

Trans-ancestry Genome-Wide Analyses in UK Biobank Yield Novel Risk Loci for Major Depression

Madhurbain Singh^{1,2*}, Chris Chatzinakos^{3,4,1}, Peter B. Barr^{3,4}, Amanda Elswick Gentry¹, Tim B. Bigdeli^{3,4}, Bradley T. Webb^{5,1}, and Roseann E. Peterson^{3,4,1*}

1. Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA
2. Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA
3. Institute for Genomics in Health, SUNY Downstate Health Sciences University, Brooklyn, NY, USA
4. Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA
5. GenOmics and Translational Research Center, RTI International, Research Triangle Park, NC, USA

*Correspondence to

Madhurbain Singh, Virginia Institute for Psychiatric and Behavioral Genetics, 800 E. Leigh St - Suite 100, Richmond, VA - 23298, USA. Email: singhm18@vcu.edu.

Roseann E. Peterson, Institute for Genomics in Health, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, 450 Clarkson Ave, Brooklyn, NY 11203, USA. Email: roseann.peterson@downstate.edu

Abstract

Most genome-wide association studies (GWASs) of depression focus on broad, heterogeneous outcomes, limiting the discovery of genomic risk loci specific to major depressive disorder (MDD). Previous UK Biobank (UKB) studies had limited ability to pinpoint MDD-associated loci due to a smaller sample with strictly defined MDD outcomes and further exclusion of many participants based on ancestry or relatedness, significantly underutilizing this resource's potential for elucidating the genetic architecture of MDD. Here, we present novel genomic insights into MDD by fully utilizing existing UKB data through (1) a trans-ancestry GWAS pipeline using two complementary approaches controlling for population structure and relatedness and (2) an increased sample with MDD symptom-level data across two mental health assessments. We identified strict MDD outcomes among 211,535 participants, representing a 38% increase in eligible participants from prior studies with only one assessment. Ancestrally inclusive analyses yielded 61 genomic risk loci across depression phenotypes, compared to 47 in the analyses restricted to participants genetically similar to European ancestry. Fourteen of these loci, including five novel, were associated with strict MDD phenotypes, whereas only one locus has been previously reported in UKB. MDD-associated genomic loci and predicted gene expression levels showed little overlap with broad depression, indicating higher specificity. Notably, polygenic scores based on these results were significantly associated with depression diagnoses across ancestry groups in the *All of Us* Research Program, highlighting the shared genetic architecture across populations. While the trans-ancestry analyses, which included non-European participants, increased the number of associated loci, the discovery of non-European ancestry-specific loci was limited, underscoring the need for larger, globally representative studies of MDD. Importantly, beyond these results, our GWAS pipeline will facilitate inclusive analyses of other traits and disorders, helping improve statistical power, representation, and generalizability in genomic studies.

Introduction

Major depressive disorder (MDD) is a common, heterogeneous, and debilitating mental health problem and a leading cause of morbidity and mortality worldwide^{1–3}. Twin and family studies indicate that additive genetic influences account for approximately 37% of the variance in MDD susceptibility⁴. Over the past decade, large-scale genome-wide association studies (GWASs) have provided valuable insights into the polygenic architecture of MDD and related depression outcomes^{5–15}.

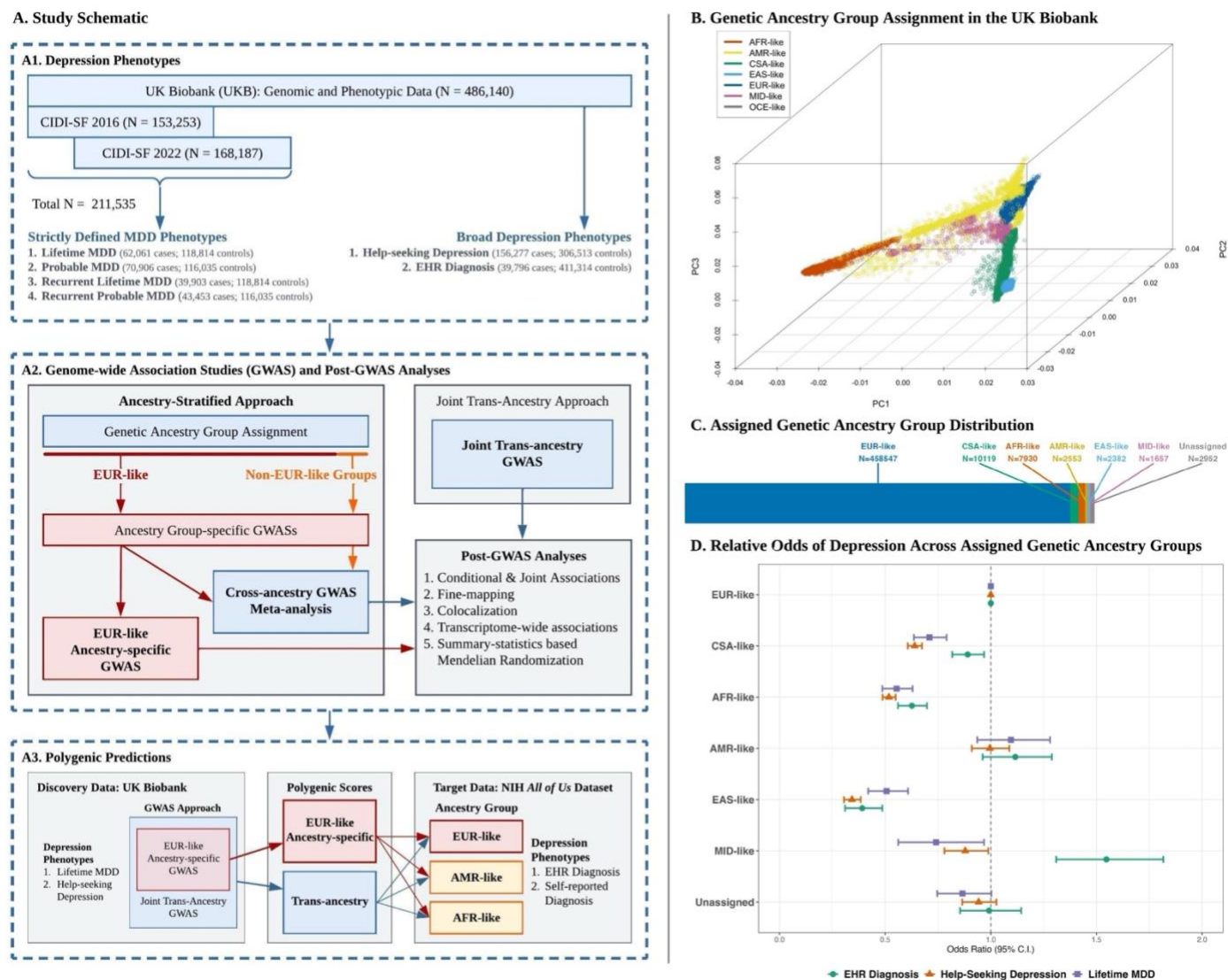
Genomic discovery for complex polygenic traits, like MDD, relies on ever-increasing sample sizes, typically achieved by meta-analyzing results across heterogeneously defined outcomes^{9,10,13–15}. Structured clinical interviews are the gold standard for diagnosing MDD but have high financial and time costs¹⁶. Self-reported questionnaires with retrospective symptom-level information provide an alternative, relatively scalable approach to deriving strictly defined MDD phenotypes^{16,17}. However, due to their wider availability, most data analyzed in the current large-scale studies^{13–15} come from broad, minimally defined depression phenotypes, such as electronic health records (EHR) and self-reported help-seeking or diagnosis. Prior research in UK Biobank¹⁸ (UKB) highlighted that GWAS of broad depression phenotypes primarily implicates genomic loci associated with general susceptibility to psychological distress rather than specific to MDD¹⁷.

In UKB, strict MDD phenotypes may be derived from self-reported retrospective symptoms assayed through the Composite International Diagnostic Interview Short Form (CIDI-SF)¹⁹ questionnaire. CIDI-SF was initially assessed in only around a third of the UKB participants (N~153k in 2016–17). Previous GWASs were further restricted to unrelated participants either of self-reported White-British ethnicity¹⁷ or genetically similar to European ancestry²⁰, thus excluding many potential participants due to genetic ancestry or relatedness. CIDI-SF data were reassessed in 2022–23 in a broader set of participants (N~168k) but have yet to be analyzed in the GWAS of MDD. Extant GWASs in UKB have reported just one genomic risk locus for MDD²⁰ and are yet to leverage this resource's full potential for genomic discovery.

Furthermore, most data in current GWASs of depression come from participants genetically similar to European ancestry^{9,13,21}. While strides have been made in increasing genetic diversity in studies of depression in recent years^{10,11,14,15}, these studies meta-analyzed heterogeneous depression outcomes across available cohorts of different genetic ancestry groups. While we expect genetic architectures to be largely similar across populations, some genetic associations of complex traits may vary across ancestry groups due to differences in genetic background (e.g., genetic drift, allele frequencies, and linkage disequilibrium [LD]) and potential environmental moderators²¹. MDD and depression phenotypes show moderate genetic correlations across populations (e.g., 0.33–0.41 between European and East-Asian samples)^{11,22} and limited cross-population replicability of specific associations^{11,14,22}. Thus, critical gaps remain in our understanding of MDD-specific genetic architecture across populations.

To address these limitations, we performed trans-ancestry GWAS of strictly defined MDD¹⁷ and broad depression⁸ outcomes in UKB (**Fig. 1A**). We assigned the UKB participants to ancestry groups based on genetic similarity to an expanded population reference panel. We combined the two CIDI-SF assessments to create a broader set of participants with strict MDD phenotypes. Building on the recommendations from the Psychiatric Genomics Consortium's *Cross-Population Analyses Working Group*²¹, we employed two complementary approaches for conducting ancestrally inclusive GWAS: the *ancestry-stratified* meta-analysis and the *joint trans-ancestry* mixed-model GWAS. The results demonstrate the enhanced power achieved by this ancestrally inclusive GWAS pipeline, which revealed novel genomic loci associated with MDD and may facilitate additional trans-ancestry work in UKB and other datasets.

Figure 1: Study Overview, Genetic Ancestry Groups Assignment, and Differences in Depression Outcome Rates between Ancestry Groups



A. Study schematic presenting (A1) an overview of the depression outcomes, (A2) GWAS and post-GWAS analyses in UKB, and (A3) follow-up analyses in the All of Us Research Program. **B.** Genetic principal components analysis (PCA) plot showing the relative positions of UKB participants along the first three genetic PCs. Using PCA, the participants were assigned to “genetic ancestry groups” based on their genetic similarity to major global populations in the 1000 Genomes Project Phase-3 and Human Genome Diversity Project: European/EUR-like, Central and South Asian/CSA-like, African/AFR-like, Admixed American/AMR-like, East Asian/EAS-like, and Middle Eastern MID-like. **C.** Sample size distribution of the assigned genetic ancestry groups. **D.** Relative odds of different depression outcomes in the non-EUR-like groups relative to the EUR-like group, controlling for age, sex, and UKB assessment center.

Methods

Depression Phenotypes in UK Biobank

UKB is an ongoing population-based prospective study of ~500,000 participants from the UK, aged 40-69 at recruitment (2006–2010), with rich phenotypic, genetic, and linked EHR data¹⁸. We operationalized strict MDD phenotypes in UKB based on self-reported retrospective symptoms aggregated across the two CIDI-SF¹⁹ assessments in partially overlapping subsets of participants (**Supplementary Methods, Supplementary Table 1**). Building on prior research¹⁷, *LifetimeMDD* cases were defined as those meeting the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition)²³ MDD criteria of symptoms, episode length, and accompanying distress or functional impairment. We additionally derived *ProbableMDD* cases requiring at least four self-reported MDD symptoms (compared to five symptoms in DSM-5), reflecting the CIDI-SF criterion for epidemiological screening of MDD¹⁹. Further, recurrent cases of both phenotypes (*RecLifetimeMDD*, *RecProbableMDD*) were identified based on the reported lifetime number of such episodes. For comparison with prior studies^{8,17} using broad depression outcomes assessed in nearly all participants, we defined Help-seeking Depression (*HelpSeekingDep*) based on participants' self-reports of seeking medical help for psychological distress symptoms. *EHR Diagnosis (EHRdep)* was derived by aggregating depression-related diagnostic entries (primary or secondary) across primary care, hospital inpatient, and death register records.

Genetic Similarity-Based Ancestry Group Assignment

For ancestry-stratified analyses, we used the *POP-MaD* (population grouping by Mahalanobis distance) pipeline²⁴ to assign the UKB participants to ancestry groups²⁵ based on genetic similarity to seven global population groups in the 1000 Genomes Project Phase-3 and the Human Genome Diversity Project²⁶ (African/AFR, Admixed American/AMR, Central and South Asian/CSA, East Asian/EAS, European/EUR, Middle-Eastern/MID, and Oceania/OCE; **Supplementary Methods**). The derived groups were assigned labels AFR-like, AMR-like, CSA-like, EAS-like, EUR-like, MID-like, and OCE-like, reflecting the label of the reference

population with which the group clustered. Potential “outliers” (Mahalanobis distance more than three standard deviations from the group median) were excluded from the ancestry-stratified analyses to further reduce within-group genetic heterogeneity.

