

1 **Supplementary Materials**

2 Khan et al., Combining Transdiagnostic and Disorder-Level GWAS Enhances Precision of
3 Psychiatric Genetic Risk Profiles in a Multi-Ancestry Sample

4

5 **Supplementary Materials 3**

6 *MTAG*.....3

7 *Procedures for Summary Statistics in GenomicSEM*.....3

8 *African Ancestry Reference Panels*.....4

9 *LD Clumping & Identification of Novel Lead SNPs*6

10 *SNP-Level PheWAS*.....7

11 **Supplementary Figures..... 8**

12 *Supplementary Figure 1. Common and independent pathway models to identify factor specific Q_{SNPs}*8

13 *Supplementary Figure 2. Genetic correlations of input GWAS in European ancestry individuals*.....9

14 *Supplementary Figure 3. Genetic correlations of input GWAS in African ancestry individuals*10

15 *Supplementary Figure 4. PheWAS plots of novel SNPs for the mood disorders common factor*11

16 *Supplementary Figure 5. Regional annotation plot for rs75174029, a novel SNP identified by the European*
17 *ancestry mood/anxiety disorders GWAS.*12

18 *Supplementary Figure 6. Regional annotation plot for rs7652704, a novel SNP identified by the European*
19 *ancestry mood/anxiety disorders GWAS.*13

20 *Supplementary Figure 7. Results of MAGMA tissue expression analysis of EUR substance use disorders factor*
21 *.....*14

22 *Supplementary Figure 8. Results of MAGMA tissue expression analysis of EUR psychotic disorders factor*...15

23 *Supplementary Figure 9. Results of MAGMA tissue expression analysis of EUR mood disorders factor*16

24 *Supplementary Figure 10. Manhattan plot for substance use disorders factor in AFR ancestry individuals*17

25 *Supplementary Figure 11. Results of MAGMA tissue expression analysis of AFR ancestry substance use*
26 *disorders factor*18

27 *Supplementary Figure 12. Manhattan plot for psychiatric disorders factor in AFR ancestry individuals*.....19

28 *Supplementary Figure 13. Results of MAGMA tissue expression analysis of AFR ancestry psychiatric*
29 *disorders factor*20

30 *Supplementary Figure 14. EUR ancestry second order common factor model*.....21

31 *Supplementary Figure 15. Manhattan plot for second-order common factor representing overlap between*
32 *substance use and psychotic disorders in EUR ancestry individuals*22

33 *Supplementary Figure 16. Manhattan plot for second-order common factor representing overlap between*
34 *substance use and mood/anxiety disorders in EUR ancestry individuals*23

35 *Supplementary Figure 17. Results of MAGMA tissue expression analysis of EUR ancestry second-order*
36 *substance use and psychotic disorders factor*.....24

37	<i>Supplementary Figure 18. Results of MAGMA tissue expression analysis of EUR ancestry second-order</i>	
38	<i>substance use and mood disorders factor</i>	25
39	<i>Supplementary Figure 19. AFR ancestry second order common factor model</i>	26
40	<i>Supplementary Figure 20. Manhattan plot for second-order common factor representing overlap between</i>	
41	<i>substance use and psychiatric disorders in AFR ancestry individuals</i>	27
42	<i>Supplementary Figure 21. Results of MAGMA tissue expression analysis of AFR ancestry second-order</i>	
43	<i>substance use and psychiatric disorders factor</i>	28
44	<i>Supplementary Figure 22. Significant SNPs identified in TUD Independent GWAS</i>	33
45	<i>Supplementary Figure 23. Significant SNPs identified in SCZ Independent GWAS</i>	38
46	<i>Supplementary Figure 24. Significant SNPs identified in BD Independent GWAS</i>	43
47	<i>Supplementary Figure 25. Protein-protein interaction network plot for TUD Independent.</i>	44
48	<i>Supplementary Figure 26. Protein-protein interaction network plot for SCZ Independent.</i>	45
49	<i>Supplementary Figure 27. Protein-protein interaction network plot for BD Independent.</i>	46
50	<i>Supplementary Figure 28. Genetic correlation results for the EUR substance use disorders factor</i>	47
51	<i>Supplementary Figure 29. Genetic correlation results for the EUR psychotic disorders factor</i>	48
52	<i>Supplementary Figure 30. Genetic correlation results for the EUR mood disorders factor</i>	49
53	<i>Supplementary Figure 31. Genetic correlations between AFR ancestry common factors and psychiatric and</i>	
54	<i>substance use phenotypes</i>	50
55	<i>Supplementary Figure 32. Genetic correlations between the AFR ancestry second-order common factor and</i>	
56	<i>psychiatric and substance use traits</i>	51
57	<i>Supplementary Figure 33. PheWAS results for the EUR substance use disorders factor in Penn Medicine</i>	
58	<i>BioBank</i>	52
59	<i>Supplementary Figure 34. PheWAS results for the EUR psychotic disorders factor in Penn Medicine BioBank</i>	
60	53
61	<i>Supplementary Figure 35. PheWAS results for the EUR mood disorders factor in Penn Medicine BioBank</i> ..	54
62	<i>Supplementary Figure 36. PheWAS results for EUR ancestry second-order common factor representing</i>	
63	<i>overlap in substance use and psychotic disorders in Penn Medicine BioBank</i>	55
64	<i>Supplementary Figure 37. PheWAS results for EUR ancestry second-order common factor representing</i>	
65	<i>overlap in substance use and mood/anxiety disorders in Penn Medicine BioBank</i>	56
66	<i>Supplementary Figure 38. Hudson plot of PheWAS results for tobacco use disorders GWAS-by-subtraction in</i>	
67	<i>Penn Medicine BioBank</i>	57
68	<i>Supplementary Figure 39. Hudson plot of PheWAS results for schizophrenia GWAS-by-subtraction in Penn</i>	
69	<i>Medicine BioBank</i>	58
70	<i>Supplementary Figure 40. Hudson plot of PheWAS results for bipolar disorder GWAS-by-subtraction in Penn</i>	
71	<i>Medicine BioBank</i>	59
72	<i>Supplementary Figure 41. PheWAS results for AFR ancestry substance use disorders factor in Penn Medicine</i>	
73	<i>BioBank</i>	60
74	<i>Supplementary Figure 42. PheWAS results for AFR ancestry psychiatric disorders factor in Penn Medicine</i>	
75	<i>BioBank</i>	61
76	<i>Supplementary Figure 43. PheWAS results for AFR ancestry second-order common factor representing</i>	
77	<i>overlap in substance use and psychiatric disorders in Penn Medicine BioBank</i>	62

78

79

Supplementary Materials

80

81 **MTAG**

82

83 Using Multi-Trait Analysis of GWAS (MTAG),¹ we leveraged the genetic effects from a
84 study of Lifetime Anxiety Disorder² and a study of GAD-2 questionnaire scores³ to enhance the
85 statistical power of a GWAS for a broad spectrum of anxiety disorders⁴ in European ancestry
86 individuals. We chose to enhance the power of the summary statistics from Otowa, et al., 2016,
87 because they included the most diverse array of anxiety disorders among the three anxiety
88 GWAS. This choice was supported by the strong genetic correlations between the generalized
89 anxiety GWAS⁴ and both GWAS of lifetime anxiety disorder ($r_g = 0.7429$) and GAD-2 scores (r_g
90 $= 0.7309$)

91

92 Effective sample sizes were calculated as the sum of $4/(1/n_{\text{case}} + 1/n_{\text{control}})$ for each
93 cohort in each of the two case-control GWAS. For the GAD-2 score GWAS, the total sample
94 size was used as the input for MTAG because GAD-2 is a continuous trait. As quality control
95 measures included in the MTAG software, SNPs with $MAF < 0.01$ were excluded from analysis,
96 along with duplicate SNPs and those with missing values. Following MTAG analysis, the
97 effective sample size for follow-up analyses was calculated using the formula described by
98 Turley, et al., 2018.¹

99

100 **Procedures for Summary Statistics in GenomicSEM**

101

102 All summary statistics and analyses were conducted on the NCBI hg19/GRCh37 genome
103 assembly. For traits with continuous outcomes (i.e., GAD-2 score), the total sample size was
104 used for LDSC and GenomicSEM computations. For traits with a binary outcome (i.e., case-
105 control), the effective sample size column contained within the GWAS summary statistics was
106 used. When no effective sample size column was present, effective sample size was calculated
107 for each set of summary statistics using the formula described by Grotzinger, et al, 2023:⁵

108

$$N_{eff} = \sum_{k=1}^N 4 * v_k * (1 - v_k) * n_k$$

109

110 Where v and n are the sample prevalence and sample total, respectively, for the k^{th} cohort
111 of a GWAS meta-analysis of N cohorts. Summary statistics were then prepared for GWAS using
112 the following options in GenomicSEM: The “se.logit” flag was set to “TRUE” when the standard
113 error column reflected the standard error of a logistic beta, the “OLS” flag was set to “TRUE”
114 when the phenotype reflected a continuous outcome, and the “linprob” flag was set to “TRUE”
115 when the phenotype was of a binary outcome but with only Z-statistics present as a measure of
116 effect in the GWAS summary statistics. SNPs were then filtered based on $MAF > 0.01$ and 0.6.
117 Following preparation of summary statistics, 2,083,079 SNPs remained for analysis in the
118 European-ancestry subset, and 6,350,709 SNPs remained for analysis in the African-ancestry
119 subset.

119

120 **African Ancestry Reference Panels**

121 To determine the optimal linkage disequilibrium (LD) score reference panel for use in the
122 African ancestry gSEM models, we compared three sets of references: (1) 1000 Genomes Phase
123 3, (2) PanUKB, and (3) Million Veteran Program (MVP). We used publicly available 1000
124 Genomes¹¹ and PanUKB¹² LD scores. MVP LD scores were generated from 1000 randomly
125 selected African ancestry MVP participants using covariate-adjusted LD score regression (cov-
126 LDSC),¹³ which is a method that has shown improved performance among admixed populations,
127 such as African Americans. As recommended to further account for population stratification, the
128 first ten ancestry-specific principal components (PCs) were computed within the sample and
129 included as covariates when generating LD scores.

130 To ensure the accuracy of and prevent bias in estimates derived from the LD scores, we
131 restricted LD score regression (LDSC) analyses to well-imputed, biallelic autosomal SNPs that
132 are outside of the MHC region. The set of SNPs meeting this criteria varied for each LD
133 reference panel. For the 1000 Genomes Phase 3 panel, we used the list of 1,217,312 HapMap3
134 SNPs provided in the reference files prepared by Finucane et al. (2015)¹⁴ for LDSC. For the
135 PanUKB reference, we retained all 1,190,983 SNPs, as LD scores were computed only for SNPs
136 that met the aforementioned criteria and passed additional quality control, including having
137 imputation quality (R^2) > 0.90 and minor allele frequency > 0.01 (see [https://pan-
138 dev.ukbb.broadinstitute.org/docs/ld/index.html](https://pan-dev.ukbb.broadinstitute.org/docs/ld/index.html)). For the MVP reference, we restricted our
139 analyses to SNPs that met the same criteria as those used by the Broad Institute to prepare the
140 PanUKB reference files. Thus, a total of 2,388 SNPs were removed due to low MAF, and 8,707
141 were removed due to low imputation quality, leaving 1,516,281 SNPs in MVP.

142 In comparing the performance of the three sets of reference panels, we evaluated: (1) the
143 number of SNPs retained following filtering and munging the input summary statistics, (2)
144 liability scale SNP-based heritability, (3) confounding evidenced by inflated values on the LDSC
145 intercept, and (4) the length and distribution of resulting LD blocks. Results are presented below:
146
147

1000G reference, 1000G SNPlist				
trait	# snps	heritability	SE	intercept
AUD	423441	0.0806	0.0133	1.0304
TUD	691706	0.0445	0.008	1.0257
ODU	250024	0.0668	0.0229	1.0236
CanUD	604363	0.0616	0.0116	1.0306
GAD2	869992	0.0282	0.0365	1.0076
MDD	869312	0.0415	0.0188	1.0184
SCZ	891719	0.1204	0.0294	1.0587
BIP	891833	0.1417	0.0642	1.0344

148
149

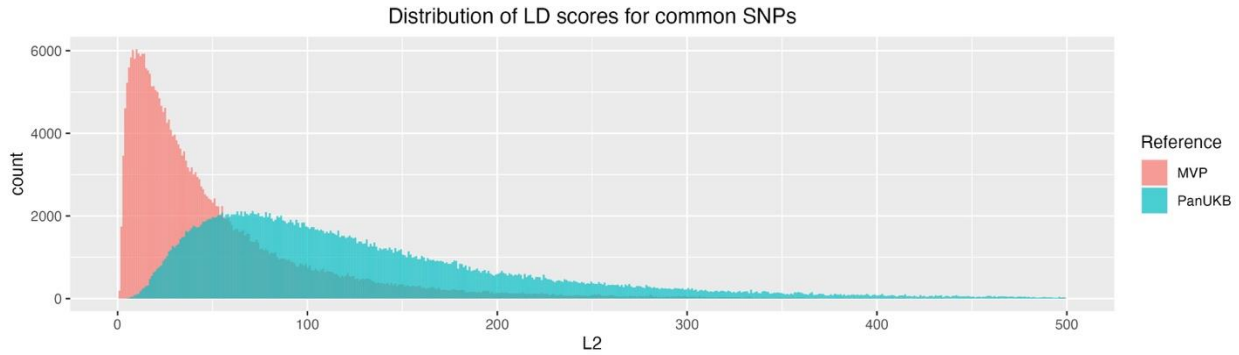
150
151

MVP reference, MVP SNPlist				
trait	# snps	heritability	SE	intercept
AUD	1508956	0.0427	0.0062	1.0615
TUD	1515026	0.0434	0.0065	1.0631
ODD	1512615	0.0218	0.0092	1.0347
CanUD	1507988	0.021	0.0056	1.0464
GAD2	1476669	0.0391	0.0217	1.0056
MDD	1480787	1.00E-03	0.0082	1.0297
SCZ	1512163	0.0496	0.0155	1.0649
BIP	1511435	0.0554	0.0328	1.0324

UKBB reference, Pan-UKBB SNPlist				
trait	# snps	heritability	SE	intercept
AUD	613531	0.0885	0.0153	1.0352
TUD	979504	0.064	0.0083	1.0235
ODD	429964	0.0662	0.0206	1.034
CanUD	897996	0.068	0.0116	1.0289
GAD2	1152886	0.0619	0.0394	1.0026
MDD	1144306	0.0397	0.017	1.0193
SCZ	1184566	0.1661	0.0278	1.05
BIP	1183486	0.2238	0.0631	1.0254

152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167

Using the 1000 Genomes LD reference panel and SNP list resulted in the fewest number of SNPs remaining after performing LDSC on the input summary statistics, including as few as 250,024 SNPs for ODD. As LDSC accuracy decreases as the number of SNPs decreases,¹⁵ we chose not to progress with the 1000 Genomes reference panels due to the potential for unreliable genetic correlations upon which gSEM models are based. On the other hand, the reference panels generated in MVP resulted in the largest number of remaining SNPs but tended to produce lower heritability estimates than the other reference panels, including a non-significant heritability estimate for MDD. MVP also consistently had the highest inflation in test statistics based on the LDSC intercept. Finally, examining the distribution of the LD scores, MVP LD scores were consistently lower than those using PanUKB. As PanUKB reference panels resulted in an adequate number of SNPs available for analyses, produced significant heritability estimates, showed low inflation in test statistics, and had a broader distribution of LD scores compared to MVP (see below), we conducted African ancestry analyses using PanUKB references.



168

Comparison across ALL SNPs in both sets		
Statistic	MVP L2	PanUKB L2
Min	0.202	2.564
25 th percentile	24.679	53.958
Median	48.104	86.065
Mean	77.451	110.042
75 th percentile	92.463	136.970
Max	3672.834	3193.00

169

170

171 LD Clumping & Identification of Novel Lead SNPs

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

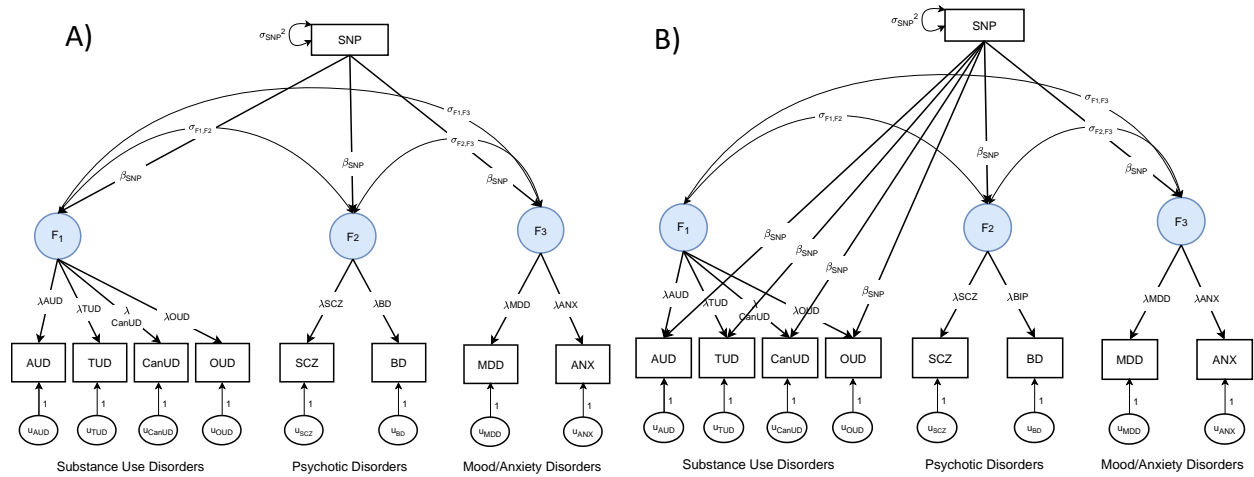
Following common factor GWAS and GWAS-by-subtraction, LD clumping of summary statistics results was performed using PLINK 1.9¹⁶ with ancestry-matched 1000 Genomes Phase 3 (for European) or PanUKB (for African) reference panels, a significance threshold of 5×10^{-8} for index SNPs, r^2 threshold of 0.10, and physical distance threshold of 3000kb. For common factor GWAS, SNPs were considered not have been identified by any input GWAS if they were not located within $\pm 1000\text{kb}$ of any lead SNP from any input study for the corresponding common factor. Lead SNPs from input studies were obtained from the supplementary materials for each input GWAS.

To determine if a lead SNP from common factor GWAS had previously been associated with any of the input traits by any previous study, a review of GWAS Catalog¹⁷ was conducted. First, common factor GWAS lead SNP chromosome and base-pair information was lifted over from NCBI assembly hg19/GRCh37 to hg38/GRCh38 using the UCSC Genome Browser's LiftOver tool.¹⁸ Then, for each lead SNP, a query of GWAS Catalog was conducted of all GWAS reporting significant SNPs in the range of $\pm 1000\text{kb}$ of the lead SNP's position. The list of trait associations was subsequently reviewed for any terms corresponding to any input traits for the common factor GWAS. If there were no matches, then the SNP was considered novel in that it had not been previously associated with any previous GWAS of the input traits for a common factor at the time the search was conducted.

193 **SNP-Level PheWAS**

194 For any novel SNPs that were identified in GWAS, we performed a SNP-level PheWAS
195 using GWAS Atlas.¹⁹ Analyses examined 4,756 publicly available GWASs and used a
196 Bonferroni corrected p-value of 1.05×10^{-5} to identify significant associations.

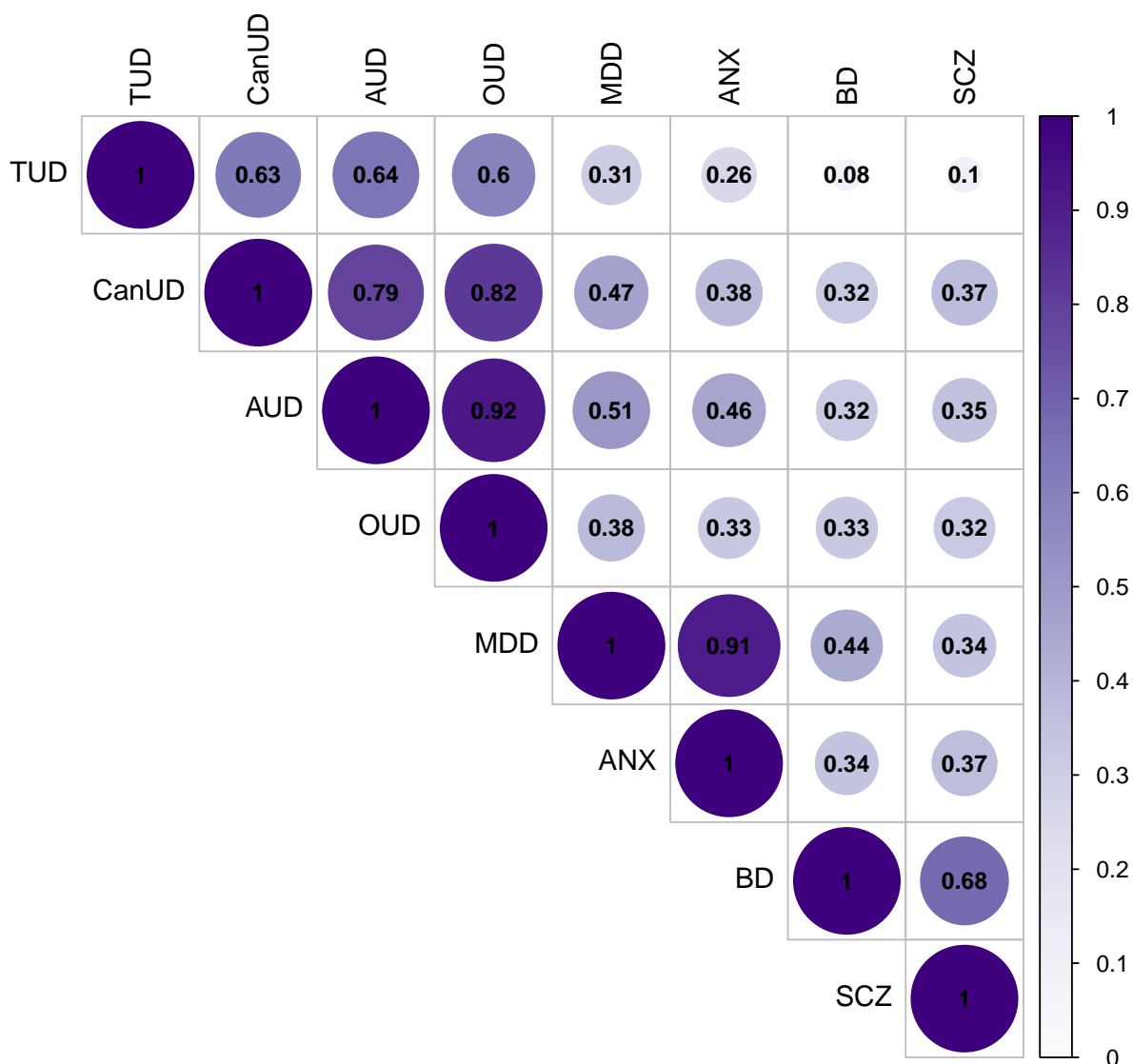
Supplementary Figures



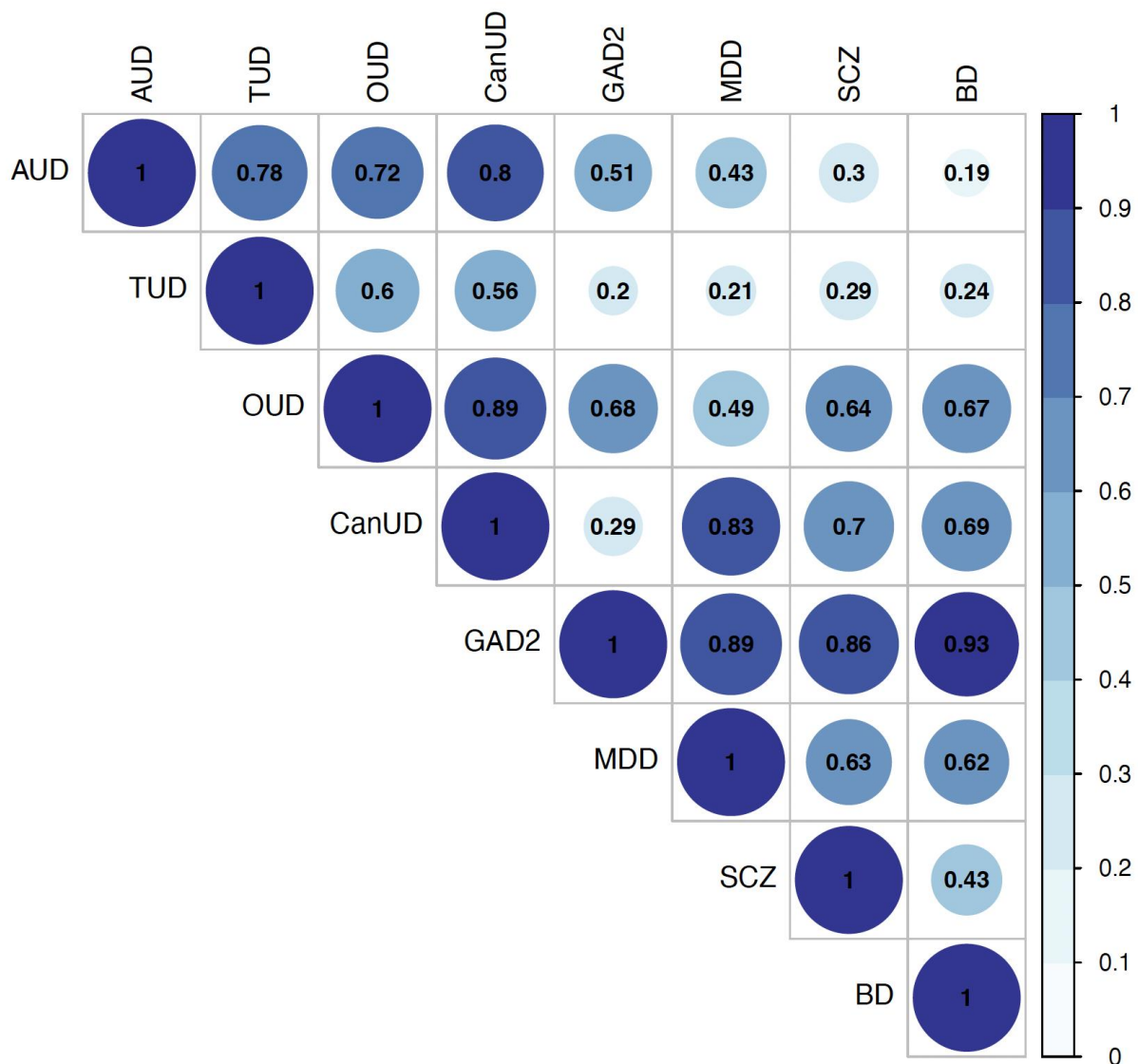
199

Supplementary Figure 1. Common and independent pathway models to identify factor specific Q_{SNPs}

202 Panel A depicts the common pathway model where a given SNP's effects operate through the
 203 factors. Panel B depicts the independent pathway model for Factor 1. In this model, each SNP
 204 predicts the indicators of Factor 1, as well as the other two factors. A χ^2 difference test was
 205 performed for the two models to determine if the SNP's effects could be explained by its
 206 association with the factor or, instead, by its association with specific indicators. Follow-up
 207 independent pathway models (as shown in Panel B) were run for the each of the other two first-
 208 order factors to identify their factor-specific Q_{SNPs} . An analogous approach was applied for the
 209 second-order factors and for African ancestry models. SNPs whose χ^2 p-value was $< 5 \cdot 10^{-8}$ were
 210 removed from summary statistics prior to performing downstream analyses.



