Title

Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses.

Journal

Nature Medicine, 29(7)

Authors

Als, Thomas

Kurki, Mitja

Grove, Jakob

et al.

Publication Date

2023-07-01

DOI

10.1038/s41591-023-02352-1

Peer reviewed

Published in final edited form as:

Nat Med. 2023 July ; 29(7): 1832–1844. doi:10.1038/s41591-023-02352-1.

Corresponding authors: Thomas D Als tda@biomed.au.dk and Anders D Børglum anders@biomed.au.dk.

Author contributions Statement

Analysis: TDA, MK, JG, GV, KT, ET, TTN, JN, KV, DL, JB, and BZ

Sample and/or data provider and processing: J.B.G., D.D., A.R., G.A., M.B.H., P.Q., G.B.W., T.T., H.S., K.L.M., V.M., L.F., J.T., B.J.V., J.J.M., M.M, S.M., iPSYCH-Broad Consortium, E.A., K.S., M.M., T.W., D.M.H., P.B.M., M.S., J.G., I.H., P.R., M.J.D., O.M., A.P., and A.D.B.

Supervision: P.Q., K.L.M., B.J.V., J.J.M, I.H., P.R., M.J.D., O.M., A.P., and A.D.B.

Writing: T.D.A. and A.D.B.

Study design and direction: A.D.B.

All authors contributed with critical revision of the manuscript.

Ethics statement

The iPSYCH study was approved by the Scientific Ethics Committee in the Central Denmark Region (Case No 1-10-72-287-12) and the Danish Data Protection Agency. In accordance with Danish legislation, the Danish Scientific Ethics Committee has, for this study, waived the need for specific informed consent in biomedical research based on existing biobanks.

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017. The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020,,THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/202, Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20) TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4th July 2019.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 9 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, Auria Biobank AB17-5154 and amendment #1 (August 17 2020), AB20-5926 and amendment #1 (April 23 2020) and it's modification (Sep 22 2021), Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 §/2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020.

Research involving MVP[24,112] in general is approved by the Veteran Affairs Central Institutional Review Board (MVP025 19-02); the project was also approved by Central Istitutional Review Board in Boston, San Diego, and West Haven.

Code availability

Software and analytical methods used in data analyses includes: Ricopili: https://sites.google.com/a/broadinstitute.org/ricopili/

Eigensoft 6.1.4: https://www.hsph.harvard.edu/alkes-price/software/ and https://github.com/chrchang/eigensoft

Eagel v2.3.5: https://github.com/poruloh/Eagle

Minimac3: https://github.com/Santy-8128/Minimac3

METAL version 2011-03-05: https://genome.sph.umich.edu/wiki/METAL_Documentation

Genotyping and imputation with the Finnish population-specific SISu v3 reference panel: https://www.protocols.io/view/genotype-imputation-workflow-v3-0-xbgfijw.

SAIGE v0.20: Scalable and Accurate Implementation of Generalized mixed model v0.20: https://github.com/weizhouUMICH/SAIGE/

LDsc v1.0.1: https://github.com/bulik/ldsc

MiXeR v1.3: https://github.com/precimed/mixer

LAVA 2022-09-29 https://ctg.cncr.nl/software/lava and https://github.com/josefin-werme/LAVA. https://github.com/josefin-werme/LAVA The method and code used for genome partitioning used for the LAVA analyses is available at https://github.com/cadeleeuw/lava-partitioning.

GCTA v1.93.2: https://yanglab.westlake.edu.cn/software/gcta/#Overview

LDpred2: https://privefl.github.io/bigsnpr/articles/LDpred2.html

FUMA v1.3.7: https://fuma.ctglab.nl/

S-PrediXcan: https://github.com/hakyimlab/MetaXcan

eCAVIAR: http://zarlab.cs.ucla.edu/tag/ecaviar/

coloc v2: https://github.com/Stahl-Lab-MSSM/coloc2

CAUSALdb-finemapping-pipeline: https://github.com/mulinlab/CAUSALdb-finemapping-pip.

Variance partition fraction for GLM with linear, logistic or probit regression used for PNC analysis: https://github.com/GabrielHoffman/misc_vp/blob/master/calcVarPart.R

Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses

Thomas D. Als[1,2,3], **Mitja Kurki**[4], **Jakob Grove**[1,2,3,5], **Georgios Voloudakis**[6,7,8,9,10,11], **Karen Therrien**[6,7,8,9,10,11,12], **Elisa Tasanko**[13], **Trine Tollerup Nielsen**[1,2,3], **Joonas Naamanka**[13], **Kumar Veerapen**[14,15,16], **Daniel Levey**[17,18], **Jaroslav Bendl**[6,7,8,9,10], **Jonas Bybjerg-Grauholm**[2,19], **Biao Zheng**[6,7,8,9,10], **Ditte Demontis**[1,2,3], **Anders Rosengren**[2,20], **Georgios Athanasiadis**[2,20], **Marie Bækved-Hansen**[2,19], **Per Qvist**[1,2,3], **G. Bragi Walters**[21], **Thorgeir Thorgeirsson**[21], **Hreinn Stefánsson**[21], **Katherine L Musliner**[2,22,23,24], **Veera M. Rajagopal**[1,2,3], **Leila Farajzadeh**[1,2,3], **Janne Thirstrup**[1,2,3], **Bjarni J. Vilhjálmsson**[2,5,25], **John J. McGrath**[25,26,27], **Manuel Mattheisen**[2], **Sandra Meier**[2], **Esben Agerbo**[2,22,23], **Kári Stefánsson**[21], **Merete Nordentoft**[2,28], **Thomas Werge**[2,20], **David M. Hougaard**[2,19], **Preben B. Mortensen**[2,22,23], **Murray B. Stein**[29,30], **Joel Gelernter**[17,18], **Iiris Hovatta**[13], **Panos Roussos**[6,7,8,9,10,11,31], **Mark J. Daly**[4,14,15,16], **Ole Mors**[2,32], **Aarno Palotie**[4], **Anders D. Børglum**[1,2,3]

[1]Department of Biomedicine, Aarhus University, Aarhus, Denmark

[2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark

[3]Center for Genomics and Personalized Medicine, Aarhus, Denmark

[4]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

[5]Bioinformatics Research Centre, Aarhus University, 8000 Aarhus C, Denmark

[6]Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, NY, USA

[7]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, NY, USA

[8]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[9]Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, NY, USA

[10]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[11]Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA

[12]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[13]Department of Psychology and Logopedics, SleepWell Research Program, University of Helsinki, Finland

Cox's proportional hazard modelling: the R-packages survival 3.4-0 (https://CRAN.R-project.org/package=survival). R v4.2.2 was used in general for statistical analyzes and plotting (https://www.Rproiect.org).

[14]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[15]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[16]Department of Medicine, Harvard Medical School, Boston, MA, USA

[17]Division of Human Genetics, Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

[18]Department of Psychiatry, Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA

[19]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

[20]Mental Health Centre Sct. Hans, Capital Region of Denmark, Institute of Biological Psychiatry, Copenhagen University Hospital, Copenhagen, Denmark

[21]deCODE genetics / Amgen, Reykjavik, Iceland

[22]NCRR - National Centre for Register-Based Research, Business and Social Sciences, Aarhus University, Aarhus, Denmark

[23]Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Aarhus, Denmark

[24]Department of Affective Disorders, Aarhus University Hospital-Psychiatry, Aarhus Denmark

[25]National Centre for Register-Based Research, Aarhus University, 8210 Aarhus V, Denmark

[26]Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Brisbane, QLD 4076, Australia

[27]Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia

[28]Mental Health Centre Copenhagen, Capital Region of Denmark, Copenhagen University Hospital, Copenhagen, Denmark

[29]Psychiatry Service, VA San Diego Healthcare System, San Diego, CA, USA

[30]Departments of Psychiatry and Herbert Wertheim School of Public Health, University of California, San Diego, La Jolla, CA, USA

[31]Center for Dementia Research, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

[32]Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Denmark

Abstract

Depression is a common psychiatric disorder and a leading cause of disability worldwide. We conducted a GWAS meta-analysis of 6 datasets including >1.3 million individuals, hereof 371,184 with depression, and identified 243 risk loci. 64 loci were novel, including genes encoding glutamate and GABA receptors that are targets for antidepressant drugs. Intersection with functional genomics data prioritized likely causal genes and revealed novel enrichment of

prenatal GABAergic neurons, astrocytes and oligodendrocyte lineages. We found depression to be highly polygenic, with ~11,700 variants explaining 90% of the SNP-heritability, estimating that >95% of risk variants for other psychiatric disorders (anxiety, schizophrenia, bipolar disorder and ADHD) were influencing depression risk when both concordant and discordant variants were considered, and nearly all depression risk variants influenced educational attainment. Additionally, depression genetic risk was associated with impaired complex cognition domains. We dissected the genetic and clinical heterogeneity, revealing distinct polygenic architectures across subgroups of depression and demonstrating significantly increased absolute risks for recurrence and psychiatric comorbidity among depression cases with the highest polygenic burden, with considerable sex-differences. The risks were up to 5- and 32-fold higher than cases with the lowest polygenic burden and the background population, respectively. These results deepen the understanding of the biology underlying depression, its disease progression and inform precision medicine approaches to treatment.

INTRODUCTION

Depression is a genetic and phenotypic complex psychiatric disorder with a lifetime prevalence of 15-20%[1-3]. It is often recurrent and accompanied by considerable morbidity and co-morbidity, excess mortality, increased risk of suicide and substantial costs worldwide[4-8]. Individuals diagnosed with depression have an increased risk of developing practically all other types of mental disorders, in particular anxiety, bipolar disorder, schizophrenia and substance use disorder (SUD)[9,10]. This relationship is generally bi-directional (i.e. a significantly increased risk of comorbidity between depression and other types of psychiatric disorders, regardless of whether depression is diagnosed before or after a comorbid psychiatric disorder).

Heritability estimates based on twin studies (h^2=0.37) have indicated that familial aggregation of depression is influenced by additive genetic effects[11], and several previous studies have documented a considerable genetic overlap between depression and multiple psychiatric as well as somatic disorders and traits[12-17]. Despite the substantial heritability of depression and other psychiatric disorders, the potential for translating genetic insights into precision psychiatry has yet to be fulfilled, including demonstrating clinical utility of polygenic risk scores (PRS)[18-20].

Major advancement in understanding the genetic architecture of depression has been achieved primarily via genome-wide association studies (GWAS) led by the Psychiatric Genomics Consortium (PGC). In 2021, GWAS results from the PGC[12], UK Biobank (UKB)[14,21], FinnGen[24] and 23andMe, Inc.[24] were combined with data from the Million Veteran Program (MVP) in a large meta-analysis[25], identifying 178 risk loci. The identified loci explain a small fraction of the overall heritability of depression[12,21,25] and even larger GWASs are needed to further elucidate genetic factors contributing to the risk of developing depression and advance genetically informed patient stratification and outcome prediction towards clinical utility.