Association Analyses

For the ancestry-stratified meta-analysis²¹ (**Fig. 1, Panel A2**), we first performed genetic ancestry group-specific GWASs (including the EUR-like ancestry-specific GWAS typically conducted in UKB), followed by a cross-ancestry fixed-effect meta-analysis²⁷ (**Supplementary Methods**). In the joint trans-ancestry approach, the GWAS is conducted across all participants using linear mixed models (LMMs) to account for genetic similarities between participants. In both approaches, we performed GWAS using REGENIE²⁸, which implements a whole-genome regression model closely related to the LMM and helps to control for population structure and genetic relatedness between samples. Thus, we also included related samples in both the ancestry-stratified and the joint trans-ancestry GWAS. For each set of results, independent genome-wide significant (GWS; $p < 5 \times 10^{-8}$) single nucleotide polymorphisms (SNPs) were identified using conditional-and-joint GCTA-COJO²⁹ analyses, and the overlap of GWS loci (COJO-index SNP $\pm 250\text{kb}$) was then examined between phenotypes and analytical approaches (**Supplementary Methods**).

SNP Heritability and Genetic Correlations

We used linkage disequilibrium score regression (LDSC)³⁰ to estimate the SNP-based heritability (phenotypic variance explained by genome-wide SNPs) of each depression phenotype and the genetic correlations (r_G) between phenotypes (**Supplementary Methods**). As LDSC requires relatively homogeneous ancestry samples, these analyses were limited to the EUR-like ancestry-specific GWAS, as the other groups were not large enough for ancestry-specific post-GWAS analyses.

Fine-mapping and Colocalization

At each genomic locus associated with any of the six depression phenotypes, we performed statistical fine-mapping (using *SuSieR*³¹) to identify credible sets of SNPs likely driving the locus

association (**Supplementary Methods**). We then performed colocalization analyses (using *coloc*³²) to examine whether the different phenotypes share putative causal SNPs at each locus, even if the locus does not reach the GWS threshold for a phenotype in GWAS.

Post-GWAS Gene-Level Analyses in the Brain

We performed transcriptome-wide association study (TWAS)³³ analyses using *JEPEGMIX2-P*³⁴ to predict brain region-specific gene expression levels associated with depression phenotypes, leveraging external data on expression quantitative loci (eQTL) in 13 brain regions from GTEx³⁵ (**Supplementary Methods**). Further, we performed summary-statistics-based Mendelian Randomization (SMR³⁶) analyses with the GWAS associations and the eQTL data to examine whether gene expression levels in these brain regions putatively mediate the SNP-depression association (**Supplementary Methods**). In both analyses, a within-tissue false discovery rate <0.05 was considered statistically significant after multiple-testing correction. As well-powered eQTL data were only available for EUR-like ancestry groups, these analyses were limited to the EUR-like ancestry-specific GWAS³⁷.

Polygenic Score Testing in the *All of Us* Research Program

To test the individual-level replicability and cross-ancestry transferability of the genetic associations discovered in UKB, we tested polygenic scores (PGSs) in the *All of Us* Research Program (AoU)³⁸. We derived PGSs based on EUR-like ancestry-specific and joint trans-ancestry GWAS summary statistics of *LifetimeMDD* and *HelpSeekingDep*, using PRS-CS³⁹ (**Supplementary Methods**). PGS associations with EHR-based and self-reported depression diagnoses in the EUR-like (N=112,005), AFR-like (N=41,149), and AMR-like (N=32,958) unrelated subsets of AoU participants were estimated using logistic regression, controlling for age, gender, and the first 10 genetic principal components (PCs).

Results

Extended Sample with Strictly Defined MDD Phenotypes

We identified 62,061 cases and 118,814 controls of *LifetimeMDD* (**Fig 1A**) out of 211,535 participants with at least one assessment of self-reported MDD symptoms across the two CIDI-SF surveys (**Supplementary Table 2**). This aggregated study sample yielded a 38% increase in eligible participants for GWAS of MDD, compared to N~153k participants evaluated in the first assessment, of which previous studies examined ~67k (self-identified “White-British”)¹⁷ to ~93k (EUR-like genetic ancestry)²⁰ unrelated participants.

Expanded Genetic Ancestry Group Assignments

Updated genetic ancestry group assignment using *POP-MaD* yielded subsets of 458,547 EUR-like, 10,119 CSA-like, 7,930 AFR-like, 2,553 AMR-like, 2,382 EAS-like, and 1,657 MID-like participants, representing >5% non-EUR-like participants, while 2,952 participants were classified as ancestry-group “outliers” (**Figures 1B-C, Supplementary Tables 3, Supplementary Figure 1**). Compared to previous assignments based on a random-forest algorithm used in the “Pan-UKBB” initiative⁴⁰, *POP-MaD* increased the available sample size by 36,621 (8.2%) for ancestry-stratified meta-analysis, including 33,069 (7.9%) for EUR-like ancestry-specific analyses (**Supplementary Tables 4**). The genetic ancestry groups showed significant mean differences in case prevalences of both strict MDD and broad depression outcomes (**Fig. 1D, Supplementary Table 5**), with CSA-like, AFR-like, EAS-like, and MID-like participants being less likely to meet the case criteria than EUR-like participants.

Importantly, it should be noted that genetic similarity-based subsets are created to reduce genetic heterogeneity for statistical analyses (e.g., to minimize population stratification) and should not be considered synonymous with social constructs of racial and ethnic identities²⁵. In the joint trans-ancestry analyses, all participants were included without considering such groupings while statistically accounting for their genetic similarity.

Inclusive GWAS Approaches Maximize the Power for Genomic Discovery and Uncover Novel Loci for MDD

We identified genome-wide significant (GWS; $p < 5 \times 10^{-8}$) genetic variants in three sets of results: EUR-like ancestry-specific GWAS, cross-ancestry meta-analysis, and joint trans-ancestry GWAS. Ancestry-specific GWASs in CSA-like, AFR-like, AMR-like, EAS-like, and MID-like subsets did not yield any GWS loci, likely because the sample sizes were <1-2% of the EUR-like group. In total, these analyses identified 61 independent genomic loci across depression phenotypes (**Fig. 2, Supplementary Tables 6-12, Supplementary Figure 2**). Including non-EUR-like participants (via cross-ancestry (*stratified*) and trans-ancestry (*joint*) GWAS approaches) resulted in 14 additional GWS loci despite a modest sample size increase of ~6%. Conversely, three loci in the EUR-like ancestry group-specific analyses were no longer significant in the cross-ancestry GWASs, though with largely overlapping confidence intervals across approaches. Fourteen loci (including five novel loci; **Table 1, Supplementary Figures 3-4**) were associated with one or more strict MDD phenotypes, compared to only one locus previously reported for *LifetimeMDD* in the unrelated EUR-like subset²⁰. Consistent with prior findings, the highest number of GWS loci (48) were identified in the analyses of *HelpSeekingDep*, likely reflecting its largest sample size.

To examine potential confounding due to relatedness or population structure, we compared the LDSC intercepts of GWASs performed as sensitivity analyses in the EUR-like subset by excluding related participants or ancestry-group “outliers” defined with varying stringency (**Supplementary Methods**). Notably, these analyses showed negligible non-systematic differences in their LDSC intercept, supporting the robustness of GWAS to these sources of confounding (**Supplementary Table 13, Supplementary Figure 5**).

For *LifetimeMDD*, nine significant loci, including three novel (**Table 1**), were identified. Two loci were discovered only upon including non-EUR-like participants (**Fig. 2B**), comprising one novel (rs7515936 upstream of *PAPPA2*) and one previously reported (rs4388849 in *CNTN5*)¹⁴ locus. One novel locus (rs7568777 upstream of *RTN4*) showed a significant association with *LifetimeMDD* in the EUR-like subset only but was associated with *RecLifetimeMDD* in all three

GWAS approaches. Another novel locus in *FOXP1* (**Fig. 2C**) had intronic SNPs associated with *LifetimeMDD*, *RecLifetimeMDD*, and *RecProbableMDD* in all three approaches. The remaining two novel MDD loci were tagged by intronic index SNPs in *RAPGEF4* and *ZFPM2* and associated with recurrent MDD phenotypes.

Ancestrally inclusive GWAS approaches also improved fine-mapping resolution at the associated loci (**Supplementary Tables 14-19**). Notably, the ancestry-stratified meta-analysis yielded the smallest median credible set size (fewest fine-mapped SNPs per credible set) and identified the most sets containing higher confidence fine-mapped SNPs with a posterior inclusion probability >0.50 (**Supplementary Figure 6**).

centered at the index SNPs in *FOXP1* (at 3p13), one of the novel risk loci for Lifetime MDD. The X-axis shows the SNP positions on chromosome 3 (aligned to GRCh37), and the Y-axis shows the $-\log_{10}(\text{p-value})$. The estimates from the three GWASs are shown in different colors and overlaid on each other. **D.** Number of independent genomic loci associated at the GWS level with each depression phenotype across different GWAS approaches. **E.** Upset plot showing the overlap of GWS loci between the three main sets of GWAS results, pooled across depression phenotypes. The matrix rows represent GWAS approaches, with the horizontal bars showing the number of GWS loci per approach. The columns represent the overlap between the indicated approaches, with the corresponding vertical bar depicting the number of overlapping loci.

Table 1*Novel Risk Loci Genome-wide Significantly Associated with Strictly Defined MDD Phenotypes*

Locus Nr.	Cytogenic Region	Mapped Gene	Phenotype	GWAS Approach	Index SNP	Chr	Position	Alt. Allele	Effect Allele	Effect Allele Freq.	logOR	S.E.	p-value
1	1q25.2	PAPPA2	Lifetime MDD	Joint Trans-ancestry GWAS	rs7515936	1	176376432	G	T	0.582	0.041	0.007	4.74E-08
2	2p16.1	RTN4	Lifetime MDD	EUR-like Ancestry GWAS	rs7568777	2	55383427	T	A	0.192	0.052	0.009	3.91E-08
			Rec. Lifetime MDD	EUR-like Ancestry GWAS	rs7568777	2	55383427	T	A	0.192	0.065	0.011	4.25E-09
				Cross-ancestry Meta-analysis	rs7568777	2	55383427	T	A	0.193	0.063	0.011	9.41E-09
				Joint Trans-ancestry GWAS	rs7568777	2	55383427	T	A	0.193	0.060	0.011	3.42E-08
3	2q31.1	RAPGEF4	Rec. Lifetime MDD	Joint Trans-ancestry GWAS	rs2113808	2	173730969	G	A	0.670	0.052	0.009	1.77E-08
			Rec. Probable MDD	EUR-like Ancestry GWAS	rs2676507	2	173729572	A	G	0.678	0.052	0.009	7.96E-09
				Cross-ancestry Meta-analysis	rs2676507	2	173729572	A	G	0.676	0.053	0.009	3.53E-09
				Joint Trans-ancestry GWAS	rs2113808	2	173730969	G	A	0.670	0.054	0.009	1.71E-09
4	3p13	FOXP1	Lifetime MDD	EUR-like Ancestry GWAS	rs35397914	3	71010071	C	G	0.341	0.048	0.008	2.37E-09
				Cross-ancestry Meta-analysis	rs34371285	3	71010069	T	C	0.336	0.048	0.008	2.53E-09
				Joint Trans-ancestry GWAS	rs35397914	3	71010071	C	G	0.338	0.048	0.008	1.09E-09
			Probable MDD	EUR-like Ancestry GWAS	rs34371285	3	71010069	T	C	0.341	0.048	0.008	3.84E-10
				Cross-ancestry Meta-analysis	rs34371285	3	71010069	T	C	0.336	0.047	0.008	2.53E-09
				Joint Trans-ancestry GWAS	rs35397914	3	71010071	C	G	0.338	0.048	0.008	1.59E-10
			Rec. Probable MDD	EUR-like Ancestry GWAS	rs564737701	3	70954761	T	TA	0.302	0.053	0.009	1.88E-08
				Cross-ancestry Meta-analysis	rs564737701	3	70954761	T	TA	0.302	0.052	0.009	2.40E-08

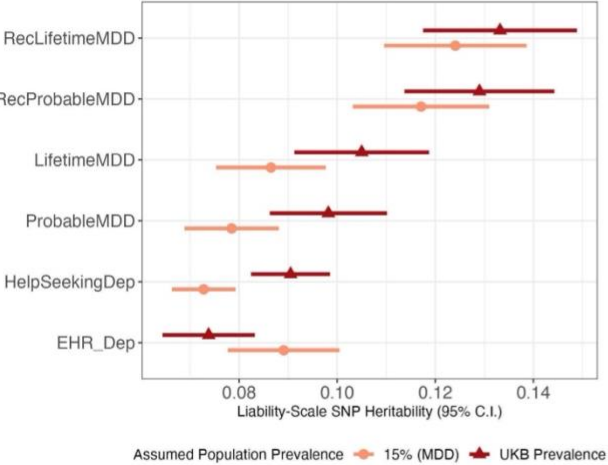
				Joint Trans-ancestry GWAS	rs564737701	3	70954761	T	TA	0.302	0.051	0.009	4.01E-08
5	8q23.1	ZFPM2	Rec. Lifetime MDD	Joint Trans-ancestry GWAS	rs4565433	8	106497543	T	G	0.153	0.066	0.012	3.28E-08
			Probable MDD	Joint Trans-ancestry GWAS	rs200626614	8	106380204	G	T	0.181	0.050	0.009	4.60E-08
			Rec. Probable MDD	EUR-like Ancestry GWAS	rs4565433	8	106497543	T	G	0.155	0.064	0.012	3.64E-08
				Cross-ancestry Meta-analysis	rs4565433	8	106497543	T	G	0.153	0.065	0.012	2.23E-08
				Joint Trans-ancestry GWAS	rs4565433	8	106497543	T	G	0.153	0.066	0.012	1.10E-08

Note. The SNP association statistics (logOR, S.E., and p-value) are estimated conditional on age, sex, assessment center, genotyping microarray platform, and the first 20 genetic PCs. The effect allele frequency in the cross-ancestry meta-analysis is the sample size-weighted average of the ancestry group-specific allele frequencies. **Supplementary Figures 3-4** present the Forest Plots and the regional association plots, respectively, of these loci.