211
 212 **Supplementary Figure 2. Genetic correlations of input GWAS in European ancestry**
 213 **individuals**
 214 AUD = alcohol use disorder, CanUD = cannabis use disorder, TUD = tobacco use disorder,
 215 OUD = opioid use disorder, MDD = major depressive disorder, BD = bipolar disorder, ANX =
 216 anxiety disorders, SCZ = schizophrenia. Traits are ordered based on hierarchical clustering.
 217

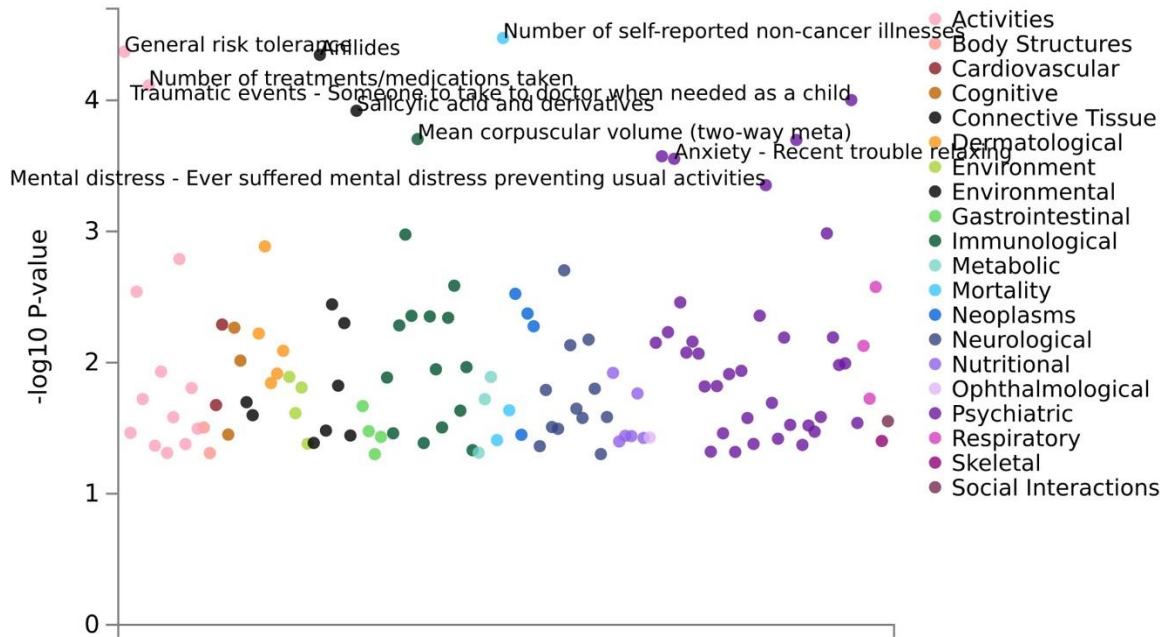


218

219 **Supplementary Figure 3. Genetic correlations of input GWAS in African ancestry**
 220 **individuals**

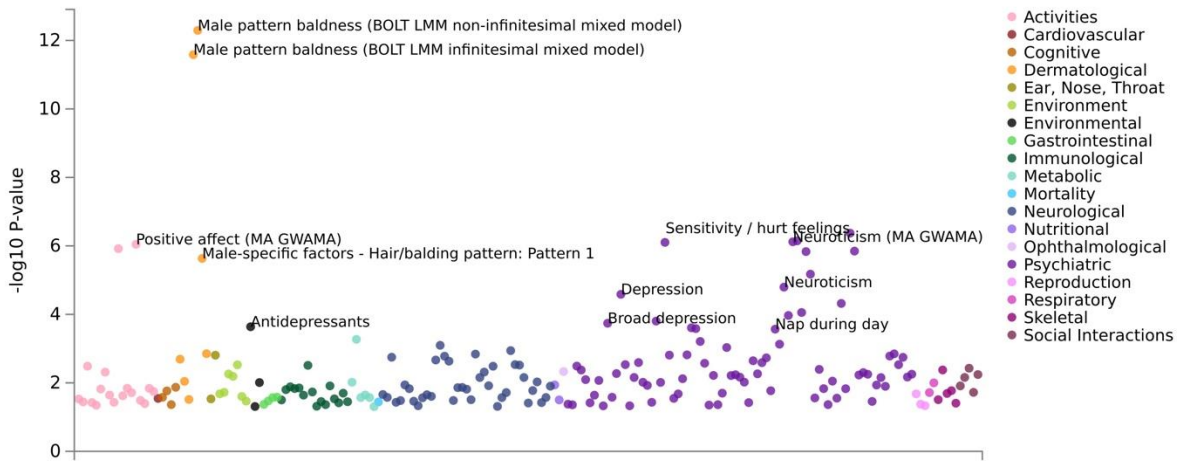
221 MDD = major depressive disorder, BD = bipolar disorder, GAD-2 = Generalized Anxiety
 222 Disorder-2 scores, SCZ = schizophrenia, AUD = alcohol use disorder, TUD = tobacco use
 223 disorder, CanUD = cannabis use disorder, OUD = opioid use disorder. Traits are ordered based
 224 on hierarchical clustering.

rs75174029



225
226

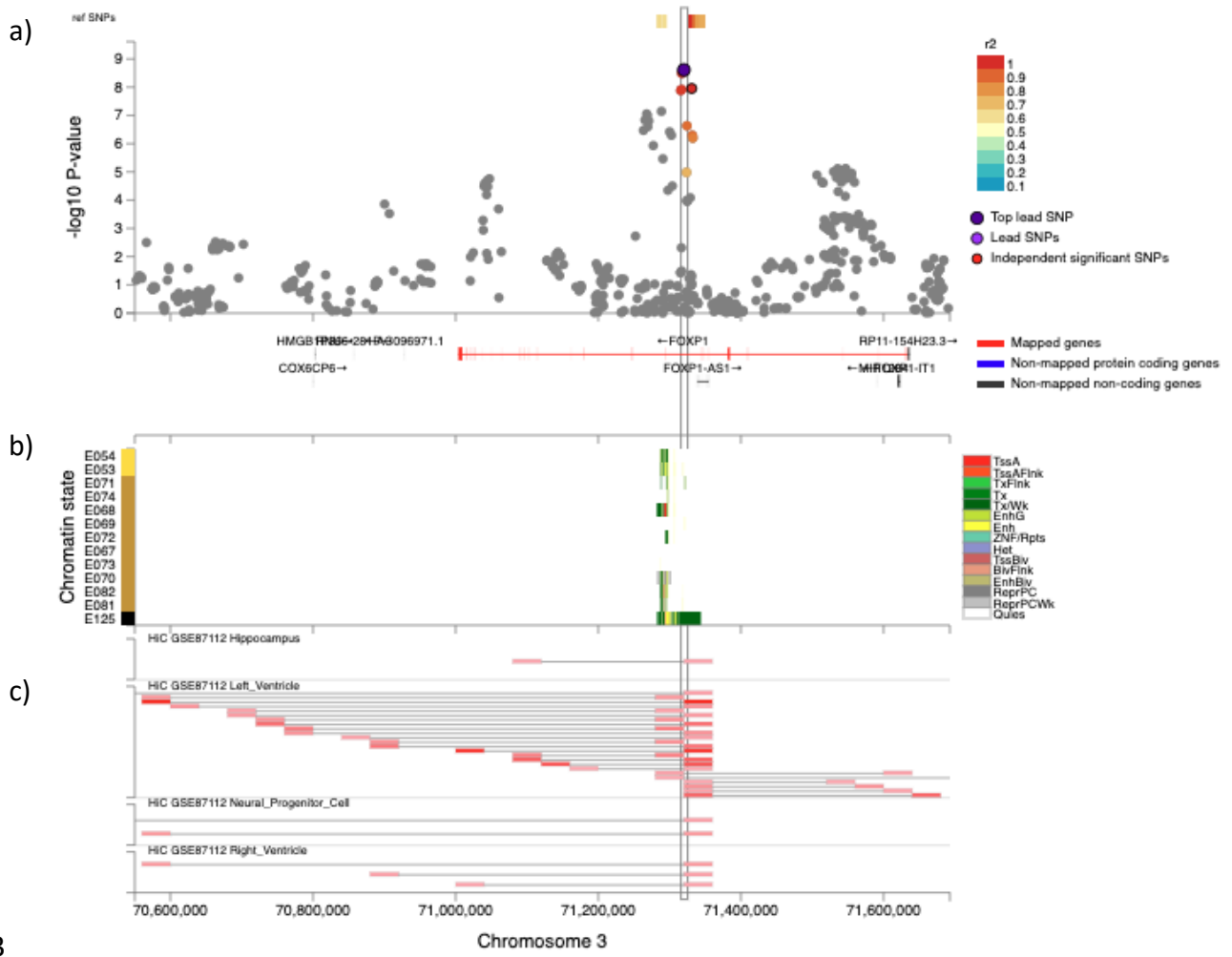
rs7652704



227
228
229
230
231
232

Supplementary Figure 4. PheWAS plots of novel SNPs for the mood disorders common factor

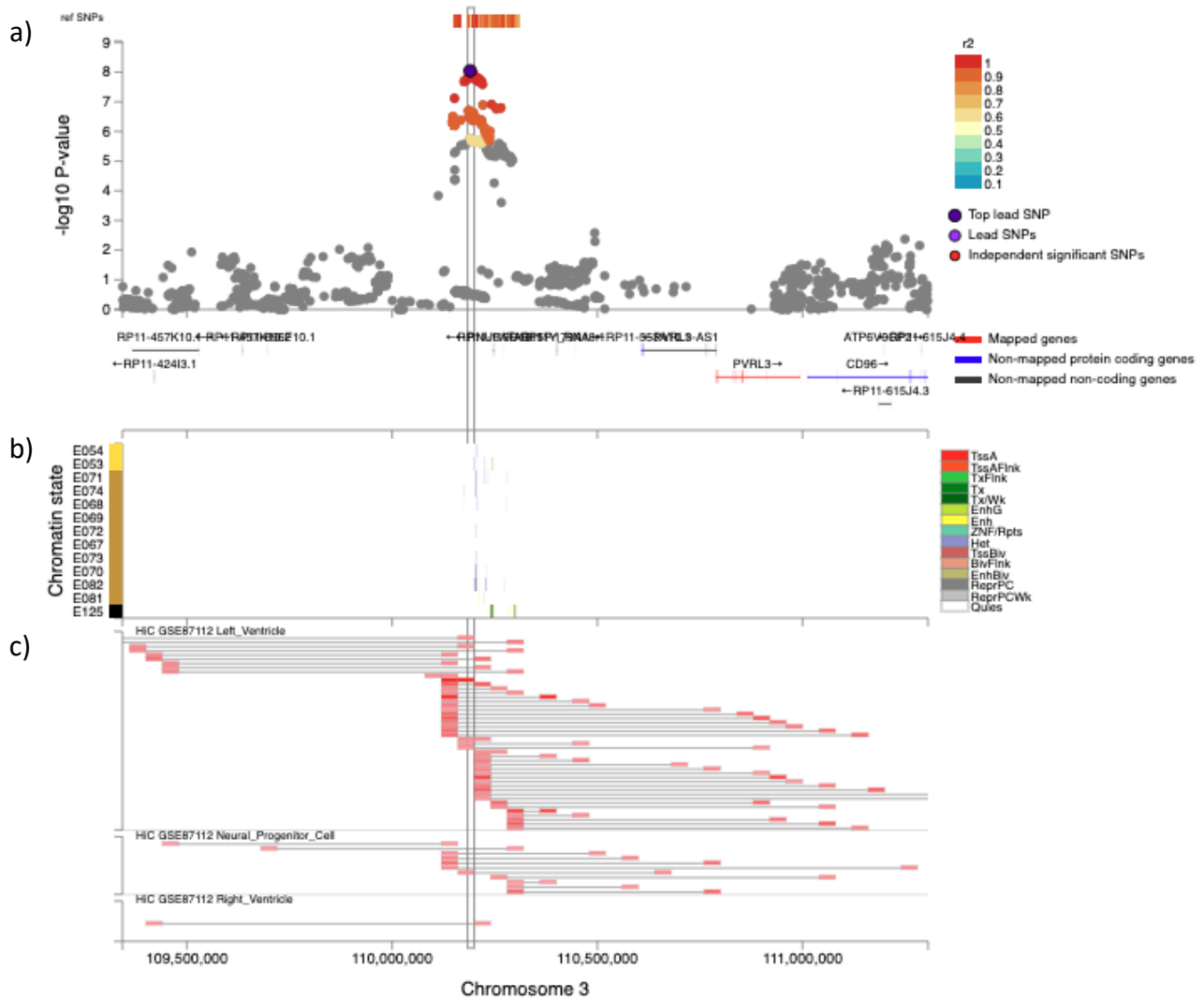
PheWAS plots were produced using GWAS Atlas.



233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249

Supplementary Figure 5. Regional annotation plot for rs75174029, a novel SNP identified by the European ancestry mood/anxiety disorders GWAS.

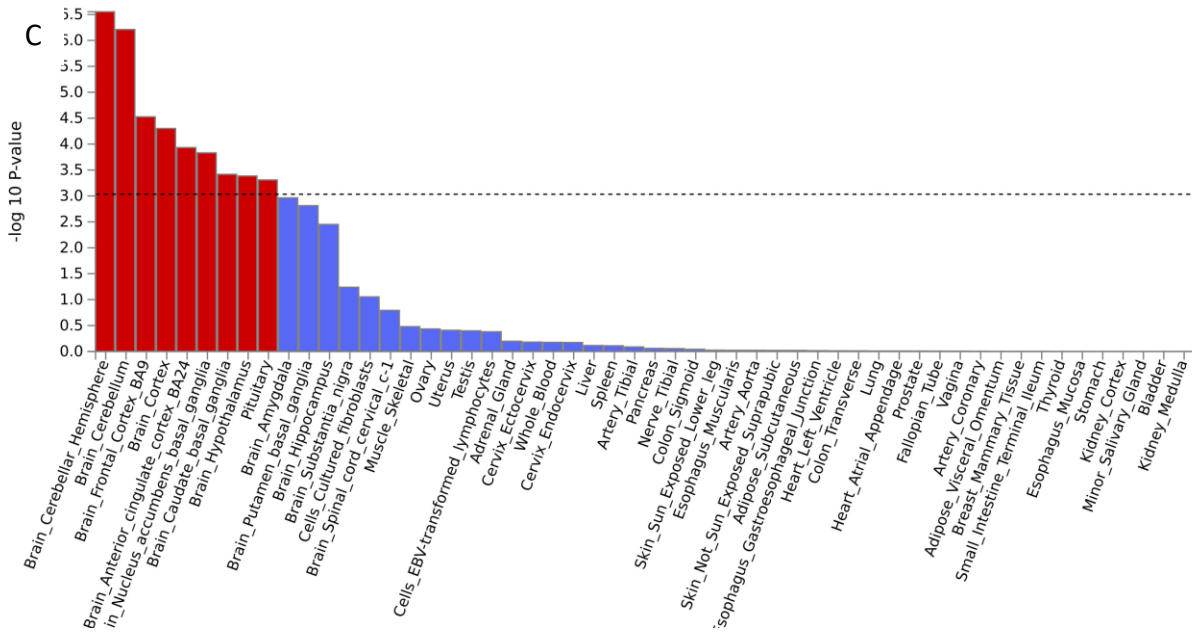
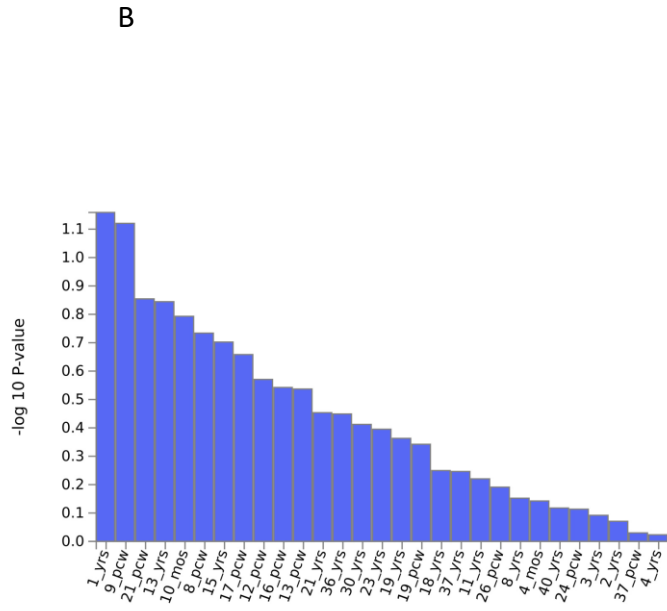
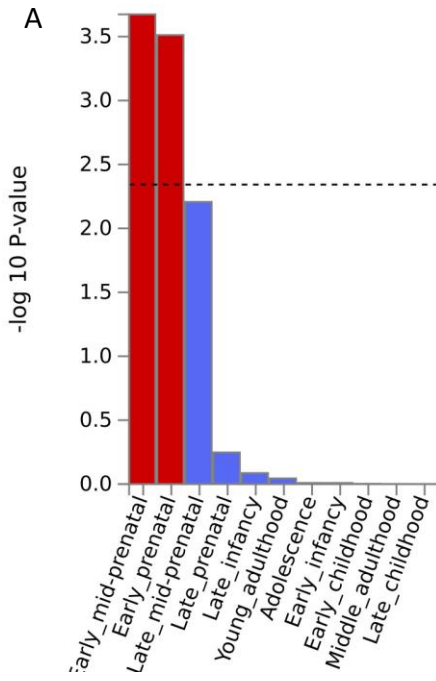
(a) rs75174029 (in purple), its linked SNPs, and their position relative to genes. rs75174029's predicted genomic target *FOXP1* is shown in red. (b) Colocalization of rs75174029 with ROADMAP 15 core chromatin states (right-hand key) in 15 brain tissues (left hand key). E054 = ganglion eminence-derived neurospheres, E053 = cortex-derived neurospheres, E071 = hippocampus, E074 = substantia nigra, E068 = anterior caudate, E069 = cingulate gyrus, E072 = inferior temporal lobe, E067 = angular gyrus, E073 = dorsolateral prefrontal cortex, E070 = germinal matrix, E082 = female fetal brain, E081 = fetal male brain, E125 = NH-A astrocytes. TssA = Active Transcription Start Site, TsAFlnk = flanking active TSS, TxFlnk = transcribed at gene 5' and 3', Tx = strong transcription, TxWk = weak transcription, EnhG = genic enhancers, Enh = enhancers, ZNF/Rpts = ZNF genes and repeats, Het = heterochromatin, TssBiv = bivalent/poised TSS, BivFlnk = Flanking bivalent TSS/Enh, EnhBiv = bivalent enhancer, ReprPC = repressed PolyComb, PreprPCWk = weak repressed PolyComb, Quies = quiescent/low. (c) Colocalization with Hi-C signal in brain tissues. Each line represents an interaction, with the two red regions representing the loci which make contact.



250

251 **Supplementary Figure 6. Regional annotation plot for rs7652704, a novel SNP identified by**
 252 **the European ancestry mood/anxiety disorders GWAS.**

253 (a) rs7652704 (in purple), its linked SNPs, and their position relative to genes. rs7652704's
 254 predicted genomic target *PVRL3* (*NECTIN3*) is shown in red. (b) Colocalization of
 255 rs7652704 with ROADMAP 15 core chromatin states (right-hand key) in 15 brain tissues (left
 256 hand key). E054 = ganglion eminence-derived neurospheres, E053 = cortex-derived
 257 neurospheres, E071 = hippocampus, E074 = substantia nigra, E068 = anterior caudate, E069 =
 258 cingulate gyrus, E072 = inferior temporal lobe, E067 = angular gyrus, E073 = dorsolateral
 259 prefrontal cortex, E070 = germinal matrix, E082 = female fetal brain, E081 = fetal male brain,
 260 E125 = NH-A astrocytes. TssA = Active Transcription Start Site, TssAFlnk = flanking active
 261 TSS, TxFlnk = transcribed at gene 5' and 3', Tx = strong transcription, TxWk = weak
 262 transcription, EnhG = genic enhancers, Enh = enhancers, ZNF/Rpts = ZNF genes and repeats,
 263 Het = heterochromatin, TssBiv = bivalent/poised TSS, BivFlnk = Flanking bivalent TSS/Enh,
 264 EnhBiv = bivalent enhancer, ReprPC = repressed PolyComb, PreprPCWk = weak repressed
 265 PolyComb, Quies = quiescent/low. (c) Colocalization with Hi-C signal in brain tissues. Each line
 266 represents an interaction, with the two red regions representing the loci which make contact.



267

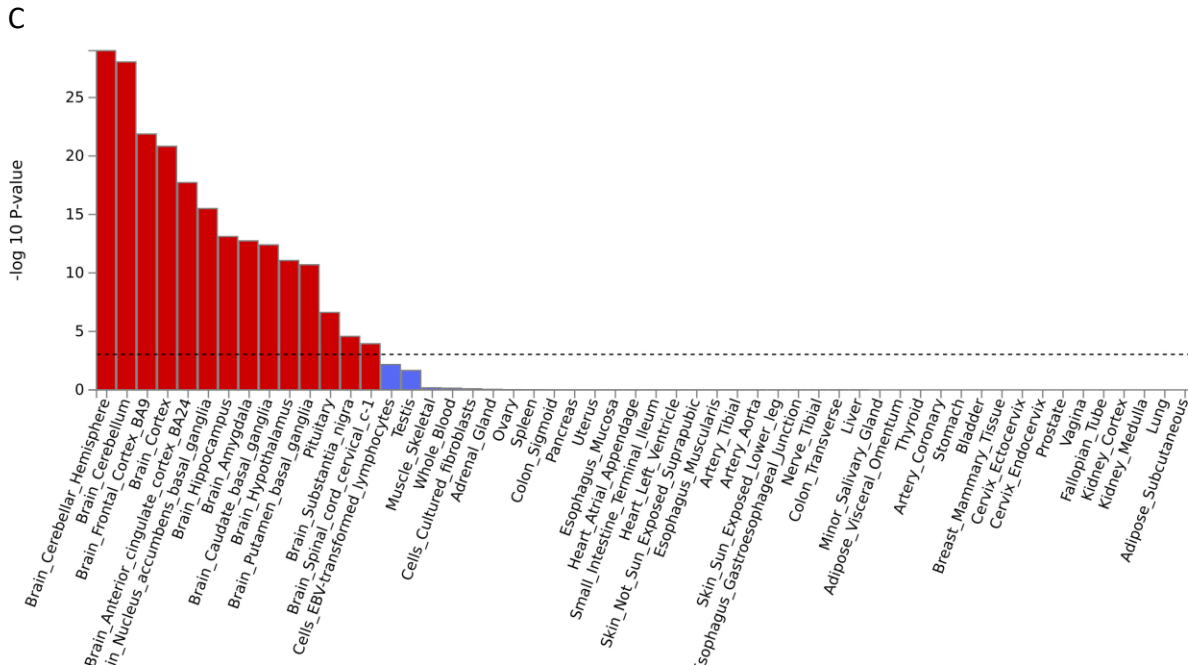
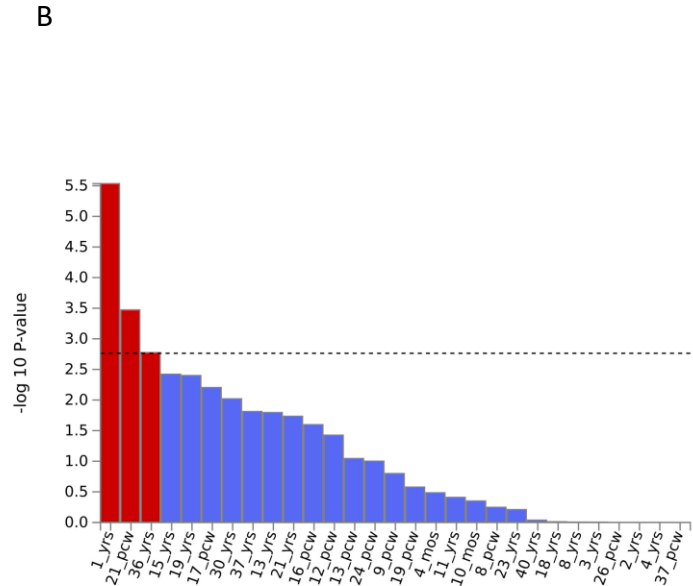
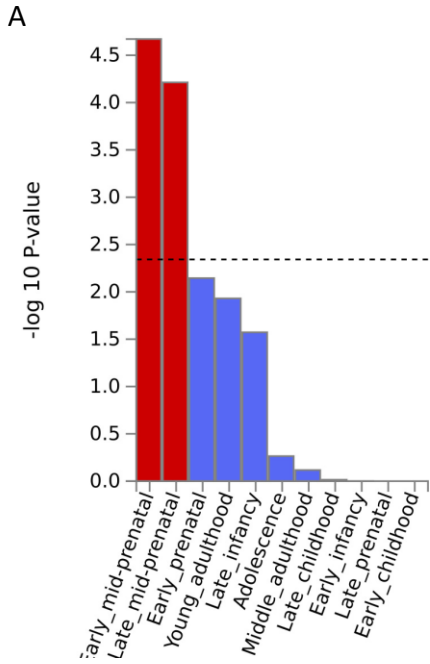
268

269

270 **Supplementary Figure 7. Results of MAGMA tissue expression analysis of EUR substance**
 271 **use disorders factor**

272 Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are
 273 shown in Panel C. Dashed line represents significance threshold.