Here, we present a large GWAS meta-analysis of depression which included expanded iPSYCH[26] and FinnGen[23] cohorts, as well as PGC, UKB, 23andMe and MVP data[25],

revealing numerous novel risk loci and infer pathophysiologic implications of associated variants by intersecting with functional genomics data. We refine the genetic architecture of depression and case subgroups and demonstrate the impact of depression genetic risk on domains of cognitive performance. Leveraging nation-wide longitudinal health data on the Danish iPSYCH cohort, we dissect the genetic architecture of single-episode and recurrent depression as well as individuals who have developed anxiety, bipolar disorder, schizophrenia and SUD comorbidities. Furthermore, to inform precision psychiatry approaches, we calculate time-dependent absolute risks and hazard rate ratios for developing recurrent depression, anxiety, bipolar disorder, schizophrenia and SUD depending on different polygenic burdens of depression

RESULTS

Genome-wide association

We analyzed data from the large population-based case-cohort of iPSYCH[26,27], which include genotypes from all individuals born in Denmark between 1981 and 2008 who have received treatment for depression in hospitals and outpatient clinics (ICD-10 codes F32-F33 in the Danish Psychiatric Central Research Register[28]). Compared to the latest GWAS of depression (from May 2021)[25], which included 18,629 cases and 17,841 controls from the initial iPSYCH2012 cohort[12,27], data from 11,710 individuals with depression and 18,410 controls were included from the expanded iPSYCH2015 cohort[26,27], which leads to a total of 30,618 individuals with depression and 38,200 controls after relatedness pruning and removal of ancestry outliers. In addition, we included an updated dataset consisting of 28,098 individuals with depression and 228,817 controls from the FinnGen[23] study. When combining these with data from previously published samples from the PGC, UKB, 23andMe and MVP, the number of samples added up to a total of 371,184 individuals with depression and 978,703 controls (Supplementary Table S1).

We performed a variance weighted fixed effects meta-analysis using METAL[29], testing the effects of 6,037,120 single nucleotide polymorphisms (SNPs) common across all datasets. This revealed a total of 303 genome-wide significant linkage disequilibrium (LD)-independent (r^2<0.1) lead variants located in 243 distinct loci. A conditional association analysis using GCTA-COJO[30,31] retained 251 independent SNPs in the 243 loci. Manhattan plots are shown in Figure 1A,B, regional plots are provided in Supplementary Figure S1 (X chromosome in Supplementary Figure S2) and details on lead variants are provided in Supplementary Table S2A, while GCTA-COJO independent variants are listed in Supplementary Table S2B. No statistically significant heterogeneity was observed between the datasets (Supplementary Figure S3). 64 of the 243 loci were novel, i.e., not overlapping with the two most recent depression meta-analyses[14,25] (Supplementary Figure S4 and Table S2A). The three most significant loci were located near *NEGR1*, in *SORCS3* and in the *HIST1* histone cluster, respectively. Among the novel loci, the three strongest associations were in *BPTF, LINGO1* and *GRIA1* (Table 1). All three genes have been associated with monogenic forms of neurodevelopmental disorders[32-34] and *GRIA1*, encoding glutamate ionotropic receptor AMPA type subunit 1 (GluA1), implicates a role of AMPA receptors in the etiology of depression. We also note that the seventh-strongest novel locus is located in

GABRA1, which suggests a role of GABA receptors in developing depression. Both AMPA and GABA receptors are targets for antidepressants[35,36].

Genetic correlations and heritability

Analyzing the GWAS meta-analysis by LD score regression (LDsc)[37] produced a genomic inflation factor (λ_{GC}) estimate of 1.89 with an intercept of 1.06 (SE=0.01) and an attenuation ratio of 0.047 (SE=0.012) (Supplementary Table S3), indicating that 95% of the observed inflation of the test statistics (Supplementary Figure S5) is due to a polygenic signal rather than population structure. The individual GWASs (iPSYCH2015, Howard et al. 2019[14], Wray et al. 2018[12], Levey et al. 2020[25] and FinnGen[23]) showed significant pairwise genetic correlations, ranging from r_G=0.77 to r_G=0.95 (Supplementary Table S4 and Figure S6), thus supporting that the GWAS results could be combined in a meta-analysis. We note that the SNP-heritability estimate[37] for the iPSYCH cohort ($h^2_{SNP} = 0.167$, SE=0.014, prevalence=0.2) was significantly higher (range of h^2_{SNP} difference: 0.057 - 0.098) compared to the other cohorts (Supplementary Table S3 and Figure S7). This may reflect that the sample from the iPSYCH cohort is more homogeneous and includes a relatively young population, with early onset and severe cases of depression who have been treated in hospitals[38].

We investigated genetic correlations between depression and other phenotypes available at LD Hub[39] and in-house, including 258 published GWASs and 597 GWAS results for UKB traits. Depression was significantly correlated ($P<2 \times 10^{-4}$) with 364 phenotypes representing several domains and confirmed previous observations[12-15,40-45] (Supplementary Table S5A and Figure S8A). Among psychiatric disorders, depression showed significant correlation with e.g., ADHD (r_G=0.56, SE=0.022, $P=1 \times 10^{-135}$), autism (r_G=0.35, SE=0.033, $P=6.5 \times 10^{-24}$), bipolar disorder (r_G=0.31, SE=0.033, $P=3 \times 10^{-18}$), schizophrenia (r_G=0.33, SE=0.021, $P=4 \times 10^{-53}$), anxiety (r_G=0.79, SE=0.017, $P=3.2 \times 10^{-193}$), alcohol dependence (r_G=0.65, SE=0.097, $P=4.3 \times 10^{-9}$) and cannabis use disorder (r_G=0.44, SE=0.036, $P=2.2 \times 10^{-31}$). In UKB data, the three strongest genetic correlations were seen for the phenotypes: “Seen doctor (GP) for nerves_ anxiety_ tension or depression”, “Mood swings” and “Miserableness” (r_G=0.96, 0.71, 0.71, respectively; Supplementary Table S5B and Figure S8B).

To dissect the genetic overlap observations further, we used uni- and bivariate gaussian mixture modeling, as implemented in MiXeR[46], to quantify the actual number of variants that (i) explain 90% of the SNP-heritability of depression, and (ii) overlap between depression and depression-correlated phenotypes. For comparison, we also included two phenotypes showing low genetic correlations with depression (height and Alzheimer’s disease) and one with moderate correlation (epilepsy). 11,750 (SE=310) common variants were estimated to confer liability to depression, suggesting that depression is the most polygenic of the major psychiatric disorders evaluated, showing number of risk variants ranging between 7,000-10,500[46,47] (Figure 2, Supplementary Figure S9-1 and Supplementary Table S6A). MiXeR considers all variants irrespective of the direction of correlation (i.e. both variants with same and opposite direction of effects, hereafter collectively referred to as “influencing” variants). The vast majority of variants conferring

risk to the other psychiatric disorders investigated were found to influence depression (range 87-99%; Figure 2; Supplementary Table S6A), most pronounced for anxiety, schizophrenia, bipolar disorder and ADHD[47] with 95-99% of their risk variants also influencing depression. The other correlated traits also showed substantial overlap with depression (range 87-97%), while height and Alzheimer's disease did not. We note that nearly all (99%) of the depression risk variants were found to influence educational attainment. The fraction of variants affecting both traits in the same (concordant) direction varied considerably; lowest for educational attainment (42%) and highest for SUD (86%) and anxiety (89%).

Local genetic correlations in 2,495 loci analyzed using LAVA[48] (Supplementary Figure S9-2 and Table S6B) generally supported the results of the MiXeR analyses. The proportion of loci with positive genetic correlations was significantly correlated with the proportion of MiXeR-estimated influencing variants with concordant effects (Pearson's correlation r=0.76, P=0.0069, supplementary figure S9-3). Although LAVA estimates generally appear to be more extreme than MiXeR, these findings support the validity of MiXeR estimates of mixed effect directions (Supplementary Figure S9-1 and Supplementary Table S6A).

Gene-wise and pathway analysis

A genome-wide gene-based association study conducted in Multi-Marker Analysis of GenoMic Annotation (MAGMA)[49], mapped the GWAS SNPs to 17,840 protein coding genes and revealed 411 genes significantly associated with depression, after Bonferroni correction for the number of genes tested ($P<2.8 \times 10^{-6}$). A total of 314 significant genes were located in 141 of the 242 identified GWAS loci, while the remaining 97 significant genes were located outside these loci (Supplementary Table S7D). The most significant gene within each of the 141 genomic risk loci is labeled in Figure 1C.

To investigate enrichment of biological pathways, we analyzed 8,664 gene-sets derived from GO Biological Process (N=7,658) and GO Cellular Components (N= 1,006) ontology in the MSigDB database, identifying 479 significant gene-sets after correction for multiple testing (FDR<0.05, Supplementary Table S8), which included several gene-sets that have not previously shown significant enrichment in depression[14,25]. The majority of the top-ranking sets were related to neuronal development and function, including the five most significant gene-sets ($P_{adj}<10^{-14}$): GO_SYNAPSE, GO_NEURON PART, GO_SYNAPSE ORGANIZATION, GO_SYNAPSE ASSEMBLY and GO_NEURON DIFFERENTIATION.

Transcriptome-wide association and GWAS-eQTL prioritization

To identify and prioritize putative causal genes, we performed a transcriptome-wide association study (TWAS), imputing the genetically regulated gene expression using EpiXcan[50] and models trained on expression data from the PsychENCODE Consortium[51,52] for genes and isoforms detected in the dorsolateral prefrontal cortex (DLPFC). Among 34,646 transcripts (genes and isoforms) tested, we identified 2,541 transcripts at FDR<0.05 and, after Bonferroni correction of all transcripts tested, 324 transcripts from 201 genes showed significant differential imputed gene expression (Bonferroni P-value threshold $P=1.44x10^{-10}$) in DLPFC between depression and control groups (Supplementary Table S9). The Bonferroni significant transcripts were located in 88 independent regions[53]. The

top gene/isoform is labeled in Figure 1D and regional TWAS/GWAS Miami-plots for each of the 88 regions are shown in Supplementary Figure S10, six highlights are shown in Extended Data Figure 1. In 38 of the 88 regions, the top transcript was >100 times more significant than the second-most associated gene/isoform in the region (Supplementary Table S9B), appointing those as plausible causal candidates.

To further prioritize likely causal genes and variants, we performed co-localization analyses integrating fine-mapped GWAS results, using the CAUSALdb pipeline (Methods), and eQTL data from a meta-analysis of three brain datasets[54] applying a fixed-effect model. First, we adopted the Coloc method[55,56], which revealed 13 genes with strong evidence for both GWAS-association, eQTL-association and co-localization (i.e., with a posterior probability of PPH4>0.8; Supplementary Table S10). The three top-ranked genes were: *FURIN, NEGR1* and *CKS2*. Secondly, we conducted the eQTL and GWAS CAusal Variants Identification in Associated Regions (eCAVIAR) approach[57], in which both eQTL and GWAS were fine-mapped and the product of posterior probability (CLPP) was calculated, prioritizing variants with at least a single variant with CLPP >= 0.01. The eCAVIAR approach revealed five prioritized variants located in three genes: *FURIN, GPR27* and *TCTA* (Supplementary Table S11 and Figure S11).

Tissue and cell type enrichment

We next tested whether the GWAS results were enriched with respect to the transcriptomic profiles of human tissues. At the specific tissue level, we found significant enrichment exclusively in brain tissues, including all the brain tissues analyzed (Bonferroni corrected P-value threshold P=0.00093, with P-values ranging from $P=1.17x10^{-4}$ to $P=8.36x10^{-13}$ for the brain tissues, Supplementary Figure S12). Cell-type enrichment analyses revealed experiment-wide significant association (across all 13 datasets tested) of primarily neuronal cell-types, including dopaminergic and GABAergic neurons (P-values ranging from $P=1.3x10^{-4}$ to $P=5.8x10^{-11}$ for the significant associations after Bonferroni correction $P= 1.87x10^{-4}$, Figure 3A and Supplementary Figure S13). Both GABAergic neurons and oligodendrocyte progenitor cells of the human prefrontal cortex were already enriched at prenatal stages.

To further evaluate cell-type enrichment, we intersected the GWAS results with two recent epigenomic maps of cell-specific open chromatin[58,59] using an LD score partitioned heritability approach[60]. Again, we observed a clear contrast between the enrichments in the brain and non-brain tissue (Figure 3B). Consistent with our FUMA-based results and prior reports[12,14], the strongest enrichments were measured for neuronal cell types, phenotypically manifested by severe synaptic loss and deficits in functional connectivity[61,62]. Conversely, the enrichments of genetic depression risk in astrocytes and oligodendrocyte lineages have not been reported previously, albeit having support in behavioral and postmortem studies[63-65].