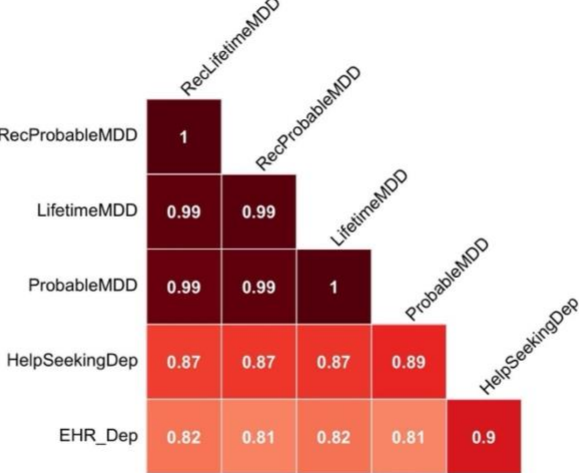
Strict MDD Phenotypes are Associated with Genomic Loci Not Shared with Broad Depression

Consistent with prior literature¹⁷, strict MDD phenotypes had higher SNP-based heritability than broad depression ones, ranging from 0.074 (S.E. 0.005) for *EHRdep* to 0.133 (S.E. 0.008) for *RecLifetimeMDD* in the EUR-like subset (**Fig. 3A, Supplementary Tables 20-21**). The four MDD phenotypes showed near-perfect r_G with each other (range: 0.99–1.00; in the EUR-like subset) but slightly lower r_G with *HelpSeekingDep* (0.87–0.89) and lower still with *EHRdep* (0.81–0.82), indicating that MDD shares approximately 79% ($r^2 \times 100$) genetic variance with *HelpSeekingDep* and 67% with *EHRdep* (**Fig. 3B**). Of the 14 genomic risk loci identified across strict MDD phenotypes, only four loci were also GWS with *HelpSeekingDep* or *EHRdep* (despite their much larger sample sizes), suggesting higher specificity for MDD (**Fig. 3C**). Importantly, none of the five novel MDD loci were GWS in the analyses of broad depression phenotypes. In colocalization analyses applied across all depression-associated loci in each GWAS approach, the MDD phenotypes shared putative causal SNPs at 35–44% of the loci (**Fig. 3D, Supplementary Tables 22-25**). In contrast, the MDD phenotypes shared putative causal SNPs with *HelpSeekingDep* at 20–22% of loci and *EHRdep* at only 7–13%.

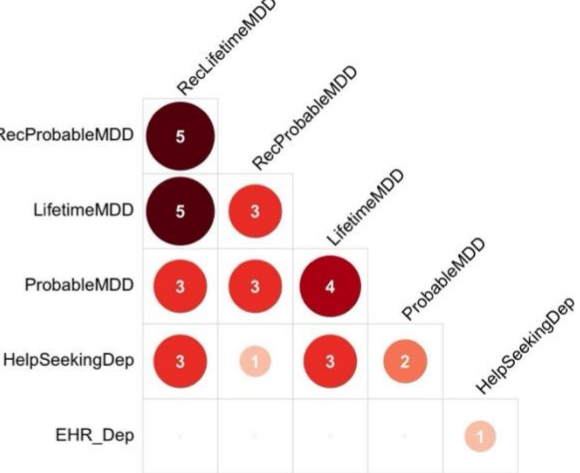
A. SNP-Heritability of Depression Phenotypes in EUR-like Ancestry Group



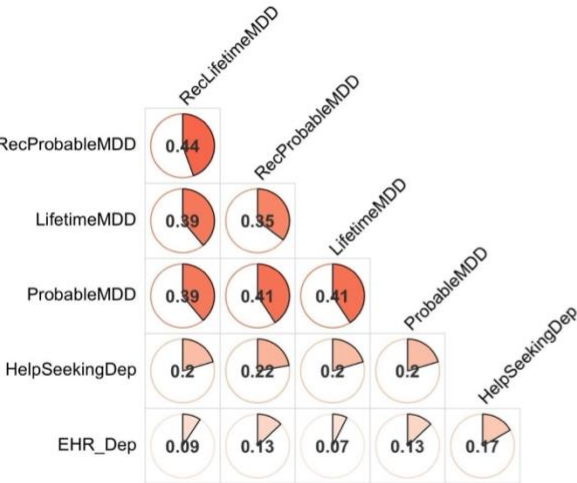
B. Genetic Correlations Between Depression Phenotypes in EUR-like Ancestry Group



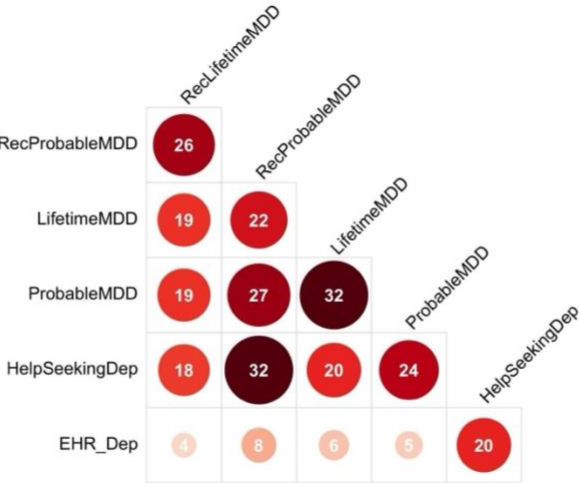
C. Number of Shared Genome-wide Significant Loci Between Phenotypes



D. Proportion of Credible Sets of SNPs with Colocalization Between Phenotypes in EUR-like Ancestry Group



E. Number of Shared Brain-Expressed Genes Associated with Phenotypes in TWAS in EUR-like Ancestry Group



F. Overlap of Brain-Expressed Genes with Putative Effects on Depression in SMR Analyses in EUR-like Ancestry Group

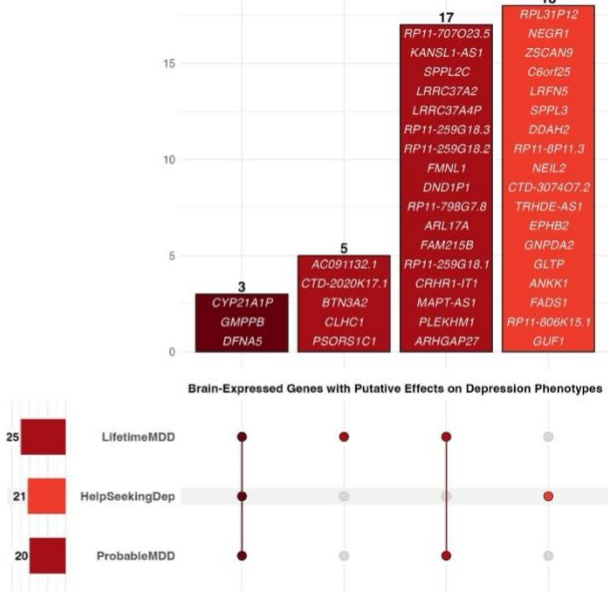


Figure 3: Overlap between Depression Phenotypes at Genome-wide and Locus-level Resolution.

A. Genome-wide SNP-based heritabilities in the EUR-like subset estimated using LDSC analyses. The plot shows the point estimate and the 95% confidence interval. Two sets of liability-scale heritabilities are shown, with different assumed population prevalence rates of the depression outcomes: 15% (prevalence of MDD) and the phenotype-specific overall prevalence observed in UKB. **B.** Genome-wide SNP-based genetic correlations in the EUR-like subset estimated using LDSC analyses and assuming UKB prevalence to be the population prevalence rate. **C.** Number of genome-wide significant risk loci overlapping between phenotypes. The number of loci per phenotype is aggregated across the three GWAS results and then examined for overlap with the other phenotypes. **D.** Proportion of genomic risk loci showing significant (pair-wise) colocalization between phenotypes based on the EUR-like ancestry-specific GWAS. Similar overlaps were seen with the other GWAS results (**Supplementary Table 25**). **E.** Number of overlapping unique genes (in 13 brain regions) significantly associated with different phenotypes in the EUR-like ancestry-specific TWAS. **F.** Upset plot showing the overlap of unique genes (across 13 brain regions) with significant putative causal effects on depression phenotypes in the EUR-like ancestry-specific SMR analyses. The matrix rows represent depression phenotypes, with the corresponding horizontal bar showing the number of causally linked genes. The columns represent the overlap between the indicated phenotypes, with the corresponding vertical bar showing the number and names of the overlapping genes.

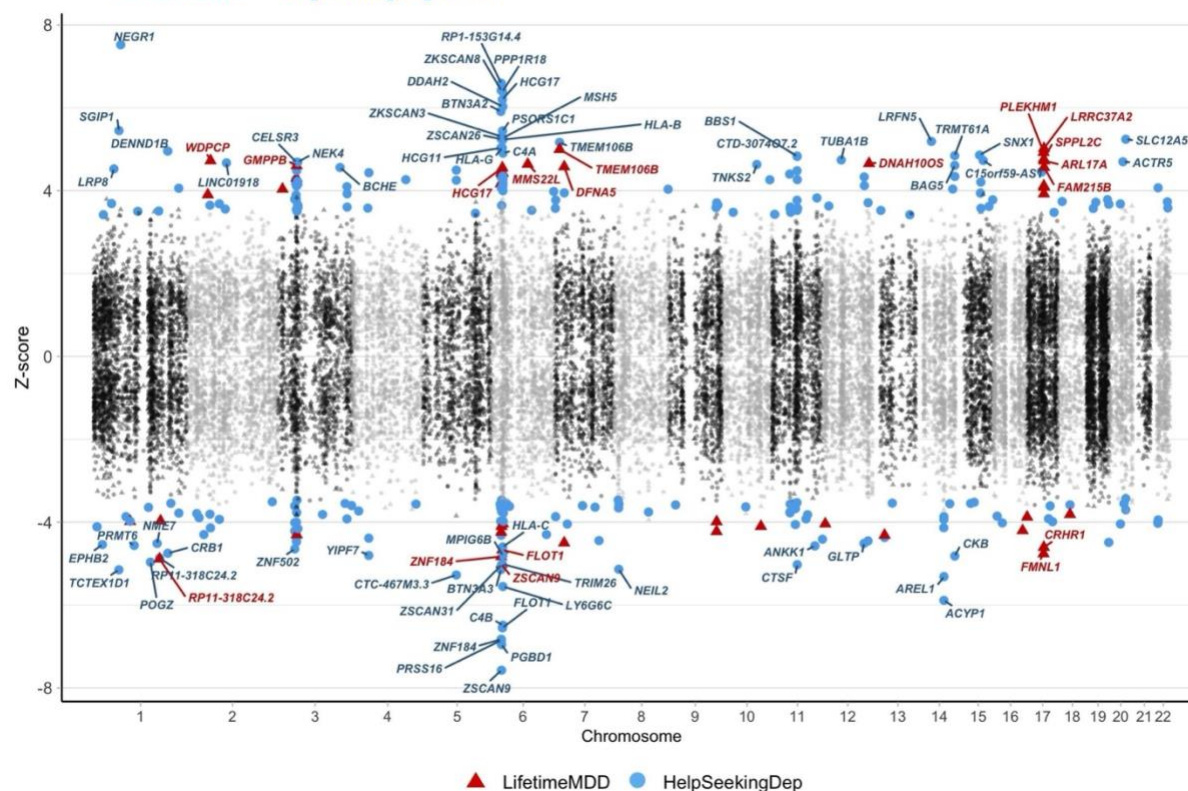
HPA Axis Genes are Implicated in Lifetime MDD but not Broad Depression

TWAS analyses revealed 299 unique genes with brain region-specific expression levels associated with the six depression phenotypes, of which 78 were associated with one or more strict MDD phenotypes (**Supplementary Tables 26-31**). Consistent with the GWAS results, *HelpSeekingDep* was associated with considerably more genes (244) than the other phenotypes: 44 with *LifetimeMDD*, 47 with *ProbableMDD*, 33 with *RecLifetimeMDD*, 46 with *RecProbableMDD*, and 34 with *EHRdep*. The MDD phenotypes shared around half of the associated genes with each other and with *HelpSeekingDep* (**Fig. 3E**), indicating unique gene-level associations discovered with strict MDD phenotypes despite their smaller sample size. To illustrate, only 20 genes were shared between *LifetimeMDD* and *HelpSeekingDep* (**Fig. 4A**). Here, two notable clusters of genes in high-LD chromosomal regions emerged: the genes in the major histocompatibility complex region associated with both *LifetimeMDD* and *HelpSeekingDep*, and the genes at 17q21.31 associated almost exclusively with *LifetimeMDD*. The latter region includes genes involved in the HPA (hypothalamus-pituitary-adrenal) axis and stress response, including the *CRHR1* (corticotropin-releasing hormone receptor 1) gene, not associated with *HelpSeekingDep*. The 17q21.31 locus contains a common inversion polymorphism in EUR-like populations⁴¹, and the mapped gene cluster has previously been implicated in post-traumatic stress disorder⁴², autism spectrum disorder⁴³, and neurodegenerative disorders⁴¹.

