

# **Mendelian randomization integrating GWAS and eQTL data revealed genes pleiotropically associated with major depressive disorder**

Huarong Yang<sup>1\*</sup>, Di Liu<sup>2\*</sup>, Chuntao Zhao<sup>3</sup>, Bowen Feng<sup>4</sup>, Wenjin Lu<sup>5</sup>, Xiaohan Yang<sup>6</sup>, Minglu Xu<sup>6</sup>, Weizhu Zhou<sup>7</sup>, Huiquan Jing<sup>6</sup>, Jingyun Yang<sup>8,9</sup>

<sup>1</sup>Department of Neurology, The Second Affiliated Hospital of Guizhou Medical University, Kaili, Guizhou, China

<sup>2</sup>Beijing Key Laboratory of Clinical Epidemiology, School of Public Health, Capital Medical University, Beijing, China

<sup>3</sup>Brain Tumor Center, Cancer & Blood Diseases Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

<sup>4</sup>Odette School of Business, University of Windsor, Windsor, ON, Canada

<sup>5</sup>Department of Mathematics, University College London, London, United Kingdom

<sup>6</sup>Department of Health Management and Policy, School of Public Health, Capital Medical University, Beijing, China

<sup>7</sup>Department of Epidemiology and Health statistics, School of Public Health , Shenyang Medical College

<sup>8</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

<sup>9</sup>Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

\*The two authors contributed equally to this paper and share first authorship.

Correspondence to  
Huiquan Jing  
Department of Health Management and Policy  
School of Public Health  
Capital Medical University

10 Youanmen Xitoutiao, Fengtai District

Beijing, 100069

China

Tel: 86-10-83911576

E-mail: [hqjing@ccmu.edu.cn](mailto:hqjing@ccmu.edu.cn)

or

Jingyun Yang

Rush Alzheimer's Disease Center

Rush University Medical Center

1750 W Harrison ST, STE 1000

Chicago, IL 60612, USA

Tel: 1-312-563-9035

E-mail: [jingyun\\_yang@rush.edu](mailto:jingyun_yang@rush.edu)

**Running title** Genes pleiotropically associated with MDD

## Abstract

**Objectives:** To prioritize genes that are pleiotropically or potentially causally associated with the risk of MDD.

**Methods:** We applied the summary data-based Mendelian randomization (SMR) method integrating GWAS and expression quantitative trait loci (eQTL) data in 13 brain regions to identify genes that were pleiotropically associated with the risk of MDD. In addition, we repeated the analysis by using the meta-analyzed version of the eQTL summary data in the brain (brain-eMeta).

**Results:** We identified multiple significant genes across different brain regions that may be involved in the pathogenesis of MDD. The prime-specific gene *BTN3A2* (corresponding probe: ENSG00000186470.9) was the top hit showing pleiotropic association with MDD in 9 of the 13 brain regions and in brain-eMeta, after correction for multiple testing. Many of the identified genes are located in the human major histocompatibility complex (MHC) region on chromosome 6 and are mainly involved in immune response.

**Conclusions:** Our SMR analysis revealed that multiple genes showed pleiotropic association with MDD across the brain regions. These findings provide important leads to a better understanding of the mechanism of MDD, and reveals potential therapeutic targets for the prevention and effective treatment of MDD.

**Key words:** Major depressive disorder; expression quantitative trait loci; summary Mendelian randomization; pleiotropic association

## Introduction

Major depressive disorder (MDD), also known as clinical depression, is a significant medical condition impacting individual's mood, behavior as well as appetite and sleep, even thoughts of suicide [1]. MDD is a leading cause of disability and morbidity worldwide [2], with an estimated lifetime prevalence of around 15% [3]. MDD is a complex multi-factorial disorder, with contributions from both genetic and environmental factors [4]. However, the exact etiology of MDD remains to be unclear, and there is pressing urgency to further explore the pathological mechanisms underlying MDD to facilitate the design and implementation of efficient prevention strategies or novel treatments.

