactions constitute the molecular machinery of cells, and the proper assembly and binding of proteins are fundamental to the functional integrity of molecular machines. Conversely, aberrant interactions often lead to dysregulated cellular processes and can ultimately cause disease (Pawson et al., 2003). To facilitate the investigation of protein interactions, we collected protein interaction datasets from eight public resources for human (Stark et al., 2006; Li et al., 2017; Franz et al., 2018; Meyer et al., 2018; Luck et al., 2020; Huttlin et al.,

2021; Kamburov et al., 2021; Kim et al., 2022). We integrated these datasets into a unique resource and obtained a total of 9,021,673 non-redundant protein-protein interactions. This dataset will provide crucial insights for researchers to elucidate the interaction of proteins encoded by SCZ risk genes.

***Prioritized drug target genes*:** Although antipsychotic medications are currently available, many exhibit limited efficacy and are associated with severe side effects, highlighting an urgent need for novel therapeutic strategies. Genetic studies improve the success rate of clinical translation (Nelson et al., 2015). To this end, we previously conducted a comprehensive analysis and prioritized SCZ risk genes by integrating multiple lines of evidence, including PWAS, TWAS, MR, and colocalization analyses. By using actionable drug targets from the study of Gaziano et al. (Gaziano et al., 2021), we performed MR using protein quantitative trait loci (pQTL) (Wingo et al., 2021) and eQTL (Qi et al., 2022), and identified 108 druggable SCZ risk genes and 3,713 potential therapeutic compounds, including FDA-approved drugs. These potential drug targets and associated drugs were incorporated into SZDB3.0.

Gene Prioritization

SZDB3.0 is a comprehensive database that aggregates multi-dimensional data. To identify the most plausible candidate genes for SCZ, we conducted a systematic analysis to prioritize SCZ risk genes. Firstly, independent SNPs significantly associated with SCZ were extracted from GWAS summary statistics using PLINK (v1.9) (Purcell et al., 2007). These SNPs were then mapped to putative risk genes via eQTL that obtained from BrainMeta v2 (Qi et al., 2022). Subsequently, we integrated additional lines of evidence, including DEGs, as well as risk genes and proteins identified through TWAS, PWAS, and MR analyses. Finally, genes consistently supported by multiple lines of evidence were prioritized to establish a high-confidence set of SCZ risk genes.

References

Bennett, D. A., Buchman, A. S., Boyle, P. A., Barnes, L. L., Wilson, R. S., & Schneider, J. A., 2018. Religious orders study and rush memory and aging project. J Alzheimers Dis 64, S161–s189.

Brandes, N., Linial, N., & Linial, M., 2020. PWAS: proteome-wide association study-linking genes and phenotypes by functional variation in proteins. Genome Biol. 21, 173.

Collado-Torres, L., Burke, E. E., Peterson, A., Shin, J., Straub, R. E., Rajpurohit, A., Semick, S. A., Ulrich, W. S., Price, A. J., Valencia, C., et al., 2019. Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. Neuron 103, 203–216.e208.

Dai, S., Dang, X., & Gong, D., 2026. Molecular regulatory mechanisms of schizophrenia-associated functional non-coding variants. Mol. Psychiatry. doi: 10.1038/s41380-026-03566-8.

Dang, X., Teng, Z., Yang, Y., Li, W., Liu, J., Hui, L., Zhou, D., Gong, D., Dai, S.-S., Li, Y., et al., 2025. Gene-level analysis reveals the genetic aetiology and therapeutic targets of schizophrenia. Nat. Hum. Behav. 9, 609–624.

Dong, S., Zhao, N., Spragins, E., Kagda, M. S., Li, M., Assis, P., Jolanki, O., Luo, Y., Cherry, J. M., Boyle, A. P., et al., 2023. Annotating and prioritizing human non-coding variants with RegulomeDB v.2. Nat. Genet. 55, 724–726.

Emani, P. S., Liu, J. J., Clarke, D., Jensen, M., Warrell, J., Gupta, C., Meng, R., Lee, C. Y., Xu, S., Dursun, C., et al., 2024. Single-cell genomics and regulatory networks for 388 human brains. Science 384, eadi5199.

Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G. D., & Morris, Q., 2018. GeneMANIA update 2018. Nucleic Acids Res. 46, W60–w64.

Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., Ruderfer, D. M., Oh, E. C., Topol, A., Shah, H. R., et al., 2016. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. 19, 1442–1453.

Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al., 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362, eaat8127.

Gao, B., & Zhou, X., 2024. Mesusie enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. Nat. Genet. 56, 170-179.

Gaziano, L., Giambartolomei, C., Pereira, A. C., Gaulton, A., Posner, D. C., Swanson, S. A., Ho, Y. L., Iyengar, S. K., Kosik, N. M., Vujkovic, M., et al., 2021. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. Nat. Med. 27, 668–676.

Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., Troakes, C., Turecki, G., O'Donovan, M. C., Schalkwyk, L. C., et al., 2016. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat. Neurosci. 19, 48–54.

Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., Baltier, K., Gebreab, F., Gygi, M. P., Thornock, A., Zarraga, G., Tam, S., et al., 2021. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184, 3022–3040.e3028.

Jaffe, A. E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T. M., Weinberger, D. R., & Kleinman, J. E., 2016. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. Nat. Neurosci. 19, 40–47.

Kamburov, A., & Herwig, R., 2021. ConsensusPathDB 2022: molecular interactions update as a resource for network biology. Nucleic Acids Res. 50, D587–D595.

Kim, C. Y., Baek, S., Cha, J., Yang, S., Kim, E., Marcotte, E. M., Hart, T., & Lee, I., 2022. HumanNet v3: an improved database of human gene networks for disease research. Nucleic Acids Res. 50, D632–D639.

Kurki, M. I., Karjalainen, J., Palta, P., Sipila, T. P., Kristiansson, K., Donner, K. M., Reeve, M. P., Laivuori, H., Aavikko, M., Kaunisto, M. A., et al., 2023. Finngen provides genetic insights from a well-phenotyped isolated population. Nature 613, 508–518.

Lam, M., Chen, C. Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B. C., et al., 2019. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat. Genet. 51, 1670–1678.

Lee, S., McAfee, J. C., Lee, J., Gomez, A., Ledford, A. T., Clarke, D., Min, H., Gerstein, M. B., Boyle, A. P., Sullivan, P. F., et al., 2025. Massively parallel reporter assay investigates shared genetic variants of eight psychiatric disorders. Cell 188, 1409–1424.

Li, T., Wernersson, R., Hansen, R. B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C. T., Rigina, O., Rapacki, K., Stærfeldt, H. H., et al., 2017. A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods 14, 61–64.

Liu, J., Li, S., Li, X., Li, W., Yang, Y., Guo, S., Lv, L., Xiao, X., Yao, Y. G., Guan, F., et al., 2021. Genome-wide association study followed by trans-ancestry meta-analysis identify 17 new risk loci for schizophrenia. BMC Med. 19, 177.

Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F. J., Charloteaux, B., et al., 2020. A reference map of the human binary protein interactome. Nature 580, 402–408.

McAfee, J. C., Lee, S., Lee, J., Bell, J. L., Krupa, O., Davis, J., Insigne, K., Bond, M. L., Zhao, N., Boyle, A. P., et al., 2023. Systematic investigation of allelic regulatory activity of schizophrenia-associated common variants. Cell Genom. 3, 100404.

Meyer, M. J., Beltrán, J. F., Liang, S., Fragoza, R., Rumack, A., Liang, J., Wei, X., & Yu, H., 2018. Interactome INSIDER: A structural interactome browser for genomic studies. Nat. Methods 15, 107–114.

Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., Kaul, R., et al., 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710.

Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P. C., Li, M. J., Wang, J., et al., 2015. The support of human genetic evidence for approved drug indications. Nat. Genet. 47, 856–860.

Ng, B., White, C. C., Klein, H. U., Sieberts, S. K., McCabe, C., Patrick, E., Xu, J., Yu, L., Gaiteri, C., Bennett, D. A., et al., 2017. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat. Neurosci. 20, 1418–1426.

Pawson, T., & Nash, P., 2003. Assembly of cell regulatory systems through protein interaction domains. Science 300, 445–452.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., et al., 2007. Plink: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Qi, T., Wu, Y., Fang, H., Zhang, F., Liu, S., Zeng, J., & Yang, J., 2022. Genetic control of RNA splicing and its distinct role in complex trait variation. Nat. Genet. 54, 1355–1363.