274



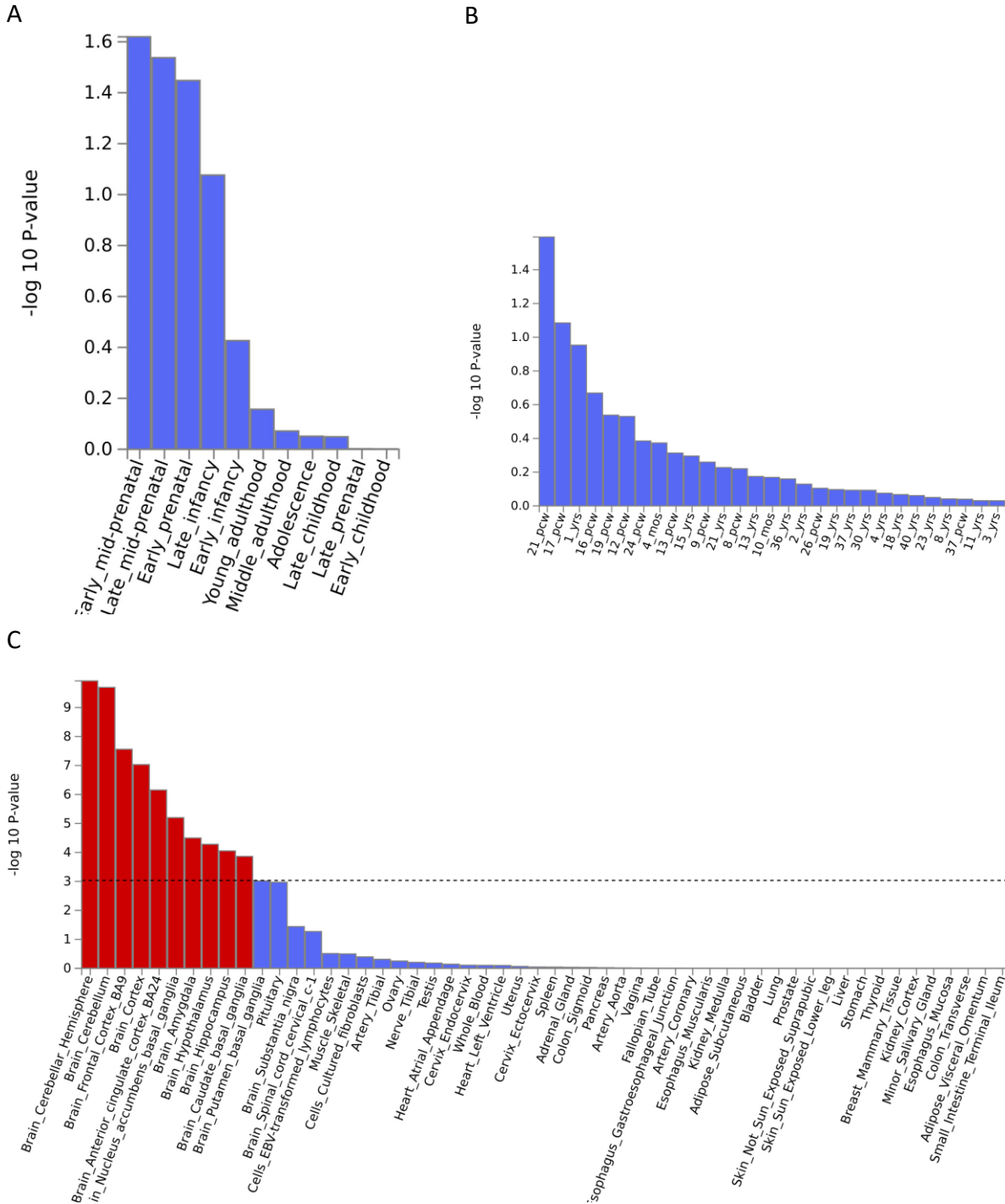
275
276

277
278
279
280
281
282
283

Supplementary Figure 8. Results of MAGMA tissue expression analysis of EUR psychotic disorders factor

Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are shown in Panel C. Dashed line represents significance threshold.

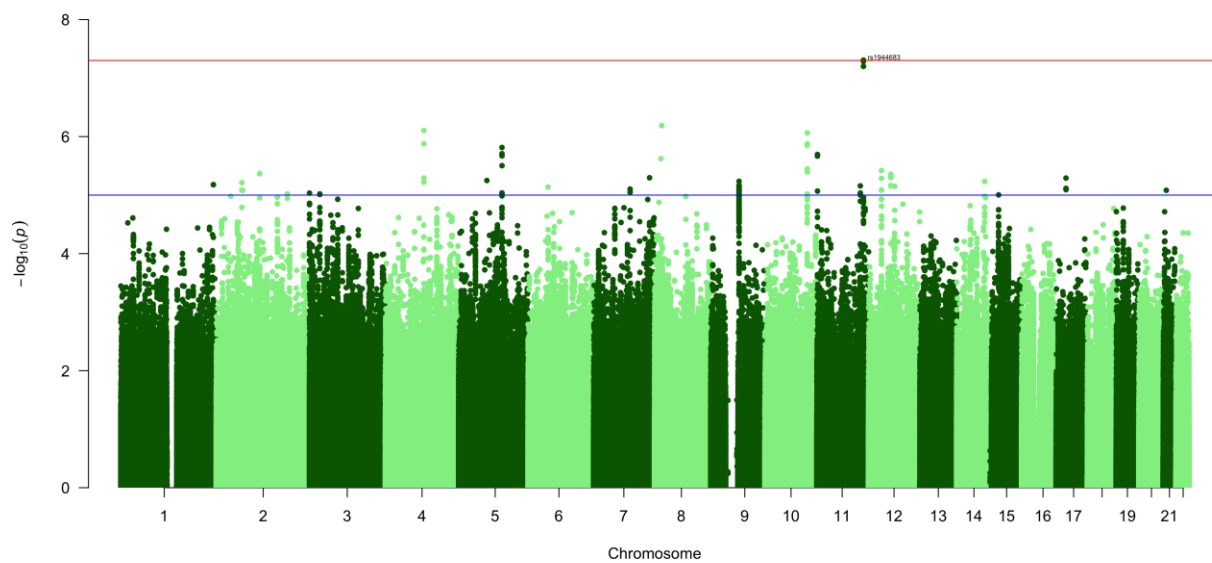
284
285



286
287
288
289
290

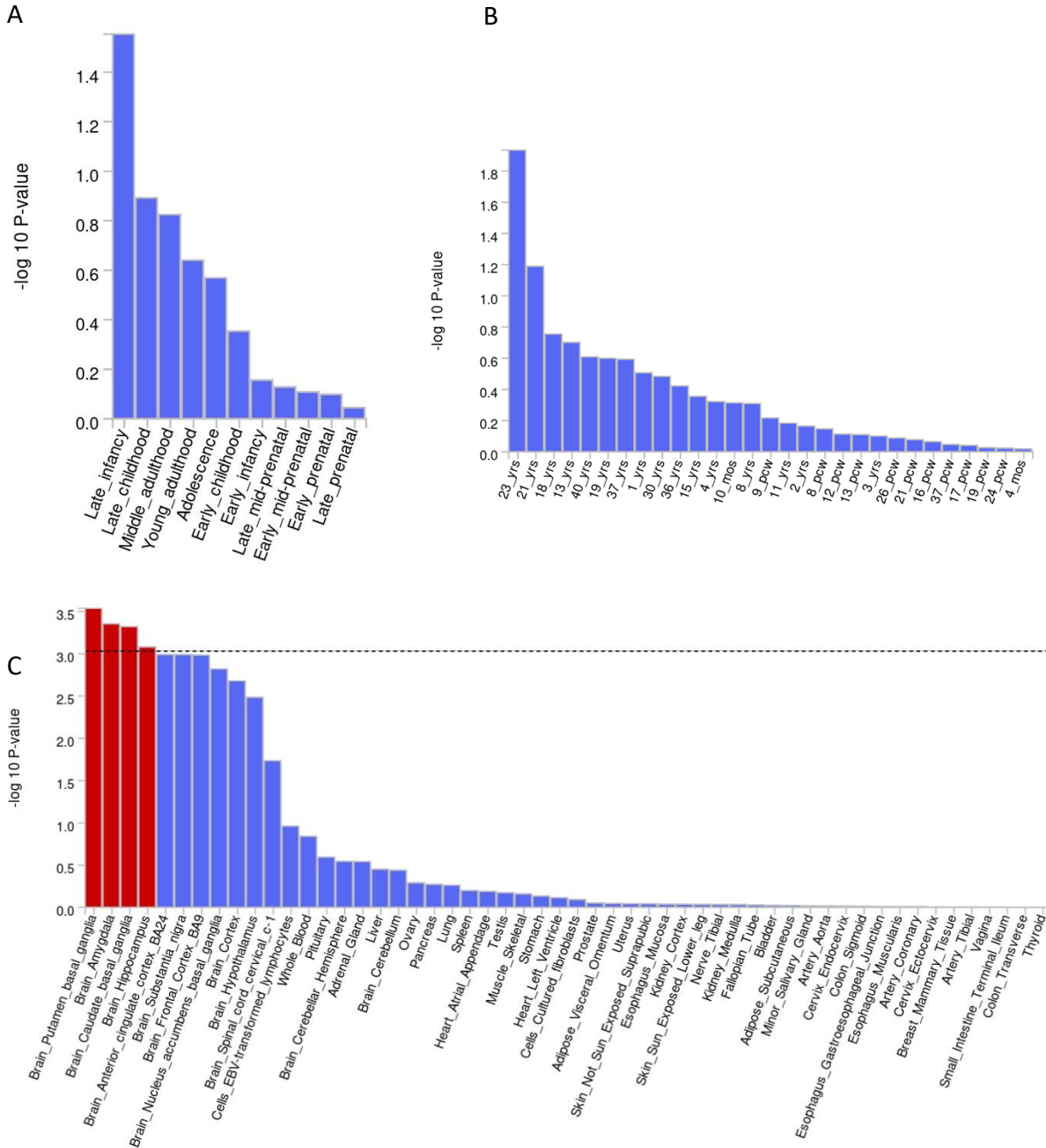
Supplementary Figure 9. Results of MAGMA tissue expression analysis of EUR mood disorders factor

Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are shown in Panel C. Dashed line indicates significance threshold.



291
292
293
294

Supplementary Figure 10. Manhattan plot for substance use disorders factor in AFR ancestry individuals

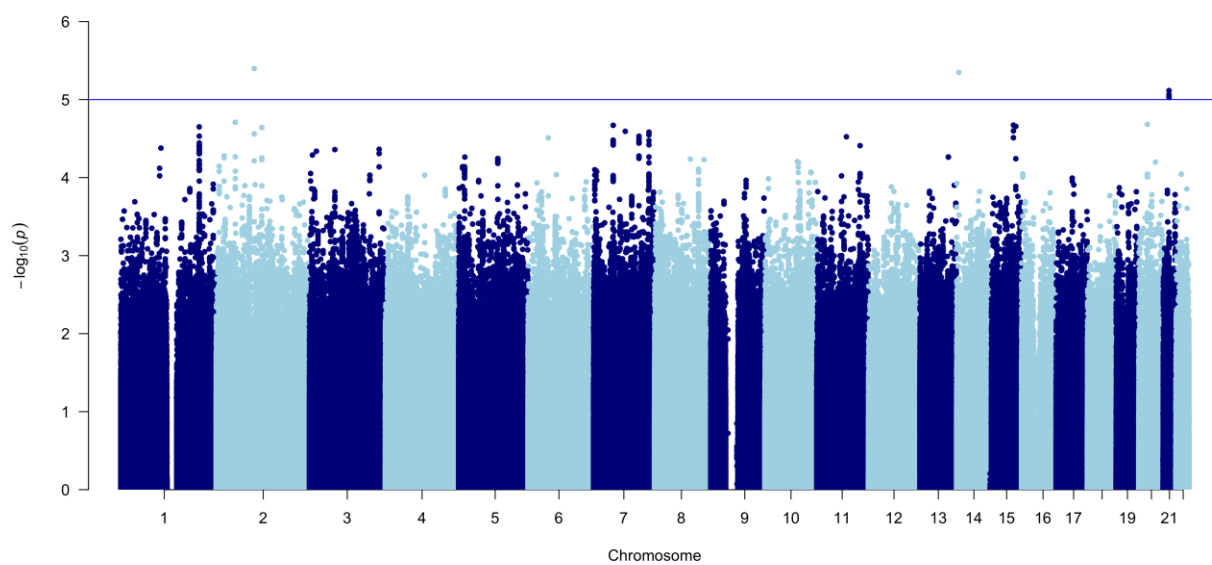


295

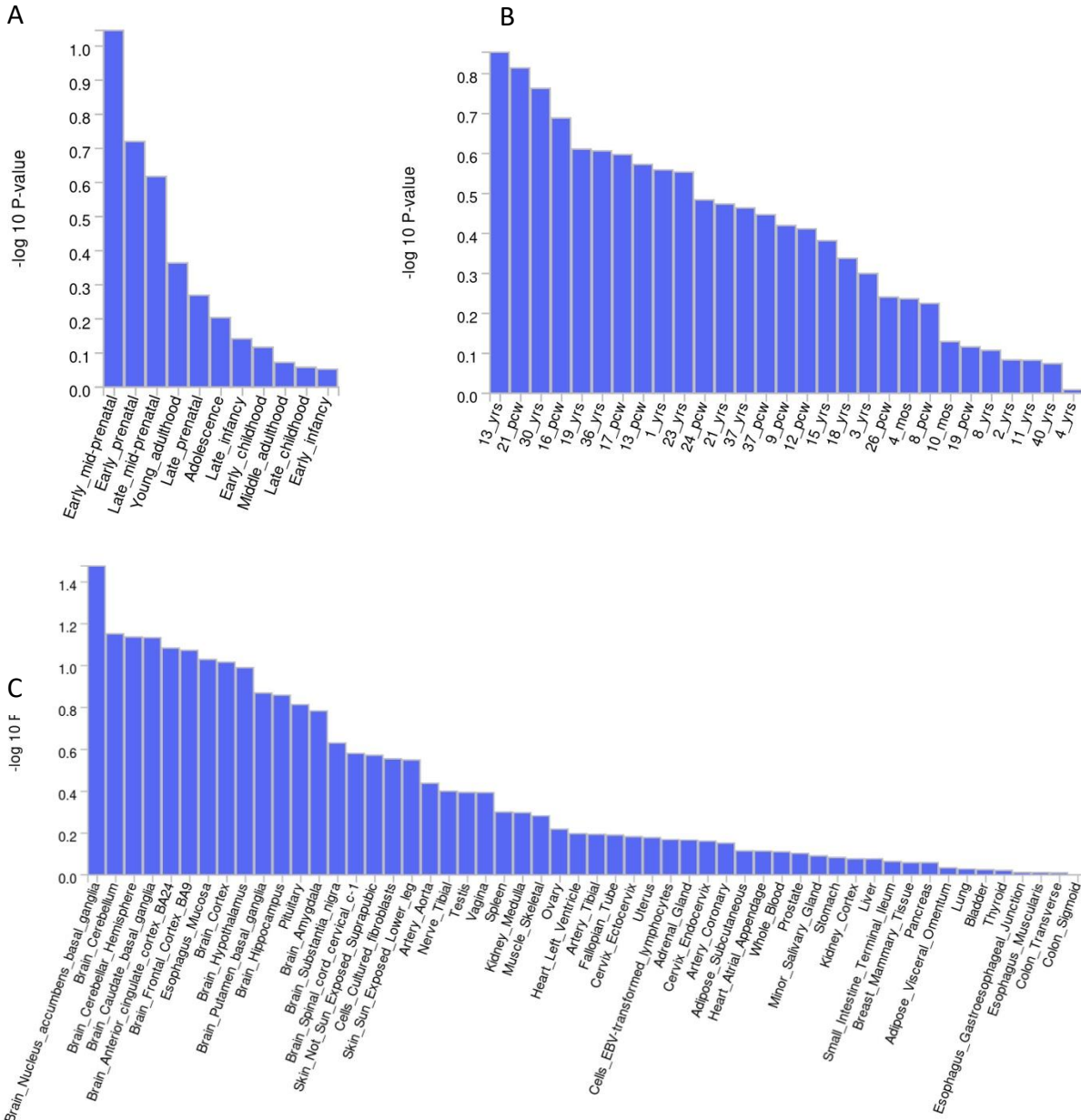
296

297 **Supplementary Figure 11. Results of MAGMA tissue expression analysis of AFR ancestry**
 298 **substance use disorders factor**

299 Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are
 300 shown in Panel C. Dashed line indicates significance threshold.



301
302 **Supplementary Figure 12. Manhattan plot for psychiatric disorders factor in AFR ancestry**
303 **individuals**
304



305

306

307

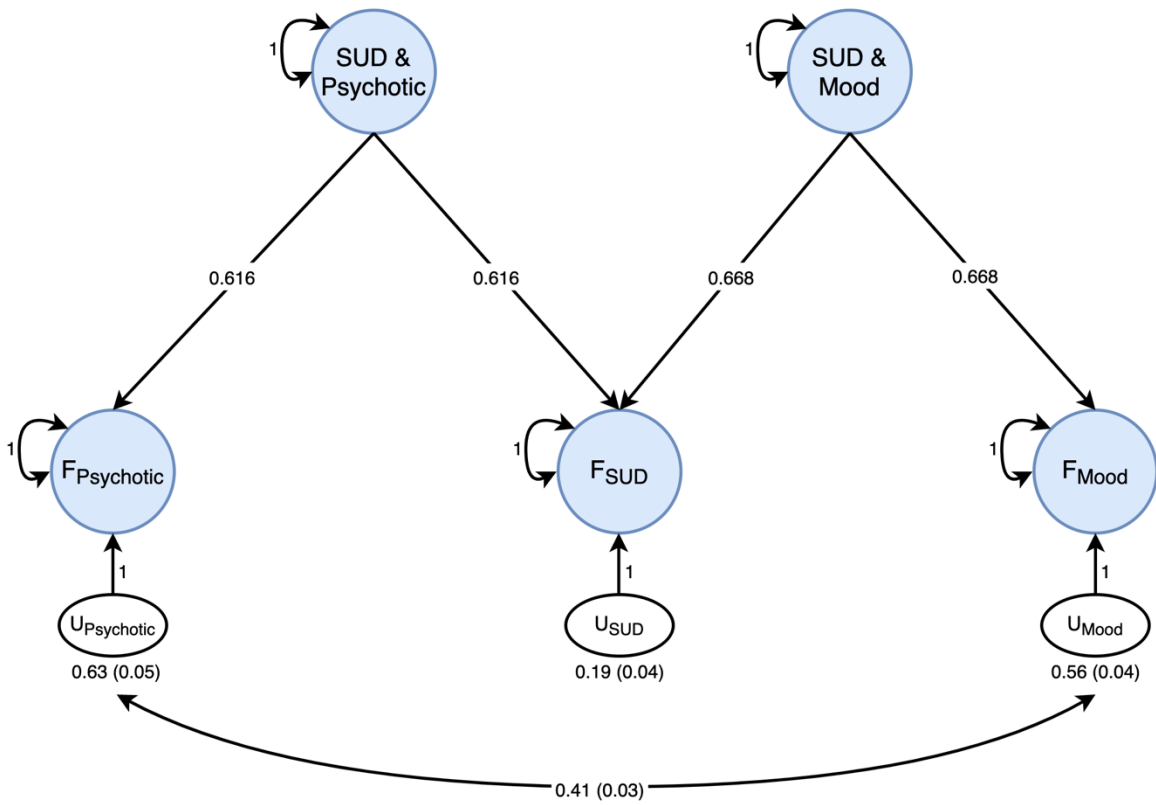
308

309

310

Supplementary Figure 13. Results of MAGMA tissue expression analysis of AFR ancestry psychiatric disorders factor

Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are shown in Panel C.



311

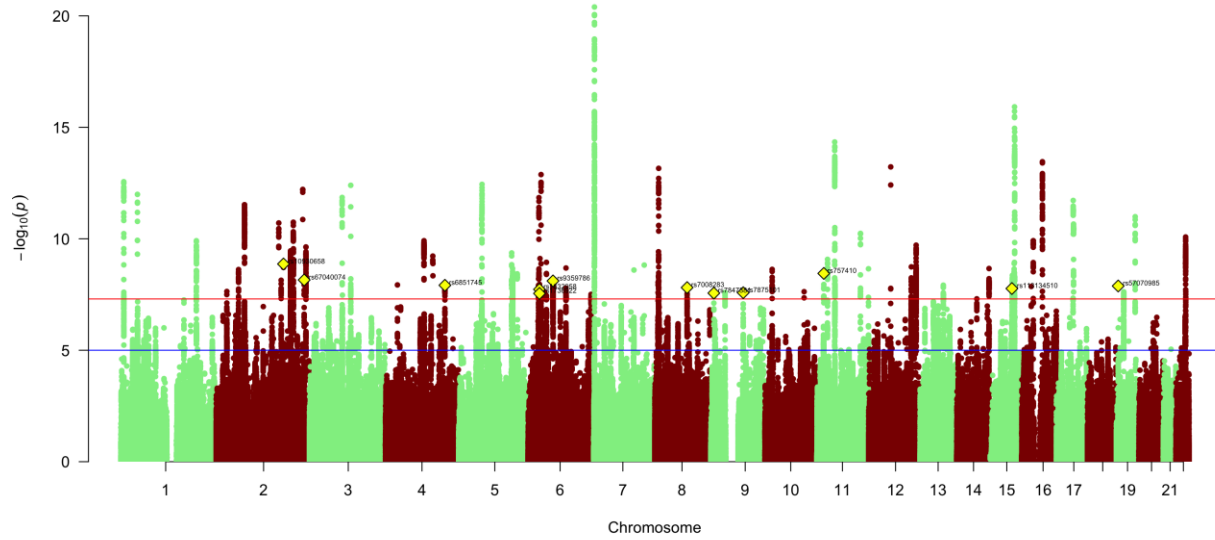
312

Supplementary Figure 14. EUR ancestry second order common factor model

313

Model fit statistics: $\chi^2(2) = 57.61$, $p = 3.09 \times 10^{-13}$, AIC = 65.61, CFI = 0.91, SRMR = 0.07.

314

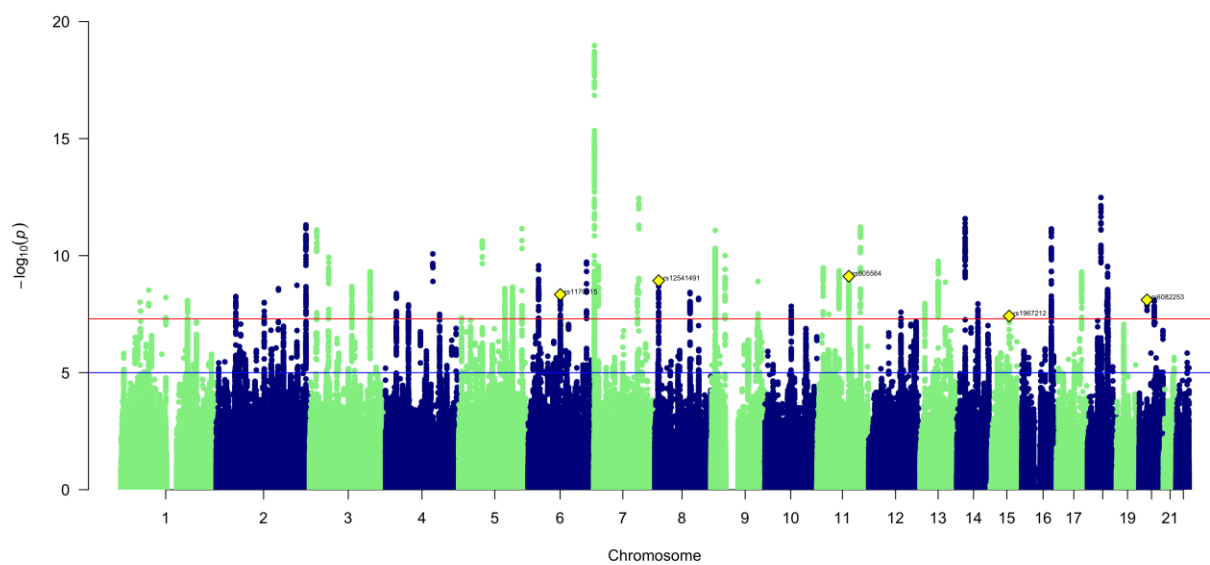


316

317 **Supplementary Figure 15. Manhattan plot for second-order common factor representing**318 **overlap between substance use and psychotic disorders in EUR ancestry individuals**

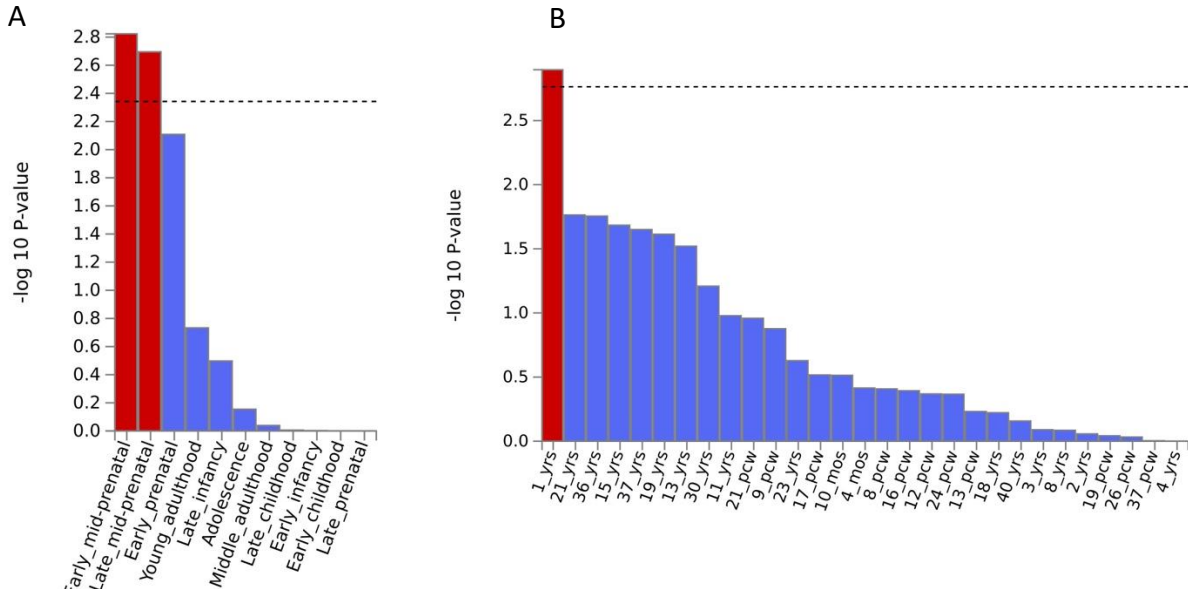
319 GWAS identified 76 lead SNPs, 12 of which were not in any of the input GWAS.