Previous twin studies found the heritability of MDD to be approximately 30%–40% [5,6]. Although genome-wide association studies (GWAS) have been successful in identifying genetic variants associated with MDD [7-11], biological interpretation of the findings remain largely unclear. It is likely that these genetic loci exert their effect on MDD via gene expression. Therefore, it is important to explore the relationship between genetic variation and gene expression to better understand the regulatory pathways underlying the pathogenesis of MDD.

Mendelian randomization (MR) is a popular method for exploring potentially causal association between an exposure (e.g., gene expression) and an outcome (e.g., MDD susceptibility) by using genetic variants as the instrumental variables (IVs) [12]. Compared with traditional statistical methods used in the association studies, MR reduces confounding and reverse causation [13,14], and has been successful in

identifying gene expressions or DNA methylation loci that are pleiotropically or potentially causally associated with various phenotypes, such as cardiovascular diseases, inflammatory bowel disease, and educational attainment [15-19].

In this study, we adopted the summary data-based MR (SMR) method integrating summarized GWAS data for MDD and expression quantitative trait loci (eQTL) data in the brain to prioritize genes that are pleiotropically or potentially causally associated with the risk of MDD. Our analysis identified a primate specific gene *BTN3A2* as a novel MDD risk gene of MDD.

## Methods

### Data sources

#### *eQTL data*

In the SMR analysis, cis-eQTL genetic variants were used as the IVs for gene expression. We performed SMR analysis for different regions in the brain. We used the Version 7 release of the eQTL summarized data from the Genotype Tissue Expression [20] project, which included 13 different regions: amygdala, anterior cingulate cortex, caudate nucleus, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, putamen, spinal cord and substantia nigra [20]. In addition, we repeated the analysis by using the meta-analyzed version of the eQTL summary data (named brain-eMeta hereafter), which included results from GTEx data of brain tissues [20], the Common Mind Consortium [21], and the Religious Orders Study and the Rush Memory and Aging

Project [22]. Results from these three studies were meta-analyzed using the MeCS method (meta-analysis of cis-eQTL in correlated samples) to increase the power of detecting brain eQTLs [23]. Only SNPs within 1 Mb distance from each individual probe are available. The eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>.

#### *GWAS data for MDD*

The GWAS summarized data for MDD were provided by the Psychiatric Genomics Consortium [10]. The results were based on three large genome-wide association studies [8,9,11], including a total of 807,553 individuals (246,363 cases and 561,190 controls, after excluding overlapping samples) and 8,098,588 genetic variants. The GWAS summarized data can be downloaded at <https://www.med.unc.edu/pgc/download-results/mdd/>.

#### **Statistical and bioinformatics analyses**

MR was carried out considering cis-eQTL as the IVs, gene expression as the exposure, and MDD as the outcome. MR analysis was performed using the method as implemented in the software SMR. Detailed information regarding the SMR method have been described previously [16]. Briefly, SMR uses the principles of MR integrating GWAS and eQTL summary statistics to test for pleiotropic association between gene expression and MDD due to a shared and potentially causal variant at a locus. The heterogeneity in dependent instruments (HEIDI) test was done to explore the existence of linkage in the observed association. Rejection of the null hypothesis (i.e.,  $P_{\text{HEIDI}} < 0.05$ ) indicates that the observed association might be due to two distinct

genetic variants in high linkage disequilibrium with each other. We adopted the default settings in SMR (e.g.,  $P_{\text{eQTL}} < 5 \times 10^{-8}$ , minor allele frequency [MAF]  $> 0.01$ , excluding SNPs in very strong linkage disequilibrium [LD,  $r^2 > 0.9$ ] with the top associated eQTL, and removing SNPs in low LD or not in LD [ $r^2 < 0.05$ ] with the top associated eQTL), and used false discovery rate (FDR) to adjust for multiple testing.

Annotations of the transcripts were based on the Affymetrix exon array S1.0 platforms. To functionally annotate putative transcripts, we conducted functional enrichment analysis using the functional annotation tool “Metascape” for the significantly tagged genes in different brain regions and in brain-eMeta. Gene symbols corresponding to putative genes ( $P < 0.05$ ) were used as the input of the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis.