Association with cognitive performance

Educational attainment was among the traits showing strong negative genetic correlation with depression (Supplementary Table S5A). To further evaluate the impact of depression

genetic risk on cognition, we analyzed the association of depression polygenic scores (DEP-PRS) with 15 cognitive measures in the Philadelphia Neurodevelopmental Cohort (PNC, N=4,973)[66,67]. Cognitive performance was measured by the Computerized Neurocognitive Battery[68], including 14 tests in 5 domains: executive-control, episodic memory, complex cognitive processing, social cognition, and sensorimotor speed. In addition, the Wide Range Achievement Test (WRAT-4)[69] was used as a proxy measure for overall IQ[67]. The depression-PRS was negatively associated with abstraction and mental flexibility (β=−0.039, SE=0.014, FDR=0.030, Figure 4, Supplementary Table S12A). Applying a narrower definition of depression (see Methods, Supplementary Table S13 and Figure S14) yielded moderately stronger negative associations, with attention, abstraction, and mental flexibility in the executive-control domain, and verbal reasoning in the reasoning domain surpassing statistical significance (Figure 4, Supplementary Table S12B). These results are consistent with observational studies reporting that individuals with depression display lower performance in cognitive domains such as executive function, memory, language and attention[70-72], and demonstrate that genetic depression risk is associated with attenuated functioning in specific cognitive domains in a community cohort of youths (aged 8-21 at enrollment).

Polygenic architecture and co-morbidity rates for recurrency

To dissect the polygenic architecture of single-episode and recurrent depression, we used dates of diagnosis in the Danish Psychiatric Central Research Register to group the total of 30,618 individuals with depression from the iPSYCH cohort into single episode (N=24,101) and recurrent depression (N=6,517) case groups (see Methods and Supplementary Table S14 for details). We conducted three GWASs: 1) single-episode vs control groups, 2) recurrent depression vs. control groups and 3) recurrent vs. single-episode depression groups (excluding all controls). SNP-heritability estimates were similar for recurrent and single-episode depression and, for the case-only analysis, not significantly different from zero (Supplementary Table S3 and Figure S7). Likewise, the genetic correlations with other phenotypes showed similar patterns for single-episode and recurrent depression (Supplementary Figure S15 and Table S15).

We next investigated the polygenic load for depression in the recurrent and single-episode case groups, using a multivariate polygenic risk score (mvPRS) approach[15] and the depression GWAS meta-analysis[14,25] without iPSYCH samples for training. This analysis showed significant association of single-episode (β=0.36, SE=0.0081, $P<10^{-7}$) and recurrent depression (β=0.43, SE=0.014, $P<10^{-7}$) case groups with PRS for depression (DEP-PRS), but with a significantly larger effect size for recurrent than for single-episode depression ($P=4.8 \times 10^{-6}$, Supplementary Figure S16-1A and Supplementary Table S16A). This observed increased polygenic load among recurrent cases reinforces previous observations[11,12,73,74].

To further dissect the genetic architecture of the two subgroups, we investigated the PRS load for psychiatric disorders and traits showing strong genetic correlation with depression including anxiety, bipolar disorder, schizophrenia, autism, ADHD, SUD, substance use and neuroticism (Supplementary Figure S16-1 and Supplementary Table S16A). The results

showed an overall pattern of increased PRS among recurrent depression compared to single-episode case groups (P=0.00075), primarily driven by significantly different burdens of depression, anxiety, bipolar disorder, and neuroticism PRS.

To complement these analyses, we compared the rates of comorbid psychiatric disorders between individuals with single-episode and recurrent depression, using data from the Danish Psychiatric Central Research Register, and found highly significant increase in comorbidity among those with recurrent depression (Extended Data Table 1). This is consistent with well-established correlations between depression symptom severity, recurrence, chronicity and comorbidities[75,76]. The most compelling differences were observed for anxiety (OR=1.64, P=2 x 10^{-53}), bipolar disorder (OR=1.86, P=3.1 x 10^{-27}), schizophrenia (OR=1.76, P=9.6 x 10^{-48}) and SUD (OR=1.64, P=3.7 x 10^{-32}).

In summary, compared to individuals with single-episode depression, those with recurrent depression showed significantly increased polygenic load particularly for depression, anxiety, bipolar disorder and neuroticism-PRS as well as higher rates of comorbid psychiatric disorders.

Risk prediction of recurrent depression

To investigate whether PRS can prospectively predict recurrence among individuals with first-onset depression, we performed Cox's regression analyses estimating hazard rate ratio (HRR) and absolute risk of developing a second episode of depression over time among individuals having a first diagnosis of depression in a hospital. This was done for groups with higher polygenic load, in PRS deciles. We found that the HRR of developing a second-episode generally increased with higher depression-PRS, most significantly for the 10^{th} DEP-PRS decile with $HRR_{10/1}$=1.33 (SE=0.06, P=2.5 x 10^{-6}) when compared to the first decile and $HRR_{10/middle}$=1.14 (SE=0.04, P=0.001) or compared with the middle (2^{nd}-9^{th}) deciles (Figure 5, Supplementary Figure S17-1 and Table S17A). Similarly, the absolute risk of a second diagnosis of depression increased with time since the first diagnosis for all depression-PRS deciles, with the trajectories for the 10^{th} decile reaching an absolute risk of 38% compared to 29% and 33% for lowest and middle deciles, respectively (Figure 5).

Prompted by the results of the mvPRS analysis, we performed the same analysis for PRSs from other psychiatric disorders, which yielded significant results only for BP-PRS (HRR_{10}=1.10, SE=0.04, P=0.021, Supplementary Figure S17-1 and Table S17A). The HRR for developing single-episode and recurrent depression in the general population can be found in Supplementary Figure S18.

Polygenic architecture of psychiatric comorbidities

Individuals diagnosed with depression are at an increased risk of developing other psychiatric disorders as documented in Extended Data table 1 and a large body of studies [9,10], but little is known about the genetic constitution of those that develop these conditions. We examined the polygenic architecture of individuals with depression who have also developed anxiety, bipolar disorder, schizophrenia and/or SUD, using mvPRS analyses and PRSs from depression and eight psychiatric phenotypes genetically correlated with depression (Supplementary Table S5A and Figure S8A). This analysis showed that the

subgroup of individuals with depression with a co-diagnosis of anxiety had a significantly increased PRS for 5 out of 9 psychiatric phenotypes tested compared to those with depression without anxiety, i.e. DEP-PRS ($P=4.9 \times 10^{-10}$), ANX-PRS ($P=1.8 \times 10^{-14}$), SZ-PRS ($P=7.4 \times 10^{-05}$), ASD-PRS (P=0.021) and Neuroticism-PRS ($P=1.5 \times 10^{-13}$) (Extended Data Figure 2, see Supplementary Table S18A for differences in effect sizes between depression-subtypes). Likewise, individuals with depression who had transitioned to bipolar disorder had a significantly increased PRS for 3 out of 9 psychiatric phenotypes tested compared to those with depression without a later bipolar disorder diagnosis, including BP-PRS ($P=6.1 \times 10^{-17}$), SZ-PRS ($P=3.7 \times 10^{-12}$) and DEP-PRS ($P=8.6 \times 10^{-5}$) (Extended Data Figure 3, see Supplementary Table S18D for differences in effect sizes between depression subtypes). Individuals with depression and a co-diagnosis of schizophrenia showed increased PRS load for all 9 psychiatric phenotypes except autism (Extended Data Figure 4, see Supplementary Table S12G for differences in effect sizes between depression subtypes), most significantly for SZ-PRS ($P=1.1 \times 10^{-14}$), BP-PRS ($P=3.0 \times 10^{-8}$), substance use (SU) PRS ($P=1.8 \times 10^{-5}$) and DEP-PRS ($P=1.9 \times 10^{-5}$). SUD comorbidities showed highly increased PRS loads for 8 out of 9 psychiatric phenotypes (Extended Data Figure 5, see Supplementary Table S18J for differences in effect sizes between depression subtypes and Supplementary Figure S19 for sex-stratified analysis), most significantly for SU-PRS ($P=7.5 \times 10^{-86}$), ADHD-PRS ($P=3.9 \times 10^{-32}$) and SUD-PRS ($P=1.8 \times 10^{-30}$).

Overall, we demonstrated that all four comorbid case groups (depression with anxiety, bipolar disorder, schizophrenia or SUD) have increased polygenic burdens of common risk variants for several psychiatric phenotypes compared to the non-comorbid depression case groups (overall mvPRS p-values: $P_{DEP\text{-}ANX}=1.2 \times 10^{-21}$, $P_{DEP\text{-}BP}=1.5 \times 10^{-15}$, $P_{DEP\text{-}SZ}=1.7 \times 10^{-15}$, $P_{DEP\text{-}SUD}=6.4 \times 10^{-97}$), revealing an overall PRS pattern that distinguishes the comorbid case groups from their non-comorbid counterparts. We also note that all three comorbid subgroups showed an increased load of DEP-PRS compared to the depression-only case groups.

Risk prediction of comorbid psychiatric disorders

To assess the utility of PRS in prediction of developing comorbid disorders, we performed Cox's regression analysis on the depression case group with and without additional diagnosis of anxiety, bipolar disorder, schizophrenia and/or SUD. In addition to PRSs for the individual disorders, we included 4 aggregate scores, one for each disorder, combining all PRSs into a single score weighted by their association with anxiety, bipolar disorder, schizophrenia, and SUD respectively (see Methods).

For all 4 co-diagnoses, we found that the HRR increased with higher polygenic loads and was statistically significant for the 10^{th} PRS decile compared to the middle deciles (Figure 5 Extended Data Figure 6-9 and Supplementary Table S19-S22). This was the case for both the single-disorder PRSs and (more strongly) for the aggregate (SUM)PRSs. Similarly, the absolute risk trajectories showed significant differences across deciles for all co-diagnoses, most considerably for bipolar disorder and SUD with absolute risks for the 10^{th} aggregate

PRS decile reaching 13% and 21%, respectively, representing a 3.25 (bipolar disorder) and 5.25 (SUD) fold increase in risk compared to the least burdened (1st decile) case groups.

Performing sex-specific analyses revealed substantial differences in absolute risks between the sexes (Supplementary Figures S20-S23), most considerably for developing SUD and schizophrenia. For SUD, male risks reached 35%, 15% and 6% for PRS decile 10, middle and 1, respectively, while female risks were 15%, 7% and 3%. For schizophrenia, the risks for males were 25%, 13% and 13% compared to 9%, 8% and 8% for females.

The overall most differentiating absolute risks for developing a comorbid disorder were seen for males in the top aggregate PRS decile, reaching risks of 35% and 25% for developing SUD and schizophrenia, respectively, in contrast to the lowest decile case groups (both sexes) with risks of 4% (SUD) and 8% (schizophrenia) and the background population without a depression diagnosis with risks of 2% (SUD) and 1% (schizophrenia) for the middle deciles.

In comparison to the background population without any depression diagnoses, the absolute risks for the top decile case groups were between 7 and 30 times higher. Particularly for SUD risk, the top decile of the background population surpassed the risk of the lowest decile case group (Figure 5J).

DISCUSSION

We performed a depression GWAS meta-analysis of more than 1.3 million individuals, identifying 251 independent risk variants in 243 genomic loci, of which 64 are novel. We prioritized likely causal genes and revealed novel enrichments of neuro-developmental and functional pathways, prenatal GABAergic neurons, astrocytes and oligodendrocyte lineages. Dissecting the genetic and clinical heterogeneity, we identified distinct polygenic architectures across subgroups of depression and demonstrated increased, sex-dependent, risks for recurrence and psychiatric comorbidity among depression cases with the highest polygenic burden, informing precision psychiatry approaches.