320

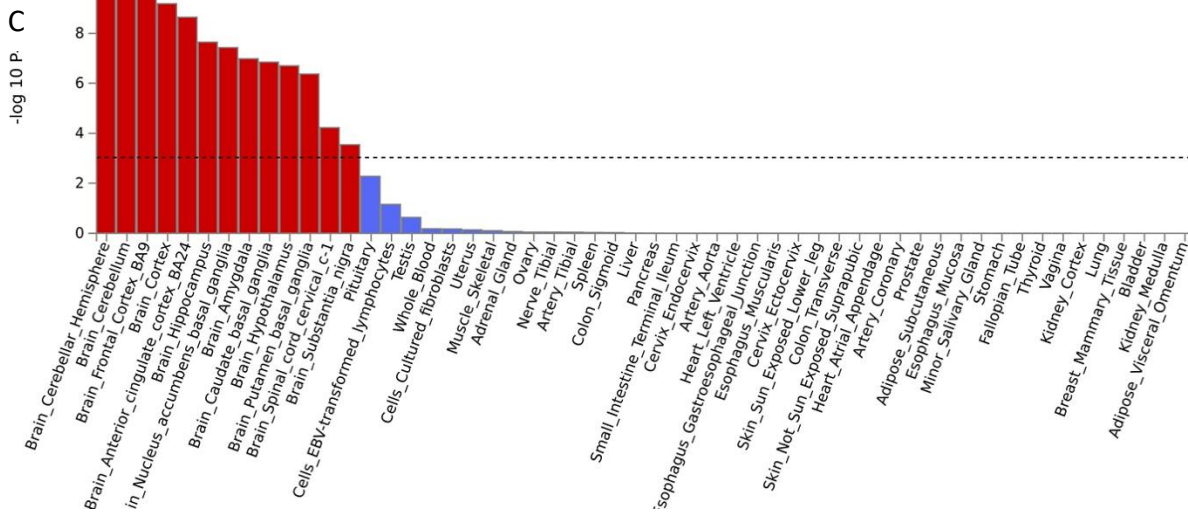


321
322 **Supplementary Figure 16. Manhattan plot for second-order common factor representing**
323 **overlap between substance use and mood/anxiety disorders in EUR ancestry individuals**
324 **GWAS identified 63 lead SNPs, 5 of which were not in any of the input GWAS.**
325

326



327

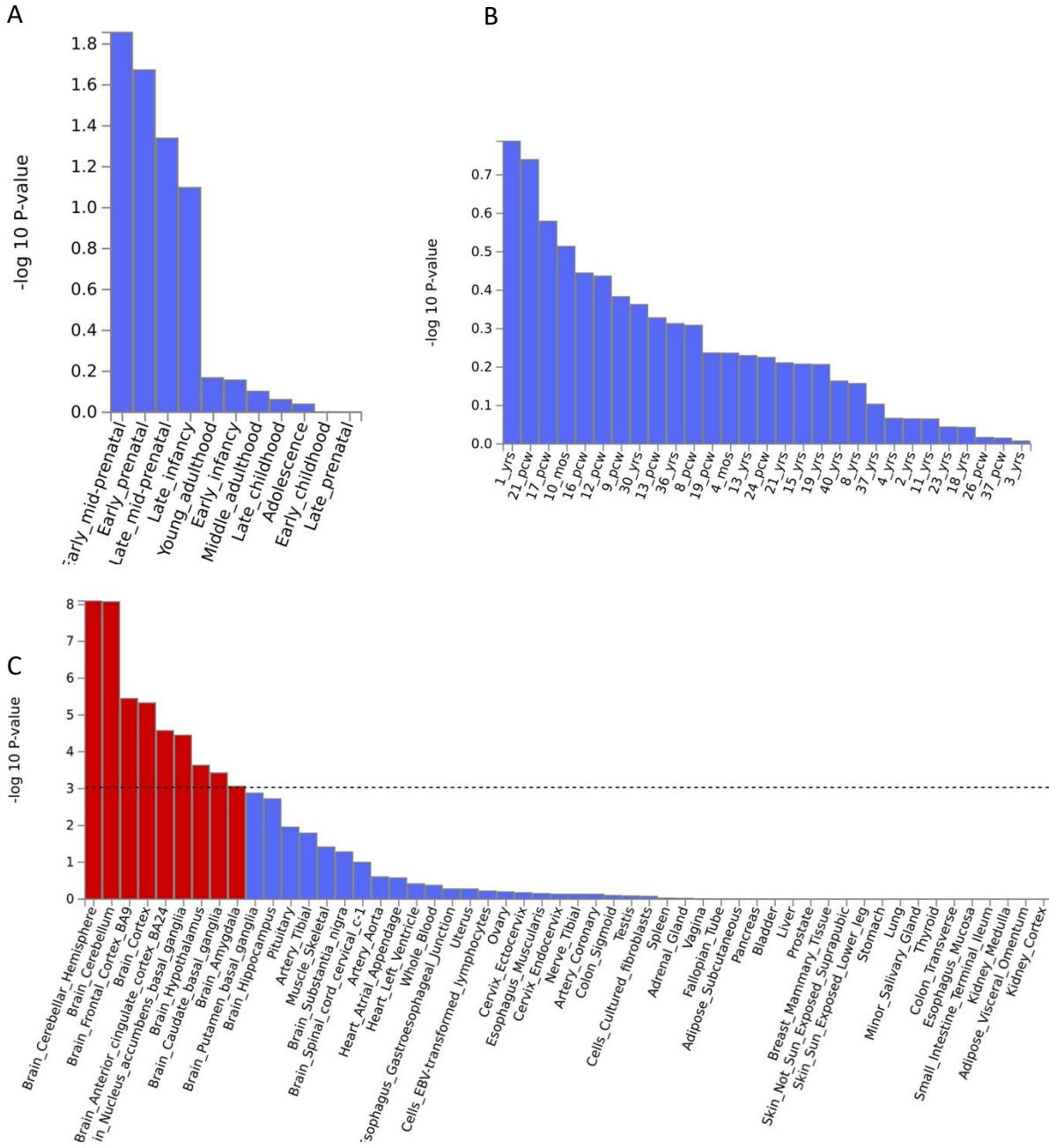


332

Supplementary Figure 17. Results of MAGMA tissue expression analysis of EUR ancestry second-order substance use and psychotic disorders factor

Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are shown in Panel C. Dashed line indicates significance threshold.

333



334

335

336

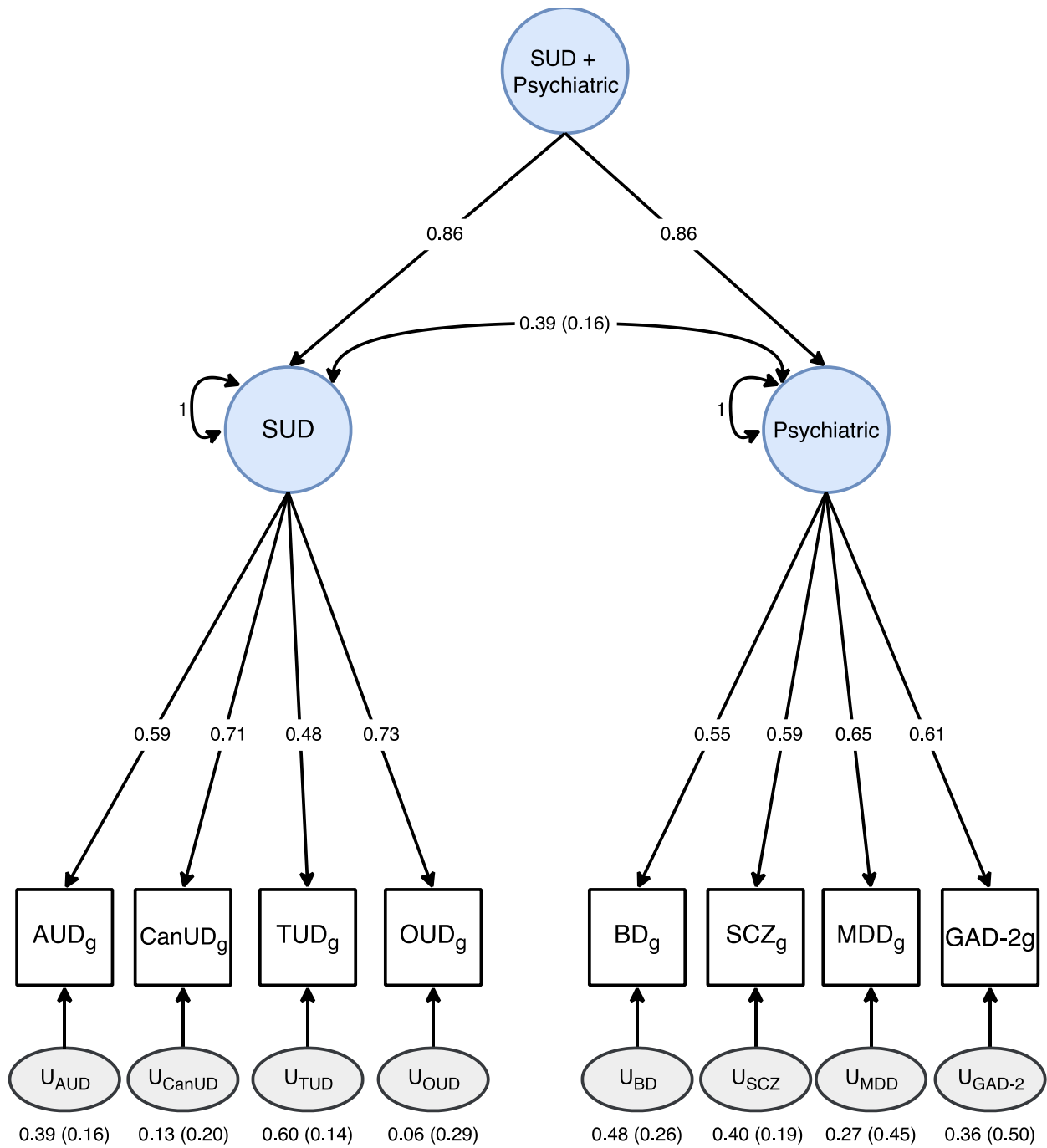
337

338

339

Supplementary Figure 18. Results of MAGMA tissue expression analysis of EUR ancestry second-order substance use and mood disorders factor

Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are shown in panel C. Dashed line indicates significance threshold.



340

341

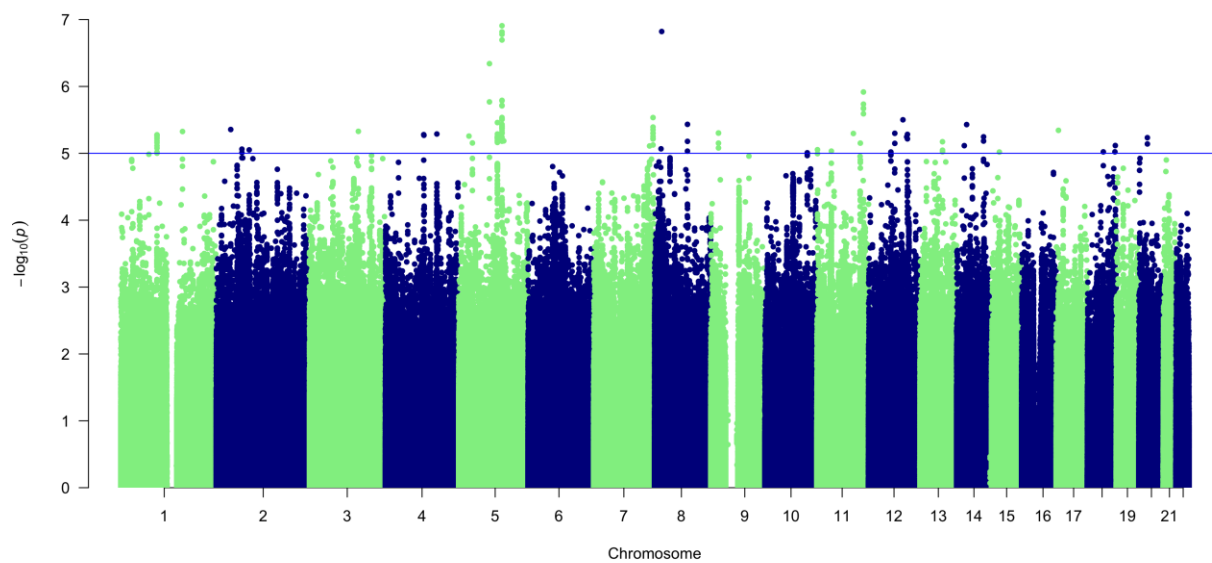
Supplementary Figure 19. AFR ancestry second order common factor model

342

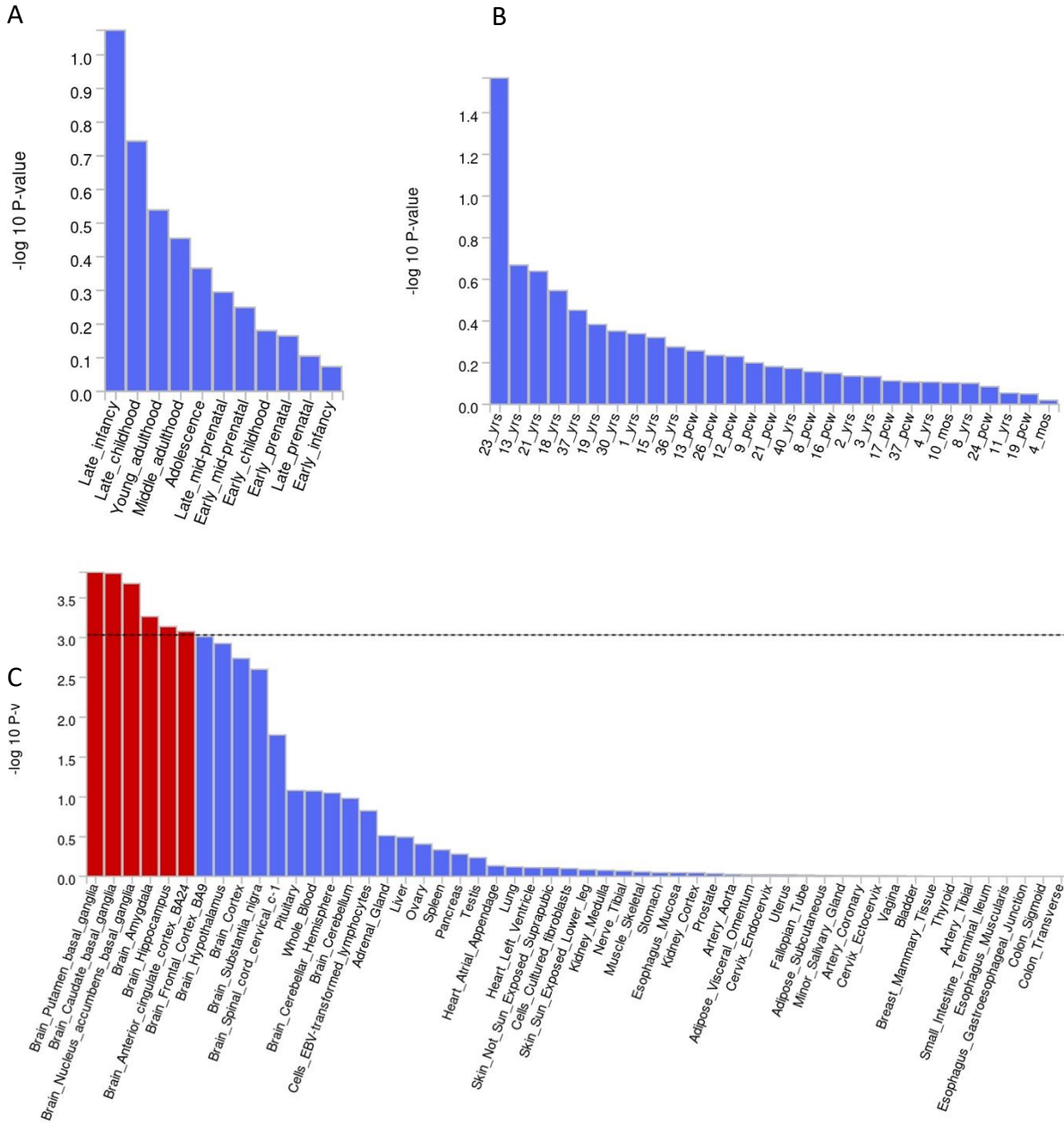
Model fit statistics: $\chi^2(19) = 21.49$, $p = 0.31$, $AIC = 55.49$, $CFI = 0.99$, $SRMR = 0.10$.

343

344



345
346 **Supplementary Figure 20. Manhattan plot for second-order common factor representing**
347 **overlap between substance use and psychiatric disorders in AFR ancestry individuals**
348
349



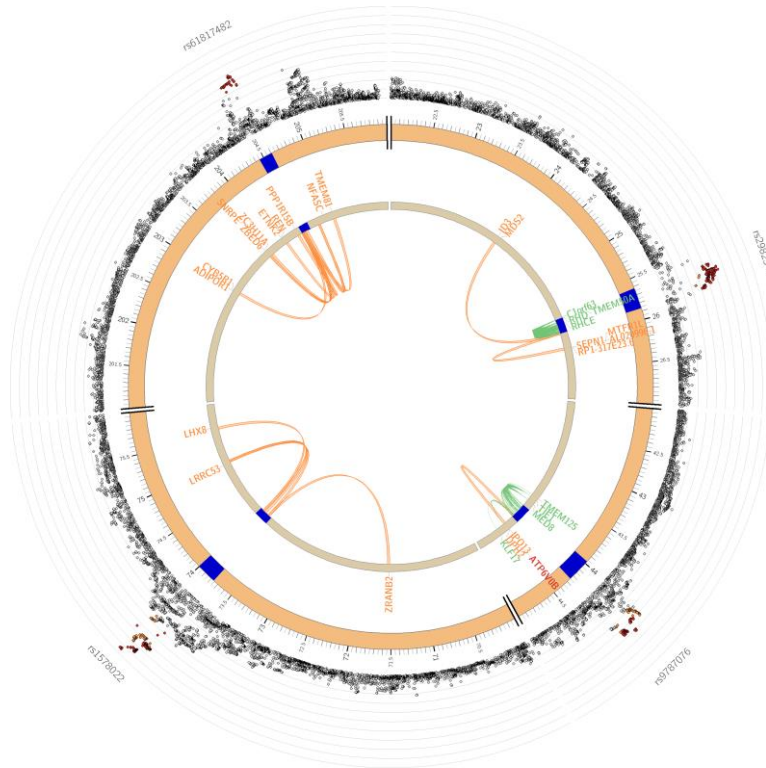
350

351

352 **Supplementary Figure 21. Results of MAGMA tissue expression analysis of AFR ancestry**
 353 **second-order substance use and psychiatric disorders factor**

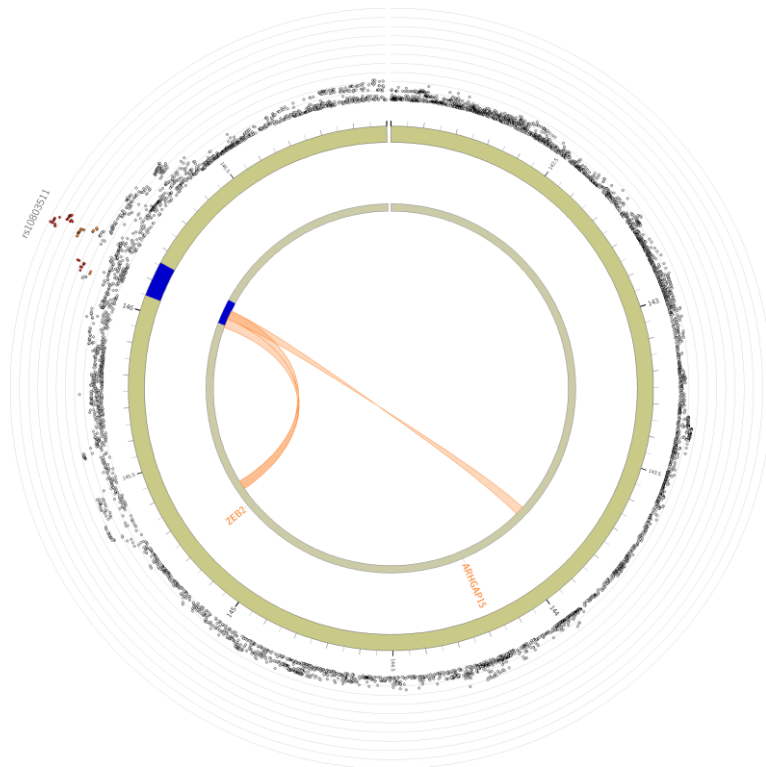
354 Results for the BrainSpan database are shown in panels A and B, and results for GTEx v8 are
 355 shown in panel C. Dashed line indicates significance threshold.