Data cleaning and statistical/bioinformatical analysis was performed using R version 4.0.2 (<https://www.r-project.org/>), PLINK 1.9 (<https://www.cog-genomics.org/plink/1.9/>) and SMR (<https://cnsgenomics.com/software/smr/>).

## Results

### *Basic information of the summarized data*

The number of participants used for generating the eQTL data varied across the brain regions, ranging from 114 to 209, so did the number of eligible probes involved in the final SMR analysis, ranging from 814 to 2,786. The brain-eMeta analysis involved more subjects ( $n=1,194$ ) and more probes ( $n=7,421$ ). The GWAS meta-analysis data

involved roughly 800,000 subjects. The detailed information was shown in **Table 1**.

### *SMR analysis in the 13 brain regions*

Out of the 13 brain regions, the human major histocompatibility complex (MHC) gene *BTN3A2* (ENSG00000186470.9) was the top hit showing pleiotropic association with MDD in 9 regions, after correction for multiple testing. Each of the other two genes, *RPL31P12* (ENSG00000227207.2) and *RPI-265C24.5* (ENSG00000219392.1), was the top gene pleiotropically associated with MDD in two brain regions (**Figure S1**).

Specifically, for *BTN3A2*, the most significant associations with MDD were detected in two brain regions: caudate nucleus and spinal cord ( $\beta$  [SE]=0.043 [0.008],  $P=7.76\times10^{-8}$ ;  $\beta$  [SE]=0.042 [0.008],  $P=1.72\times10^{-7}$ ; **Figure 1**). It also showed significant pleiotropic association with MDD in the four brain regions where it was not the top gene (**Table S1**). *RPL31P12* showed the most significantly pleiotropic association with MDD in cerebellar hemisphere and cerebellum ( $\beta$  [SE]=-0.037 [0.006],  $P=7.53\times10^{-11}$ ;  $\beta$  [SE]=-0.033 [0.005],  $P=1.34\times10^{-12}$ , respectively; **Figure 2**). *RPI-265C24.5* showed significant pleiotropic association in cortex and nucleus accumbens ( $\beta$  [SE]=0.036 [0.007],  $P=6.13\times10^{-08}$ ;  $\beta$  [SE]=0.036 [0.006],  $P=1.63\times10^{-08}$ , respectively; **Figure 3**).

The complement gene *C4A* (ENSG00000244731.3) was significantly associated with MDD in 7 different brain regions, after correction for multiple testing (**Table S1**).

Of note, both *BTN3A2*, *C4A* and *RPI-265C24.5* are on chromosome 6 while

*RPL31P12* is on chromosome 1. Two brain regions, cerebellar hemisphere and cerebellum, have a relatively large number of significant genes (21 genes and 30 genes, respectively; **Table 2**).

GO enrichment analysis of biological process and molecular function showed that the significant genes across the different brain regions were involved in four GO terms, including negative regulation of endopeptidase activity (GO:0010951), adaptive immune response (GO:0002250), platelet degranulation (GO:0002576) and negative regulation of defense response (GO:0031348; **Figure S2A**). Concept network analysis of the identified genes revealed multiple domains related with immune response (**Figure S2B**). More information could be found in **Table S2**.

#### *SMR analysis in the brain-eMeta*

Using brain-eMeta eQTL data, we found 75 genes that showed pleiotropic association with MDD, after correction for multiple testing. Specifically, we identified *BTN3A2* (ENSG00000186470) that showed the most significantly pleiotropic association with MDD ( $\beta$  [SE]=0.027 [0.004],  $P=3.44 \times 10^{-12}$ ; **Table S3**), followed by *RPL31P12* (ENSG00000227207,  $\beta$  [SE]=-0.039 [0.006],  $P=3.43 \times 10^{-11}$ ). We found that *C4A* and *RPI-265C24.5* also showed significant pleiotropic association with MDD ( $\beta$  [SE]=0.031 [0.005],  $P=1.58 \times 10^{-8}$  and  $\beta$  [SE]=0.047 [0.008],  $P=2.11 \times 10^{-9}$ , respectively).