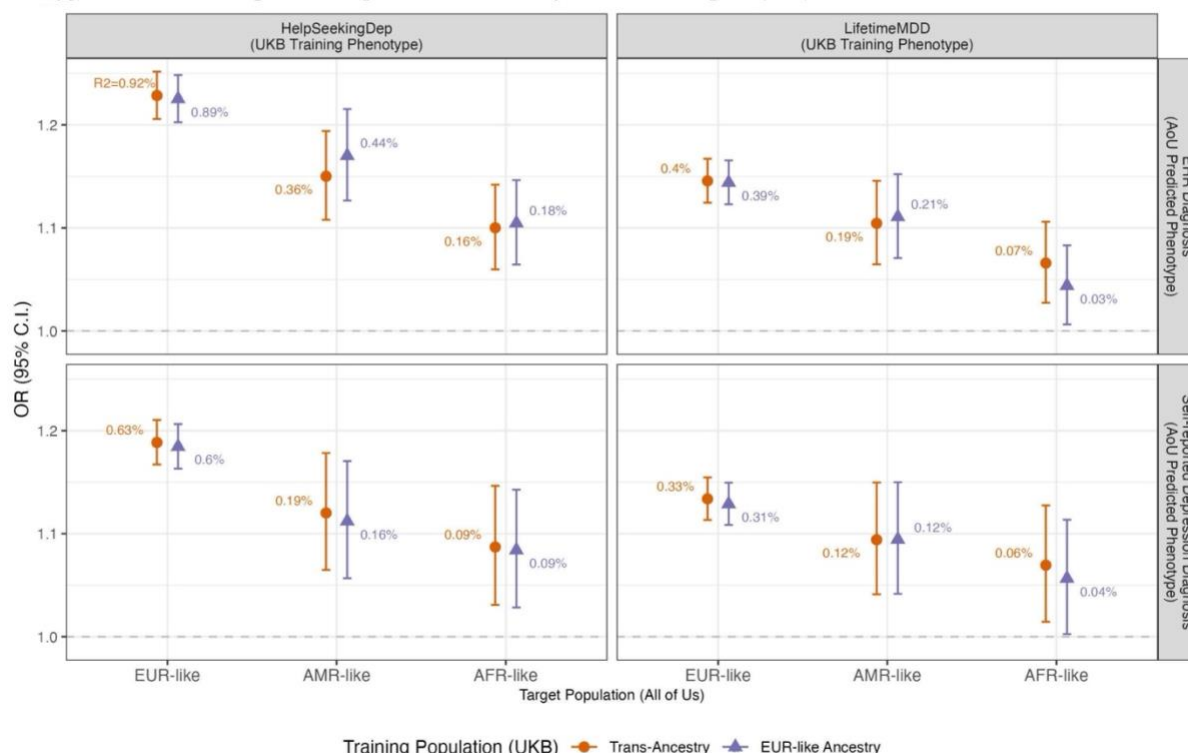
Further SMR analyses showed significant putative causal influences of brain region-specific gene expression levels on *LifetimeMDD*, *ProbableMDD*, and *HelpSeekingDep* (**Fig. 3F**, **Supplementary Tables 32-34**). The 20 genes putatively influencing *ProbableMDD* were all a subset of the 25 genes influencing *LifetimeMDD*. However, of the 21 genes with potential effects on *HelpSeekingDep*, only three were shared with *LifetimeMDD*. Thus, these analyses highlighted genes likely specific to MDD (e.g., *CRHR1-IT1* and *LRRC37A4P* [17q21.31]) versus those shared with *HelpSeekingDep* (e.g., *GMPPB* and *DFNA5*).

Figure 4: Follow-up Analyses Comparing Lifetime MDD and Help-Seeking Depression: TWAS in the Brain and Polygenic Prediction

A. TWAS of Lifetime MDD and Help-Seeking Depression



B. Polygenic Prediction of Depression Diagnosis in the NIH All of Us Research Program (AoU)



A. Miami plot showing genes predicted to be associated with Lifetime MDD and Help-Seeking Depression in TWAS analyses based on the EUR-like ancestry-specific GWAS. TWAS analyses were performed across 13 brain regions. The Y-axis indicates each gene's Z-score in the brain region with the most significant association with each phenotype. Significant associations with Lifetime MDD are shown in red, while those with Help-Seeking Depression are shown in blue. **B.** Associations of UKB-based PGSs of Lifetime MDD and Help-Seeking Depression with depression diagnoses in the EUR-like, AMR-like, and AFR-like participants from the All of Us Research Program (AoU). The vertical panels indicate the UKB depression phenotype used to compute the PGSs, and the horizontal panels indicate the depression diagnosis being predicted in AoU (EHR-based and self-reported). The error bars indicate the 95% confidence interval of the odds ratio, and the text label indicates the incremental pseudo- R^2 (in %). All PGS associations are conditional on age, gender, and the first 10 genetic PCs and are statistically significant at FDR <0.05. The PGSs shown here were based on the EUR-like ancestry-specific and the joint trans-ancestry GWASs. **Supplementary Figure 8** compares these results with PGSs computed through cross-ancestry meta-analysis in PRS-CSx.

Polygenic Scores from UKB Significantly Associated with Depression Diagnoses Across Ancestry Groups in AoU

Using EUR-like ancestry-specific GWAS summary statistics from UKB, both *LifetimeMDD* and *HelpSeekingDep* PGSs were significantly associated with EHR-derived and self-reported depression diagnoses in the EUR-like subset of AoU participants (**Fig. 4B, Supplementary Table 35**). The *HelpSeekingDep* PGS had a stronger association with depression diagnoses (O.R. = 1.23; 95% confidence interval (C.I.) = [1.20, 1.25]; incremental R^2 = 0.89% for EHR Diagnosis) than *LifetimeMDD* PGS (O.R. = 1.14; C.I. = [1.12, 1.17]; incremental R^2 = 0.39%), which prior research¹⁷ has demonstrated to be due to the former's larger GWAS sample size. The association of PGSs based on trans-ancestry GWAS results showed trivial, non-significant differences from EUR-like ancestry-based PGSs, likely because EUR-like participants comprised ~95% of the UKB trans-ancestry sample. Notably, both EUR-like ancestry-based and trans-ancestry PGSs were also significantly associated with depression diagnoses in the AMR-like and AFR-like participants, albeit with reduced effect sizes. For instance, for the association of EUR-like ancestry-based *HelpSeekingDep* PGS with EHR depression diagnosis: O.R. = 1.17; C.I. = [1.13, 1.22]; incremental R^2 = 0.44% in AMR-like participants, and O.R. = 1.10; C.I. = [1.06, 1.15]; incremental R^2 = 0.18% in AFR-like participants.

Discussion

This study presents novel MDD-specific genomic insights in UKB through two key innovations: a considerably larger set of participants with strictly defined MDD phenotypes from two repeated assessments of symptom-level data and an inclusive, trans-ancestry GWAS pipeline accommodating related participants across the genetic ancestry continuum, which maximized the statistical power for genomic discovery. While prior UKB analyses showed that broad depression outcomes are primarily associated with genomic loci not specific to MDD¹⁷, previous studies used smaller samples and could identify just one MDD-associated locus. Here, we identified 14 independent genomic risk loci, including five novel, associated with MDD phenotypes. The MDD-associated genomic loci and TWAS-predicted gene expression levels in the brain showed little overlap with broad depression, indicating higher specificity. Despite 95% of UKB participants being genetically similar to EUR ancestry, the trans-ancestry analyses uncovered 14 additional genomic loci across depression phenotypes, highlighting the value of ancestrally inclusive GWAS. This pipeline implements the ethical imperative of adopting more inclusive GWAS approaches, ensuring that, as far as possible, the analyses include all participants who contributed DNA and survey data.

Two of the five novel genomic risk loci for MDD are intergenic regions upstream of *PAPPA2* and *RTN4*. *PAPPA2* regulates the bioavailability of insulin-like growth factor⁴⁴, a neurotrophic hormone crucial for central nervous system development, maturation, and neuroplasticity⁴⁵. *RTN4* inhibits neurite outgrowth and axon regeneration and plasticity⁴⁶. Notably, high *RTN4* expression levels are postulated as a state-level blood biomarker of psychological stress⁴⁷ and hallucinations⁴⁸ and predict future hospitalizations attributed to these symptoms^{47,48}. The other three loci are intronic regions in *RAPGEF4*, *FOXP1*, and *ZFPM2*. *RAPGEF4* is expressed primarily in the brain, neuroendocrine, and endocrine tissues⁴⁹ and is involved in neuronal differentiation, neurite outgrowth, neurotransmitter release, memory, and learning⁴⁹, as well as corticotropin-releasing hormone-mediated dendritic spine loss in acute stress⁵⁰. Moreover, *RAPGEF4* knock-out mice show depression- and anxiety-like behaviors responsive to fluoxetine (selective serotonin reuptake inhibitor)⁵¹. *FOXP1* encodes a transcription factor critical to the early development of several organ systems, including the brain⁵². Deletions or mutations in this gene cause *FOXP1 Syndrome*, characterized by, among other features, language deficits, autism

spectrum disorder, intellectual disability, and congenital brain anomalies⁵². *ZFPM2* is a zinc-finger transcription factor associated with GABAergic neuron differentiation in the midbrain and hindbrain tegmental nuclei that control mood, memory, motivated behavior, and movement⁵³.

These novel loci have previously been reported to be linked with other MDD-associated traits. For instance, *PAPPA2*, *RTN4*, *FOXP1*, and *ZFPM2* are associated with body weight⁵⁴, body mass index⁵⁵, or body fat percentage⁵⁶. *RTN4*, *RAPGEF4*, *FOXP1*, and *ZFPM2* are associated with educational attainment⁵⁷ (in the opposite direction of effect). *RTN4*, *RAPGEF4*, and *FOXP1* are associated with substance use⁵⁸ or other risk-taking/externalizing behaviors⁵⁹. Notably, SNPs in *FOXP1* have also been implicated in GWASs of retrospectively self-reported childhood maltreatment⁶⁰, insomnia⁶¹, and neuroimaging-based total brain volume⁶² and sulcal depth⁶³. These pleiotropic associations may arise from horizontal pleiotropy due to shared endophenotypes and/or vertical pleiotropy due to causal influences between MDD and other traits.

PGSs based on UKB GWAS results showed significant within- and cross-ancestry associations with depression diagnoses in AoU, highlighting the shared genetic architecture of MDD across populations. However, consistent with an increasing genetic distance from EUR ancestry, the effect sizes reduced progressively in AMR-like and AFR-like participants, underscoring the need for globally representative genetic studies to prevent exacerbating health disparities. The diminished associations may also reflect differences in healthcare access and, thus, depression diagnosis between populations, further reducing the transferability of UKB GWAS results to non-EUR-like participants.

When interpreting these findings, it should be noted that the MDD phenotypes were not obtained from gold-standard clinical diagnostic interviews. Although phenotypes based on structured clinical interviews and self-reported retrospective symptoms from CIDI-SF likely lie on the same genetic liability continuum¹⁷, the respondent's current mood may prime the recall of past episodes in the latter. Secondly, prior research has shown that participation in the online mental health questionnaire in UKB is genetically correlated with several mental health traits⁶⁴, though follow-up analyses found no evidence of its influence on MDD case-control status¹⁷. UKB is

also enriched for healthy volunteers with overall better health and higher levels of education than the general UK population⁶⁵, which could introduce bias in GWAS of behavioral traits⁶⁶.

In this study, we performed both ancestry-stratified meta-analysis and joint trans-ancestry GWAS, following best practices²¹. The ancestry group-specific analyses in the former exclude the participants identified as potential ancestry group outliers but can facilitate post-GWAS analyses requiring ancestrally matched external reference data, such as LDSC analyses³⁰ for SNP-based heritability and genetic correlations, gene-level TWAS³³ and SMR³⁶ analyses, and PGS analyses³⁹. In comparison, the joint trans-ancestry approach is the most inclusive as it includes all participants across the genetic ancestry continuum. Given the predominance of EUR-like participants in UKB, the trans-ancestry genetic analyses in our study are likely skewed toward EUR ancestry, as was evidenced by the trivial differences between the predictive performance of PGSs based on EUR-like ancestry-specific and trans-ancestry GWASs. Also, the genetic associations across ancestral groups may be influenced by reduced coverage of common genetic variants in non-EUR-like groups⁶⁷. The non-EUR-like ancestry groups were considerably smaller and had less power for ancestry-specific GWAS, post-GWAS analyses, or for computing cross-ancestry PGSs by meta-analyzing ancestry-specific results with PRS-CSx⁶⁸.

Nevertheless, the increased number of genomic risk loci discovered upon including the non-EUR-like participants demonstrates the advantages of our ancestrally inclusive pipeline. Moreover, the non-EUR-like ancestry-specific results from this study may contribute to large-scale ancestry-specific and cross-ancestry GWAS meta-analyses of MDD. To achieve parity in the population-specific understanding of the genetics of MDD, continued long-term efforts to recruit currently underrepresented populations remain imperative. Likewise, it is essential to build globally representative reference datasets to facilitate post-GWAS analyses⁶⁹.

Overall, trans-ancestry GWAS of depression outcomes maximized the genomic insights from extant data and uncovered novel genomic risk loci for MDD. Our findings extend the prior genome-level findings¹⁷ of little overlap between MDD and broad depression to genomic locus and gene expression levels, re-emphasizing the need for investing in larger, globally representative datasets with strictly defined, or preferably clinically evaluated, MDD. Such

studies will be vital for bridging the gaps in our understanding of MDD and developing novel disease-specific therapies generalizable across populations. In the immediate term, our analytical pipeline may facilitate further trans-ancestry analyses of existing data to help improve global representation, statistical power, and generalizability in genomic studies.

Data Availability

The UK Biobank data used in this research is available to “bona fide researchers for health-related research in the public interest” through an application process accessible through the UK Biobank website, <https://www.ukbiobank.ac.uk/>.

The summary statistics from the genome-wide association analyses will be made available through OSF.

Conflicts of Interest

Nothing to declare.

Acknowledgments

This research has been conducted using the UK Biobank Resource application number 30782 (PI: REP). We gratefully acknowledge Dr. Kenneth S. Kendler for his valuable guidance and feedback on the MDD phenotypes derived *in silico* and analyzed in this study. The authors wish to acknowledge their respective consortia and networks including the PGC Cross-Population Analyses Working Group (REP, MS, CC, TBB, AEG, BTW; <https://pgc.unc.edu/for-researchers/working-groups/cross-population-analyses-working-group/>), and the PsycheMERGE Network Diversity Initiative (REP, MS, CC, PBB, TBB; <https://psychemerge.com/diversity-initiative-workgroup/>). This work was supported by the National Institutes of Health grants R01MH125938 (REP, MS, CC, AEG, TBB, BTW) and K01AA031748 (AEG) and the Brain Behavior Research Foundation NARSAD grant 28632 P&S Fund (REP). Computing costs in the All of Us Researcher Workbench were supported by R01MH125938 (PI: REP). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies and the National Institutes of Health.