356



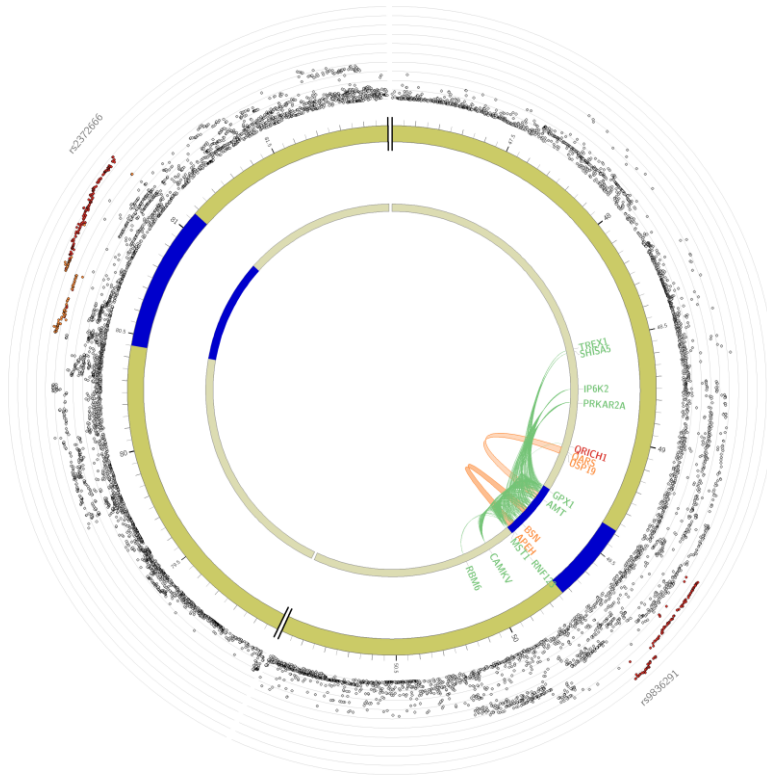
357
358

Chromosome 1



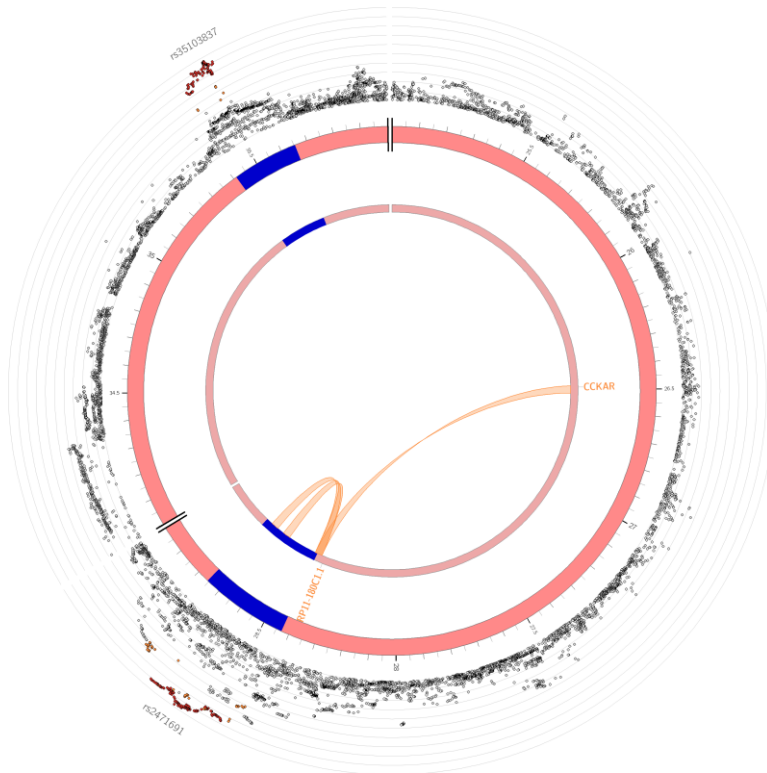
359
360

Chromosome 2



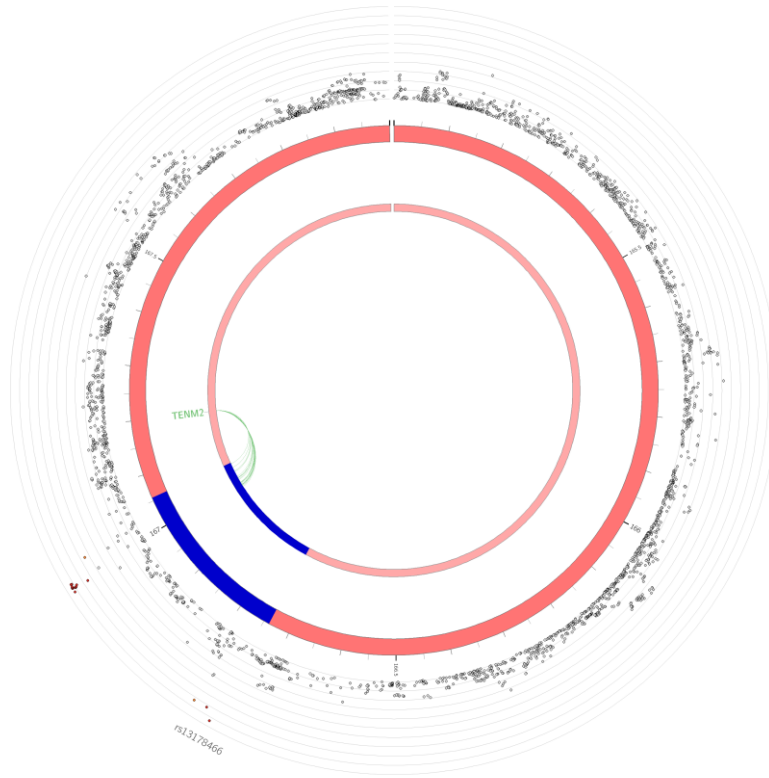
361
362

Chromosome 3



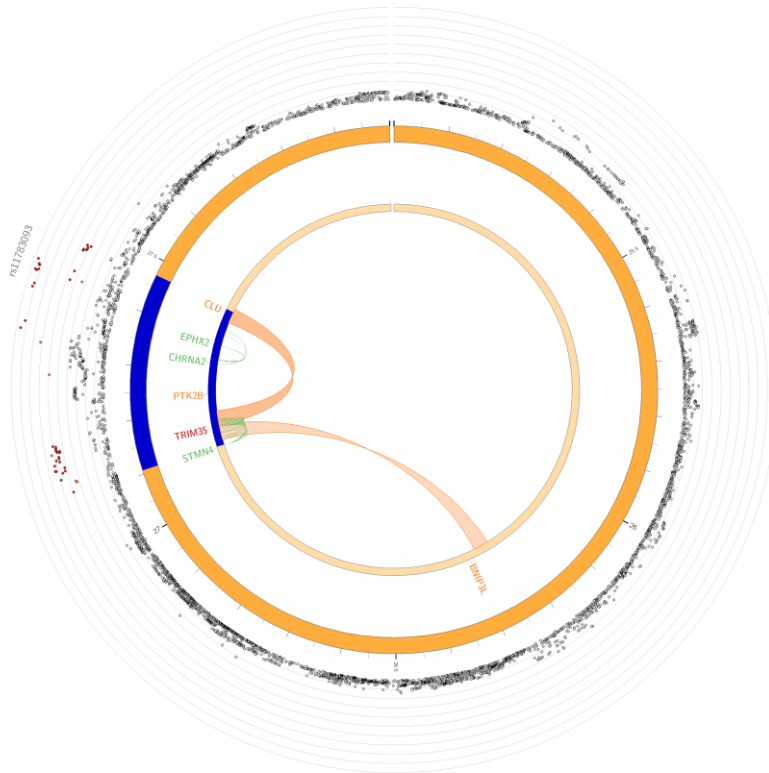
363
364

Chromosome 4



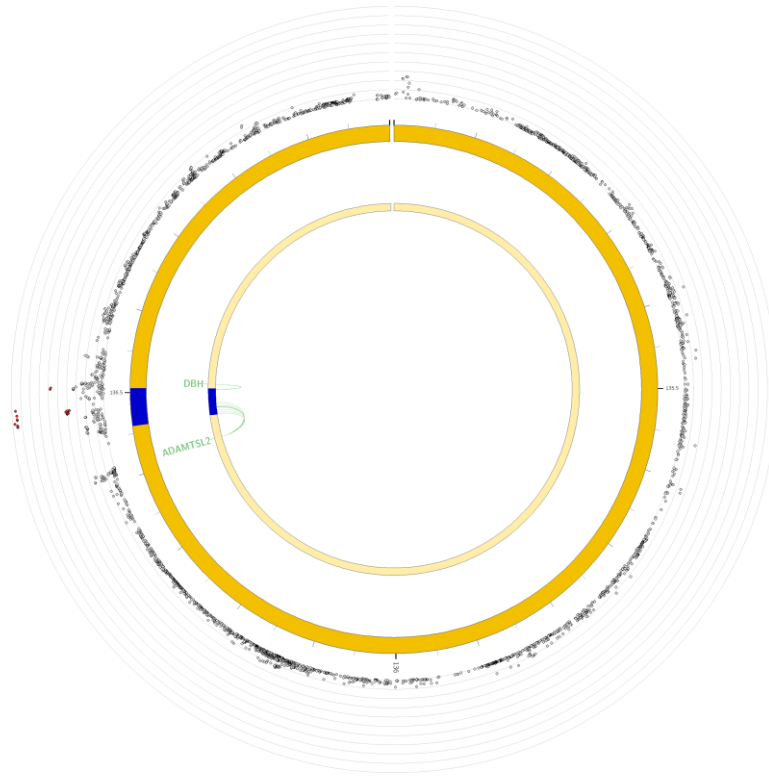
365
366

Chromosome 5



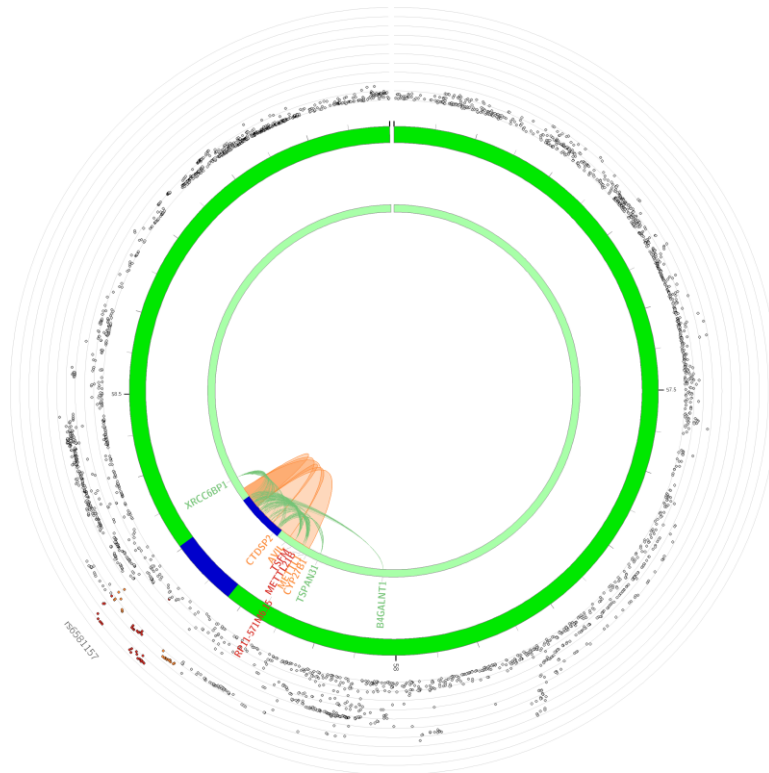
367
368

Chromosome 8



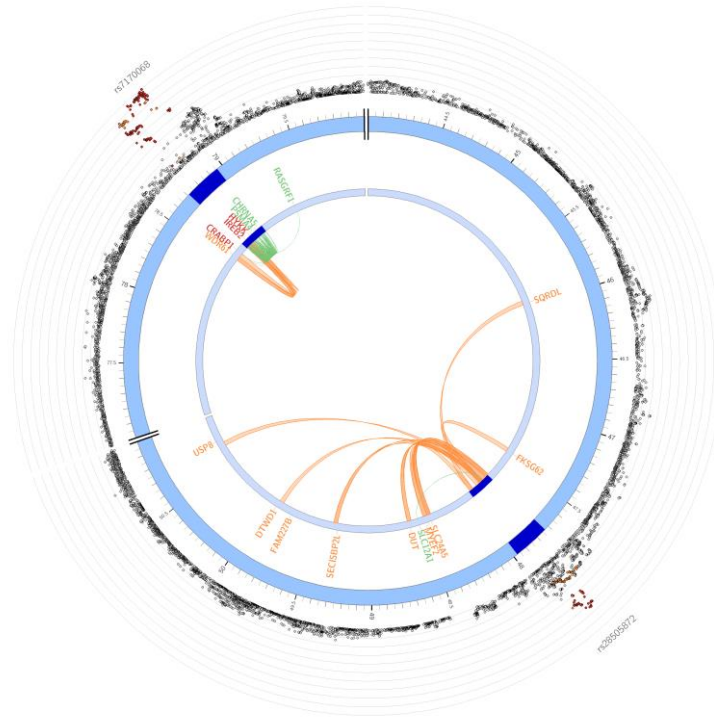
369
370

Chromosome 9



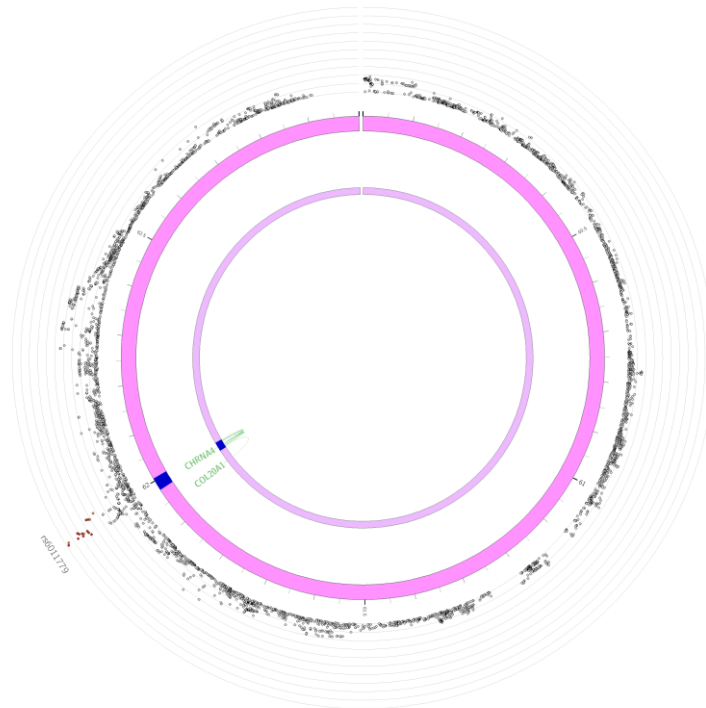
371
372

Chromosome 12



Chromosome 15

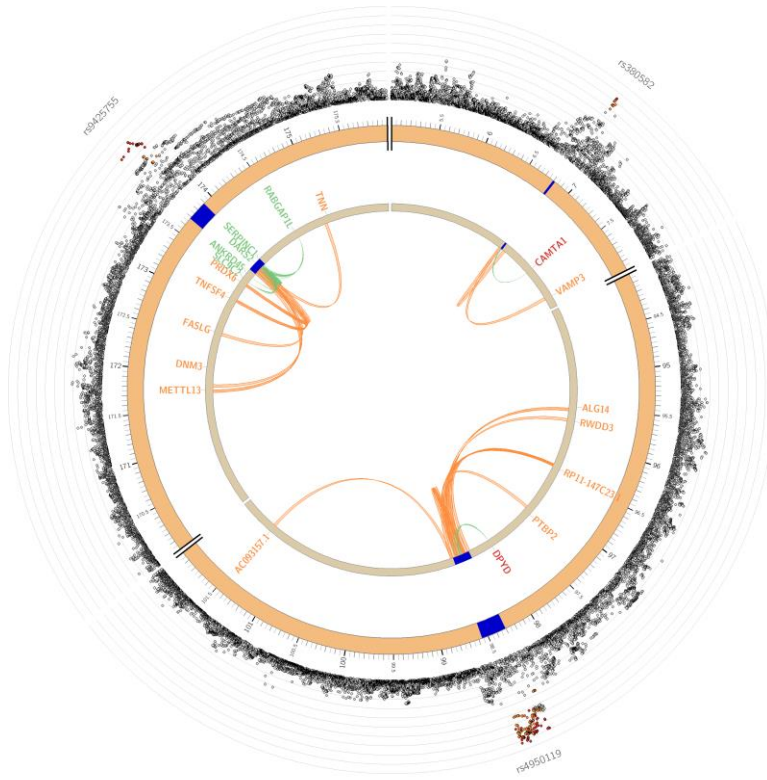
373
374



Chromosome 20

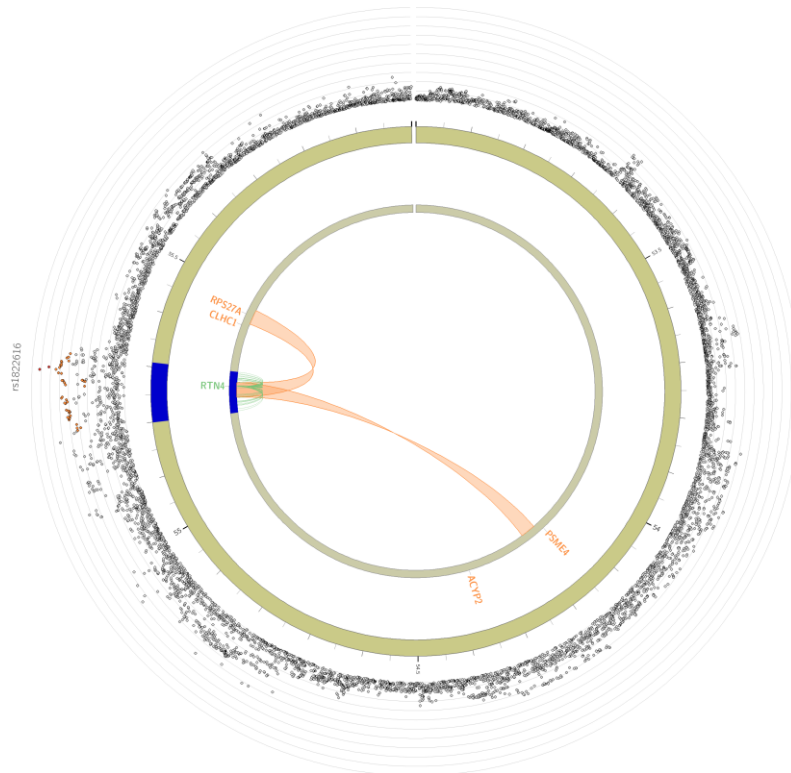
375
376

377 **Supplementary Figure 22. Significant SNPs identified in TUD Independent GWAS**
 378 The chromosome is depicted as a circle with SNPs plotted by their $-\log_{10}(p\text{-value})$, and lead
 379 SNPs annotated. Orange links indicate chromatin contact, green links indicate cis-eQTLs, and
 380 red links indicate both.