GO enrichment analysis of biological process and molecular function showed that the significant genes in brain-eMeta were involved in eight GO terms, including allograft rejection (ko05330), butyrophilin (BTN) family interactions

(R-HSA-8851680), platelet degranulation (GO:0002576), immunoregulatory interactions between a lymphoid and a non-lymphoid cell (R-HSA-198933), nuclear chromosome segregation (GO:0098813), telomere maintenance (GO:0000723), organelle localization by membrane tethering (GO:0140056) and lipid transport (GO:0006869; **Figure S3A**). Similar to the findings for the different brain regions, concept network analysis in brain-eMeta also revealed multiple domains related with immune response (**Figure S3B**). More information could be found in **Table S4**.

## Discussion

In this study, we integrated GWAS and eQTL data in the MR analysis to explore putative genes that showed pleiotropic/potentially causal association with MDD susceptibility. Across the different brain regions, we identified multiple significant genes that may be involved in the pathogenesis of MDD. The identified genes were mainly involved in immune response. Our findings provided important leads to a better understanding of the mechanisms underlying MDD and revealed potential therapeutic targets for the effective treatment of MDD.

Several of the identified genes in our study, such as *BTN3A2*, *BTN3A3*, *PRSSI6*, *HLA-C*, *C4A* and *HLA-DMA*, are located in or around the human major histocompatibility complex (MHC) region on chromosome 6. MHC represents the most complex genomic region due to its unintelligible linkage disequilibrium [24]. Many genes in or around MHC play an important role in immune response and immune regulation, and are involved in a variety of inflammatory and autoimmune

diseases [25-29]. The MHC regions can be roughly divided into three classes that are functionally distinct, with class I and II regions containing highly polymorphic human leukocyte antigen (HLA) genes associated with autoimmune disease risk [30,31] and class III region containing complement component 4 regions associated with schizophrenia risk [32]. Recent GWASs identified a number of genetic variants in the MHC region associated with depression risk, with the strongest association observed in or near the class I region [9-11].

We found that *BTN3A2* were significantly associated with MDD across the brain regions. *BTN3A2*, which encodes a member of the immunoglobulin superfamily, resides in the juxta-telomeric region (class I) of MHC [33]. The *BTN3A2* protein may be involved in adaptive immune response [34]. Previous studies showed that *BTN3A2* was a potential risk gene for Alzheimer's disease, schizophrenia and intellectual disability [35-37]. A meta-analysis of GWAS found that *BTN3A2* was associated with neuroticism [38], an important risk factor for MDD [39]. Overexpression of *BTN3A2* suppressed the excitatory synaptic activity onto CA1 pyramidal neurons, most likely through the interaction with the presynaptic adhesion molecule neuroligins [35,40]. Previous research showed that *BTN3A2* was expressed in multiple cell types in the brain, including astrocyte, neuron, oligodendrocyte, and microglia [41]. These findings, together with ours, demonstrated the important role of *BTN3A2* in the nervous system and highlighted the potential of this gene as a promising target for the prevention and treatment of MDD.

A previous GWAS of MDD highlighted the importance of the prefrontal brain regions [10]. In the prefrontal cortex, we found four significant genes, including *BTN3A2*, *C4A*, *RPI-265C24.5* and *CYP21A1P*, that were associated with MDD after correction for multiple testing. The gene *C4A* was significant in a total of seven brain regions and in brain-eMeta. *C4A* localizes to the MHC class III region and encodes the acidic form of complement factor 4. In the mouse brain, *C4A* gene is mainly expressed in astrocyte and neurons [42]. *C4A* is involved in the classical complement activation pathway [43] and was reported to be associated with schizophrenia, aging and Alzheimer's disease [32,44,45]. Moreover, genetic variants in *BTN3A2* and *C4A* were in different LD blocks, suggesting that both genes might be independent risk factors for mental disorders such as schizophrenia and MDD [35].