Qi, T., Wu, Y., Zeng, J., Zhang, F., Xue, A., Jiang, L., Zhu, Z., Kemper, K., Yengo, L., Zheng, Z., et al., 2018. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat. Commun. 9, 2282.

Sanderson, E., Glymour, M. M., Holmes, M. V., Kang, H., Morrison, J., Munafo, M. R., Palmer, T., Schooling, C. M., Wallace, C., Zhao, Q., et al., 2022. Mendelian randomization. Nat. Rev. Methods Primers 2, 6.

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M., 2006. BioGRID: A general repository for interaction datasets. Nucleic Acids Res. 34, D535–D539.

Townsley, K. G., Brennand, K. J., & Huckins, L. M., 2020. Massively parallel techniques for cataloguing the regulome of the human brain. Nat. Neurosci. 23, 1509–1521.

Trubetskoy, V., Pardinas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T. B., Bryois, J., Chen, C. Y., Dennison, C. A., Hall, L. S., et al., 2022. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604, 502–508.

Wang, D., Li, D., Dang, X., Mu, C., Liu, C., Zeng, Y., Yuan, Y., Teng, Z., Li, Y., & Luo, X. J., 2025. Mendelian randomization reveals causalities between DNA methylation and schizophrenia. Biol. Psychiatry 98, 712–723.

Wang, K., Li, M., & Hakonarson, H., 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.

Wingo, A. P., Liu, Y., Gerasimov, E. S., Gockley, J., Logsdon, B. A., Duong, D. M., Dammer, E. B., Robins, C., Beach, T. G., Reiman, E. M., et al., 2021. Integrating human brain proteomes with genome-wide association data implicates new proteins in alzheimer's disease pathogenesis. Nat. Genet. 53, 143–146.

Yang, Z., Wang, C., Liu, L., Khan, A., Lee, A., Vardarajan, B., Mayeux, R., Kiryluk, K., & Ionita-Laza, I., 2023. CARMA is a new Bayesian model for fine-mapping in genome-wide association meta-analyses. Nat. Genet. 55, 1057–1065.

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M. E., Wray, N. R., Visscher, P. M., et al., 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487.

Table S1: The related information of GWAS, differential expression, PWAS, TWAS and QTL datasets included in SZDB3.0

Data source	Sample Size (case/control)	Sample source	Publication
EUR GWAS	59,901/441,418	European ancestry	Dang et al. (2025)
EAS GWAS	29,519/44,392	East Asian ancestry	Dang et al. (2025)
Cross-ancestry GWAS	96,806/492,818	Cross-ancestry (EUR, EAS, AA and LAT)	Dang et al. (2025)
DEGs from PsychENCODE	559/936	DLPFC and PFC (EUR)	Gandal et al. (2018)
DEGs from LIBD2	153/298	DLPFC (EUR)	Collado-Torres et al. (2019)
DEGs from CMC	258/279	DLPFC (EUR)	Fromer et al. (2016)
TWAS	1,321 samples	Brain (EUR)	Gandal et al. (2018)
PWAS	152 AD patients (Banner)	DLPFC (EUR)	Wingo et al. (2021)
	376 AD patients (ROSMAP)	DLPFC (EUR)	Wingo et al. (2021)
Bulk-tissue eQTL	2,865 samples	Brain (EUR)	Qi et al. (2022)
Bulk-tissue sQTL	2,865 samples	Brain (EUR)	Qi et al. (2022)
Bulk-tissue mQTL	1,160 samples	Brain (EUR)	Qi et al. (2018)
Single-cell eQTL (Exc_neu, Inh_neu, Ast, Mic, Oligo, OPC)	388 samples	PFC (EUR)	Emani et al. (2024)

Note: LIBD2, Lieber Institute for Brain Development phase2; CMC, CommonMind Consortium; DEGs, differentially expressed genes; TWAS, transcriptome-wide association study; PWAS, proteome-wide association study; eQTL, expression quantitative trait loci; pQTL, protein quantitative trait loci; mQTL, methylation quantitative trait loci; EUR, European ancestry; EAS, East-Asian ancestry; AA, African-American ancestry; LAT, Latin ancestry; Exc_neu, excitatory neuron; Inh_neu, inhibitory neuron; Ast, astrocyte; Mic, microglia; Oligo, oligodendrocyte; OPC, oligodendrocyte progenitor cells; DLPFC, dorsomedial prefrontal cortex; PFC, prefrontal cortex.