Among the novel loci, we highlight *GRIA1* and *GABRA1*, encoding a glutamate ionotropic AMPA type 1 receptor subunit (GluA1) and a GABA receptor subunit (α1), respectively (Supplementary Table S2). The two genes also showed significant imputed differential expression in the DLPFC (Supplementary Table S9) in our TWAS[77-79]. Our pathway analysis reinforced previous reports[12,14,25] and extended the enrichment of glutamatergic and GABAergic synapses and functions, further indicating that glutamatergic and GABAergic dysfunctions are key etiologic components in depression. Along with the observed enrichment of GABAergic cell-types already present during the prenatal stage, this supports the accumulating multidisciplinary evidence that implicate excitatory/inhibitory imbalance with depression[80] and other psychiatric disorders[81-84].

Our TWAS found that *GRIA1* expression in DLPFC was decreased in the depression case group compared to control. This is consistent with findings in preclinical pharmacological studies of mood disorders showing promising results of AMPA receptor potentiators[36], as well as the convergent rapid and sustained increase in GluA1 and other synaptic

proteins associated with most fast-acting antidepressants[35]. Thus, our results provide human genetic evidence that emphasizes positive modulation of the AMPA receptor as a potential pharmacological approach.

In contrast to a reported decrease in *GABRA1* expression in the DLFPC in individuals with suicidal depression[85,86], our TWAS pointed to an increased expression in individuals with depression. $GABA_A$ receptors provide critical inhibitory control of the firing of glutamatergic excitatory neurons and they are the binding partner of several drugs in mood disorders and potential drug targets for other psychiatric disorders[87-89]. Antidepressant effects have been associated with pharmacological modulation of both synaptic[90] and extra-synaptic $GABA_A$ receptors[91], and our findings support the further study of pharmacological modulators of these receptors in the treatment of depression.

We also highlight convergent lines of evidence for a neurodevelopmental component in depression, emanating from: (i) novel loci (e.g. the top three novel loci located in genes associated with Mendelian neurodevelopmental disorders), (ii) pathway analyses showing enrichment of gene-sets related to neuronal development, (iii) enrichment of prenatal cell-types, (iv) association of depression genetic risk with attenuated cognitive performance in a community youth cohort and (v) MiXeR analyses showing substantial genetic overlap with autism, ADHD and schizophrenia. Although partly overlapping, these results implicate a role of neurodevelopmental processes in the etiology of depression.

The MiXeR bivariate Gaussian mixture modeling revealed compelling genetic overlap between depression and other traits when considering both concordant and discordant variants, suggesting that psychiatric disorders and correlated traits are substantially more intertwined than indicated by their genetic correlations. The most drastic results were observed for educational attainment, showing that 99% of the depression risk variants are also influencing educational attainment and, vice-versa, 88% of educational attainment variants are influencing depression. This almost complete overlap in influencing variants, with a majority of discordant variants (58%), is consistent with overall negative genetic correlation between the two traits (r_G=−0.23) and refines the understanding of their polygenic architecture appreciably. Although generally less pronounced, a similar picture was observed for the other mental health and behavioral traits examined. These results corroborate and extend previous results on partly older and smaller GWASs[92]. A combination of (altogether) only around 10,000-15,000 variants appear to explain 90% of the SNP heritability of the investigated major psychiatric disorders and correlated behavioral and cognitive traits. This narrows the search-space for risk variants to around 1% of the 10^6 quasi-independent SNPs genome-wide and indicates that it is mainly the size and directions of the SNP effects, rather than different SNPs/loci, that influence the development of a specific psychiatric disorder. Furthermore, the considerable overlap suggests that future fine-mapping efforts to pinpoint the causal variants could increase power by combining data from traits with overlapping influencing variants when accounting for the directional effects.

In the Danish iPSYCH cohort, we assessed PRS-based predictions for developing recurrent depression and psychiatric comorbidity over time and demonstrated statistically significant differences in hazard rates and absolute risks between individuals with depression in the

highest PRS deciles compared to other deciles for all the outcomes (developing recurrence, anxiety, bipolar disorder, schizophrenia or SUD).

The iPSYCH cohort includes relatively young individuals (mean age 23.37 years, SE=6.9, at follow-up December 31st 2016) and thus, a substantial number of those with depression will develop recurrent depression and/or comorbidities later in life. This indicates that observed risks, and likely also the differences across PRS deciles, will increase further over time. Moreover, risk prediction could probably be improved by including other risk factors such as e.g. family history[93] and clinical/phenotypic variables, as it has been shown in other complex disorders[94].

Our findings have potential clinically relevant applications. For instance, in clinical settings, a targeted effort could be envisaged that offers more frequent monitoring for development of e.g. bipolar disorder, schizophrenia or anxiety among individuals with depression or those who have the highest PRS (or, preferably, combined genetic and clinical risk) to obtain early diagnosis and initiate early treatment, which may have beneficial effects[95,96]. Similarly, identifying individuals with depression at high genetic risk for developing SUD could be an actionable and informative point of attention for both the physician and the patient and, with focused intervention measures[97], possibly preventing the subsequent development of SUD (i.e. the primary prevention of secondary comorbidities[98]). Validation in clinical settings of such potential applications is warranted in the future[18,98].

Our study has several limitations. Our analyses focused on individuals with European ancestry in order to avoid confounding from population substructure and admixture. We performed several analyses to assess the generalizability to individuals of diverse and admixed genetic ancestry but much larger samples are needed to address this important aspect. Another basic limitation is that still most of the risk loci remain to be discovered and scrutinized to inform on the pathophysiology. Additionally, studies in clinical settings are needed to assess the clinical utility of our findings.

In conclusion, we have advanced the understanding of the genetic architecture of depression, the genetically driven pathophysiology and the prediction of pertinent outcomes that should be informative for future precision psychiatry approaches in depression.

METHODS

Samples

The iPSYCH2015 sample is a nation-wide population sample extracted from a baseline cohort consisting of all children born in Denmark between May 1, 1981, and December 31, 2008, who were alive and resided in Denmark on their one-year birthday, and who have a known mother[25,26]. Individuals diagnosed with one (or more) of six major psychiatric disorders (ADHD, autism, depression, bipolar disorder, schizophrenia, anorexia) were identified via the Danish Psychiatric Central Research Register[27], which includes data on all individuals treated in Denmark at psychiatric hospitals (from 1969 onward) as well as at outpatient psychiatric clinics (from 1995 onward). A random sample from the same

birth cohort were chosen as control group. The iPSYCH2015 cohort consists of the initial case-cohort sample iPSYCH2012[26] and the recent extension iPSYCH2015i[25].

Individuals with a ICD10 F32-F33 diagnosis in 2016 or earlier were considered as depression cases and the random population-based sample excluding individuals with a depression diagnosis were used as controls. All individuals diagnosed with bipolar disorder were excluded, except for the analyses involving co-occurrence of depression and bipolar disorder diagnosis. Depression cases were divided into two groups; 1) individuals diagnosed with a single episode of depression, and 2) those fulfilling the criteria for recurrent depression. The date of the first depression episode for each individual was defined as the start date of the first contact in the Danish Psychiatric Central Research Register with an ICD-10 code of F32-F33 diagnoses at an age of 10 years or older. When defining later depression episodes only contacts with the ICD-10 codes: F32, F33.0-F33.3, F33.8-F33.9 were considered. Individuals with the ICD-10 code F33.4 “Major depressive disorder, recurrent, in remission” were omitted because these contacts are unlikely to indicate a new episode of depression. Only ICD-10: F32, F33.0-F33.3, F33.8-F33.9 diagnosis at dates later than 60 days after the end date for all previous ICD10: F32-F33 diagnosis were considered as new episodes of depression, thus categorizing that individual as having recurrent depression.

In addition to depression phenotypes, individuals with the following diagnoses were recorded in the iPSYCH sample: bipolar disorder (F30-F31), schizophrenia (F20), substance use disorders (SUD) (F10-F19, excluding F1X.0 acute intoxication), anxiety (anxiety disorders; F40-F43), autism (autism spectrum disorder; F84.0, F84.1, F84.5, F84.8 or F84), ADHD (F90.0).

In the iPSYCH2015 cohort[25], we have added 11,710 depression cases and 18,410 controls to the iPSYCH2012 sample[12,26], summarizing to a total of 34,095 cases and 45,393 controls prior to relatedness pruning and removal of ancestry outliers (Supplementary Table S14). A total of 30,618 cases and 38,200 controls were retained after relatedness pruning and removal of ancestry outliers. Of these, 6,517 had recurrent episodes of depression (see definition below), while 24,101 were classified as single-episode depression cases.

The number of cases with an additional psychiatric diagnosis (bipolar disorder, schizophrenia, anxiety, autism, ADHD, SUD) are shown in Extended Data Figure 2. When comparing these frequencies with published estimates of lifetime prevalence three main factors are important to keep in mind: (i) The investigated Danish iPSYCH cohort is young (individuals born between 1981-2008; mean age 23.37 years at follow-up), (ii) phenotypes are based on register-based hospital-diagnoses and (iii) people with early onset of affective disorders have a greater absolute risk of developing schizophrenia and related disorders, compared to those with later onsets[9]. Life-time risk of hospital-diagnosed schizophrenia and depression (ICD10 F32-F33) in the Danish population has been estimated to 1.75% (average of 1.93% in males and 1.56% in females) and 12.3% (average of 9.07% in males and 15.5% in females), respectively[3]. Extended Data Figure 2 shows that the frequencies of schizophrenia and depression in the young, randomly selected iPSYCH-population-sample are 1% and 2.7%, respectively, based on register-based hospital diagnosis. Schizophrenia

has relatively early onset, while depression has relatively late onset, hence the relatively low observed 2.7% frequency of depression in the random population-based sample of iPSYCH. Thus, the young iPSYCH sample is likely to capture relatively severe early-onset depression cases with a high risk for developing schizophrenia[9]. These factors explain the apparently high frequency of depression-cases with a schizophrenia co-diagnosis observed in the iPSYCH sample. As the iPSYCH cohort grows older, and more individuals develop depression, we will expect the frequency of schizophrenia co-diagnosis among depression cases to be reduced.

Whether recurrent depression was associated with increased comorbidity compared to single-episode depression was tested using logistic regression for each additional diagnosis while adjusting for age. Individuals were grouped in five-year age bins to construct dummy variables for the age adjustment (Extended Data Figure 2).

All analyses of the iPSYCH sample were performed at the secured national GenomeDK high-performance computing cluster in Denmark.

Genotyping, QC and imputation

iPSYCH2015 samples were linked via the unique national personal identification number to the Danish Neonatal Screening Biobank (DNSB) at Statens Serum Institute (SSI), where DNA was extracted from Guthrie cards, and whole-genome amplification was performed in triplicate, as described previously[99,100].

Genotyping in iPSYCH2012 was performed using the PsychArray V1.0 (Illumina, San Diego, California), while genotyping of iPSYCH2015i was done using the Global Screening Array v2 (Illumina, San Diego, California). Since the two samples were genotyped on different platforms, they were QCed and imputed separately.