Both MDD and schizophrenia are mental illnesses contributing substantially to the global disease burden. It was reported that depressed patients had a higher risk of developing psychosis. Moreover, even prior to the emergence of psychotic symptoms, patients with a high risk of schizophrenia had a higher risk for developing depressive symptoms [46]. In consistent with previous findings [47], some of the identified genes showing pleiotropic association with MDD were also associated with schizophrenia, such as *BTN3A2*, *BTN3A3*, *PRSSI6*, *HLA-C*, *C4A* and *HLA-DMA*, indicating a potential overlapped mechanism between schizophrenia and MDD. In addition, compared with a previous study [48], the sample size for GWAS of MDD was much large, therefore increasing the power of SMR analysis, and we performed the analysis in different brain regions.

Our study has some limitations. The number of probes used in our SMR analysis were limited for some brain regions (**Table 1**), and we may have missed some important genes. The HEIDI test was significant for some of the identified genes, indicating the possibility of horizontal pleiotropy (supplementary **Table S1** and **Table S3**), i.e., the identified association might be due to two distinct genetic variants in high linkage disequilibrium with each other. In addition, we only included study participants of European ethnicity, and our findings might not be generalized to other ethnicities. More studies are needed to validate our findings in independent populations. We adopted correction for multiple testing to reduce false positive rate; however, we may have missed important SNPs or genes. Due to a lack of individual eQTL data, we could not quantify the changes in gene expression in subjects with MDD in comparison with the control.

## **Conclusion**

Our SMR analysis revealed that multiple genes showed pleiotropic association with MDD across the brain regions. More studies are needed to explore the underlying physiological mechanisms in the etiology of MDD.

## **Acknowledgements**

The study was supported by NIH/NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG36042, U01AG61356 and 1RF1AG064312-01. Huiquan Jing's research was supported by National Key Research and Development Program of China (2018YFC2000400). Di Liu was supported by China Scholarship Council (CSC 201908110339).

The authors confirmed that all authors have reviewed the contents of the article being submitted, approved its contents, and validated the accuracy of the data.

### **Author contribution**

HJ and JY designed the study. HY, BF and WL analyzed data and performed data interpretation. DL, CZ and JY wrote the initial draft, and BF, WL, XY, MX, WZ and HJ contributed to writing the subsequent versions of the manuscript. All authors reviewed the study findings and read and approved the final version before submission.

### **Availability of data and materials**

All data generated or analyzed during this study are included in this published article and its supplementary information files.

### **Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed by the authors.

## References

- 1 Smith K. Mental health: a world of depression. *Nature*. 2014;515(7526):181.
- 2 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789-858.
- 3 Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593-602.
- 4 Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. *Nature Reviews Disease Primers*. 2016;2(1):16065.
- 5 Flint J, Kendler KS. The genetics of major depression. *Neuron*. 2014;81(3):484-503.
- 6 Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *The American journal of psychiatry*. 2000;157(10):1552-62.
- 7 Lewis CM, Ng MY, Butler AW, Cohen-Woods S, Uher R, Pirlo K, et al. Genome-wide association study of major recurrent depression in the U.K. population. *The American journal of psychiatry*. 2010;167(8):949-57.
- 8 Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European

- descent. *Nat Genet.* 2016;48(9):1031-6.
- 9 Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018;50(5):668-81.
- 10 Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci.* 2019;22(3):343-52.
- 11 Howard DM, Adams MJ, Shirali M, Clarke TK, Marioni RE, Davies G, et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun.* 2018;9(1):1470.
- 12 Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):R89-98.
- 13 Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. *Jama.* 2009;301(22):2386-8.
- 14 Burgess S, Timpson NJ, Ebrahim S, Davey Smith G. Mendelian randomization: where are we now and where are we going? *Int J Epidemiol.* 2015;44(2):379-88.
- 15 Pavlides JM, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* 2016;8(1):84.
- 16 Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.

- Nat Genet. 2016;48(5):481-7.
- 17 Porcu E, Rüeger S, Lepik K, Santoni FA, Reymond A, Kutalik Z. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat Commun. 2019;10(1):3300.
  - 18 Richardson TG, Haycock PC, Zheng J, Timpson NJ, Gaunt TR, Davey SG, et al. Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. Hum Mol Genet. 2018;27(18):3293-304.
  - 19 Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. Nat Commun. 2019;10(1):4267.
  - 20 Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550(7675):204-13.
  - 21 Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nature Neuroscience. 2016;19(11):1442-53.
  - 22 Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nature Neuroscience. 2017;20(10):1418-26.
  - 23 Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nature Communications. 2018;9(1):2282.