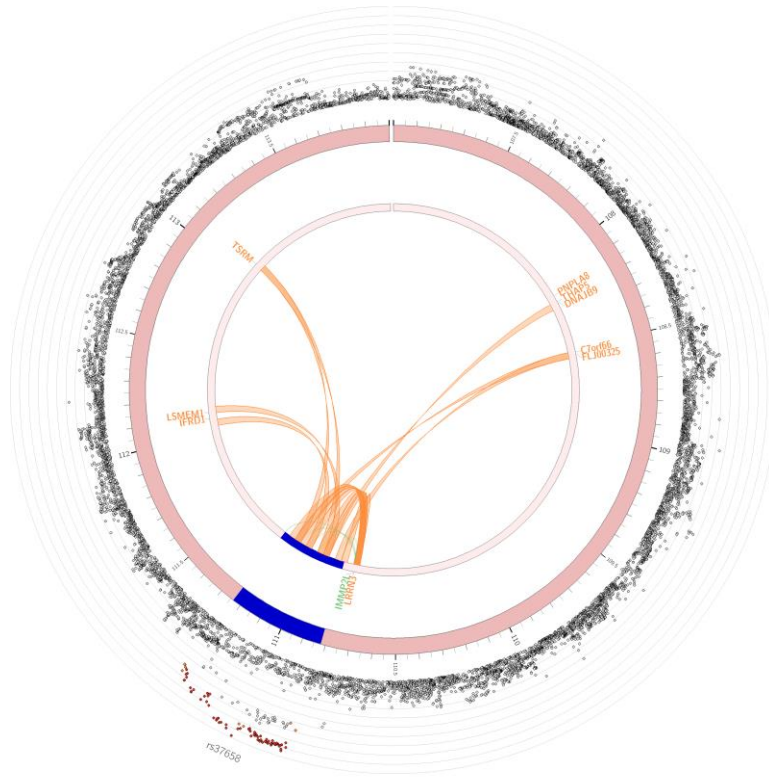
381
382

Chromosome 1



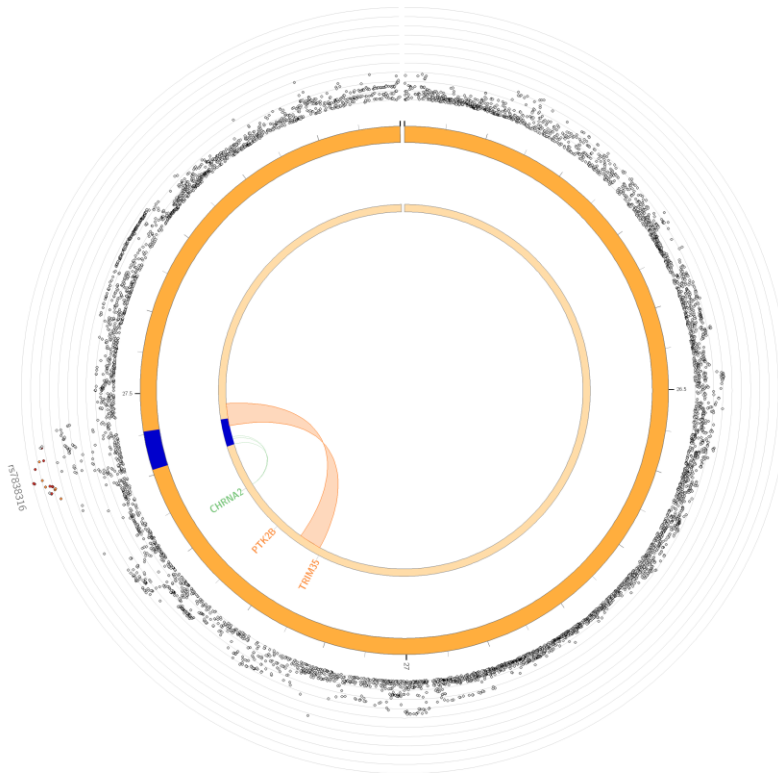
383
384

Chromosome 2



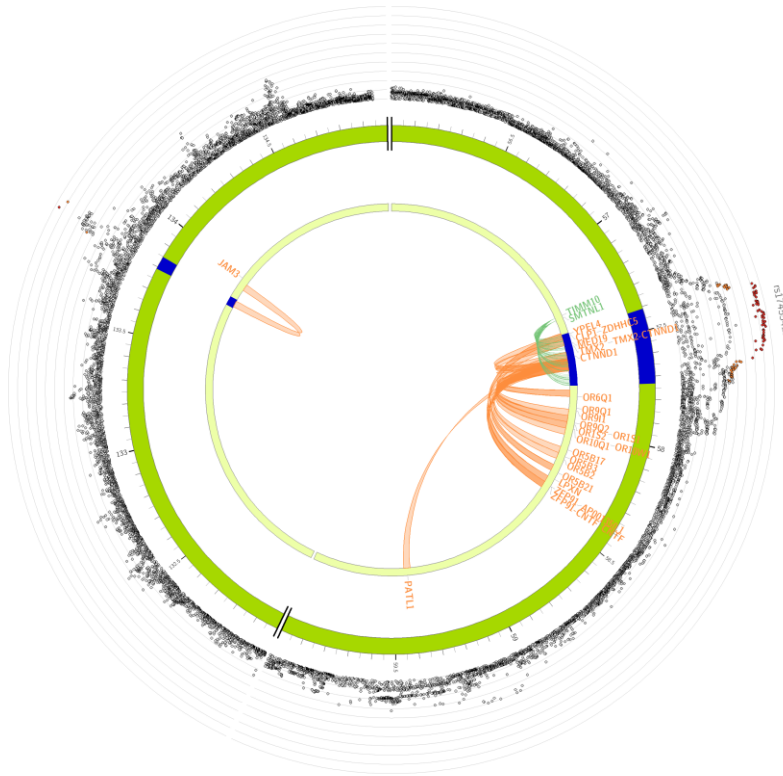
389
390

Chromosome 7



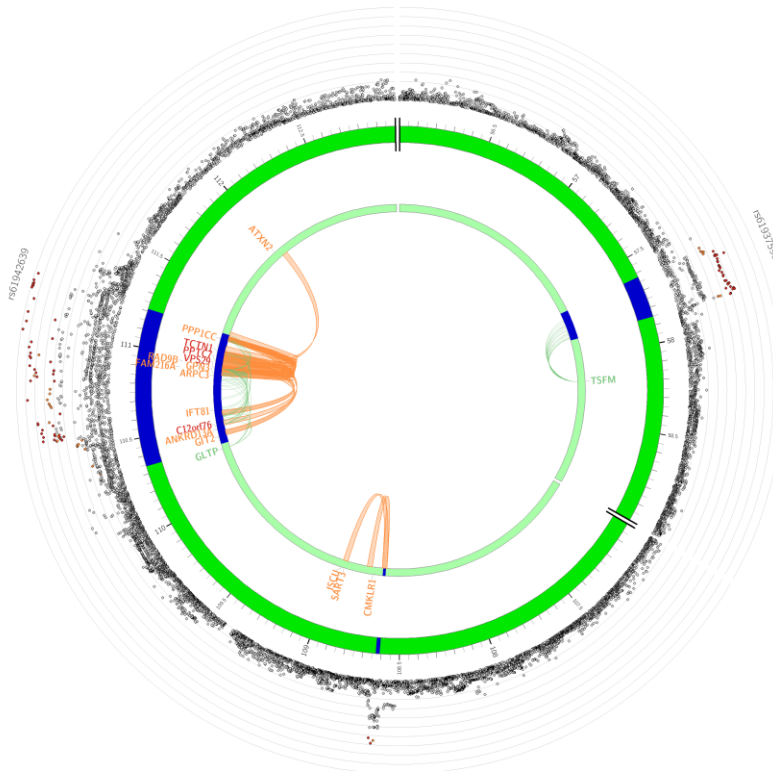
391
392

Chromosome 8



393
394

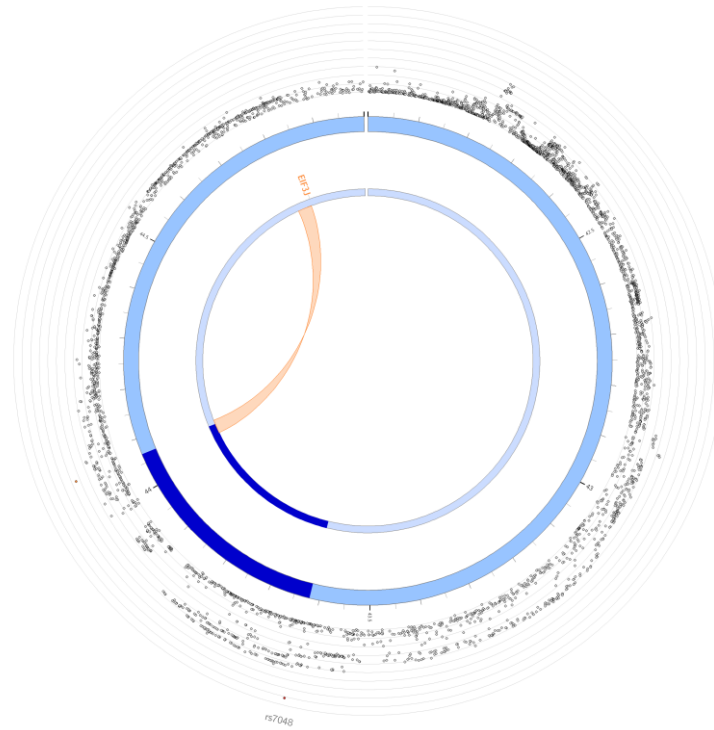
Chromosome 11



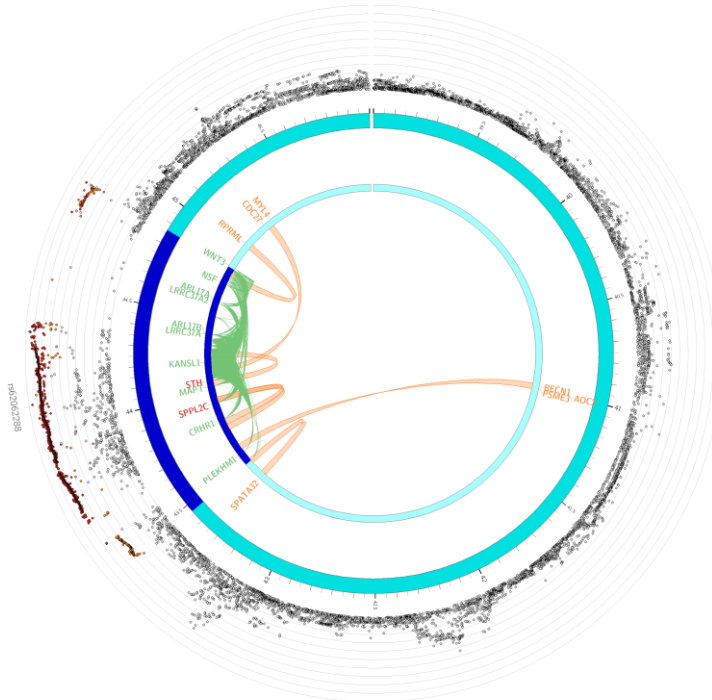
395
396

Chromosome 12

397
398



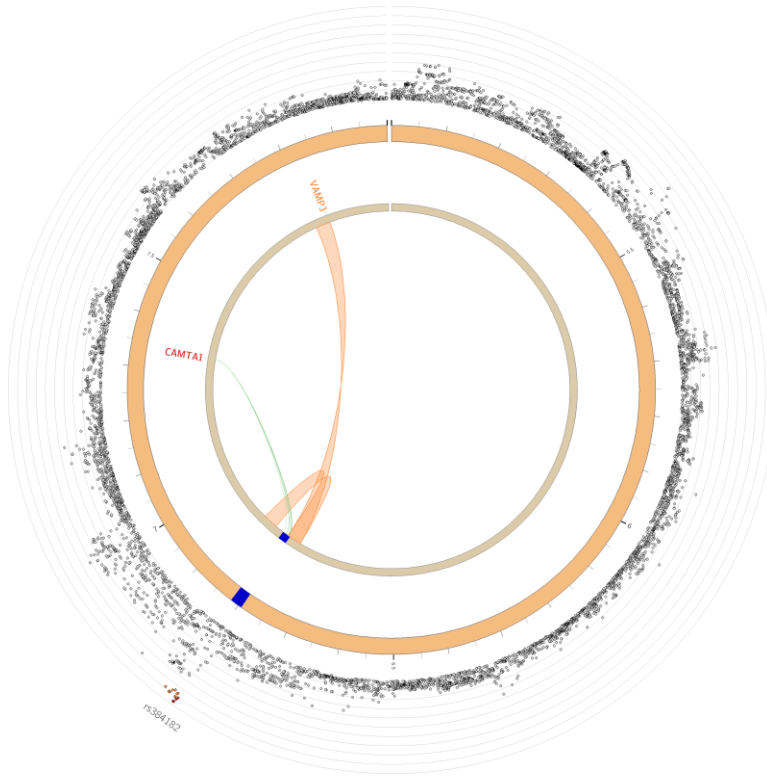
Chromosome 15



Chromosome 17

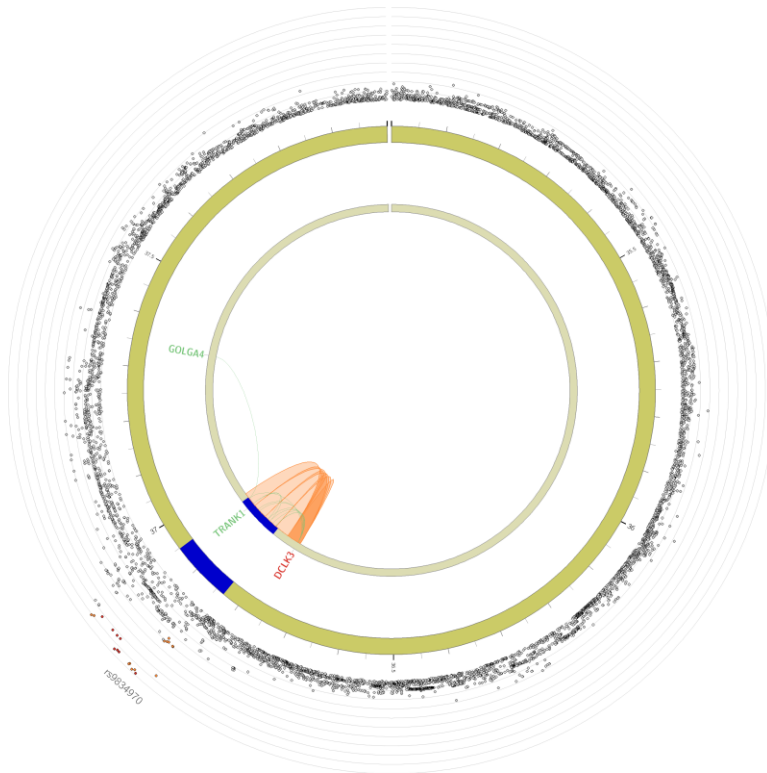
399
400

401 **Supplementary Figure 23. Significant SNPs identified in SCZ Independent GWAS**
402 The chromosome is depicted as a circle with SNPs plotted by their $-\log_{10}(\text{p-value})$, and lead
403 SNPs annotated. Orange links indicate chromatin contact, green links indicate cis-eQTLs, and
404 red links indicate both.



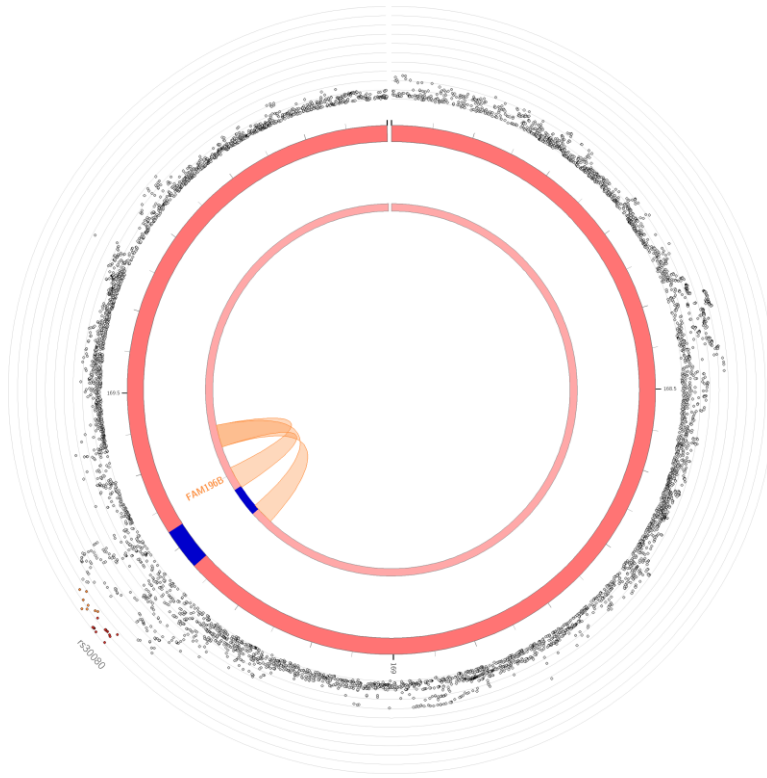
405
406

Chromosome 1



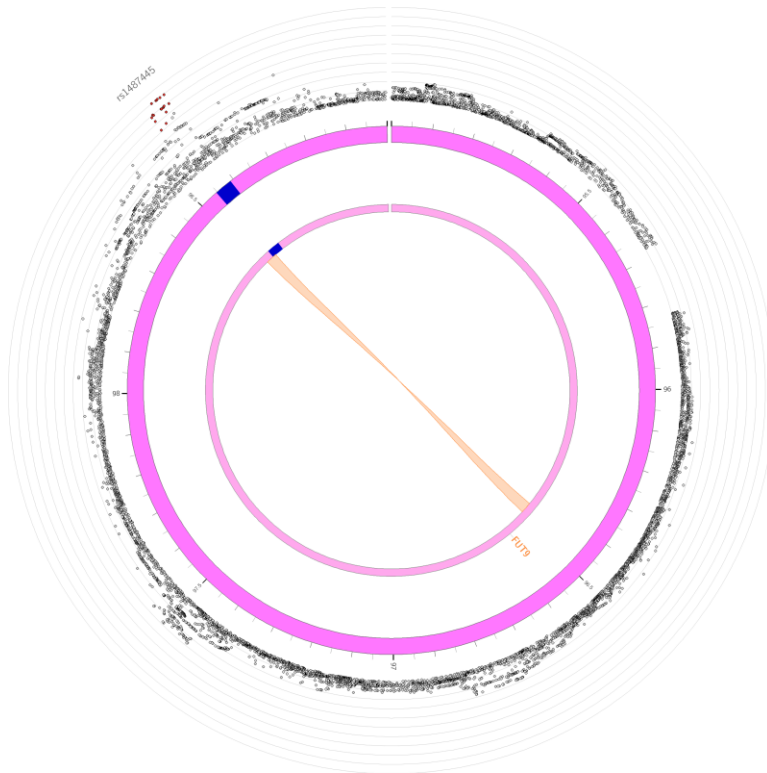
407
408

Chromosome 3



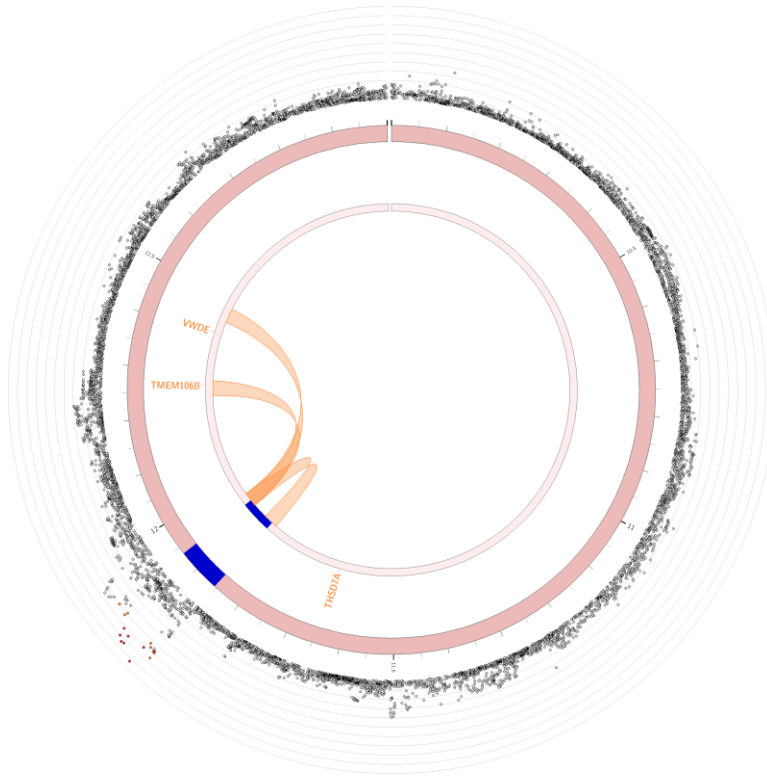
409
410

Chromosome 5



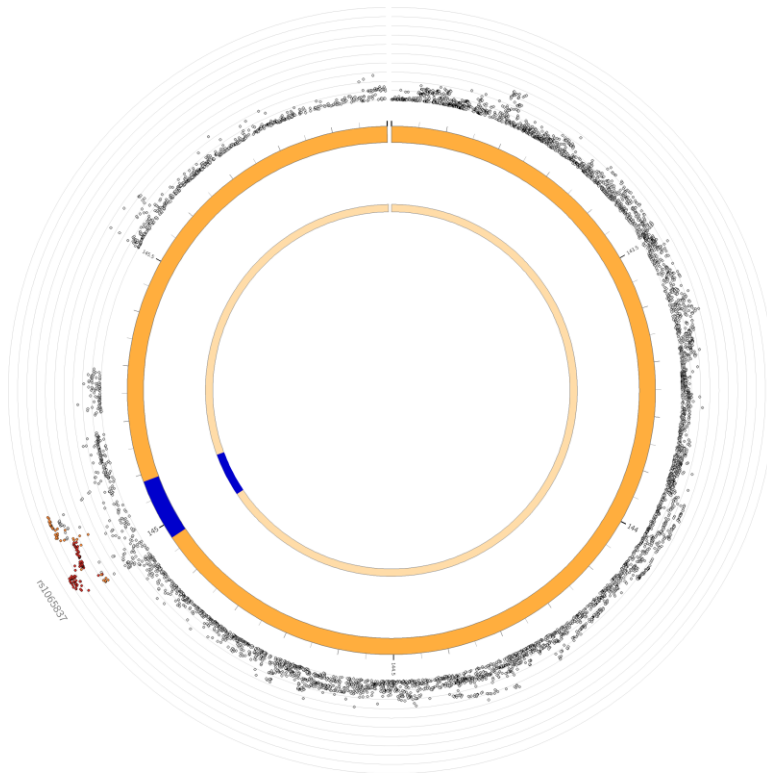
411
412

Chromosome 6



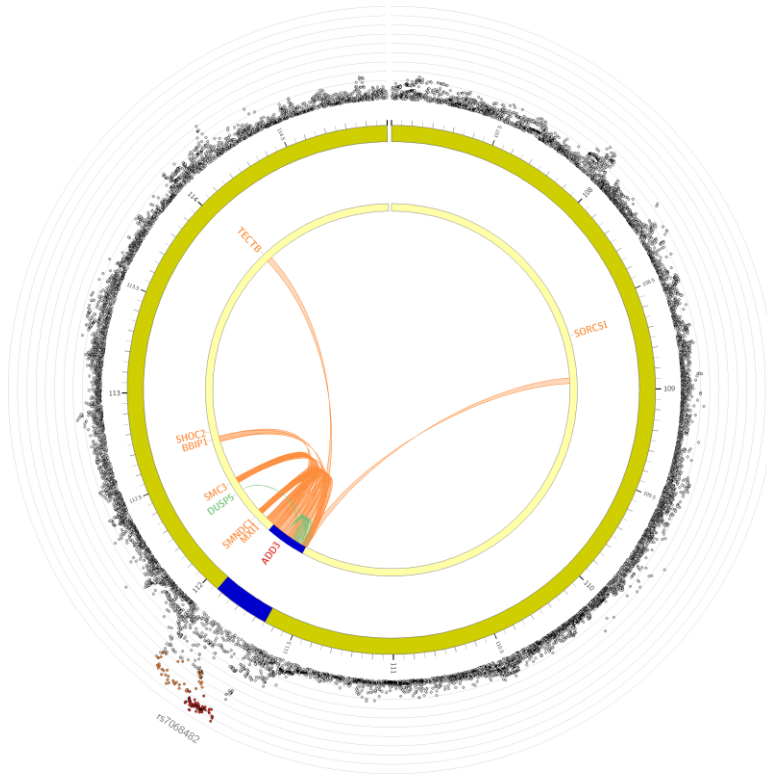
413
414

Chromosome 7



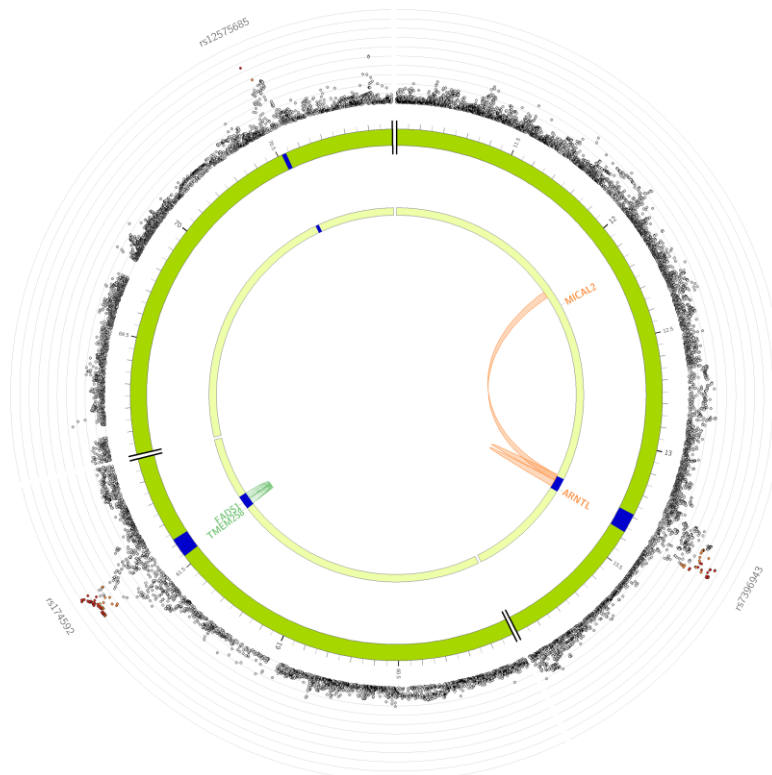
415
416

Chromosome 8



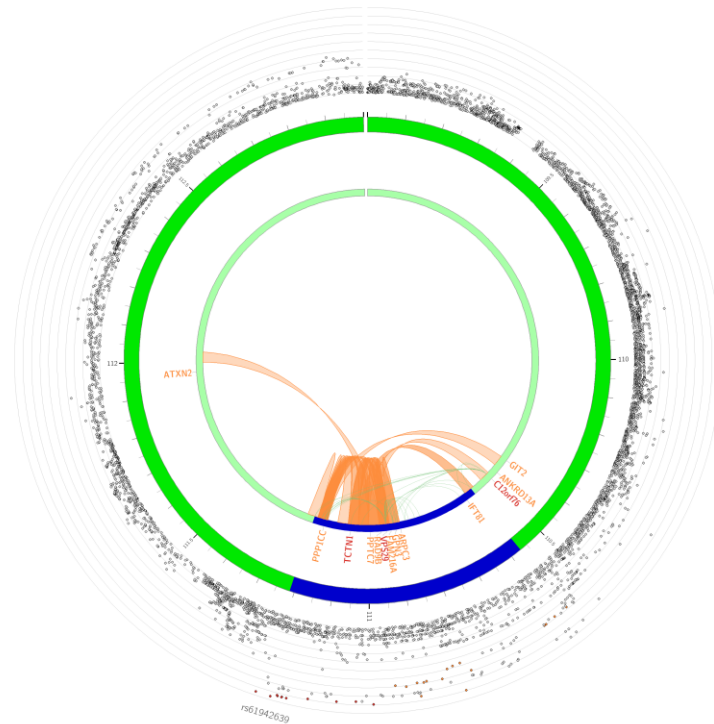
417
418

Chromosome 10



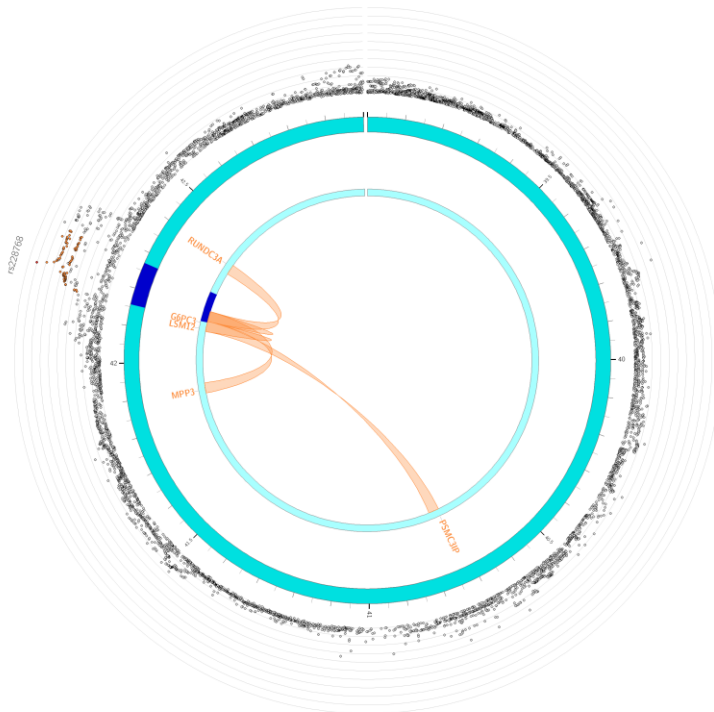
419
420

Chromosome 11



Chromosome 12

421
422

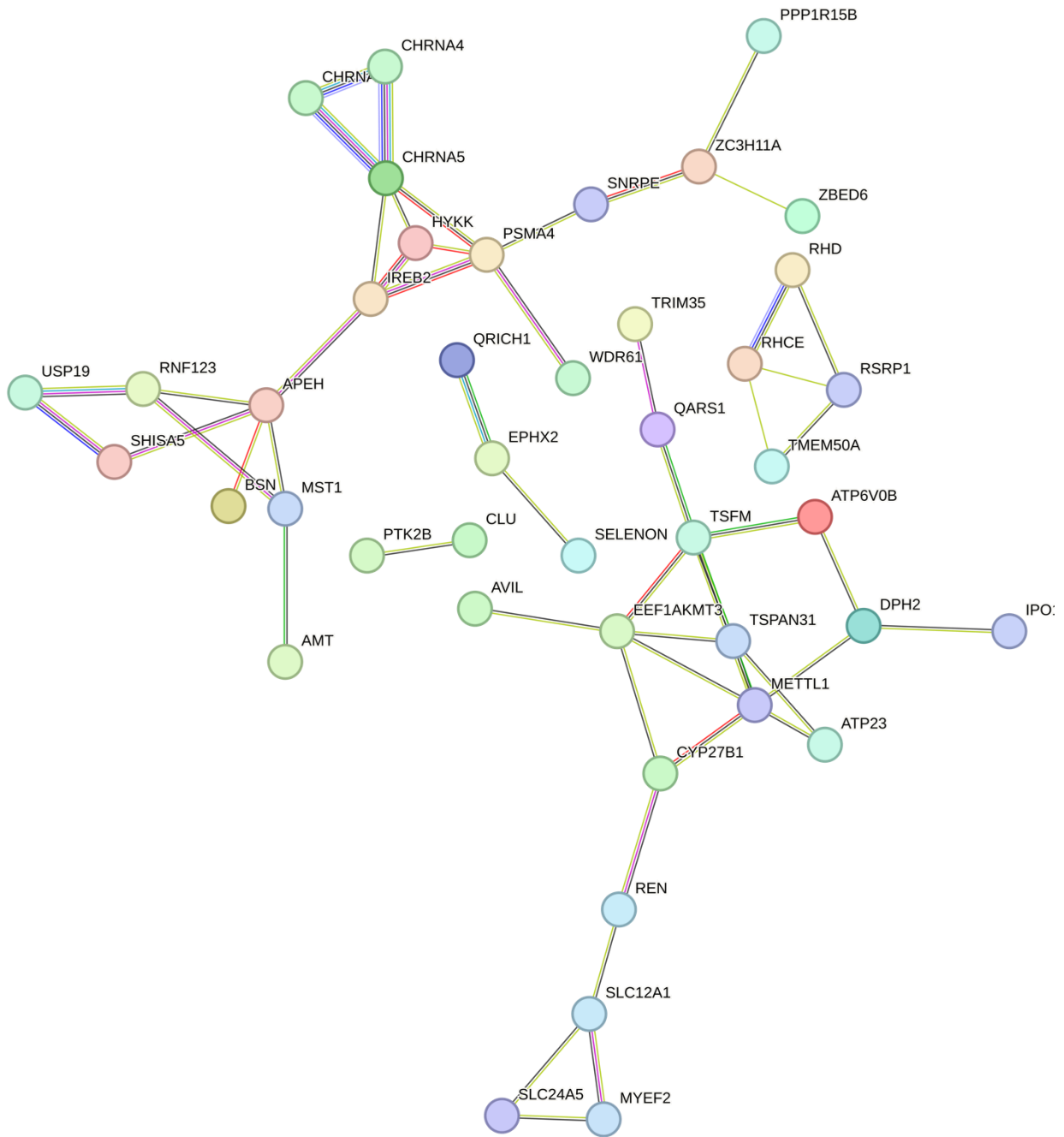


Chromosome 17

423
424

Supplementary Figure 24. Significant SNPs identified in BD Independent GWAS

425 The chromosome is depicted as a circle with SNPs plotted by their $-\log_{10}(p\text{-value})$, and lead
 426 SNPs annotated. Orange links indicate chromatin contact, green links indicate cis-eQTLs, and
 427 red links indicate both.
 428



429

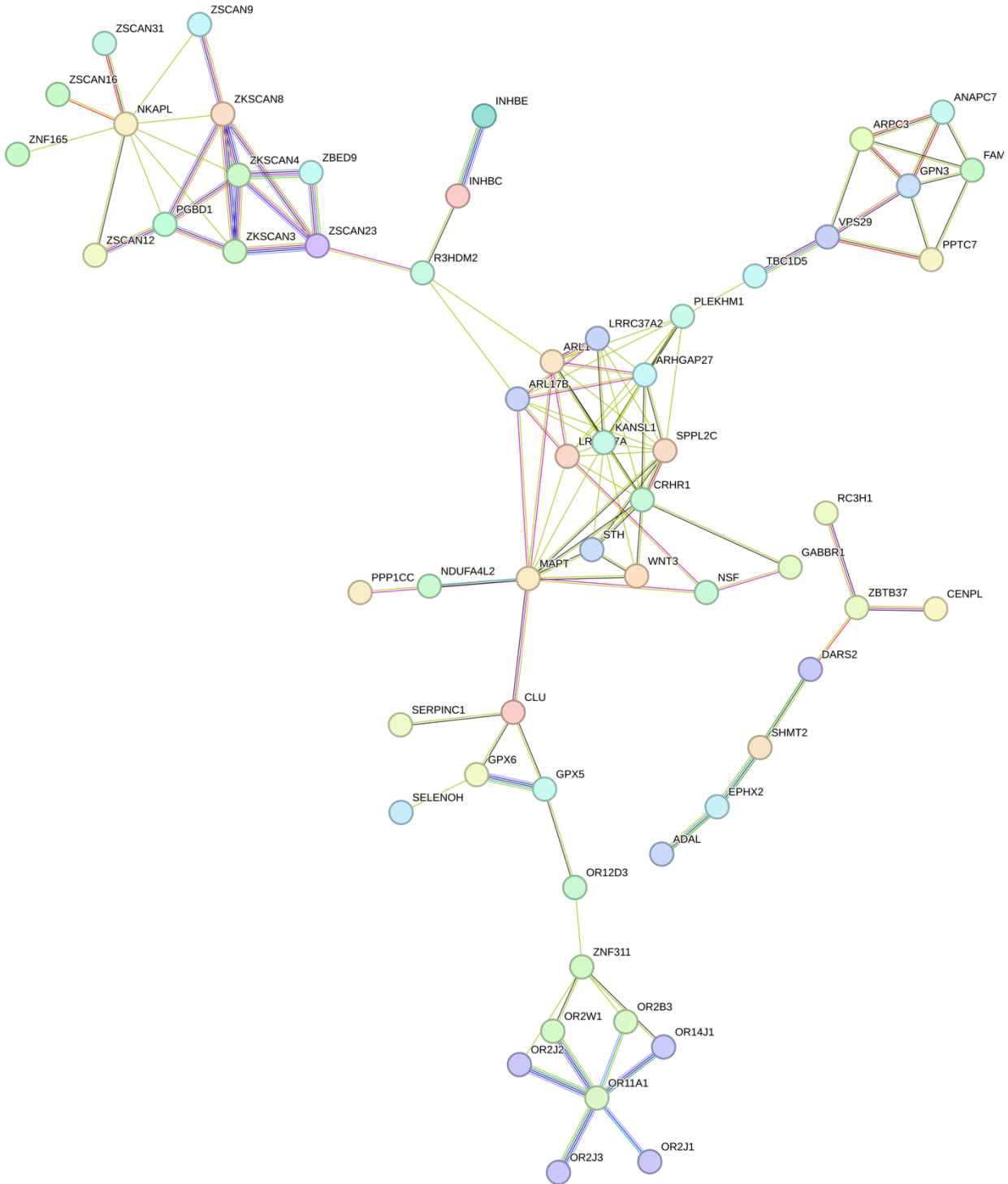
430 **Supplementary Figure 25. Protein-protein interaction network plot for TUD Independent.**