References

1. Marx W, Penninx BWJH, Solmi M, et al. Major depressive disorder. *Nat Rev Dis Primer*. 2023;9(1):44. doi:10.1038/s41572-023-00454-1
2. Malhi GS, Mann JJ. Depression. *The Lancet*. 2018;392(10161):2299-2312. doi:10.1016/S0140-6736(18)31948-2
3. James SL, Abate D, Abate KH, et al. 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. *The Lancet*. 2018;392(10159):1789-1858. doi:10.1016/S0140-6736(18)32279-7
4. Sullivan PF, Neale MC, Kendler KS. Genetic Epidemiology of Major Depression: Review and Meta-Analysis. *Am J Psychiatry*. 2000;157(10):1552-1562. doi:10.1176/appi.ajp.157.10.1552
5. Converge Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*. 2015;523(7562):588-591. doi:10.1038/nature14659
6. Peterson RE, Cai N, Bigdeli TB, et al. The Genetic Architecture of Major Depressive Disorder in Han Chinese Women. *JAMA Psychiatry*. 2017;74(2):162-168. doi:10.1001/jamapsychiatry.2016.3578
7. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3
8. Howard DM, Adams MJ, Shirali M, et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-03819-3
9. Howard DM, Adams MJ, Clarke TK, 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):343-352. doi:10.1038/s41593-018-0326-7
10. Levey DF, Stein MB, Wendt FR, 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*. 2021;24(7):954-963. doi:10.1038/s41593-021-00860-2
11. Giannakopoulou O, Lin K, Meng X, et al. The Genetic Architecture of Depression in Individuals of East Asian Ancestry. *JAMA Psychiatry*. 2021;78(11):1258. doi:10.1001/jamapsychiatry.2021.2099
12. Bigdeli TB, Voloudakis G, Barr PB, et al. Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia, Bipolar Disorder, and Depression Among Adults in the US Veterans Affairs Health Care System. *JAMA Psychiatry*. 2022;79(11):1092-1101. doi:10.1001/jamapsychiatry.2022.2742
13. Als TD, Kurki MI, Grove J, et al. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat Med*. 2023;29(7):1832-1844. doi:10.1038/s41591-023-02352-1
14. Meng X, Navoly G, Giannakopoulou O, et al. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nat Genet*. 2024;56(2):222-233. doi:10.1038/s41588-023-01596-4
15. Adams MJ, Streit F, Meng X, et al. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell*. 2025;188(3):640-652.e9. doi:10.1016/j.cell.2024.12.002

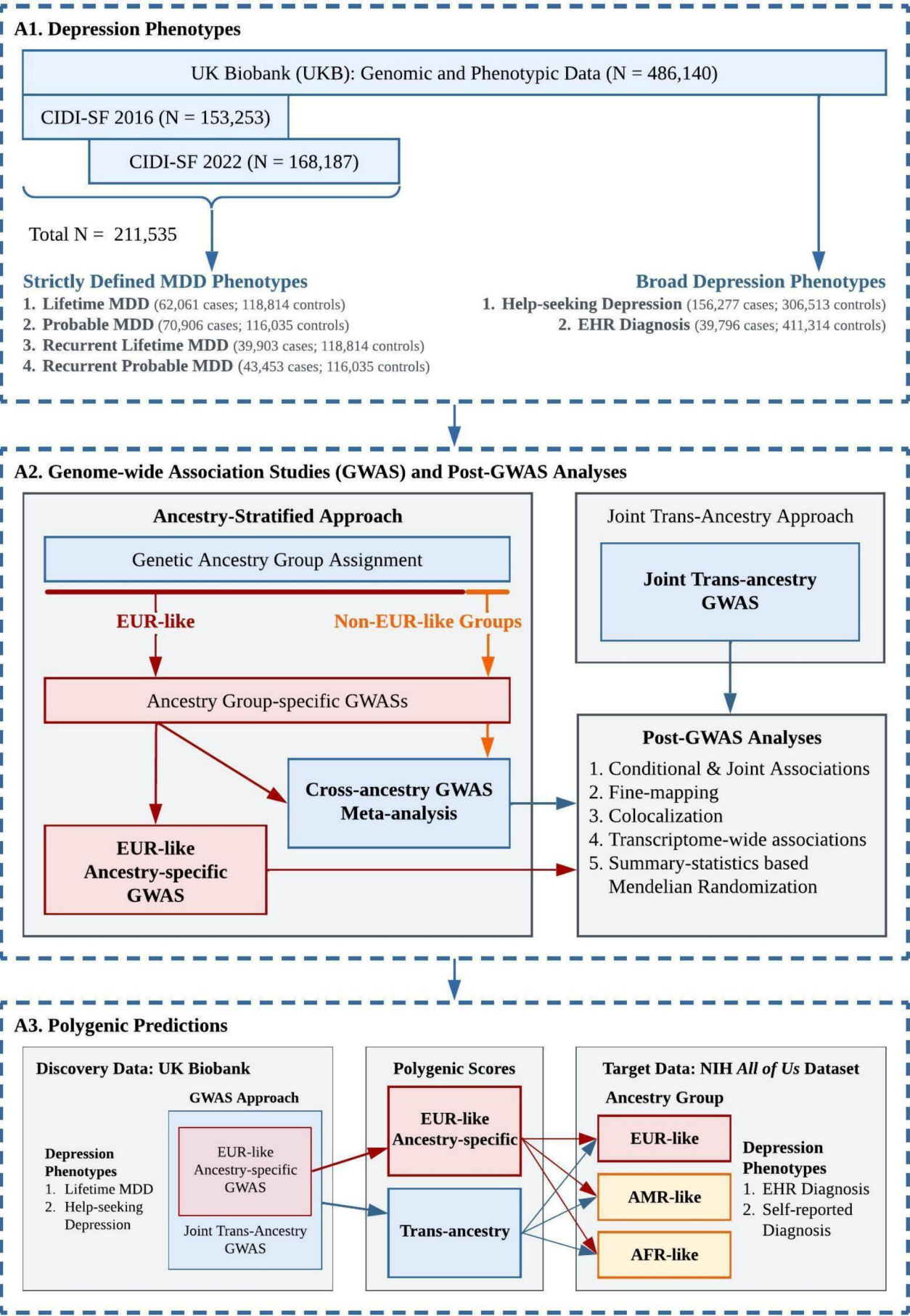
16. Flint J. The genetic basis of major depressive disorder. *Mol Psychiatry*. 2023;28(6):2254-2265. doi:10.1038/s41380-023-01957-9
17. Cai N, Revez JA, Adams MJ, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat Genet*. 2020;52(4):437-447. doi:10.1038/s41588-020-0594-5
18. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z
19. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen HU. The World Health Organization Composite International Diagnostic Interview short-form (CIDI-SF). *Int J Methods Psychiatr Res*. 1998;7(4):171-185. doi:10.1002/mpr.47
20. Coleman JRI, Peyrot WJ, Purves KL, et al. Genome-wide gene-environment analyses of major depressive disorder and reported lifetime traumatic experiences in UK Biobank. *Mol Psychiatry*. Published online 2020. doi:10.1038/s41380-019-0546-6
21. Peterson RE, Kuchenbaecker K, Walters RK, et al. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell*. 2019;179(3):589-603. doi:10.1016/j.cell.2019.08.051
22. Bigdeli TB, Ripke S, Peterson RE, et al. Genetic effects influencing risk for major depressive disorder in China and Europe. *Transl Psychiatry*. 2017;7(3):e1074-e1074. doi:10.1038/tp.2016.292
23. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Publishing; 2013.
24. Peterson RE, Edwards AC, Bacanu SA, Dick DM, Kendler KS, Webb BT. The utility of empirically assigning ancestry groups in cross-population genetic studies of addiction. *Am J Addict*. 2017;26(5):494-501. doi:10.1111/ajad.12586
25. National Academies of Sciences, Engineering, and Medicine. *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. The National Academies Press; 2023.
26. Koenig Z, Yohannes MT, Nkambule LL, et al. A harmonized public resource of deeply sequenced diverse human genomes. *Genome Res*. 2024;34(5):796-809. doi:10.1101/gr.278378.123
27. Han B, Eskin E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *Am J Hum Genet*. 2011;88(5):586-598. doi:10.1016/j.ajhg.2011.04.014
28. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet*. 2021;53(7):1097-1103. doi:10.1038/s41588-021-00870-7
29. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369-375. doi:10.1038/ng.2213
30. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211
31. 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 Stat Methodol*. 2020;82(5):1273-1300. doi:10.1111/rssb.12388

32. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLOS Genet.* 2021;17(9):e1009440. doi:10.1371/journal.pgen.1009440
33. Mai J, Lu M, Gao Q, Zeng J, Xiao J. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun Biol.* 2023;6(1):899. doi:10.1038/s42003-023-05279-y
34. Chatzinakos C, Georgiadis F, Lee D, et al. TWAS pathway method greatly enhances the number of leads for uncovering the molecular underpinnings of psychiatric disorders. *Am J Med Genet B Neuropsychiatr Genet.* 2020;183(8):454-463. doi:10.1002/ajmg.b.32823
35. The GTEx Consortium, Aguet F, Anand S, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318-1330. doi:10.1126/science.aaz1776
36. Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun.* 2018;9(1). doi:10.1038/s41467-017-02317-2
37. Bhattacharya A, Hirbo JB, Zhou D, et al. Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: Lessons from the Global Biobank Meta-analysis Initiative. *Cell Genomics.* 2022;2(10):100180. doi:10.1016/j.xgen.2022.100180
38. Bick AG, Metcalf GA, Mayo KR, et al. Genomic data in the All of Us Research Program. *Nature.* 2024;627(8003):340-346. doi:10.1038/s41586-023-06957-x
39. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun.* 2019;10(1). doi:10.1038/s41467-019-09718-5
40. Karczewski KJ, Gupta R, Kanai M, et al. Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. *medRxiv.* Published online January 1, 2024:2024.03.13.24303864. doi:10.1101/2024.03.13.24303864
41. de Jong S, Chepelev I, Janson E, et al. Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. *BMC Genomics.* 2012;13(1):458. doi:10.1186/1471-2164-13-458
42. Chatzinakos C, Pernia CD, Morrison FG, et al. Single-Nucleus Transcriptome Profiling of Dorsolateral Prefrontal Cortex: Mechanistic Roles for Neuronal Gene Expression, Including the 17q21.31 Locus, in PTSD Stress Response. *Am J Psychiatry.* 2023;180(10):739-754. doi:10.1176/appi.ajp.20220478
43. Gandal MJ, Zhang P, Hadjimichael E, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science.* 2018;362(6420):eaat8127. doi:10.1126/science.aat8127
44. Andrew M, Liao L, Fujimoto M, Khoury J, Hwa V, Dauber A. PAPP2 as a Therapeutic Modulator of IGF-I Bioavailability: in Vivo and in Vitro Evidence. *J Endocr Soc.* 2018;2(7):646-656. doi:10.1210/js.2018-00106
45. Dyer AH, Vahdatpour C, Sanfeliu A, Tropea D. The role of Insulin-Like Growth Factor 1 (IGF-1) in brain development, maturation and neuroplasticity. *Neuroscience.* 2016;325:89-99. doi:10.1016/j.neuroscience.2016.03.056
46. Konno T, Parutto P, Crapart CC, et al. Endoplasmic reticulum morphology regulation by RTN4 modulates neuronal regeneration by curbing luminal transport. *Cell Rep.* 2024;43(7):114357. doi:10.1016/j.celrep.2024.114357
47. Le-Niculescu H, Roseberry K, Levey DF, et al. Towards precision medicine for stress disorders: diagnostic biomarkers and targeted drugs. *Mol Psychiatry.* 2020;25(5):918-938.