- 24 Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al.  
  
Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet.* 2018;50(3):381-89.
- 25 Lang ML. How do natural killer T cells help B cells? Expert review of vaccines. 2009;8(8):1109-21.
- 26 Neely J, Rychkov D, Paranjpe M, Waterfield M, Kim S, Sirota M. Gene Expression Meta-Analysis Reveals Concordance in Gene Activation, Pathway, and Cell-Type Enrichment in Dermatomyositis Target Tissues. *ACR open rheumatology.* 2019;1(10):657-66.
- 27 Whitehead AS, Colten HR, Chang CC, Demars R. Localization of the human MHC-linked complement genes between HLA-B and HLA-DR by using HLA mutant cell lines. *Journal of immunology (Baltimore, Md : 1950).* 1985;134(1):641-3.
- 28 Zakharova MY, Belyanina TA, Sokolov AV, Kiselev IS, Mamedov AE. The Contribution of Major Histocompatibility Complex Class II Genes to an Association with Autoimmune Diseases. *Acta naturae.* 2019;11(4):4-12.
- 29 Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. *Nat Rev Genet.* 2004;5(12):889-99.
- 30 Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nature reviews Immunology.* 2018;18(5):325-39.
- 31 Gough SC, Simmonds MJ. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. *Current genomics.* 2007;8(7):453-65.
- 32 Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al.

- Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016;530(7589):177-83.
- 33 Afrache H, Pontarotti P, Abi-Rached L, Olive D. Evolutionary and polymorphism analyses reveal the central role of *BTN3A2* in the concerted evolution of the *BTN3* gene family. *Immunogenetics*. 2017;69(6):379-90.
- 34 Rhodes DA, Stammers M, Malcherek G, Beck S, Trowsdale J. The cluster of *BTN* genes in the extended major histocompatibility complex. *Genomics*. 2001;71(3):351-62.
- 35 Wu Y, Bi R, Zeng C, Ma C, Sun C, Li J, et al. Identification of the primate-specific gene *BTN3A2* as an additional schizophrenia risk gene in the MHC loci. *EBioMedicine*. 2019;44:530-41.
- 36 Anazi S, Maddirevula S, Fageih E, Alsedairy H, Alzahrani F, Shamseldin HE, et al. Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. *Molecular psychiatry*. 2017;22(4):615-24.
- 37 Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature*. 2009;460(7256):753-7.
- 38 Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet*. 2018;50(7):920-27.
- 39 Speed D, Hemani G, Speed MS, Børghlum AD, Østergaard SD. Investigating the causal relationship between neuroticism and depression via Mendelian

- randomization. *Acta psychiatrica Scandinavica*. 2019;139(4):395-97.
- 40 Sùdhof TC. Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. *Cell*. 2017;171(4):745-69.
- 41 Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, et al. Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron*. 2016;89(1):37-53.
- 42 Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2014;34(36):11929-47.
- 43 Chung EK, Yang Y, Rupert KL, Jones KN, Rennebohm RM, Blanchong CA, et al. Determining the one, two, three, or four long and short loci of human complement C4 in a major histocompatibility complex haplotype encoding C4A or C4B proteins. *Am J Hum Genet*. 2002;71(4):810-22.
- 44 Bennett S, Grant M, Creese AJ, Mangialasche F, Cecchetti R, Cooper HJ, et al. Plasma levels of complement 4a protein are increased in Alzheimer's disease. *Alzheimer Dis Assoc Disord*. 2012;26(4):329-34.
- 45 Cao W, Zheng D, Wang G, Zhang J, Ge S, Singh M, et al. Modelling biological age based on plasma peptides in Han Chinese adults. *Aging (Albany NY)*. 2020;12(11):10676-86.
- 46 Samsom JN, Wong AH. Schizophrenia and Depression Co-Morbidity: What We have

Learned from Animal Models. *Frontiers in psychiatry*. 2015;6:13.