Genotyping of the iPSYCH2012 sample was performed at the Broad Institute of Harvard and MIT (Cambridge, MA, USA) with PsychChip arrays from Illumina according to the manufacturer's instructions. Genotype calling of markers with MAF≥0.01 was performed by merging call sets from GenCall 1.6.2.2[101] and Birdseed 1.6[102], and less frequent variants (MAF<0.01) were called with zCall 1[103]. Genotyping and data processing were carried out in 23 waves. Genotyping of the iPSYCH2015i sample was performed at Statens Serum Institut (SSI, Copenhagen, Denmark) using the Global Screening Array v2 with a multi disease drop in (Illumina, San Diego, California) according to the manufacturer's instructions. Genotype calling of markers was performed using GenTrain V3. Preimputation quality control was performed separately for each genotyping wave using the Ricopili[104] pipeline with the specified parameters in the following order. Initially SNPs with a call rate<0.95 were removed, and subsequently all individuals with a call rate in cases or controls of<0.95 or an autosomal heterozygosity deviation F_{HET} outside the interval [−0.2;0.2] were removed. Individuals where stated sex was not consistent with sex derived from genotypes were flagged. Subsequently QC was conducted at the marker level, keeping markers with call rate ≥0.98, missing difference ≤0.02 between cases and controls, with MAF≥0.01, Hardy-Weinberg equilibrium (HWE) in controls P-value ≥$1x10^{-06}$ and Hardy-

Weinberg equilibrium (HWE) in cases P-value ≥$1x10^{-10}$ (See https://sites.google.com/a/broadinstitute.org/ricopili/preimputation-qc for further details).

The effects of three batch variables on marker genotypes were tested in iPSYCH2012 (ArrayPlate.ID, PreProc.Plate and wave) and iPSYCH2015i separately (Array.Batch, ArrayPlate.ID and PreProc.Plate). This was done using relatedness-pruned dataset with ancestry outliers removed to avoid removal of markers where batch effects were caused by population structure or cryptic relatedness rather than genotype artefacts.

Pairwise relatedness coefficients ($\hat{\pi}$) were estimated with plink using a LD pruned and MAF filtered set of SNPs (snps-only, window size=5,000, step size=300, r^2<0.05, MAF>0.05). Principal Component Analysis was conducted using the same set of LD pruned and MAF filtered SNPs, with random removal of one member of each pair with a relatedness coefficient ($\hat{\pi}$) higher than 0.2. Eigenvectors were inferred using EIGENSOFT version 6.1.4[105] on the relatedness-pruned set of individuals, and subsequently projecting all individuals onto those eigenvectors based on their genotypes. Individuals with all four grandparents born in Denmark were used as a reference for constructing a 3-dimensional ellipsoid using principal components 1, 2 and 3 with a radius of 5 standard deviations from the mean. Individuals located outside this ellipsoid were removed prior to the testing for batch effects. For each of the batch variables, each genotyped marker was tested for association with each batch versus the remaining batches pooled, using plink v1.90b4.

The exclusion of SNPs strongly associated with any of the batch-variables were based on their minimum P-value across all associations per variable. The cut-off for the wave and Array.Batch was min(P) <$2x10^{-10}$ and for PreProc. Plate and ArrayPlate.ID min(P) <$2x10^{-12}$, based on a Bonferroni correction for the number of markers tested and the number of associations done per batch-variable taking into account that batch variables are nested.

After removing SNPs falling for any of the above cut-off the remaining distribution was evaluated using QQ-plots. The expected minimum P-distribution was calculated using the inverse cumulative distribution of N independent distributions as suggested in supplementary of Schork et al.[106], N being the number batch-variable values.

iPSYCH2012 and iPSYCH2015i were imputed separately using the Ricopili pipeline[104]. All 23 genotyping waves of iPSYCH2012 were imputed together. Prephasing was done using Eagle v2.3.5[107] and the subsequent imputation was conducted using Minimac3[108], using the downloadable version of the Haplotype Reference Consortium (HRC) (accession number: EGAD00001002729)[109] as reference.

Relatedness pruning and removal of ancestry outliers

Best guess genotypes from iPSYCH2012 and iPSYCH2015i were merged, filtered and LD-pruned down to a set of roughly 30K markers, with imputation INFO score >0.8, r^2<0.075, located outside regions of long-range LD as defined by Price et al.[110], minor allele frequency >0.05 and no deviation from Hardy-Weinberg proportions (P>$1x10^{-4}$). Relatedness coefficients, based on “identity-by-state”, were estimated using plink v1.9, to

identify related (and duplicated samples), with $\hat{\pi} > 0.2$ and removing related individuals at random, but preferring cases over controls. PCA was carried out using the same set of filtered and LD-pruned SNPs as implemented in the Ricopili-pipeline[104]. A subsample of European ancestry was selected as an ellipsoid in the space of PC1-3 centered and scaled using the mean and 8 standard deviation of the subsample whose grandparents were all known to be born in Denmark.

All the analyses described below were conducted using the relatedness and ancestry pruned dataset, but some analyses were in addition also performed among ancestry outliers (see below, Supplementary Figures S24-S28 and Supplementary Tables S23-S28). A second PCA was performed on the sample excluding ancestry outliers, using EIGENSOFT[105] as implemented in the Ricopili-pipeline[104], which was used for adjusting for any remaining population substructure in downstream analyses.

FinnGen

The FinnGen[22] (https://www.finngen.fi/en) study combines genotype data with longitudinal health register data of Finland, including the causes of death, inpatient, outpatient, and drug reimbursement registers. Subjects were genotyped with Illumina and Affymetrix arrays (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa Clara, CA, USA) as described (https://www.finngen.fi/en/researchers/genotyping). Genotyping and imputation with the Finnish population-specific SISu v3 reference panel were conducted, as described (https://www.protocols.io/view/genotype-imputation-workflow-v3-0-xbgfijw). SNPs were pruned for minor allele frequency (MAF) ≥0.01 and imputation info score >0.6. GWAS was performed using the Scalable and Accurate Implementation of GEneralized mixed model (SAIGE) v0.20[111] with a kinship matrix as a random effect and age, sex, the first 10 principal components (PCs), and genotyping batch as fixed effects.

A GWAS for depression was conducted in FinnGen[22] release 6 samples containing 28098 cases (ICD-10/9: F32 or F33) and 228817 controls without manic episodes (ICD10 F30), bipolar affective disorder (ICD-10: F31, ICD-9: 296[2-7], ICD8: 296), persistent mood disorders (ICD10: F34) or other or unspecified mood disorder (ICD10: F38, F39, ICD-8: 29699). Mood (affective) disorders (ICD-10 F30-F39) from FinnGen release 2 was included in the latest depression GWAS meta-analysis by Levey et al. [24], but the current study includes depression cases (ICD-10 F32-F33) from FinnGen release 6 for the first time.

A GWAS for anxiety disorders was conducted in the FinnGen[22] release 6 sample consisting of 7,671 cases with a generalized anxiety disorder (ICD-10: F41.1, ICD-9: 3000C, ICD-8: 300,00), panic disorder (ICD-10: F41.0; ICD-9: 3000B, 3002B) or phobic anxiety (ICD-10: F40, ICD-9: 3002C, 3002D, 3002X) diagnoses and 161,438 controls without a history of any psychiatric diagnoses (ICD-10: F00-F99, ICD-9: 290-319, ICD-8: 290-315). With ICD-9 codes the first three digits signify numerical value (e.g., the code **300**2B is included in the code range 290-319). Cases with psychotic disorders (ICD-10: F2; ICD-9: 295, 297, 298; ICD-8: 295, 297, 298), autism spectrum disorders (ICD-10: F84.0, F84.1, F84.5; ICD-9: 2990; ICD-8: 299,99) or intellectual disability (ICD-10: F7, ICD-9: 317-319; ICD-8: 311-315) were excluded. Additionally, the age range of controls was adjusted to match to that of cases and a more stringent filtering on imputation info score was used for the

anxiety GWAS (INFO>0.8). This GWAS of core anxiety in the FinnGen[22] cohort was meta-analyzed with a GWAS of self-report of physician diagnosis of anxiety in the MVP[112], with the purpose of generating weights for an ANX-PRS (see below).

Million Veteran Program

Research involving MVP[24,112] in general is approved by the VA Central IRB (MVP025 19-02); the project was also approved by IRBs in Boston, San Diego, and West Haven.

Statistical analysis

GWAS analyses—Genome-wide-association analyses within iPSYCH were conducted using the Ricopili pipline[104], applying a logistic regression model using dosages of the imputed genotypes. Analyses were adjusted for PC 1-10 from the second PCA using the remaining subsample after removal of ancestry outliers and pruning for relatedness. In addition, individuals with a diagnosis of bipolar disorder were excluded.

Summary statistics from iPSYCH2015 and the following external samples were included in a fixed-effects variance-weighted meta-analysis in METAL[28]: 1) 230,118 broadly defined depression cases and 545,339 controls from Howard et al. 2019[14], including self-reported depression of 23andMe[23] and the broadly defined depression phenotype of UKB, 35,077 narrowly defined depression cases and 95,406 controls from Wray et al. 2018[12], excluding 23andMe[23]; 2) depression phenotypes from the Million Veteran Program (MVP)[24] based on ICD codes derived from electronic health records (83,810 cases and 166,405 controls); 3) 28,098 cases with ICD-10 F32 and/or F33 diagnoses and 228,817 controls from FinnGen[22]. For comparison, we also conducted a GWAS and several downstream analyses using a narrower definition of depression that excluded self-reported depression of 23andMe[12,23] and the broadly defined depression phenotype of UKB[14,21] (Supplementary Figure S14, Table S13). The results of the downstream analyses were very similar to the primary analyses without noteworthy differences, except for the DEP-PRS analysis of cognitive performance in the PNC cohort, which is mentioned in the results section.

SNP-heritability, genetic correlations and overlap with other phenotypes—We estimated SNP-heritability (h^2_{SNP}) for iPSYCH2015 cases with depression, single-episode and recurrent depression and for the external GWAS summary statistics outlined above using LD score regression[36]. Genetic correlations within iPSYCH and between iPSYCH and external GWAS summary statistics was estimated using LD score regression[36]. In addition, the genetic correlations of depression meta-analysis, single-episode and recurrent depression with other phenotypes were, evaluated using LD score regression[113] at the LD Hub[38] website. We used all available phenotypes on LD Hub, but we performed analyses for the UKB traits (N=597) and the remaining individual phenotypes (N=258) separately. To make our analyses more comparable to previous publications, and to avoid overly conservative correction of the many highly correlated traits, levels of experiment-wide significance (Bonferroni correction for number of tests applied) were also established separately within the two groups, i.e. in the UKB traits (P<8.38 x 10-5) and the remaining individual phenotypes (P<0.00019), respectively. LD Hub traits were supplemented with LD score regression[36] analyses performed locally using updated or in-house GWAS summary

statistics for traits not available at LD-hub. These include ADHD, autism[15], cannabis use disorder[39], cannabis use[114], alcohol dependence[115], drinks per week, smoking ever[116], age of initiation as a regular smoker, current versus a former smoker (smoking cessation), number of cigarettes per day, smoking initiation[117]

We applied MiXeR[45] on our depression GWAS summary statistics and a selection of additional traits (Supplementary Table S6A, Figure 2 and Supplementary Figure 9-1) to estimate (i) the number of variants explaining 90% of the SNP heritability of each trait and (ii) the genetic overlap between depression and each trait. The additional traits included in the analyses are ADHD[46], anxiety (FinnGen[22] + MVP[112]), autism[15], bipolar disorder[118], educational attainment[119], Neuroticism[120], schizophrenia[121], Smoking Initiation[117] and substance use disorder (SUD)[39,115]. For comparison we included non-psychiatric traits, i.e. Alzheimer[122], epilepsy[123] and height[124] not necessarily expected to show strong genetic correlations with depression. MiXeR analysis were conducted with default settings (https://github.com/precimed/mixer) in a two-step process: 1) a univariate model for each trait to produce estimates of the proportion of variants with non-zero additive genetic effect on the trait (i.e. “polygenicity”) and the variance of effect sizes of these non-zero variants (i.e. “discoverability”’). 2) the variance estimates obtained in the univariate analysis were applied in the bivariate model (i.e. depression vs. each of the additional traits) to obtain four estimates representing (i) zero-effect SNPs in both traits; (ii) SNPs with a specific non-zero effect on trait 1; (iii) SNPs with a specific non-zero effect on trait 2; and (iv) SNPs with a non-zero effect on both traits. Estimates of polygenic overlap and genetic correlation between pairs of traits were obtained by combining these four components.