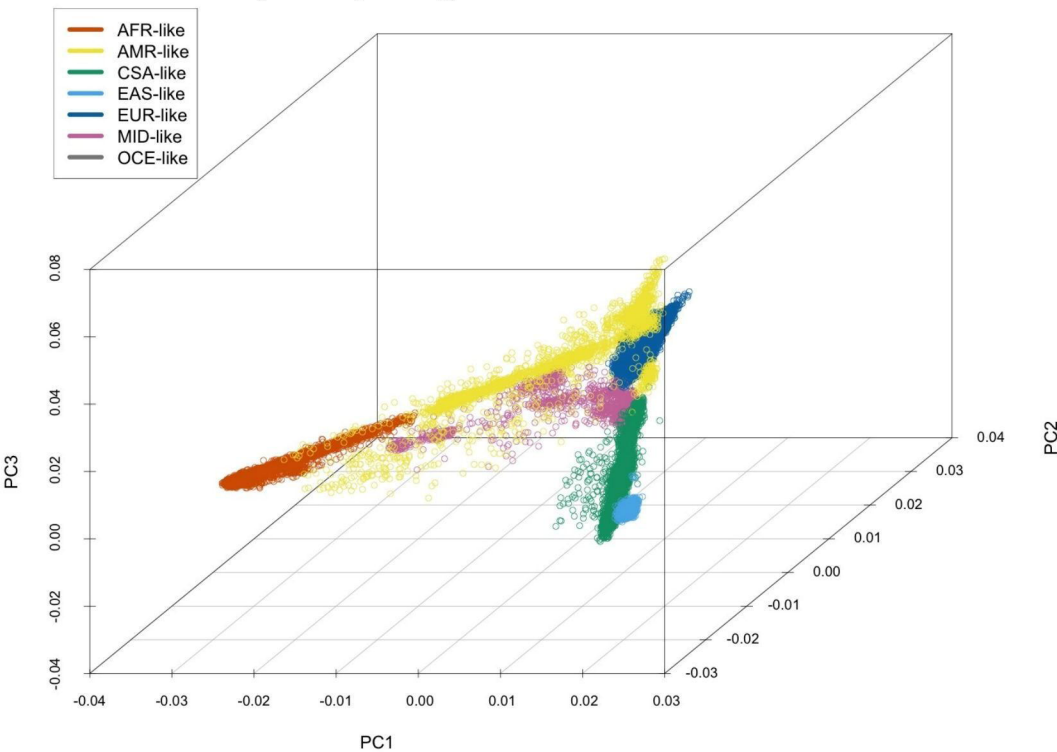
- doi:10.1038/s41380-019-0370-z
48. Hill MD, Gill SS, Le-Niculescu H, et al. Precision medicine for psychotic disorders: objective assessment, risk prediction, and pharmacogenomics. *Mol Psychiatry*. 2024;29(5):1528-1549. doi:10.1038/s41380-024-02433-8
 49. Sugawara K, Shibasaki T, Takahashi H, Seino S. Structure and functional roles of Epac2 (Rapgef4). *Gene*. 2016;575(2, Part 3):577-583. doi:10.1016/j.gene.2015.09.029
 50. Xie Z, Penzes P, Srivastava DP. Exchange protein directly activated by cAMP 2 is required for corticotropin-releasing hormone-mediated spine loss. *Eur J Neurosci*. 2019;50(7):3108-3114. doi:10.1111/ejn.14487
 51. Zhou L, Ma SL, Yeung PKK, et al. Anxiety and depression with neurogenesis defects in exchange protein directly activated by cAMP 2-deficient mice are ameliorated by a selective serotonin reuptake inhibitor, Prozac. *Transl Psychiatry*. 2016;6(9):e881-e881. doi:10.1038/tp.2016.129
 52. Lozano R, Gbokie C, Siper PM, et al. FOXP1 syndrome: a review of the literature and practice parameters for medical assessment and monitoring. *J Neurodev Disord*. 2021;13(1):18. doi:10.1186/s11689-021-09358-1
 53. Morello F, Borshagovski D, Survila M, et al. Molecular Fingerprint and Developmental Regulation of the Tegmental GABAergic and Glutamatergic Neurons Derived from the Anterior Hindbrain. *Cell Rep*. 2020;33(2):108268. doi:10.1016/j.celrep.2020.108268
 54. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415-1424. doi:10.1038/s41588-021-00931-x
 55. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019;28(1):166-174. doi:10.1093/hmg/ddy327
 56. Martin S, Cule M, Bastly N, et al. Genetic Evidence for Different Adiposity Phenotypes and Their Opposing Influences on Ectopic Fat and Risk of Cardiometabolic Disease. *Diabetes*. 2021;70(8):1843-1856. doi:10.2337/db21-0129
 57. Okbay A, Wu Y, Wang N, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet*. 2022;54(4):437-449. doi:10.1038/s41588-022-01016-z
 58. Saunders GRB, Wang X, Chen F, et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. 2022;612(7941):720-724. doi:10.1038/s41586-022-05477-4
 59. Karlsson Linnér R, Mallard TT, Barr PB, et al. Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction. *Nat Neurosci*. 2021;24(10):1367-1376. doi:10.1038/s41593-021-00908-3
 60. Dalvie S, Maihofer AX, Coleman JRI, et al. Genomic influences on self-reported childhood maltreatment. *Transl Psychiatry*. 2020;10(1):38. doi:10.1038/s41398-020-0706-0
 61. Watanabe K, Jansen PR, Savage JE, et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat Genet*. 2022;54(8):1125-1132. doi:10.1038/s41588-022-01124-w
 62. Zhao B, Luo T, Li T, et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet*. 2019;51(11):1637-1644. doi:10.1038/s41588-019-0516-6
 63. van der Meer D, Kaufmann T, Shadrin AA, et al. The genetic architecture of human

- cortical folding. *Sci Adv.* 2021;7(51):eabj9446. doi:10.1126/sciadv.abj9446
64. Adams MJ, Hill WD, Howard DM, et al. Factors associated with sharing e-mail information and mental health survey participation in large population cohorts. *Int J Epidemiol.* 2020;49(2):410-421. doi:10.1093/ije/dyz134
 65. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol.* 2017;186(9):1026-1034. doi:10.1093/aje/kwx246
 66. Schoeler T, Speed D, Porcu E, Pirastu N, Pingault JB, Kutalik Z. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav.* 2023;7(7):1216-1227. doi:10.1038/s41562-023-01579-9
 67. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet.* 2006;38(6):659-662. doi:10.1038/ng1801
 68. Ruan Y, Lin YF, Feng YCA, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet.* 2022;54(5):573-580. doi:10.1038/s41588-022-01054-7
 69. Long E, García-Closas M, Chanock SJ, Camargo MC, Banovich NE, Choi J. The case for increasing diversity in tissue-based functional genomics datasets to understand human disease susceptibility. *Nat Commun.* 2022;13(1):2907. doi:10.1038/s41467-022-30650-8

A. Study Schematic



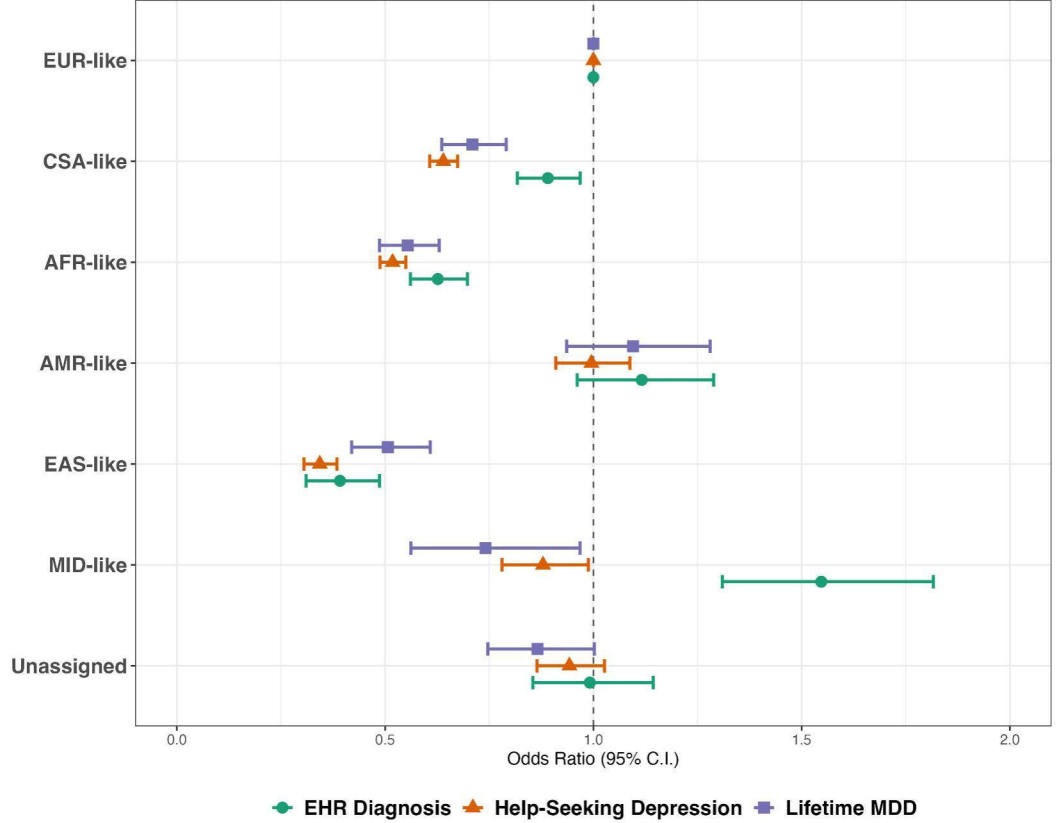
B. Genetic Ancestry Group Assignment in the UK Biobank