47 O'Brien HE, Hannon E, Hill MJ, Toste CC, Robertson MJ, Morgan JE, et al.

Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol*. 2018;19(1):194.

48 Wang X, Cheng W, Zhu J, Yin H, Chang S, Yue W, et al. Integrating genome-wide association study and expression quantitative trait loci data identifies NEGR1 as a causal risk gene of major depression disorder. *Journal of affective disorders*. 2020;265:679-86.

## Figure legends

### Figure 1. Prioritizing genes around *BTN3A2* in association with MDD.

A) Caudate nucleus; B) Spinal cord

Top plot, grey dots represent the  $-\log_{10}(P \text{ values})$  for SNPs from the GWAS of MDD, and rhombuses represent the  $-\log_{10}(P \text{ values})$  for probes from the SMR test with solid rhombuses indicating that the probes pass HEIDI test and hollow rhombuses indicating that the probes do not pass the HEIDI test. Middle plot, eQTL results for the probe ENSG000001864770.9 tagging *BTN3A2*. Bottom plot, location of genes tagged by the probe. Highlighted in maroon indicates probes that pass the SMR threshold.

GWAS, genome-wide association study; MDD, major depressive disorder; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci

### Figure 2. Prioritizing genes around *RPL31P12* in association with MDD.

A) Cerebellar hemisphere; B) Cerebellum

Top plot, grey dots represent the  $-\log_{10}(P \text{ values})$  for SNPs from the GWAS of MDD, and rhombuses represent the  $-\log_{10}(P \text{ values})$  for probes from the SMR test with solid rhombuses indicating that the probes pass HEIDI test and hollow rhombuses indicating that the probes do not pass the HEIDI test. Middle plot, eQTL results for the probe ENSG00000227207.2 tagging *RPL31P12*. Bottom plot, location of genes tagged by the probe. Highlighted in maroon indicates probes that pass the SMR

threshold.

GWAS, genome-wide association studies; MDD, major depressive disorder; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci

### **Figure 3. Prioritizing genes around *RPI-265C24.5* in association with MDD.**

A) Cortex; B) Nucleus accumbens

Top plot, grey dots represent the  $-\log_{10}(P \text{ values})$  for SNPs from the GWAS of MDD, and rhombuses represent the  $-\log_{10}(P \text{ values})$  for probes from the SMR test with solid rhombuses indicating that the probes pass HEIDI test and hollow rhombuses indicating that the probes do not pass the HEIDI test. Middle plot, eQTL results for the probe ENSG00000219392.1 tagging *RPI-265C24.5*. Bottom plot, location of genes tagged by the probe. Highlighted in maroon indicates probes that pass the SMR threshold.

GWAS, genome-wide association studies; MDD, major depressive disorder; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci

### **Figure S1. Distribution of genes show significant pleiotropic association with MDD.**

A) A lobe view of the distribution of significant genes; B) a sagittal view of the distribution of significant genes

Plots were generated using R package cerebroViz. Results for cerebellar hemisphere, cortex, nucleus accumbens and spinal cord were not plotted because they were not covered in the package.

AMY, amygdala; CAU, caudate; CB, cerebellum; CNG, anterior cingulate cortex; FL, frontal cortex; HIP, hippocampus; HTH, hypothalamus; MDD, major depressive disorder; PUT, putamen; SN, substantia nigra.

**Figure S2. Functional enrichment and gene concept network analysis based on the identified genes in different brain regions.**

A) Enriched GO terms based on the identified genes in different brain regions; B)