As a complementary to MiXeR we estimated local genetic correlations (r_g) of depression and the additional traits listed above using LAVA[47]. LAVA estimates local rg across 2,495 semi-independent genetic loci of approximately equal size (~1 Mb) in a two-step process: 1) univariate analysis: the observed h^2_{SNP} is estimated for each locus for each trait. 2) bivariate analysis: the local genetic covariance is estimated for each locus using the method of moments. Genetic loci included in the bivariate analysis were filtered according to their local h^2_{SNP} using a significance threshold of $P<10^{-4}$, consistent with previous usage of LAVA[47,91]. Sample overlap was controlled using linkage disequilibrium score regression[36]. Significance testing was performed using simulation-based P-values and we used the false discovery rate (FDR) to adjust for multiple testing, reporting loci with FDR<0.05.

While the genetic correlation analysis provides a summary measure of the genomewide correlation of SNP effect sizes, the bivariate MiXeR analysis estimates the number of overlapping SNPs/loci between the two traitsregardless of the direction of effects (i.e. both concordant and discordant effects) and calculates the fraction of concordant variants. For instance, a genetic correlation of zero can be seen for traits that have no overlapping risk SNPs as well as for traits that have completely overlapping risk SNPs with 50% concordant and 50%discordant effect directions. So, as the huge overlap observed between depression and educational attainment (Supplementary TableS6A and Figure S9-1) is based on a majority of discordant variants (58%) (and fewer concordant variants (42%)), it is consistent with the overall negative correlation observed between the two traits.

As additional validation of MiXeR results, we tested the hypothesis that MiXeR- and LAVA-derived measures of mixed effect directions were correlated, we calculated the Pearson correlation coefficient for 1) the proportion of shared "causal" variants with concordant effects (i.e causal variants affecting both traits in the same direction) estimated using MiXeR and 2) the proportion of significantly correlated genetic loci with positive correlation estimated using LAVA[91](see Supplementary Figure S9-3).

Conditional analysis and finemapping—We identified potential independent genome-wide significant lead variants for each of the broadly defined genome-wide-significant loci of our GWAS meta-analysis results[104] using the following approach: For each genome-wide significant SNP we defined the associated LD-region by recording the left and rightmost variant with $r^2 \geq 0.1$, assigning the SNP with the lowest P-value within each LD-clump as the index SNP. Index SNPs were considered independent when $r^2 < 0.1$ within 3 Mb windows. To define GWAS loci, a 50kb window was added on each side of the LD-region and overlapping LD-regions were combined into a single locus, assigning the index SNP with the lowest P-value as lead SNP. Only a single SNP was kept from within the MHC region, due to extended linkage disequilibrium and a strong association signal of the MHC region (chr 6:25-35 Mb).

To identify additional independent index variants, we performed a stepwise model selection procedure to select independently associated SNPs implemented in GCTA-COJO[29,30]. We assigned posterior probabilities (PP) of being causal to SNPs and constructed credible sets of SNPs that cumulatively capture 95% of the regional posterior probability[125] using PAINTOR[126-128], CAVIARBF and FINEMAP[129], using the CAUSALdb-finemapping-pipeline (https://github.com/mulinlab/CAUSALdb-finemapping-pip). We applied a conservative approach and assumed one causal variant for each locus.

Co-localization analyses were performed to evaluate the extent of overlap between eQTL and GWAS signatures in depression and to identify putative causal genes from GWAS associations. Considering only the 95% credible SNPs from the fine-mapped depression-meta-analysis, we integrated GWAS results and eQTL data from a previous meta-analysis integrating signals among three brain datasets[53] applying a fixed-effect model. The eQTL data originates from eQTL meta-analysis on RNA-sequenced gene expression data from the dorsolateral prefrontal cortex from PsychENCODE[50] and ROSMAP[130], and from 13 brain regions from GTEx[131].

Using the Coloc method[54,55], we extracted eQTL signals of genes within 200 kb distance to significant GWAS variants ($P<5 \times 10^{-8}$) using effect sizes (β-values) and standard errors from eQTL and GWAS as input. Four partially hierarchical hypotheses[55] were tested: H0, no association; H1, GWAS association only; H2, eQTL association only; H3, both GWAS and eQTL association but no co-localization; H4, both GWAS and eQTL association and co-localization, considering posterior probability for H4 (PPH4) >0.8 as strong evidence from both GWAS, eQTL and co-localization. In addition, we conducted the eQTL and GWAS CAusal Variants Identification in Associated Regions (eCAVIAR) approach[56], in which, both eQTL and GWAS were fine-mapped, and the product of posterior probability (CLPP) was calculated, prioritizing genes with at least a single variant with CLPP>=0.01.

Gene-wise and pathway analysis—We used a number of different approaches and data including those available via the FUMA v1.3.6a[48] website (http://fuma.ctglab.nl) for downstream annotation and functional characterization of significant loci.

We performed a genome-wide gene-based association analysis using the Multi-Marker Analysis of GenoMic Annotation (MAGMA) tool, as implemented in FUMA version 1.3.6a[48]. Using this we mapped SNPs from the depression GWAS meta-analysis to 17,840 protein coding genes, and performed Bonferroni correction for the total number of protein-coding genes ($P<2.8 \times 10^{-6}$, Supplementary Table S7D).

Protein coding genes were mapped if they were located within 10Kb up- or downstream from index variants or if a credible variant was annotated to the gene based on eQTL data or chromatin interaction data from human brain (data sets used in the mapping can be found in Supplementary Table S7A). No additional variant filtering by functional annotation was applied in the eQTL and chromatin interaction mapping. This analysis identified 411 depression risk genes, which were used in a gene-set enrichment analysis within the GENE2FUNC module of FUMA, where we analyzed 8,664 gene-sets derived from GO Biological Process (N=7,658; 393 gene-sets) and GO Cellular Components (N=1,006; 86 gene-sets) ontology in the MSigDB database.

Tissue and cell type enrichment—We also used FUMA to perform tissue expression analyses on data available through their website, by testing whether the identified genetic associations for both the primary and narrow depression phenotype definitions were enriched regarding transcriptome profiles of human tissues using summary statistics based on all SNPs. Finally, we used FUMA to perform cell-type enrichment analyses[48] based the depression GWAS summary statistics. We use MAGMA gene-property analysis to test cell type specificity of the depression phenotype using GWAS summary statistics. MAGMA gene-property analysis with scRNA-seq: The gene-property analysis aims to test for relationships between cell specific gene expression profiles and disease-gene associations. In all the above-mentioned analyses implemented in FUMA default settings were applied.

To further evaluate whether the genomic loci implicated in depression are enriched in any particular cell type, we intersected common depression risk variants with two recent epigenomic maps of cell-specific open chromatin[57,58] using a LD score partitioned heritability approach[59] (Figure 3B).

Transcriptome-wide association study—In addition, we performed a transcriptome-wide association study (TWAS), imputing the genetically regulated gene expression using EpiXcan[49] and using models trained on PsychENCODE Consortium (PEC)[50,51] expression data for genes and transcripts detected in the dorsolateral prefrontal cortex (DLPFC), with the aim of identifying and prioritizing putative causal loci for the broad depression phenotype definition. A total of 34,646 genes/transcripts were tested (transcripts with prediction performance $R^2>0.01$ and prediction performance q-value <0.05 with the Benjamini-Hochberg method were retained), and applying a significance threshold of $P<1.44\times10^{-6}$ (corresponding to Bonferroni correction of all genes and isoforms tested; Figure 1D, Extended Data Figure 1, Supplementary Figure S10 and Table S9),

using information on approximately independent linkage disequilibrium blocks in human populations[52] to identify independent genomic regions with genes/transcripts showing significant differential gene expression between depression cases and controls.

Polygenic risk scores—Using available summary statistics form published GWAS as training datasets, we calculated polygenic risk scores (PRS) for individuals in the iPSYCH2015 sample using LDpred2[132]. Summary statistics were filtered for an imputation info score INFO≥0.9 if available. When using external summary stats not processed using Ricopili or imputed using different imputation references, we excluded all ambiguous markers to avoid potential strand conflicts. To improve performance of the scores and avoid including artefacts from batch effects, we restricted the summary stats to include only SNPs known to be present in both the iPSYCH2012 and iPSYCH2015i data at a reasonable quality (info score 0.6 and MAF 0.05). This step also checked for allele flips.

To derive a PRS for depression within the iPSYCH2015 sample we used the depression meta-analysis[12,14,23,24] excluding all iPSYCH samples as training. We also calculated PRS based on GWAS summary statistics from bipolar disorder[118], schizophrenia[121], ADHD[13] and autism[15]. The PRS for ADHD was based on a combination of “external” GWAS sumstats (without iPSYCH data) and “internal” training in the iPSYCH sample. The iPSYCH sample was split into 10 groups, and in 10 iterations of leave-one-out, the PRS was calculated for individuals in each of the 10 groups using a GWAS performed on the remaining 9 groups as training. A fixed-effect variance-weighted meta-analysis, implemented in METAL[28], of two GWAS for anxiety, one based on self-report of physician diagnosis of anxiety in the MVP[112] and the other being core anxiety in the FinnGen cohort[22] (see above), served as weights for generating an ANX-PRS summarizing genetic risk of anxiety. To derive a PRS for neuroticism we used GWAS summary stats of a weighted neuroticism sum-score, constructed by adding up ten individual item responses (Fed-up, Guilt, Irr, Miss, Mood, Tense, Nerv feel, Suf Nerv, worry emb and worry) by Nagel et al.[120]. In addition, we calculated PRSs using the following GWAS summary stats for traits related to substance-use and substance-use-disorder as training datasets. These include *cannabis use disorder*[39](excluding iPSYCH samples, i.e., training only based on data from PGC and data from decode genetics), *cannabis use*[114], *alcohol dependence*[115], *drinks per week, smoking ever*[116], *age of initiation* as a regular smoker, current versus a former smoker (*smoking cessation*), *number of cigarettes per day*, *smoking initiation*[117]. All summary statistics were combined in a pseudo-meta-analysis, using the a fixed-effects variance-weighted meta-analysis implemented in METAL[28], to generate a SUD-PRS summarizing genetic liability of substance use disorder.

With the aim at improving the PRS prediction we attempted to exploit the genetic overlap of depression with other phenotypes. We therefore combined all the above PRSs into a single score as a weighted sum[15]. We chose here to use the log(OR) for the logistic regression of sub-phenotype of interest on each score as a continuous factor as a measure of ‘importance’ in context of the sub-phenotype of interest (either recurrent vs. single-episode or depression with additional diagnosis of bipolar disorder, schizophrenia or SUD). The logistic regressions were adjusted for the first 10 PCs of the second PCA mentioned above. We added each standardized PRS weighted by its log(OR) one PRS at the time to the

aggregated score. We started with the phenotype with highest abs(log(OR)) and ended with the lowest. This way we ended up with a sequence of scores starting with S_0 and continuing with:

$$S_k = \frac{\sum_{i=1}^{k} \log(OR_{P_i}) S_{P_i} - \mu\left(\sum_{i=1}^{k} \log(OR_{P_i}) S_{P_i}\right)}{\sigma\left(\sum_{i=1}^{k} \log(OR_{P_i}) S_{P_i}\right)}$$

Where S_{P_i} is the score for phenotype P_i, OR_{Pi} the odds ratio from the logistic regression of sub-phenotype of interest on S_{Pi} as a continuous factor while adjusting for 10 PCs, and μ is the mean and σ the standard deviation.