431 Network plot was generated using STRING database v12.0. Nodes represent proteins, and edges

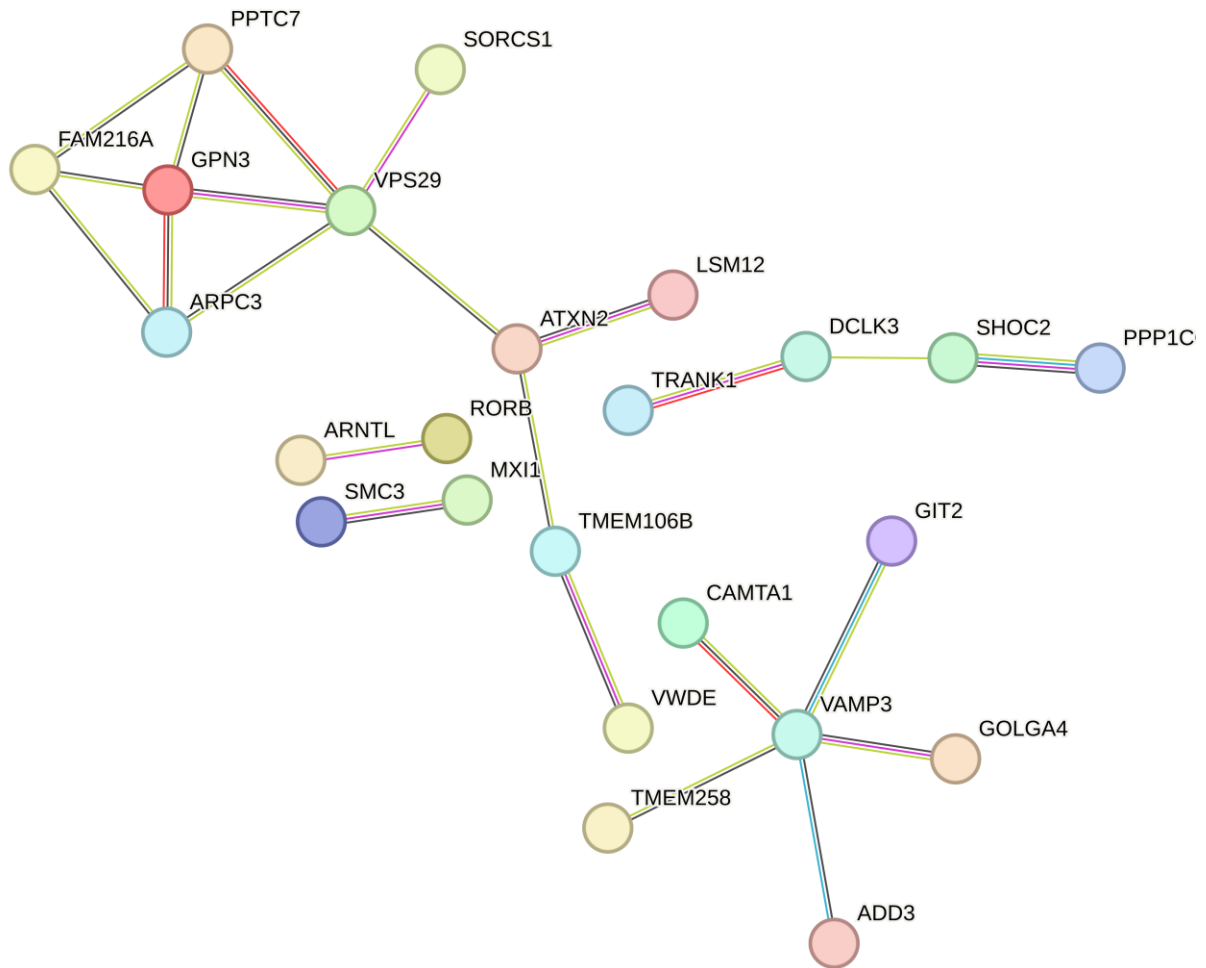
432 represent protein-protein associations. Blue and pink edges represent known interactions, while

433 green, red, and blue represent predicted interactions.

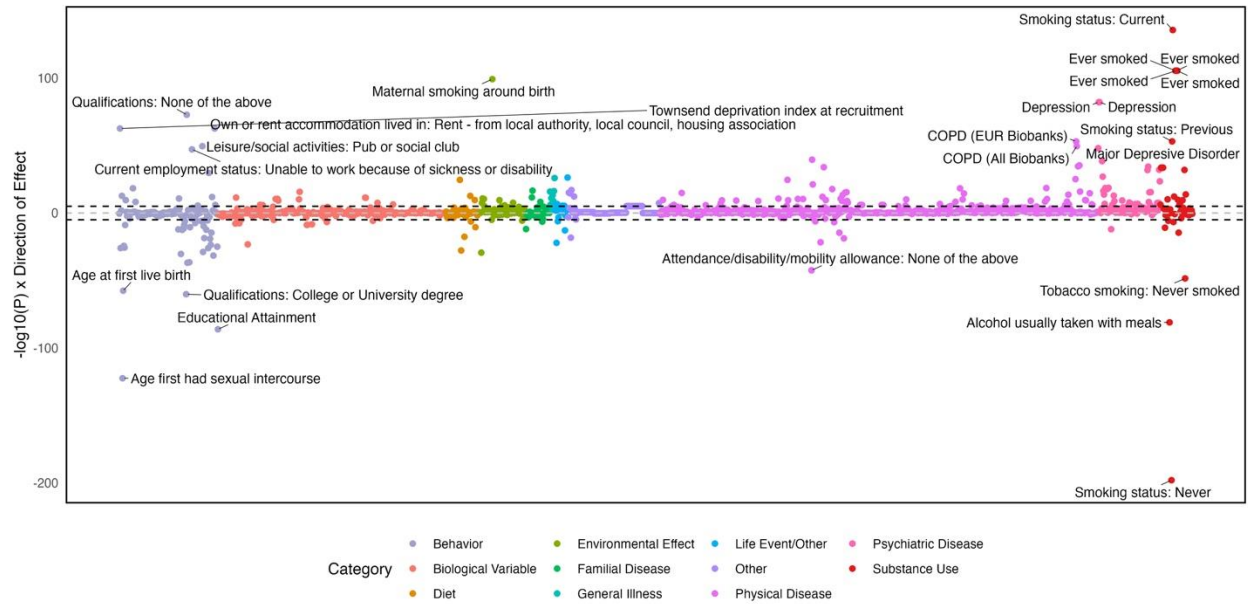
434



435
 436 **Supplementary Figure 26. Protein-protein interaction network plot for SCZ Independent.**
 437 Network plot was generated using STRING database v12.0. Nodes represent proteins, and edges
 438 represent protein-protein associations. Blue and pink edges represent known interactions, while
 439 green, red, and blue represent predicted interactions.
 440



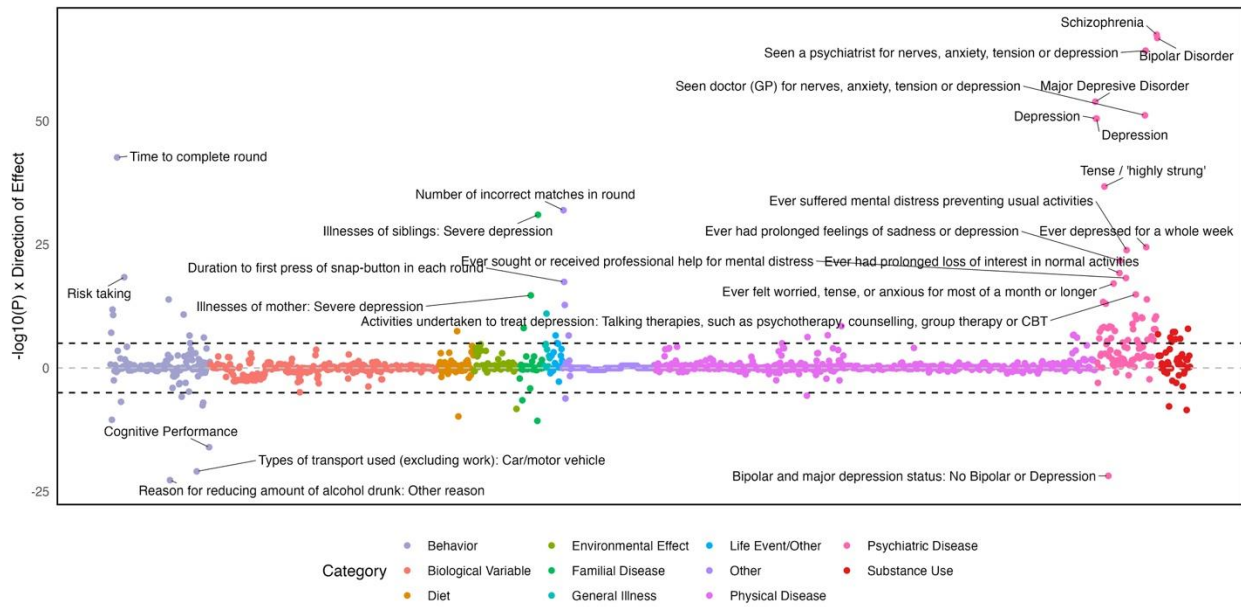
441
 442 **Supplementary Figure 27. Protein-protein interaction network plot for BD Independent.**
 443 Network plot was generated using STRING database v12.0. Nodes represent proteins, and edges
 444 represent protein-protein associations. Blue and pink edges represent known interactions, while
 445 green, red, and blue represent predicted interactions.



446
 447
 448
 449
 450
 451

Supplementary Figure 28. Genetic correlation results for the EUR substance use disorders factor

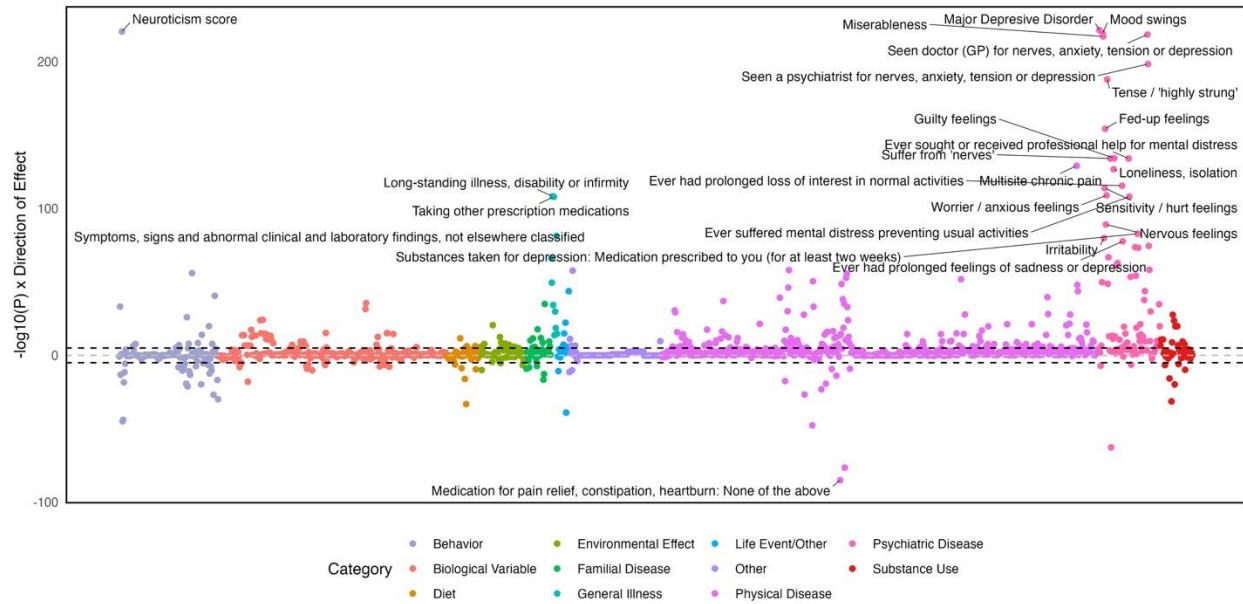
The top 25 associations are shown. Association analyses were performed using the MASSIVE pipeline.



453

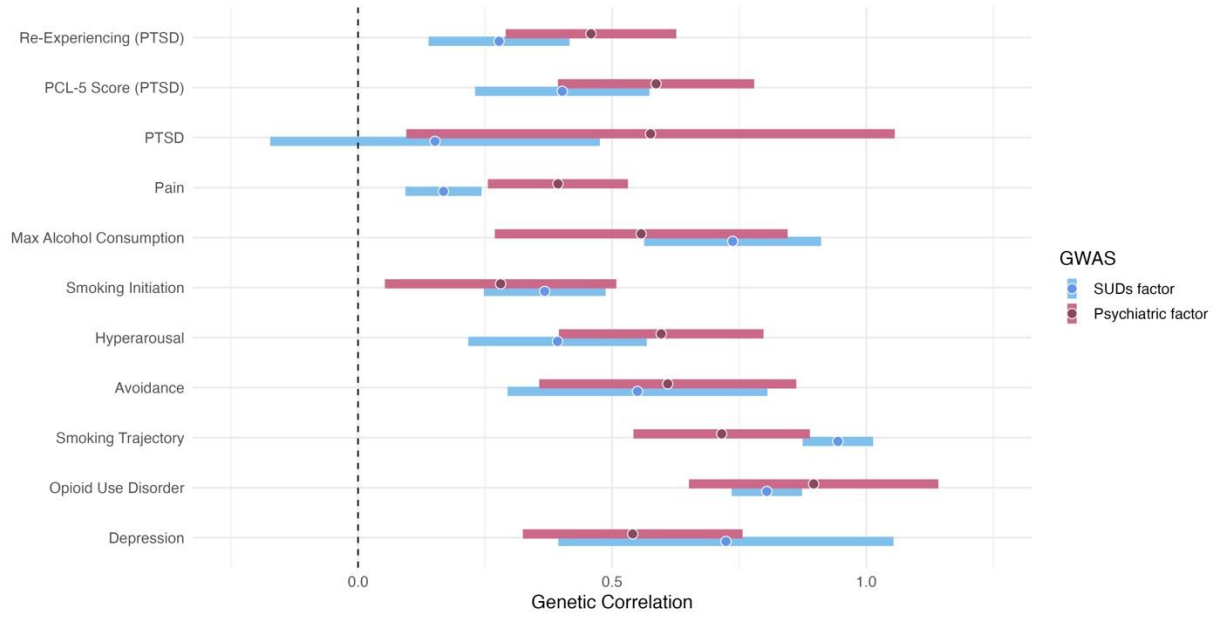
454 **Supplementary Figure 29. Genetic correlation results for the EUR psychotic disorders**
 455 **factor**

456 The top 25 associations are shown. Association analyses were performed using the MASSIVE
 457 pipeline.
 458



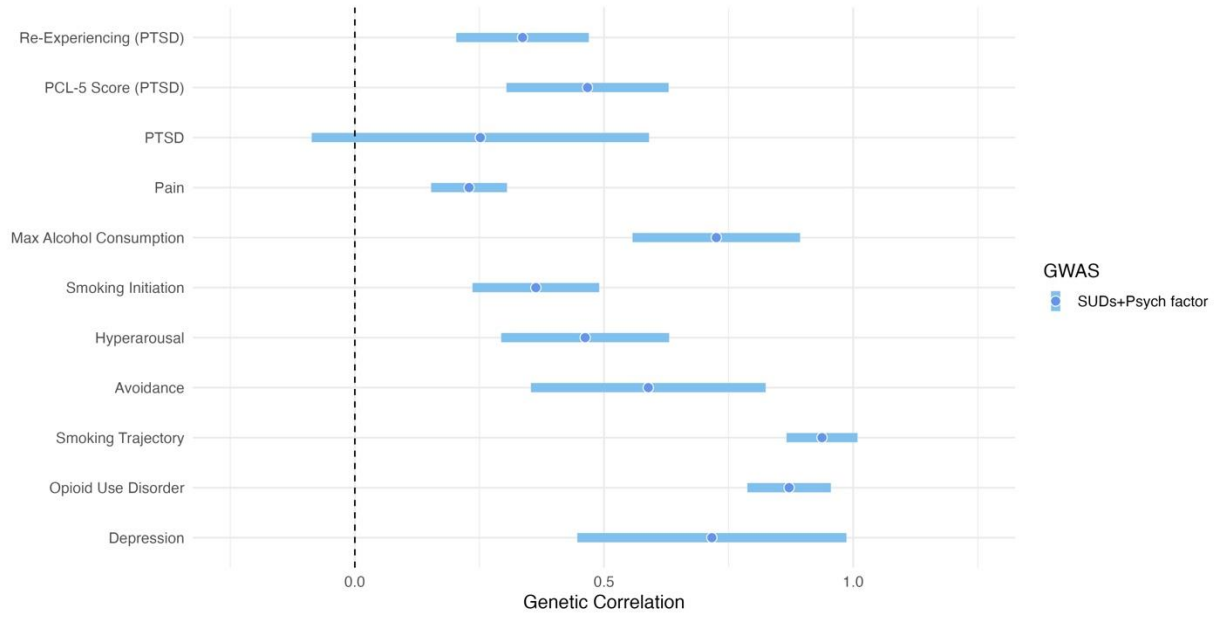
459
460
461
462
463

Supplementary Figure 30. Genetic correlation results for the EUR mood disorders factor
The top 25 associations are shown. Association analyses were performed using the MASSIVE pipeline.



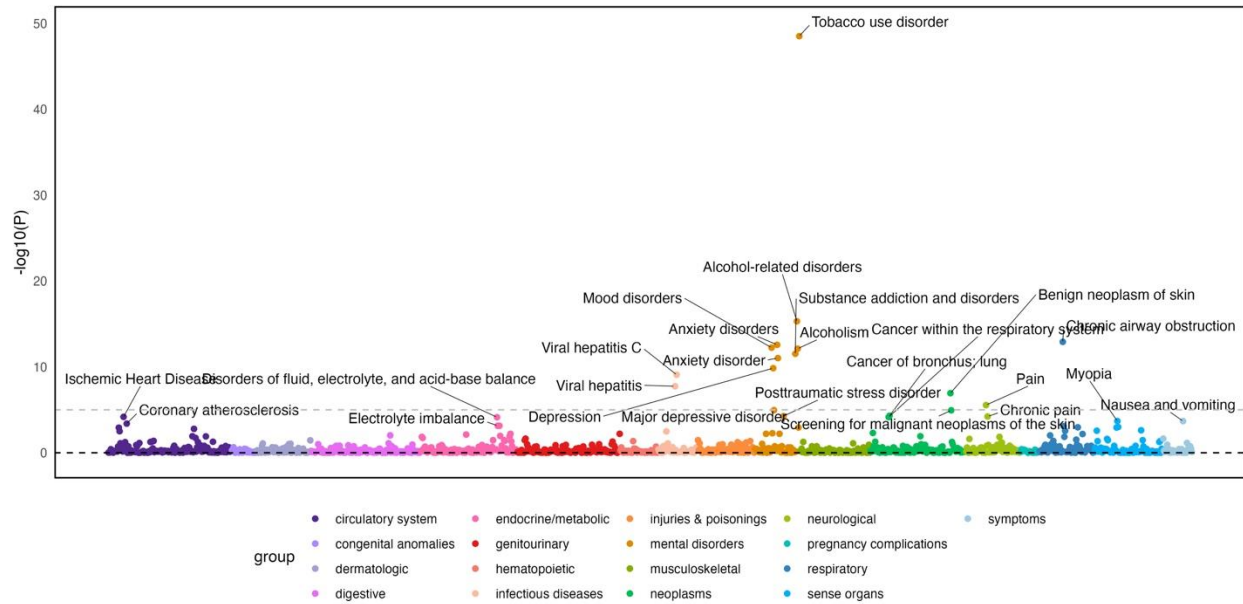
464
 465
 466
 467
 468

Supplementary Figure 31. Genetic correlations between AFR ancestry common factors and psychiatric and substance use phenotypes



469
 470
 471
 472
 473

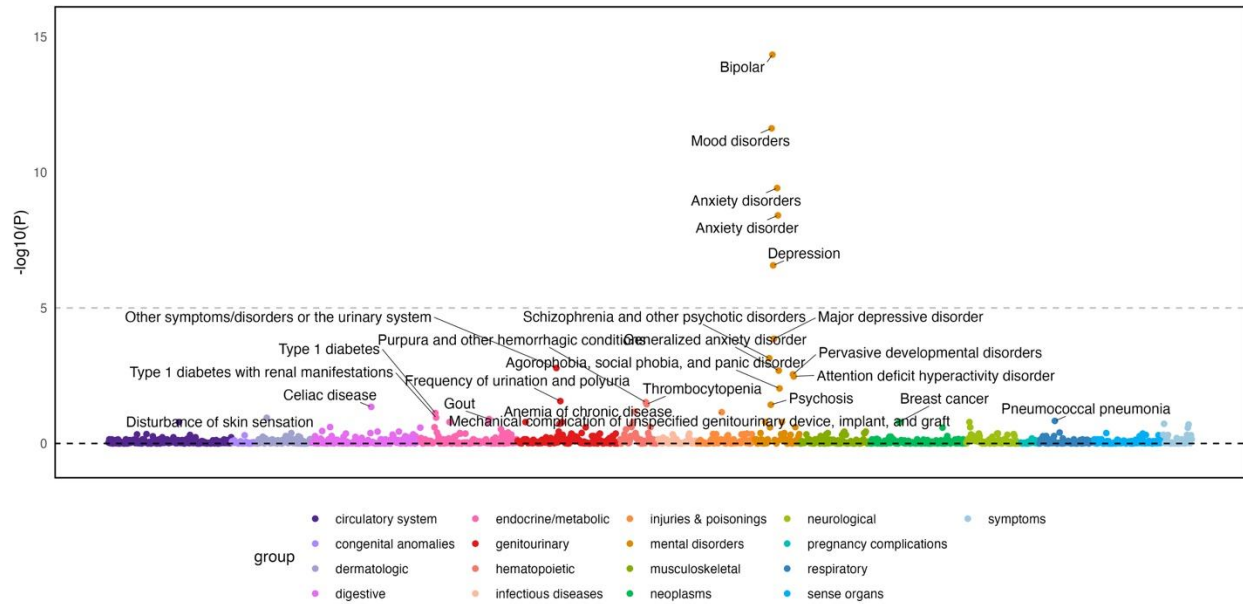
Supplementary Figure 32. Genetic correlations between the AFR ancestry second-order common factor and psychiatric and substance use traits



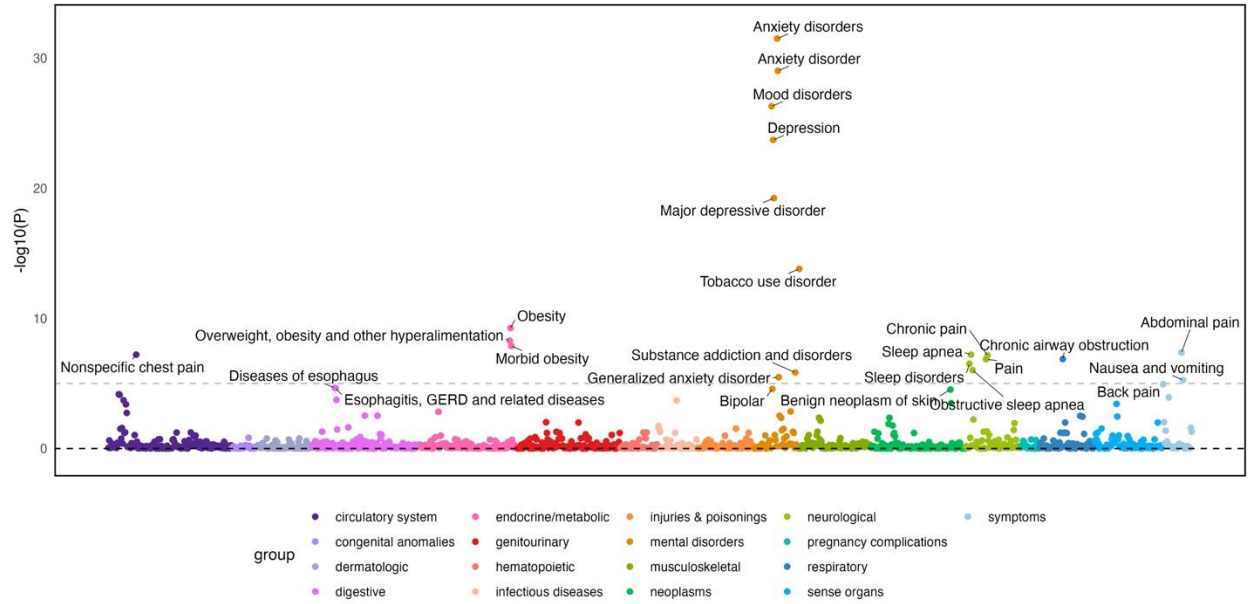
474
475
476
477
478
479

Supplementary Figure 33. PheWAS results for the EUR substance use disorders factor in Penn Medicine BioBank

The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false discovery rate (FDR) correction.