Concept network analysis of the identified genes

GO, gene ontology

**Figure S3. Functional enrichment and gene concept network analysis based on the identified genes in brain-eMeta.**

A) Enriched GO terms based on the identified genes in brain-eMeta; B) Concept

network analysis of the identified genes

GO, gene ontology

**Table 1. Basic information of the eQTL and GWAS data.**

<b>Data source</b>	<b>Total participants or cases/controls</b>	<b>Number of genetic variants or probes</b>
<b>eQTL data</b>		
<b>Amygdala</b>	129	779
<b>Anterior cingulate cortex</b>	147	1379
<b>Caudate</b>	194	2089
<b>Cerebellar hemisphere</b>	175	2615
<b>Cerebellum</b>	209	3765
<b>Cortex</b>	205	2314
<b>Frontal cortex</b>	175	1722
<b>Hippocampus</b>	165	1108
<b>Hypothalamus</b>	170	1170
<b>Nucleus accumbens</b>	202	1785
<b>Putamen</b>	170	1449
<b>Spinal cord</b>	126	915
<b>Substantia nigra</b>	114	661
<b>Brain-eMeta</b>	1,194	7,421
<b>GWAS-Meta</b>	246,363/561,190	8,098,588
<b>23andme_307k</b>	75,607/231,747	
<b>UK Biobank</b>	127,552/233763	
<b>PGC_139k</b>	43,204/95,680	

GWAS: genome-wide association studies; QTL, quantitative trait loci; PGC, Psychiatric Genomics Consortium

**Table 2. Summary of the SMR analysis across the 13 brain regions.**

Regions	N_gene	Gene_ID	Gene	CHR	top_SNP	P <sub>SMR</sub>	Q_value
<b>Amygdala</b>	5	ENSG00000186470.9	BTN3A2	6	rs9393703	$3.08 \times 10^{-07}$	$2.40 \times 10^{-04}$
<b>Anterior cingulate cortex</b>	5	ENSG00000186470.9	BTN3A2	6	rs28551159	$2.46 \times 10^{-07}$	$3.40 \times 10^{-04}$
<b>Caudate</b>	7	ENSG00000186470.9	BTN3A2	6	rs9379853	$7.76 \times 10^{-08}$	$1.62 \times 10^{-04}$
<b>Cerebellar hemisphere</b>	21	ENSG00000227207.2	RPL31P12	1	rs1460943	$7.53 \times 10^{-11}$	$1.97 \times 10^{-07}$
<b>Cerebellum</b>	30	ENSG00000227207.2	RPL31P12	1	rs1460943	$1.34 \times 10^{-12}$	$5.03 \times 10^{-09}$
<b>Cortex</b>	7	ENSG00000219392.1	RP1-265C24.5	6	rs2295594	$6.13 \times 10^{-08}$	$1.42 \times 10^{-04}$
<b>Frontal cortex</b>	4	ENSG00000186470.9	BTN3A2	6	rs9379853	$2.20 \times 10^{-07}$	$3.78 \times 10^{-04}$
<b>Hippocampus</b>	3	ENSG00000186470.9	BTN3A2	6	rs9393703	$5.32 \times 10^{-07}$	$5.90 \times 10^{-04}$
<b>Hypothalamus</b>	3	ENSG00000186470.9	BTN3A2	6	rs72841536	$5.34 \times 10^{-07}$	$6.25 \times 10^{-04}$
<b>Nucleus accumbens</b>	3	ENSG00000219392.1	RP1-265C24.5	6	rs4713135	$1.63 \times 10^{-08}$	$2.91 \times 10^{-05}$
<b>Putamen</b>	2	ENSG00000186470.9	BTN3A2	6	rs9379853	$3.41 \times 10^{-07}$	$4.95 \times 10^{-04}$
<b>Spinal cord</b>	2	ENSG00000186470.9	BTN3A2	6	rs71557332	$1.72 \times 10^{-07}$	$1.58 \times 10^{-04}$
<b>Substantia nigra</b>	1	ENSG00000186470.9	BTN3A2	6	rs28551159	$4.42 \times 10^{-06}$	$2.92 \times 10^{-03}$

Number of genes means the number of statistically significant genes in each region after correction for multiple testing using false discovery rate (Q value < 0.05); top probe and gene is the corresponding gene; top SNP is the top associated cis-eQTL in the eQTL analysis; P<sub>SMR</sub> is the P value for SMR analysis

CHR, chromosome; SNP, single-nucleotide polymorphism; SMR, summary data-based Mendelian randomization; QTL, quantitative trait loci