To examine potential polygenic heterogeneity across depression sub-phenotypes, we investigated how PRS trained on the different phenotypes described above were distributed across depression sub-phenotypes in iPSYCH using a multivariate PRS approach both in a sex-stratified and in an unstratified analysis. The method is described in detail in Grove et al.[15], and is a regression of multiple outcome variables, and in principle a linear regression for each PRS on the depression sub-phenotypes allowing for comparisons on the average PRS across sub-phenotypes for PRS from a number of phenotypes adjusting for necessary covariates. Since the variance-covariance matrix is fitted jointly and accounting for the covariance the inherent correlation between scores is adjusted for when estimating variance and in the hypothesis testing. We performed four mvPRS analyses testing for differences among depression-subtypes; depression cases who have developed (1) anxiety, (2) bipolar disorder, (3) schizophrenia or (4) SUD. Each of the four multiple regression analyses tests for individual and distinct hypotheses, one for each PRS, and the reported P-values are not corrected for multiple testing. Bipolar disorder cases were excluded from the analyses not involving bipolar disorder.

In addition, we applied a Cox's proportional hazard model to estimate hazard rate ratios and absolute risk of developing 1) a second episode of depression, 2) anxiety, 3) bipolar disorder, 4) schizophrenia and 5) SUD, among individuals already being diagnosed with their first episode of depression. Analyses were stratified by PRS deciles of the phenotypes described above and using functions from the R-packages survival (https://CRAN.R-proiect.org/package=survival)[133,134]. while correcting for batch (iPSYCH2012 and iPSYCH2015i) and PCs. The Cox's proportional hazard models were stratified into three groups based on 1st, 2nd-9th and 10th decile. The middle 2nd–9th decile group, i.e. the 80% prediction interval of the PRS, was used as reference. Two-sided 95% confidence intervals were obtained for the stratified absolute risk trajectories using Cox's proportional hazards model with covariates[135], taking the mean of each covariate (PC 1-5 and iPSYCH sample) and using the survfit.coxph function implemented in the survival R-package. All Cox's analyses were conducted for both sexes combined and for females and males separately. In addition, HRR and absolute risk of developing 1) anxiety, 2) bipolar disorder, 3) schizophrenia, and 4) SUD were estimated in the iPSYCH subcohort excluding all depression cases, to obtain estimates in the general Danish population for comparison with the iPSYCH depression-cohort.

Depression-PRS in the Philadelphia Neurodevelopmental Cohort (PNC)

Depression-PRS (DEP-PRS) were calculated for 4,973 individuals of European ancestry from the Philadelphia Neurodevelopmental Cohort (PNC)[65,66] using the primary depression GWAS meta-analysis (and narrowly defined depression GWAS meta-analysis) of the current study as training. Genotypes from the first PNC release (dbGaP phs000607.v1.p1) and neurocognitive phenotypes from the third release (dbGaP phs000607.v3.p2) were utilized. Individuals whose genotypically inferred and phenotypically reported sex did not match, those who did not meet the identity by descent (IBD) filter ($\hat{\pi} > 0.185$), those who did not meet the individual-level missingness filter of 0.05, and those with heterozygosity rates +/− three standard deviations from the mean were removed. SNPs that failed to meet the Hardy-Weinberg proportions (HWP), minor allele frequency (MAF), and SNP-level missingness filters of 0.00001, 0.01, and 0.05, respectively, were also removed. Genotype imputation was performed using the reference panel HRC r1.1 2016 on the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html#!). selecting the “Mixed population” option. Imputed SNPs with an imputation R^2<0.03 and those who failed to meet the missingness, MAF, and HWP thresholds stated above were removed. The first two Multi-Dimensional Scaling (MDS) dimensions were plotted to identify and remove outlier genotypes. The GemTools package in R[136,137] and Ward’s hierarchical clustering methods were used to identify individuals of European ancestry.

PRS-CS[138] was used to apply continuous shrinkage priors to the effect sizes from depression GWAS summary statistics. A European LD reference panel provided by the developers of PRS-CS was utilized (https://github.com/getian107/PRScs), which draws from the 1000 Genomes Project data. The following PRS-CS default settings were used: parameter a in the γ-γ prior=1, parameter b in the γ-γ prior=0.5, MCMC iterations=1,000, number of bum-in iterations=500, and thinning of the Markov chain factor=5. The global shrinkage parameter phi was set using a fully Bayesian determination method. Individual-level DEP-PRS were calculated using Plink v2.0[139]. The associations between scaled (mean=0, SD=1) DEP-PRS and 15 scaled (mean=0, SD=1) neurocognitive phenotypes in the PNC were assessed using linear regression of phenotypes on DEP-PRS.

The neurocognitive phenotypes included performance on the Computerized Neurocognitive Battery (CNB)[67], as well as results from the Wide Range Achievement Test (WRAT-4). A $\sqrt{k-x}$ transformation (where k is a constant) was performed on CNB measures of age differentiation, emotion differentiation, spatial memory, and verbal reasoning, as their distributions were left-skewed. A log10(k - x) transformation was performed on CNB measures of abstraction and mental flexibility, attention, verbal memory, and working memory, as their distributions were left-skewed. A $\log_{10}(x)$ transformation was performed on the WRAT-4 measure while a 1/x transformation was performed on the CNB measure of sensorimotor processing, as their distributions were right-skewed. For the CNB measure of emotion identification, a $\sqrt{k-x}$ transformation was performed on the PEIT-AB form of the task and a $\log_{10}(k-x)$ transformation was performed on the PEIT-C, PEIT-D, and PEIT-E forms of the task, each due to their left-skew. For the CNB measure of nonverbal reasoning, a $\sqrt{k-x}$ transformation was performed on the PMAT-18B form of the task due to its left-skew, while a $\sqrt{x}$ transformation was performed on the PMAT-24A form of the

task due to its right-skew. No transformations were performed on CNB measures of facial memory, finger tapping speed, and spatial reasoning. Neurocognitive phenotype scores were multiplied by −1, following data transformation and scaling, when needed so that higher scores always indicate better performance across the assessed tasks.

Covariates included in the analysis were age at neurocognitive testing, age squared, the first 10 MDS dimensions, sex, and genotyping batch. R^2 was used to report the total variance explained by DEP-PRS and model covariates for the 15 tested neurocognitive phenotypes. Additionally, a variance partitioning tool[140] (https://github.com/GabrielHoffman/misc_vp/blob/master/calcVarPart.R) was used to determine the variance explained by DEP-PRS and each covariate individually. FDR-adjusted p-values were reported.

Ancestry outliers

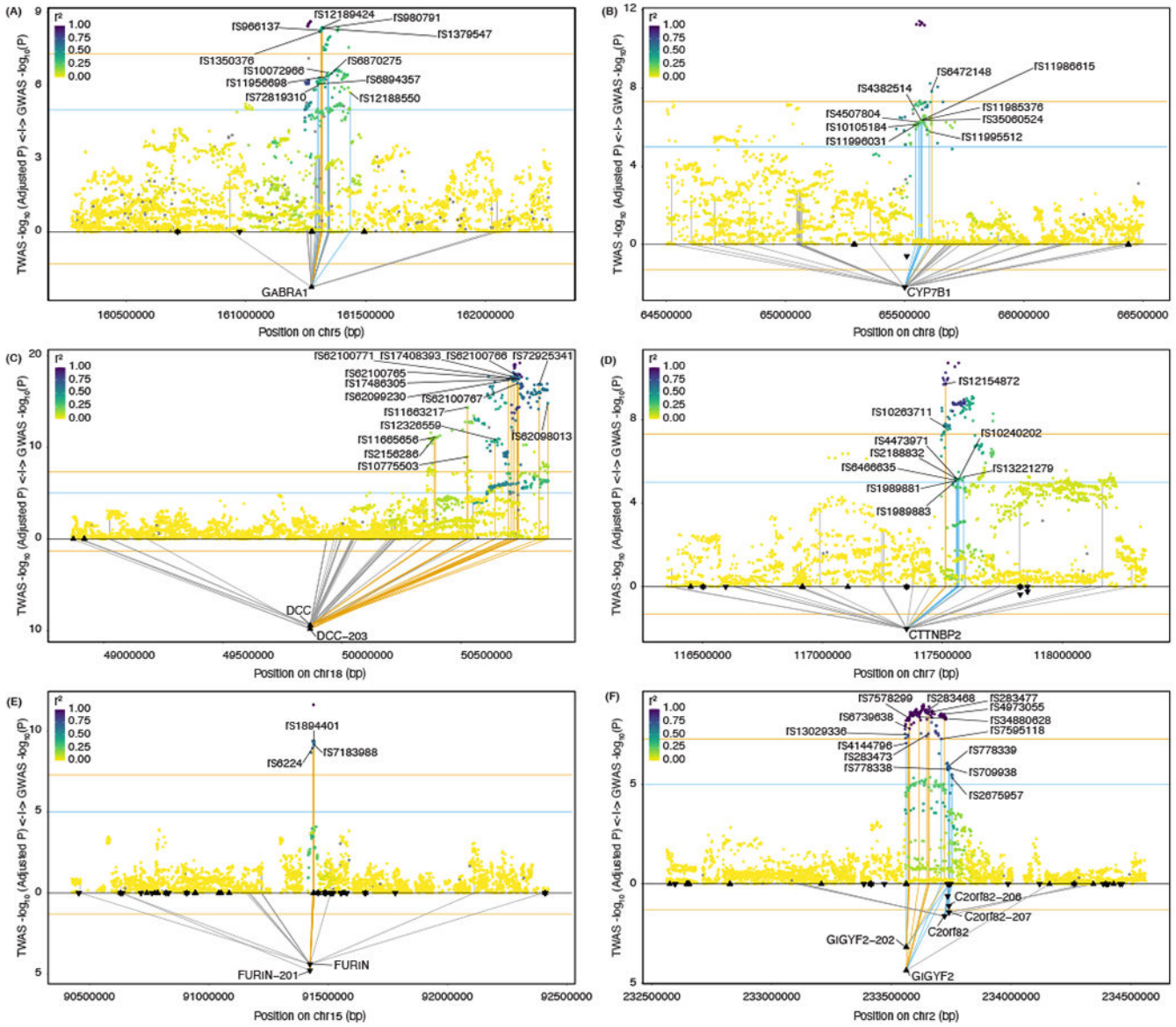
Ancestry outliers (10%) were removed to avoid confounding from population substructure and admixture. To assess generalizability of our main results, we conducted depression-GWAS of the ancestry-outliers using the Ricopili and adjusting for the first 10 principal components (Supplementary Figure S24A and S24B). This GWAS showed a high genetic correlation (r_G=0.95, SE=0.24, P=1x10–4) with our primary depression meta-analysis as shown in Supplementary Table S23. Sign-tests of overall replication of the direction of effects for various GWAS P-value thresholds using our primary meta-analysis as discovery sample and the iPSYCH2015 ancestry outliers as replication sample, showed an overall replication ratio of 0.59-0.64 (Supplementary Table S24). This suggests that many more findings would likely replicate, provided a much larger and more powerful sample, in other non-European ancestries. Previous GWAS of depression using trans-ancestral meta-analysis[24] generally supports the transferability of GWAS results across ancestries.