C. Assigned Genetic Ancestry Group Distribution

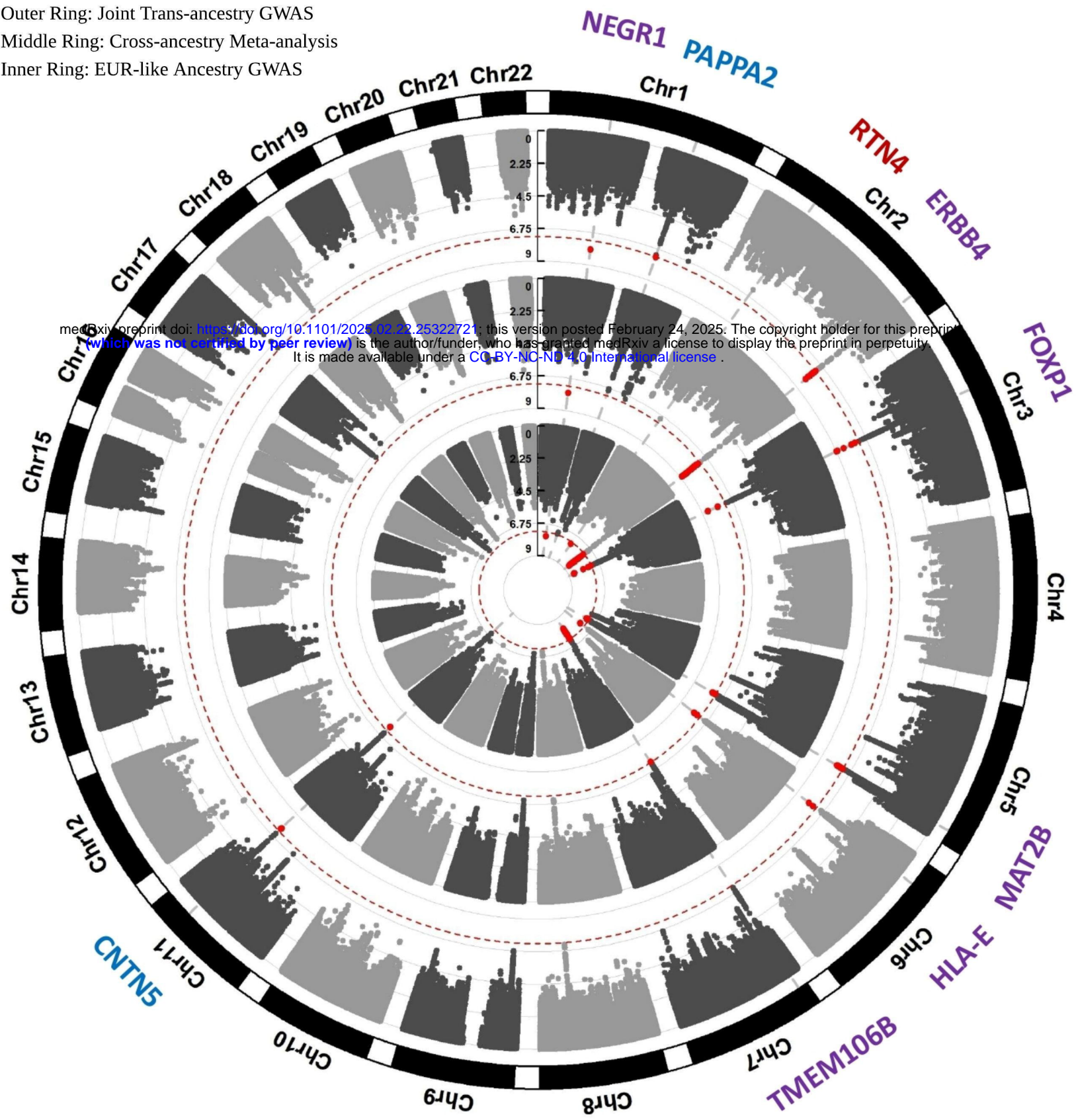


D. Relative Odds of Depression Across Assigned Genetic Ancestry Groups

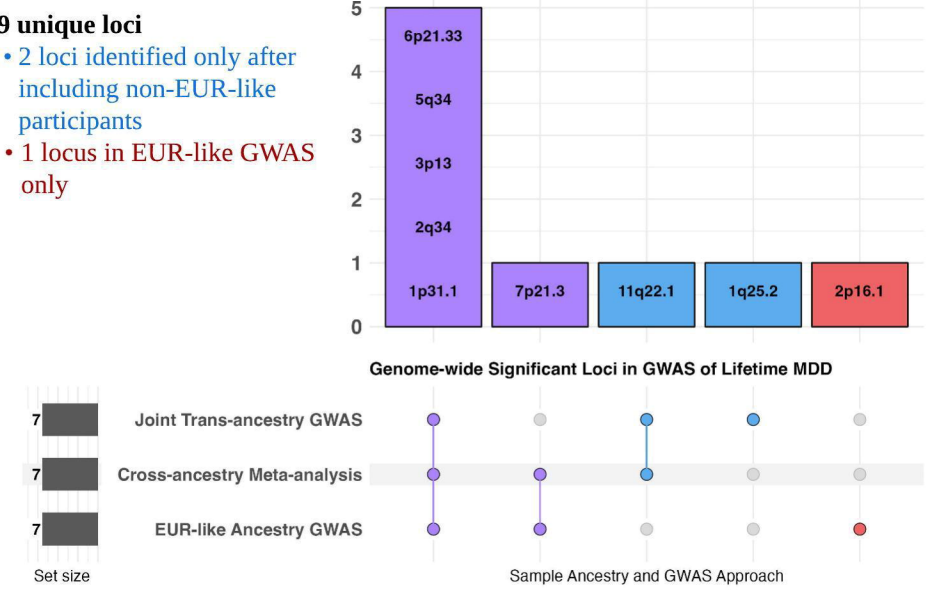


A. Circular Manhattan Plots of GWAS of Lifetime MDD

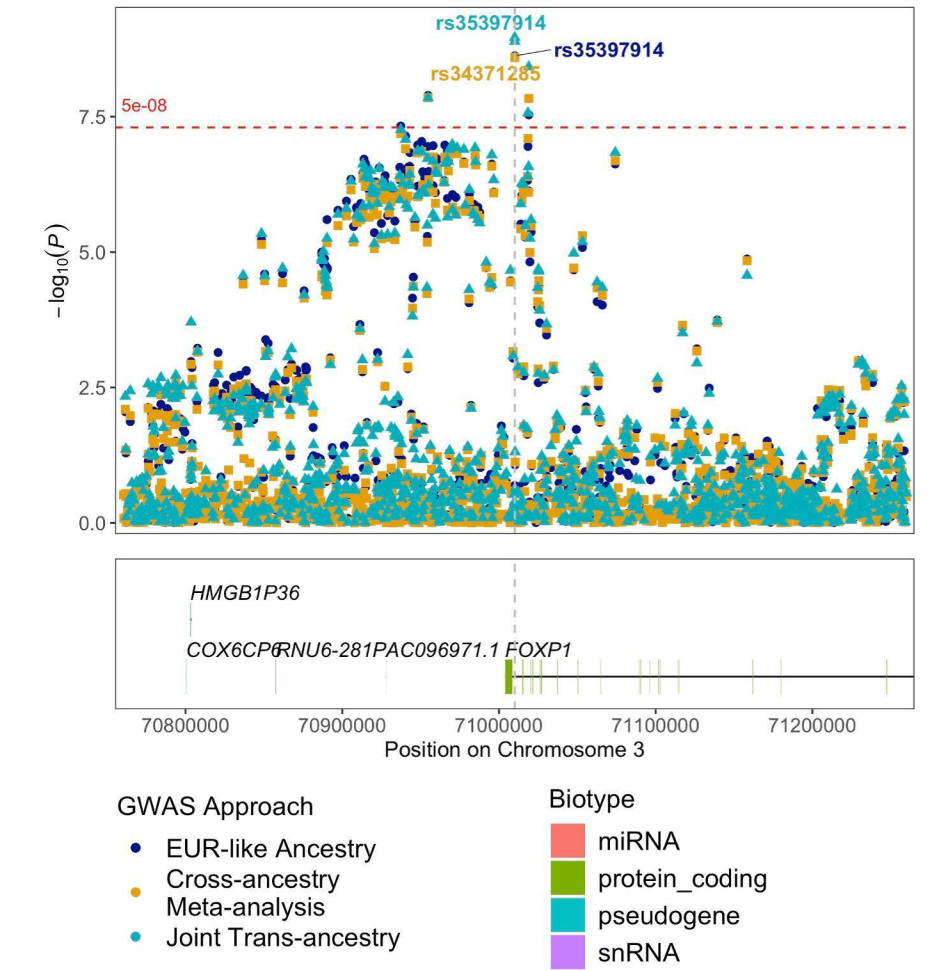
Outer Ring: Joint Trans-ancestry GWAS
Middle Ring: Cross-ancestry Meta-analysis
Inner Ring: EUR-like Ancestry GWAS



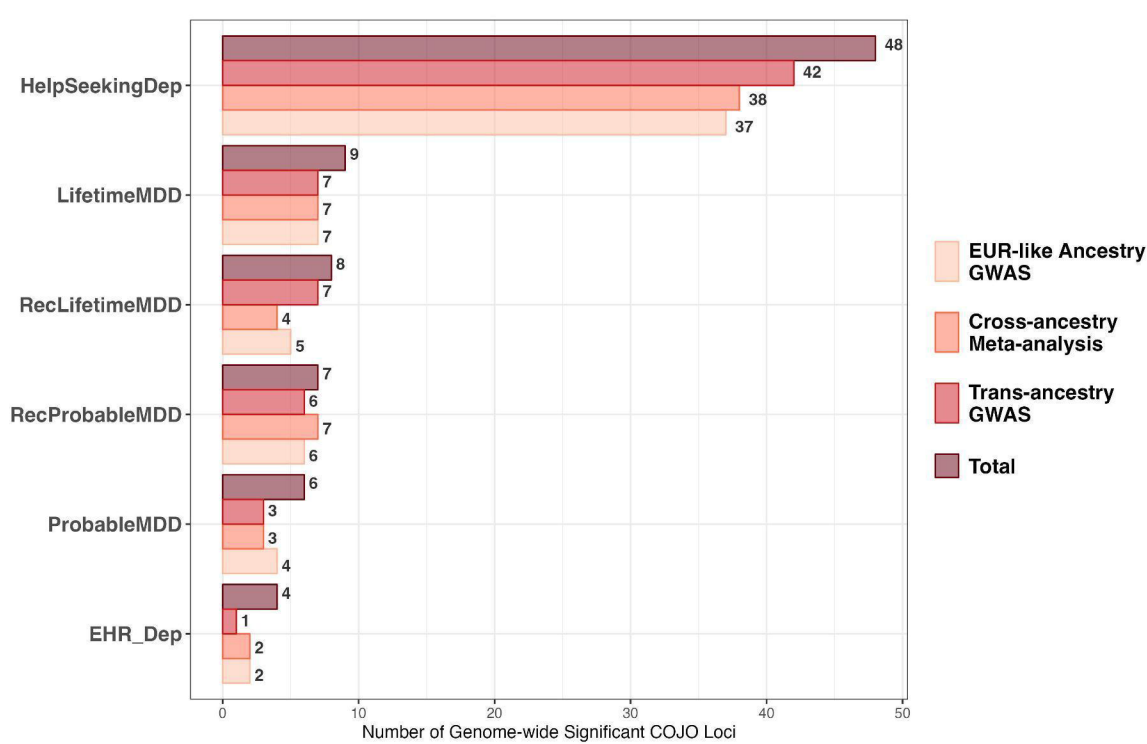
B. Genomic Loci Associated with Lifetime MDD Across Approaches



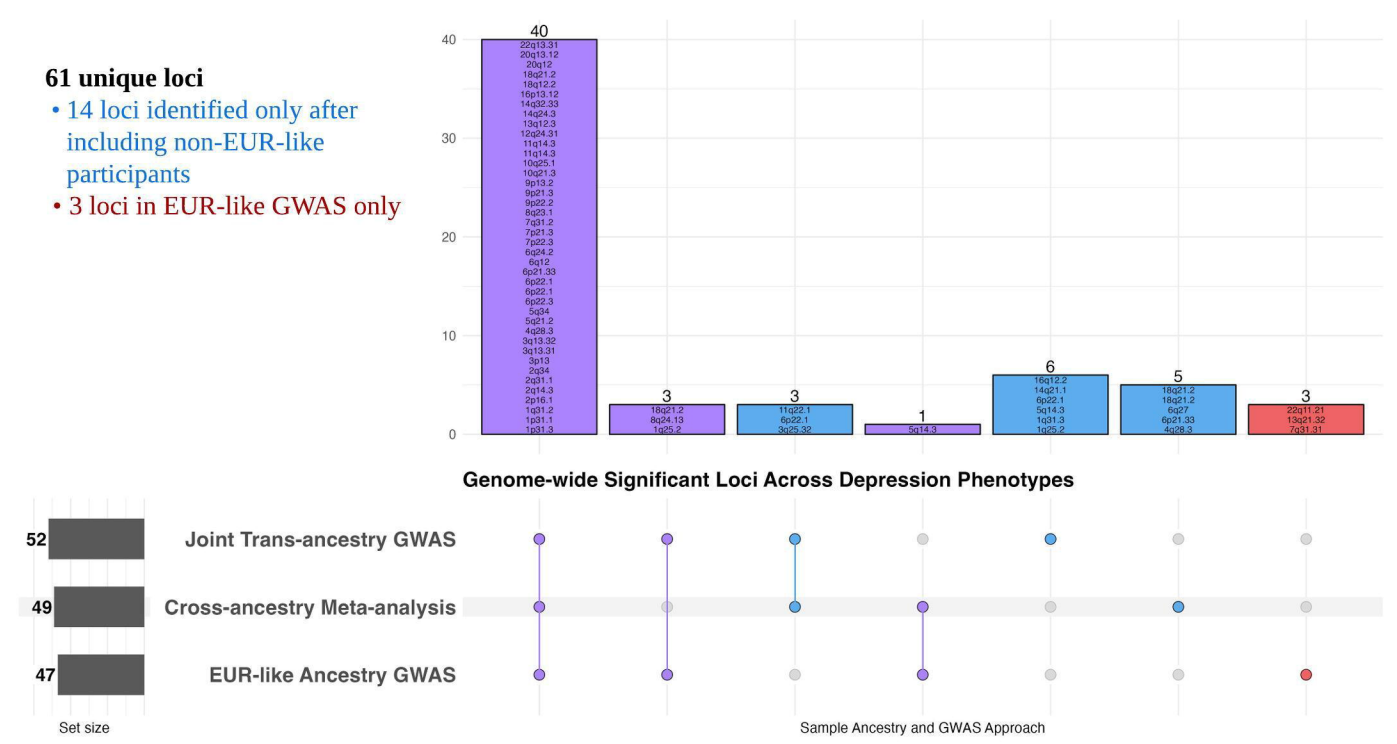
C. Regional Plot of *FOXP1* gene with SNPs associated with LifetimeMDD



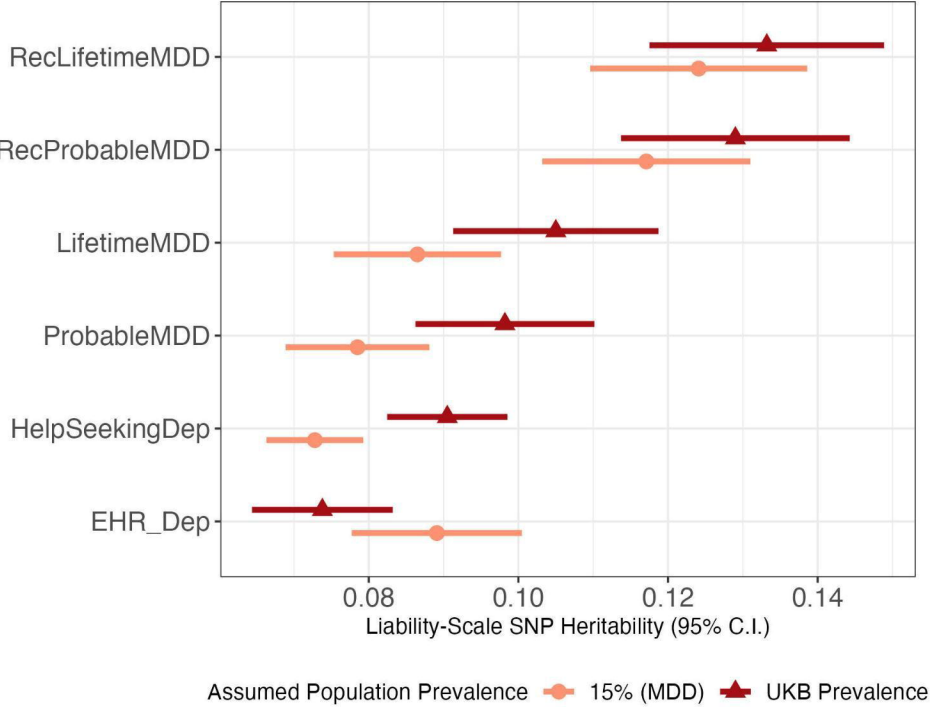
D. Genomic Loci Associated with Depression Phenotypes per GWAS Approach



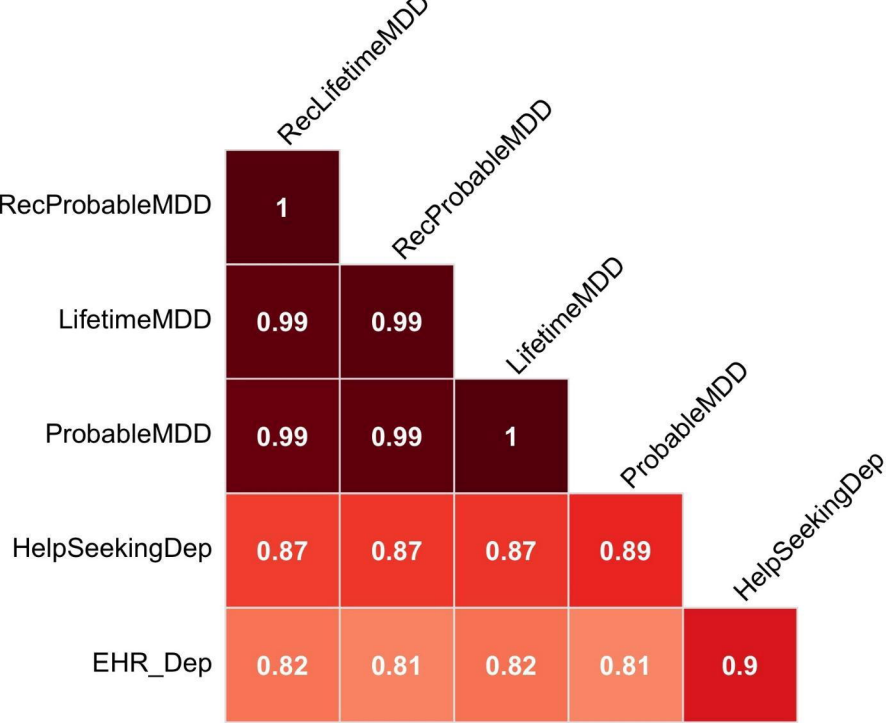
E. Genomic Loci Associated Across Depression Phenotypes and GWAS Approaches