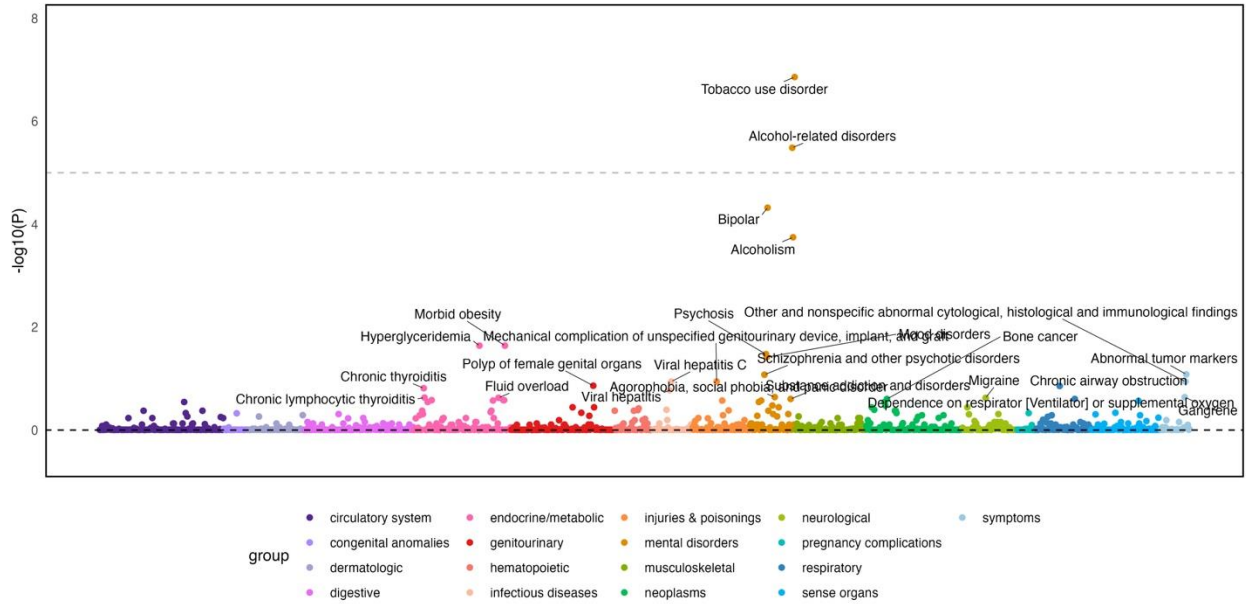
480
 481 **Supplementary Figure 34. PheWAS results for the EUR psychotic disorders factor in Penn**
 482 **Medicine BioBank**
 483 The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false
 484 discovery rate (FDR) correction.
 485



486
 487
 488
 489
 490
 491

Supplementary Figure 35. PheWAS results for the EUR mood disorders factor in Penn Medicine BioBank

The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false discovery rate (FDR) correction.



493

494

Supplementary Figure 36. PheWAS results for EUR ancestry second-order common factor representing overlap in substance use and psychotic disorders in Penn Medicine BioBank

495

The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false discovery rate (FDR) correction.

496

497

498

499

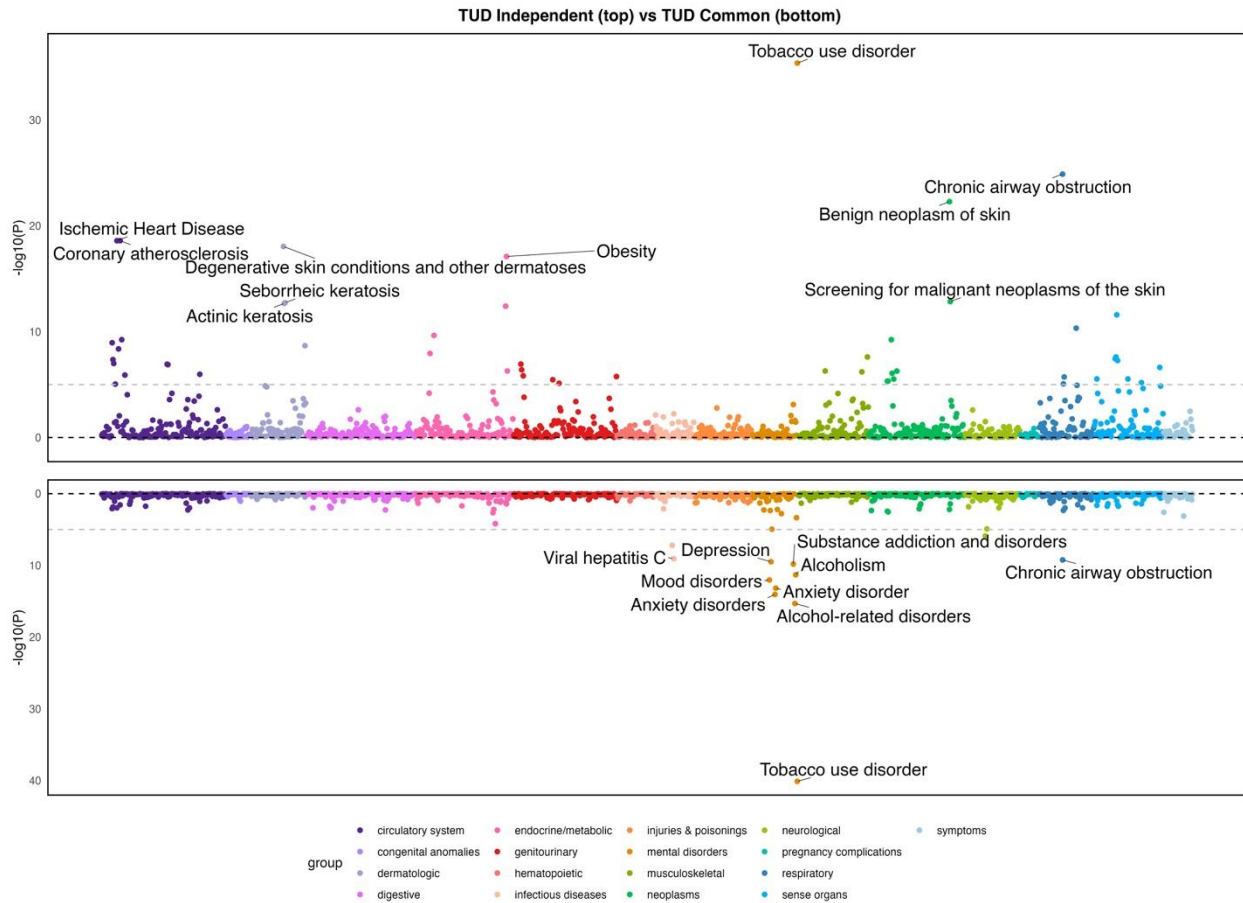


500
501
502
503
504
505
506

Supplementary Figure 37. PheWAS results for EUR ancestry second-order common factor representing overlap in substance use and mood/anxiety disorders in Penn Medicine BioBank

The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false discovery rate (FDR) correction.

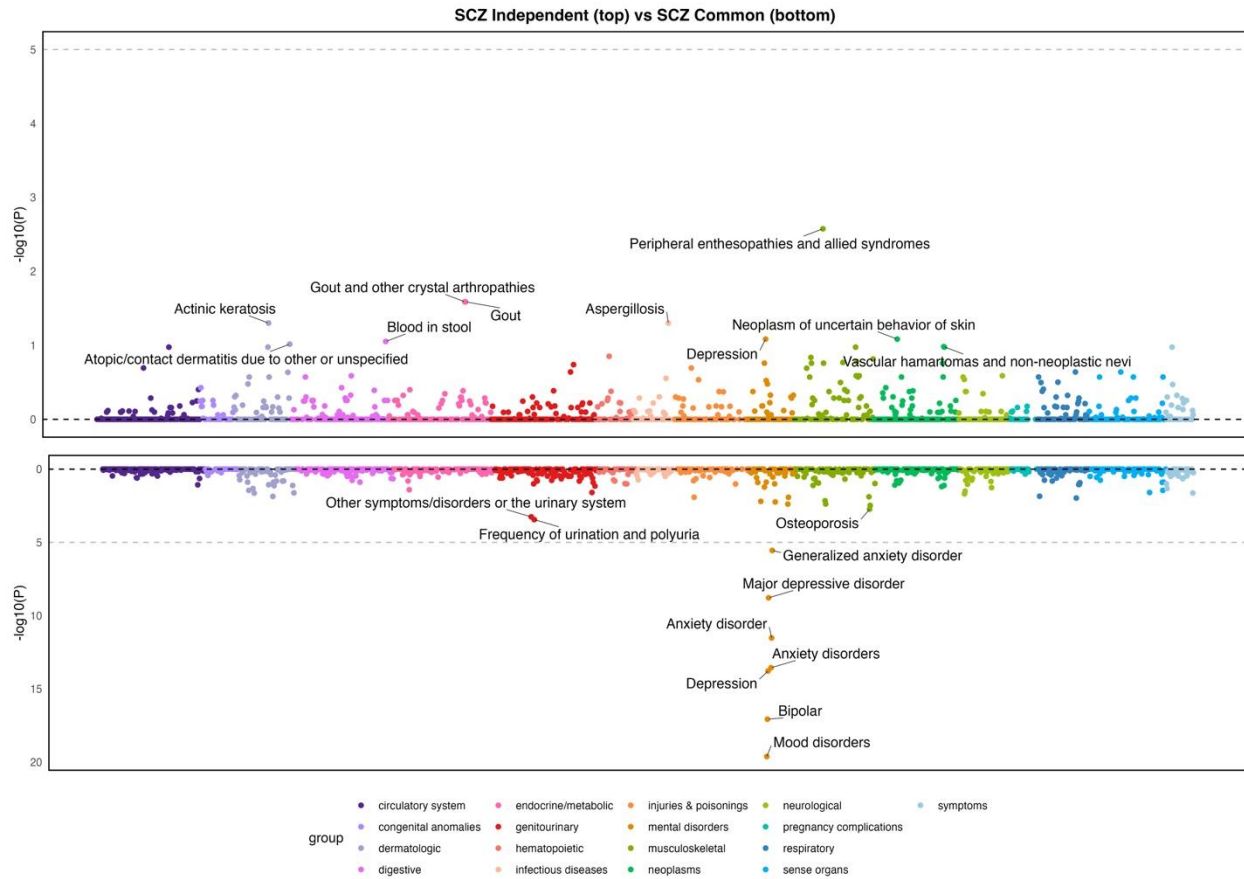
507



508

509 **Supplementary Figure 38. Hudson plot of PheWAS results for tobacco use disorders**
510 **GWAS-by-subtraction in Penn Medicine BioBank**

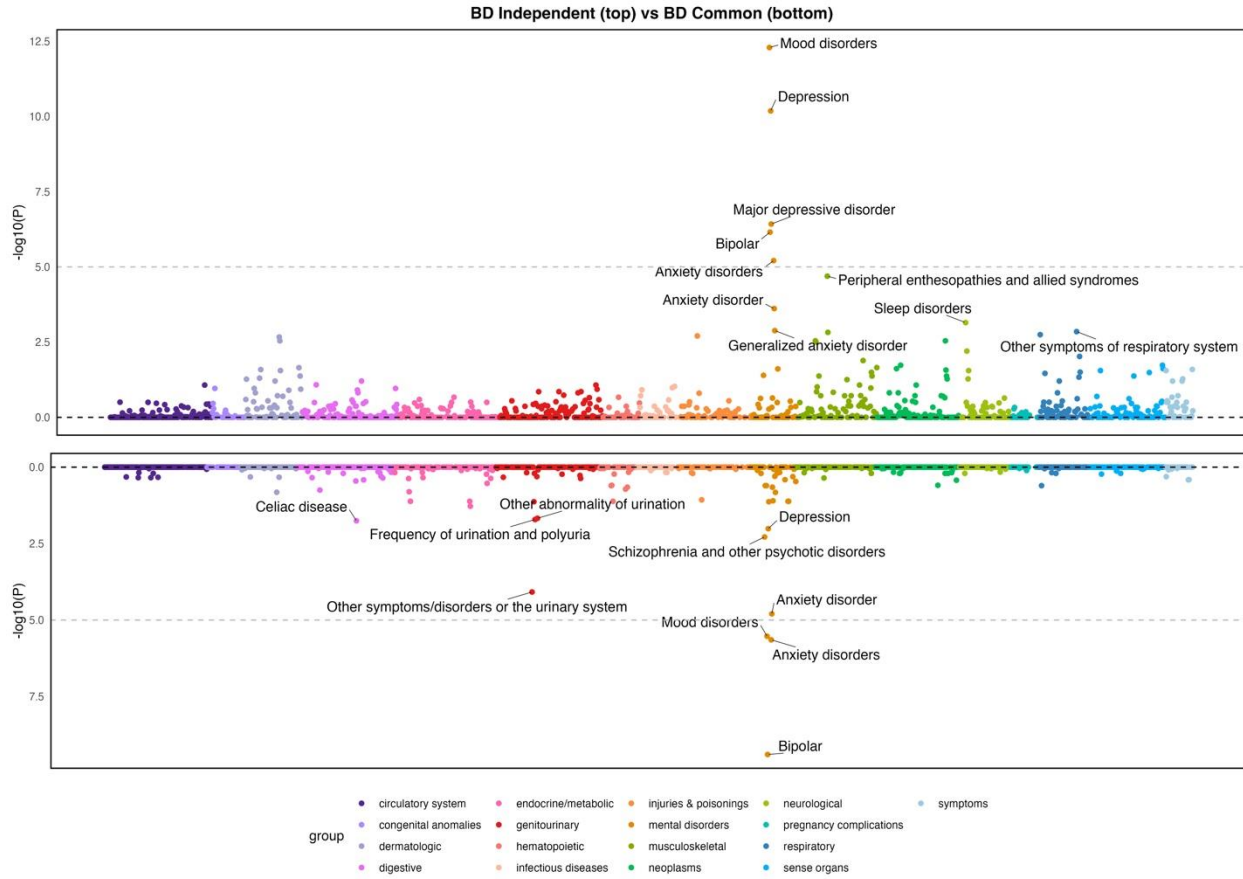
511 The top 10 associations for each GWAS are shown. All p-values were adjusted using Benjamini-
512 Hochberg false discovery rate (FDR) correction.



513

514 **Supplementary Figure 39. Hudson plot of PheWAS results for schizophrenia GWAS-by-**
 515 **subtraction in Penn Medicine BioBank**

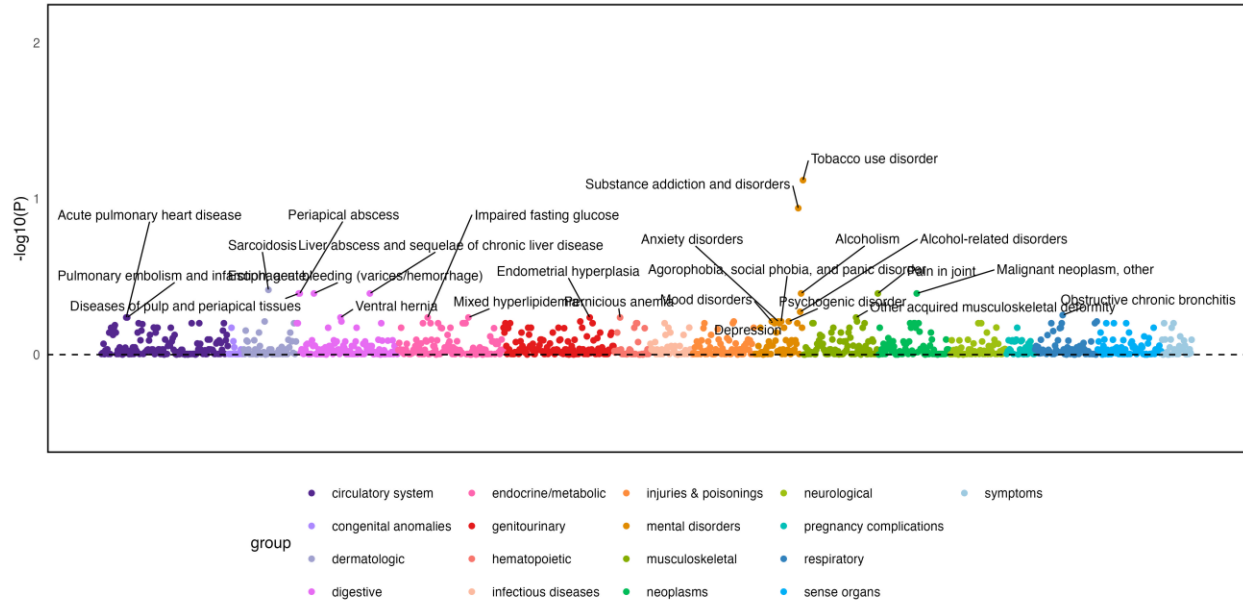
516 The top 10 associations for each GWAS are shown. All p-values were adjusted using Benjamini-
 517 Hochberg false discovery rate (FDR) correction.



518

519 **Supplementary Figure 40. Hudson plot of PheWAS results for bipolar disorder GWAS-by-**
 520 **subtraction in Penn Medicine BioBank**

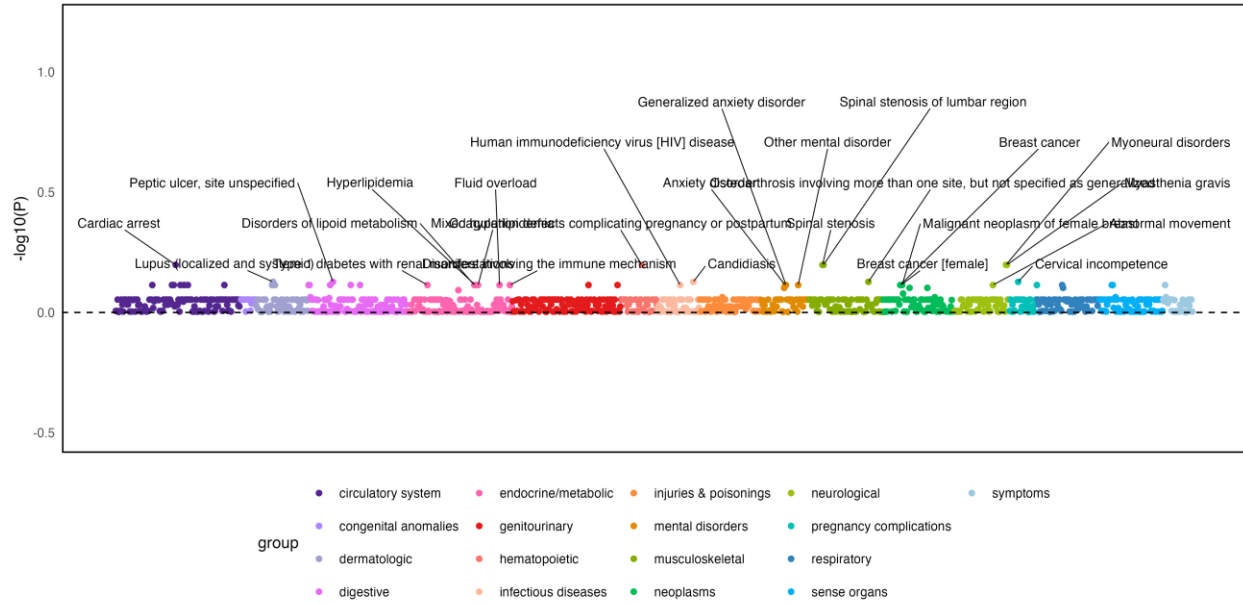
521 The top 10 associations for each GWAS are shown. All p-values were adjusted using Benjamini-
 522 Hochberg false discovery rate (FDR) correction.



523
524
525
526
527
528
529
530

Supplementary Figure 41. PheWAS results for AFR ancestry substance use disorders factor in Penn Medicine BioBank

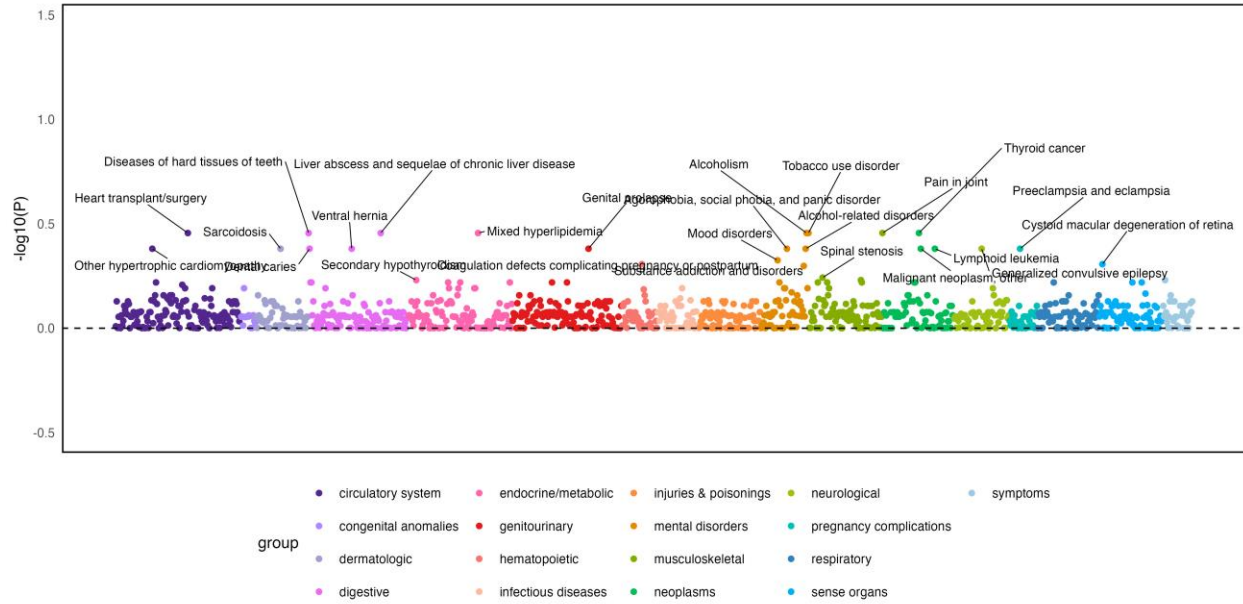
The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false discovery rate (FDR) correction.



531
 532 **Supplementary Figure 42. PheWAS results for AFR ancestry psychiatric disorders factor**
 533 **in Penn Medicine BioBank**

534 The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false
 535 discovery rate (FDR) correction.

536
 537
 538
 539



540
541
542
543
544
545

Supplementary Figure 43. PheWAS results for AFR ancestry second-order common factor representing overlap in substance use and psychiatric disorders in Penn Medicine BioBank
The top 25 associations are shown. All p-values were adjusted using Benjamini-Hochberg false discovery rate (FDR) correction.

References

- 546
547
548 1. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using
549 MTAG. *Nature Genetics* **50**, 229-237 (2018).
- 550 2. Purves, K.L. *et al.* A major role for common genetic variation in anxiety disorders.
551 *Molecular Psychiatry* **25**, 3292-3303 (2020).
- 552 3. Levey, D.F. *et al.* Reproducible genetic risk loci for anxiety: Results from ~200,000
553 participants in the Million Veteran Program. *American Journal of Psychiatry* **177**, 223-
554 232 (2020).
- 555 4. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders.
556 *Molecular Psychiatry* **21**, 1391-1399 (2016).
- 557 5. Grotzinger, A.D., Fuente, J.d.l., Privé, F., Nivard, M.G. & Tucker-Drob, E.M. Pervasive
558 downward bias in estimates of liability-scale heritability in genome-wide association
559 study meta-analysis: A simple solution. *Biological Psychiatry* **93**, 29-36 (2023).
- 560 6. Grotzinger, A.D. *et al.* Genetic architecture of 11 major psychiatric disorders at
561 biobehavioral, functional genomic and molecular genetic levels of analysis. *Nature*
562 *Genetics* **54**, 548-559 (2022).
- 563 7. Grotzinger, A.D. *et al.* Genomic structural equation modelling provides insights into the
564 multivariate genetic architecture of complex traits. *Nature Human Behaviour* **3**, 513-525
565 (2019).
- 566 8. Grotzinger, A.D. *et al.* Multivariate genomic architecture of cortical thickness and
567 surface area at multiple levels of analysis. *Nature Communications* **14**, 946 (2023).
- 568 9. Bentley, A.R., Callier, S.L. & Rotimi, C.N. Evaluating the promise of inclusion of
569 African ancestry populations in genomics. *npj Genomic Medicine* **5**, 5 (2020).
- 570 10. Bick, A.G. *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340-346
571 (2024).
- 572 11. The 1000 Genomes Project Consortium. A global reference for human genetic variation.
573 *Nature* **526**, 68-74 (2015).
- 574 12. Pan-UKB Team. <https://pan.ukbb.broadinstitute.org>. (2020).
- 575 13. Luo, Y. *et al.* Estimating heritability and its enrichment in tissue-specific gene sets in
576 admixed populations. *Human Molecular Genetics* **30**, 1521-1534 (2021).
- 577 14. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-
578 wide association summary statistics. *Nature Genetics* **47**, 1228-1235 (2015).
- 579 15. Ni, G. *et al.* Estimation of genetic correlation via linkage disequilibrium score regression
580 and genomic restricted maximum likelihood. *The American Journal of Human Genetics*
581 **102**, 1185-1194 (2018).
- 582 16. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based
583 linkage analyses. *The American Journal of Human Genetics* **81**, 559-75 (2007).
- 584 17. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition
585 resource. *Nucleic Acids Research* **51**, D977-D985 (2023).
- 586 18. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: Update 2006. *Nucleic Acids*
587 *Research* **34**, D590-D598 (2006).
- 588 19. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex
589 traits. *Nature Genetics* **51**, 1339-1348 (2019).
- 590