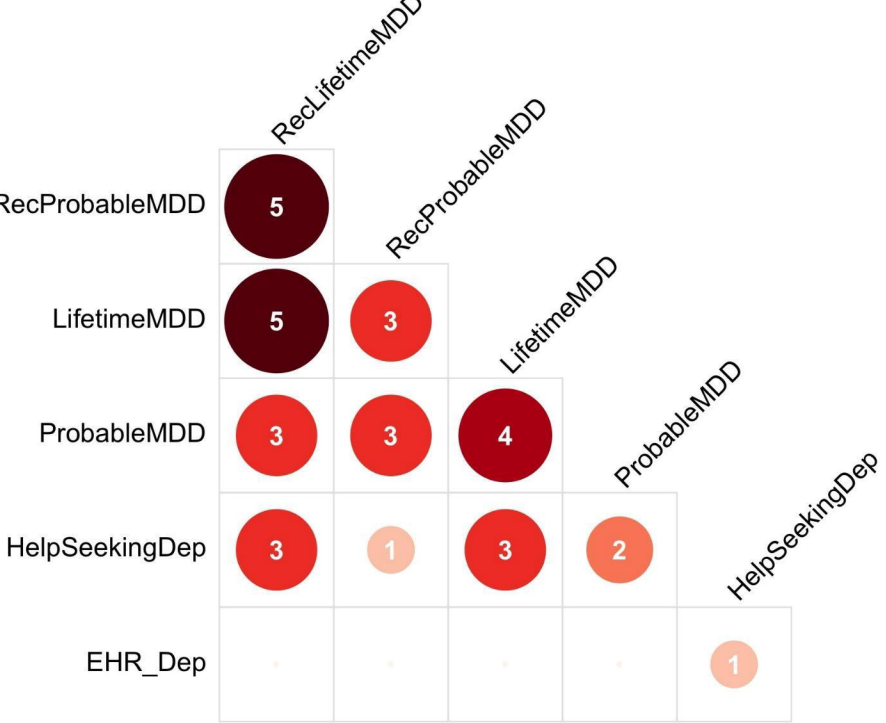
A. SNP-Heritability of Depression Phenotypes in EUR-like Ancestry Group



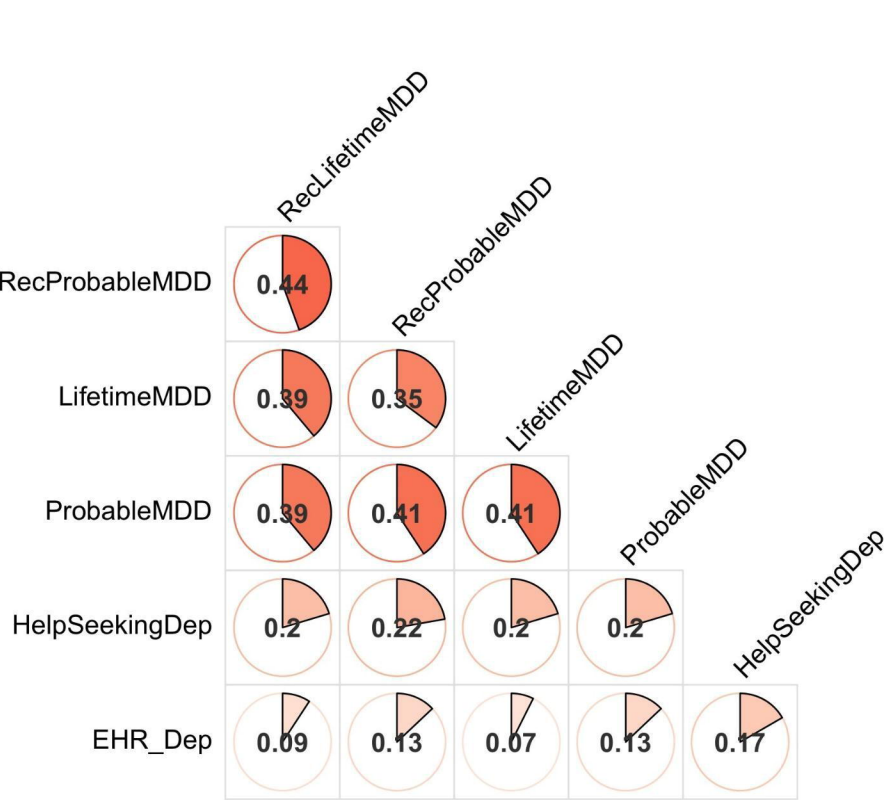
B. Genetic Correlations Between Depression Phenotypes in EUR-like Ancestry Group



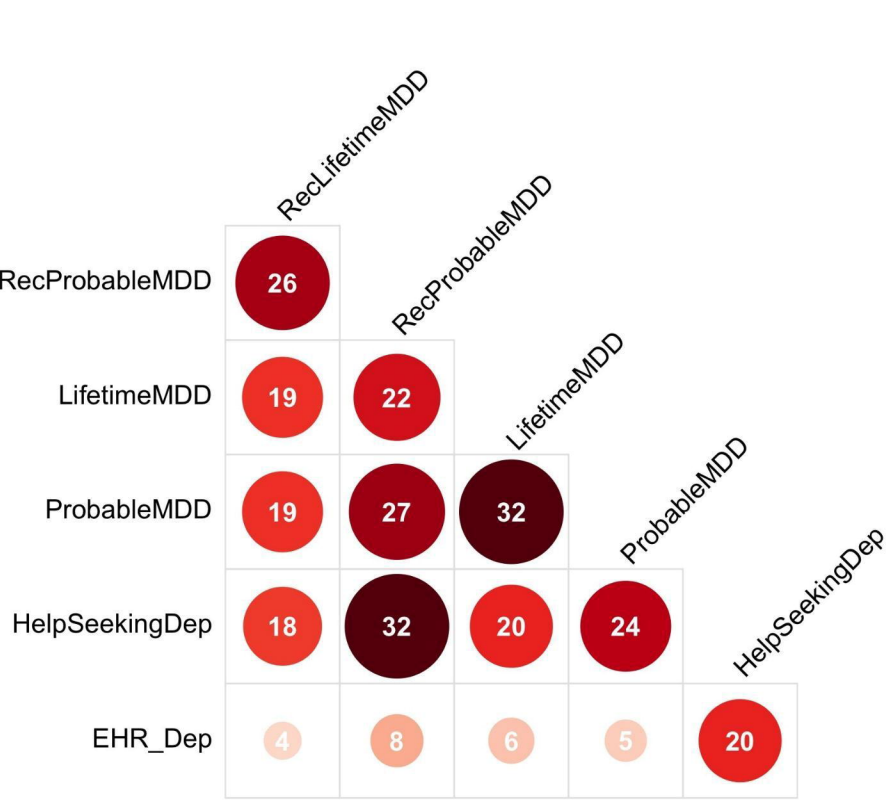
C. Number of Shared Genome-wide Significant Loci Between Phenotypes



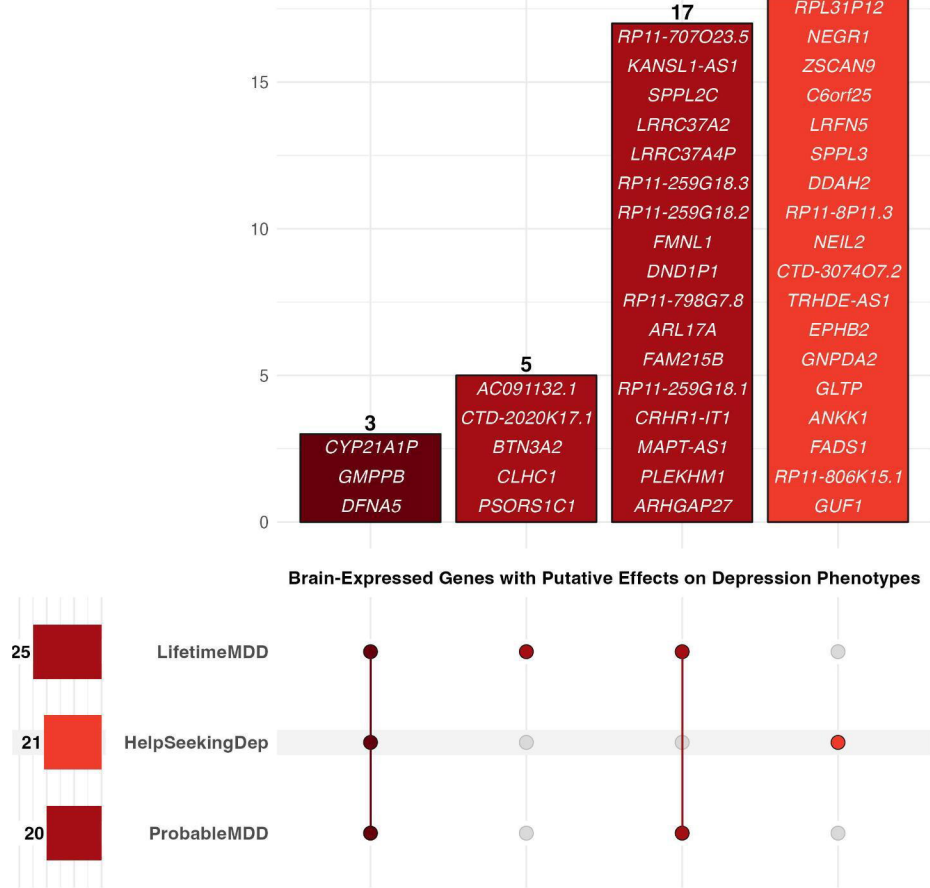
D. Proportion of Credible Sets of SNPs with Colocalization Between Phenotypes in EUR-like Ancestry Group



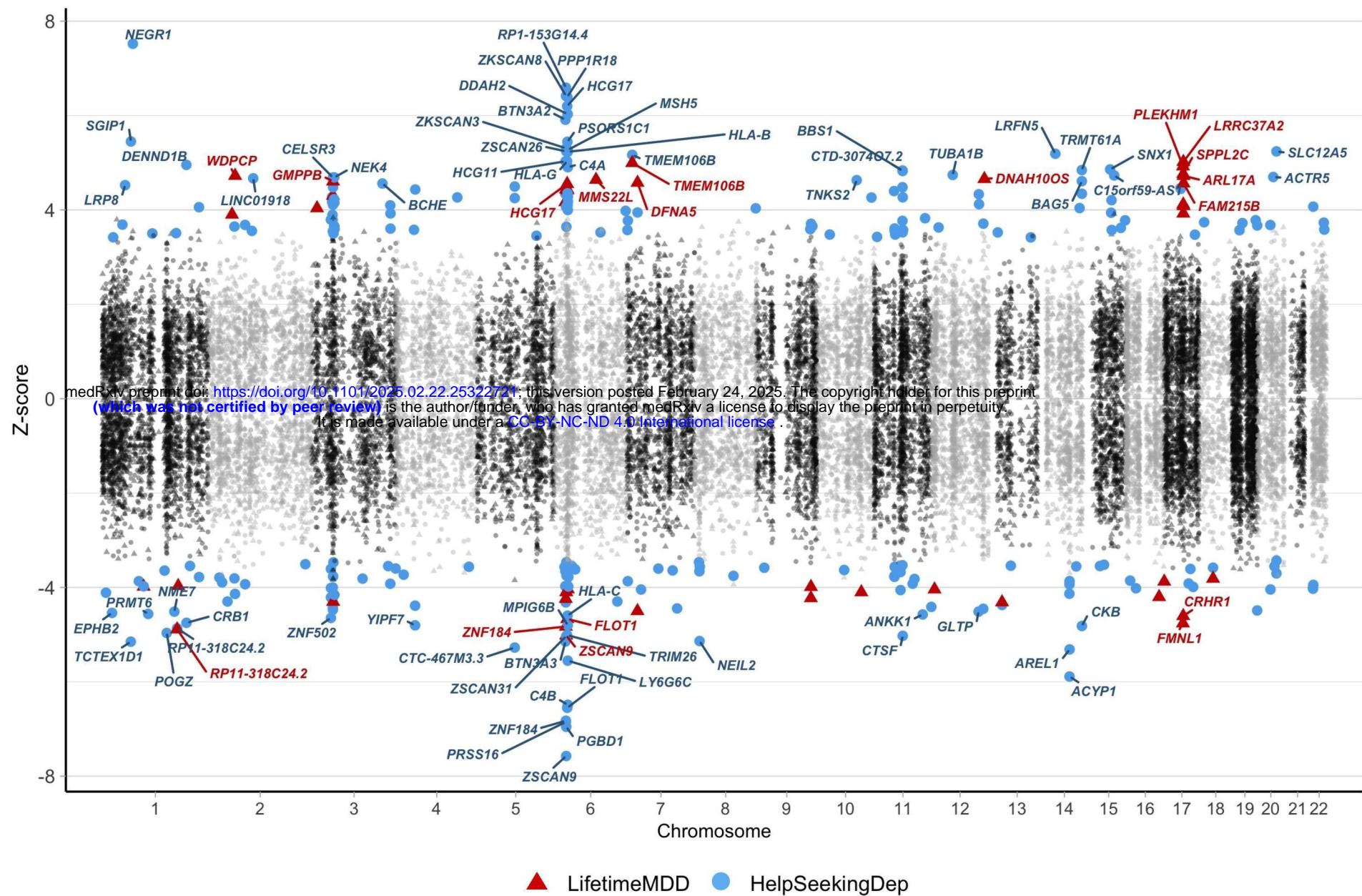
E. Number of Shared Brain-Expressed Genes Associated with Phenotypes in TWAS in EUR-like Ancestry Group



F. Overlap of Brain-Expressed Genes with Putative Effects on Depression in SMR Analyses in EUR-like Ancestry Group



A. TWAS of **Lifetime MDD** and **Help-Seeking Depression**



B. Polygenic Prediction of Depression Diagnosis in the NIH *All of Us* Research Program (AoU)