In addition, we performed the multivariate PRS and the Cox's proportional hazard analyses of recurrent depression using the ancestry outlier sample (Supplementary Figure S25-S28 and Supplementary Table S25-S28D), adjusting all analyses for the first 10 principal components and iPSYCH2012/iPSYCH2015i. The ancestry outlier mvPRS analyses (Supplementary Figure S25) showed similar directions of differences in DEP-PRS and Neuroticism-PRS for single vs. recurrent depression as in the corresponding analysis of the sample with European ancestry (see Supplementary Figure S16-1, Supplementary Table S16A), although none of the differences appeared to be significant (Supplementary Table S25). For ANX-PRS and BP-PRS, however, the observed differences in the European sample were not reproduced for the ancestry outliers. Likewise, for the Cox's regression analyses, the observed differences in HRR and absolute risk of developing a second episode of depression between10th and middle DEP-PRS and DEP-SUM-PRS deciles in the European ancestry sample (Supplementary Figure S17-1) were not reproduced in the ancestry outliers (Supplementary Figure S26). Regarding psychiatric comorbidity among depression cases a similar pattern of increased load of ANX-PRS, SZ-PRS and Neuroticism-PRS among depression cases with comorbidity for anxiety was observed in the European ancestry (Extended Data Figure 2, Supplementary Table S18A) sample and in the ancestry outlier sample, although non-significant in the latter except for Neuroticism-PRS (Supplementary Figure S27A and Table S27A). For depression cases with/without bipolar

disorder, the same direction of differences observed significantly in the European sample (Extended Data Figure 3, Supplementary TableS18D), was observed for BP-PRS and SZ-PRS in the sample of ancestry outliers, but only significantly so for the BP-PRS (P=0.013, Supplementary Figure S27B and Table S27B). The direction of differences was not reproduced for the DEP-PRS and ASD-PRS. Regarding comorbidity with schizophrenia, the same direction of differences in DEP-PRS, SZ-PRS, ADHD-PRS, ASD-PRS, Neuroticism-PRS, SU-PRS and SUD-PRS, among depression cases with / without schizophrenia, as significantly observed in the European sample (Extended Data Figure 4, Supplementary Table S18G), was also observed in the sample of ancestry outliers, while the difference appeared to be opposite for the BP-PRS (Supplementary Figure S27C and Table S27C). The direction of differences significantly observed in the European sample, with increased PRS among depression cases with SUD compared to cases without (Extended Data Figure 5, Supplementary Table S18J), was reproduced in the ancestry outlier sample except for the Neuroticism-PRS (Supplementary Figure S27D and Table S27D).

The general pattern of differences in HRR and absolute risk of developing comorbidity for schizophrenia and SUD in the 10th compared to the middle deciles, that was observed in the European sample (Extended Data Figure 8-9 and Supplementary Tables S21A and S22A) was also reflected in SZ-SUM-PRS and SUD-SUM-PRS of the ancestry outliers (See Supplementary Figure S28-3 and S28-4, and Supplementary Tables S28C and S28D). Notably, there was a tendency for a higher risk of developing SUD among depression cases inlOth decile compared to the middle deciles of the BP-PRS, SU-PRS and SUD-PRS, and less clear for the DEP-PRS, ANX-PRS and SZ-PRS (Supplementary Figure S28-4 and Table S28D). Regarding risk of comorbidity for anxiety and transitioning into bipolar disorder, the differences between 10^{th} and middle deciles observed in the European sample (Extended Data Figure 6-7 and Supplementary Table S19A and S20A) was, however, not clearly reflected in the ancestry outliers (Supplementary Figure S28-1 – S28-2 and Supplementary Table S28A and S28B).

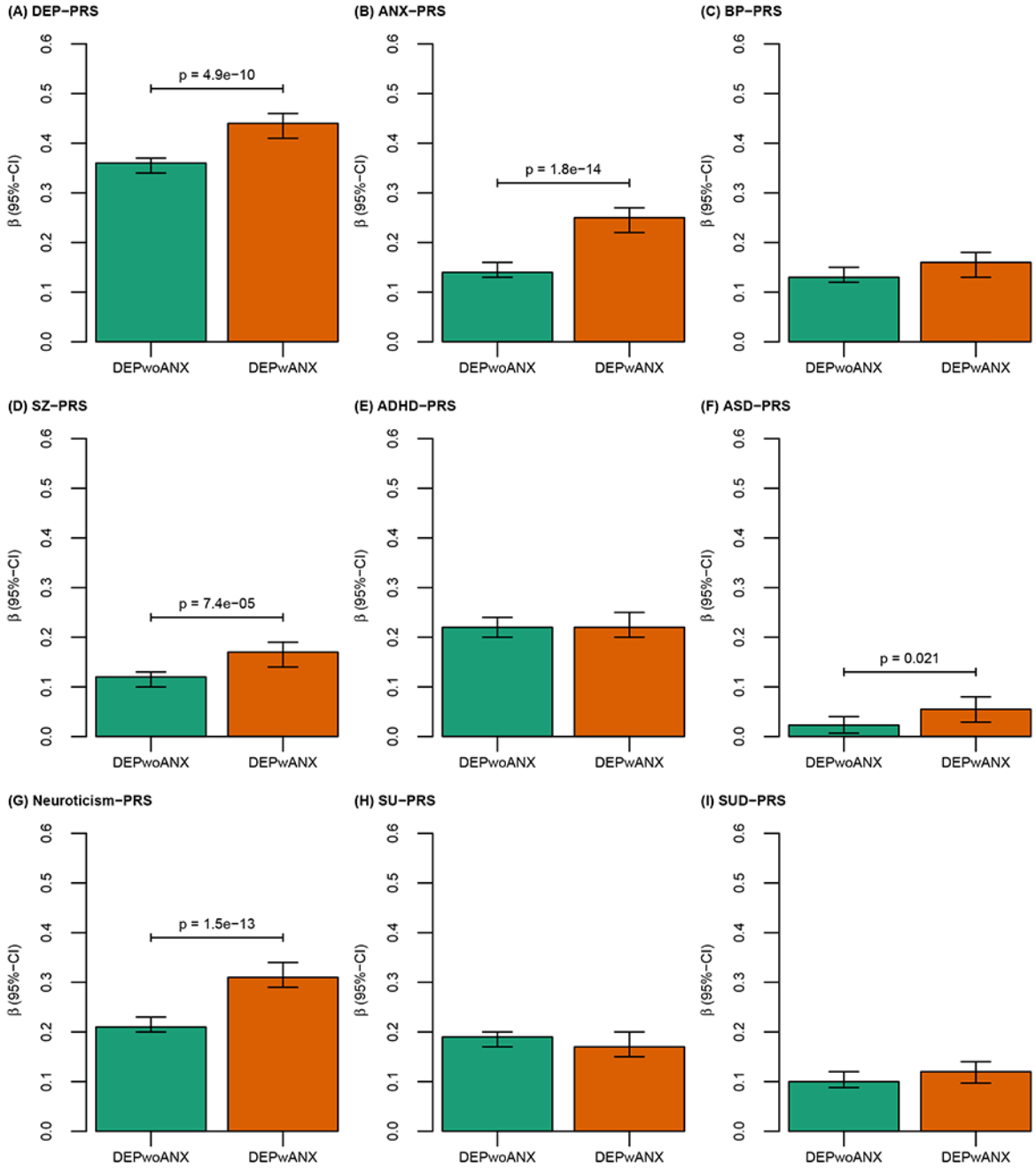
Extended Data



Extended Display Figure 1:

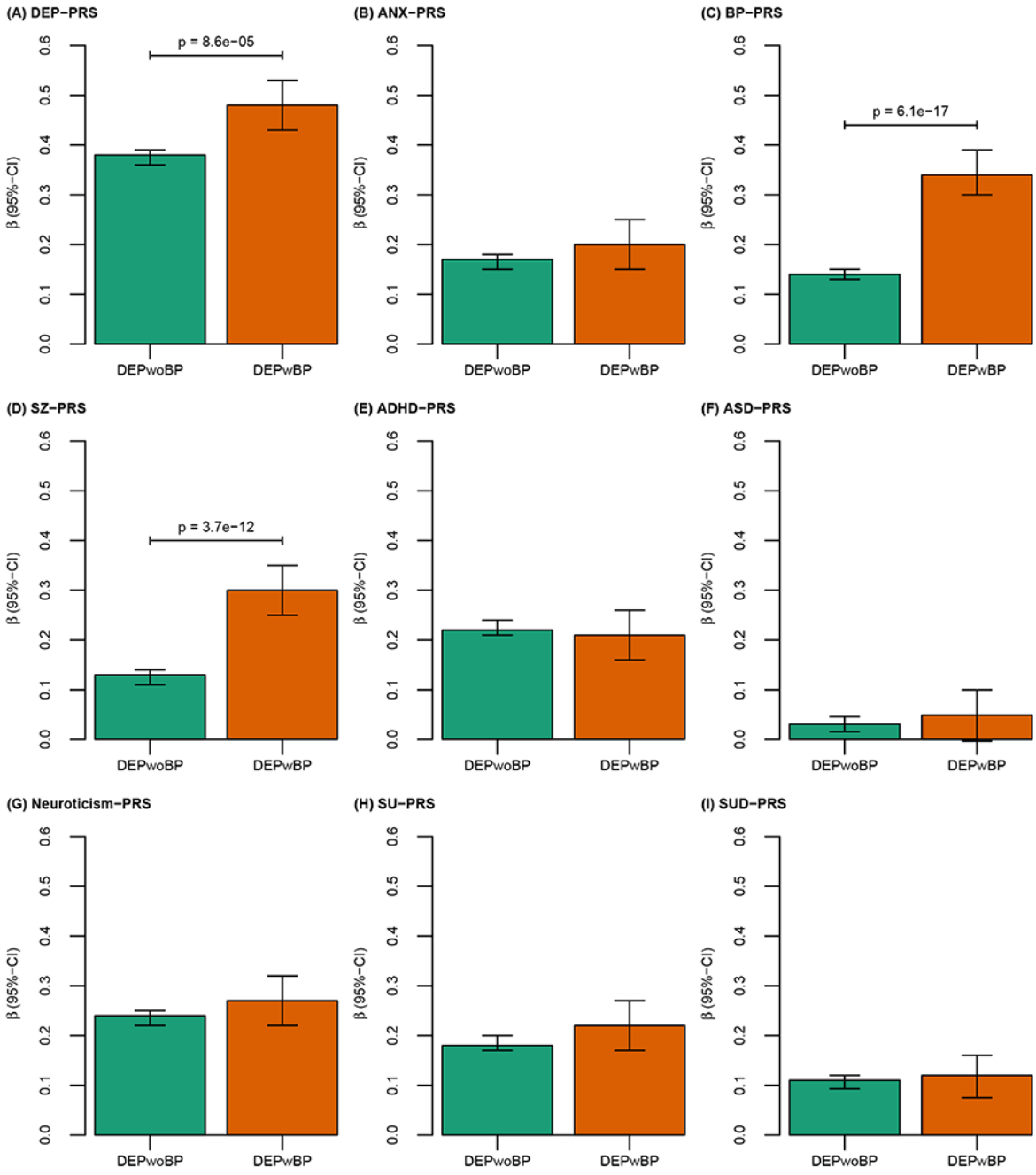
Highlighted Regional Miami plots of GWAS and TWAS results, corresponding to the genomic region of (A) GABRA1, (B) CYP7B1, (C) DCC, (D) CTTNBP2, (E) FURIN and (F) GIGYF2 genes/transcripts (1Mbp window from start site). Top panels: GWAS results (black dots): The x-axis shows genomic position, and the y-axis shows significance as $-\log_{10}$(P) of z statistics (two-sided nominal P-values); blue line corresponds to P=1×10–5, orange line to P=5×10–8 (genome-wide significance). Bottom panels: TWAS results: The x-axis shows genomic position. The y-axis shows significance as $-\log_{10}$(P) of z statistics (two-sided nominal P-values) for genes represented by both gene expression and isoform expression. Green triangles facing upwards or downwards for a positive or negative association z-score (Wald test; two-sided P-values) respectively (up- or down-regulation);

transcripts with Bonferroni-adjusted (for all reliably imputed transcripts) P-value<0.1 are labelled; orange line corresponds to Bonferroni-adjusted P=0.05. Each Bonferroni-significant transcript is connected with lines to the SNPs contributing to its transcriptomic imputation model; lines are grey when the SNPs have a P>1×10–5, blue when P<1×10–5 but orange when P<5×10–8. The SNPs that are above the blue line and contribute to the transcriptomic imputation models of significant transcripts are labelled. See Supplementary Figure S19-1 to S19-88 and Table S19.



Extended Display Figure 2:

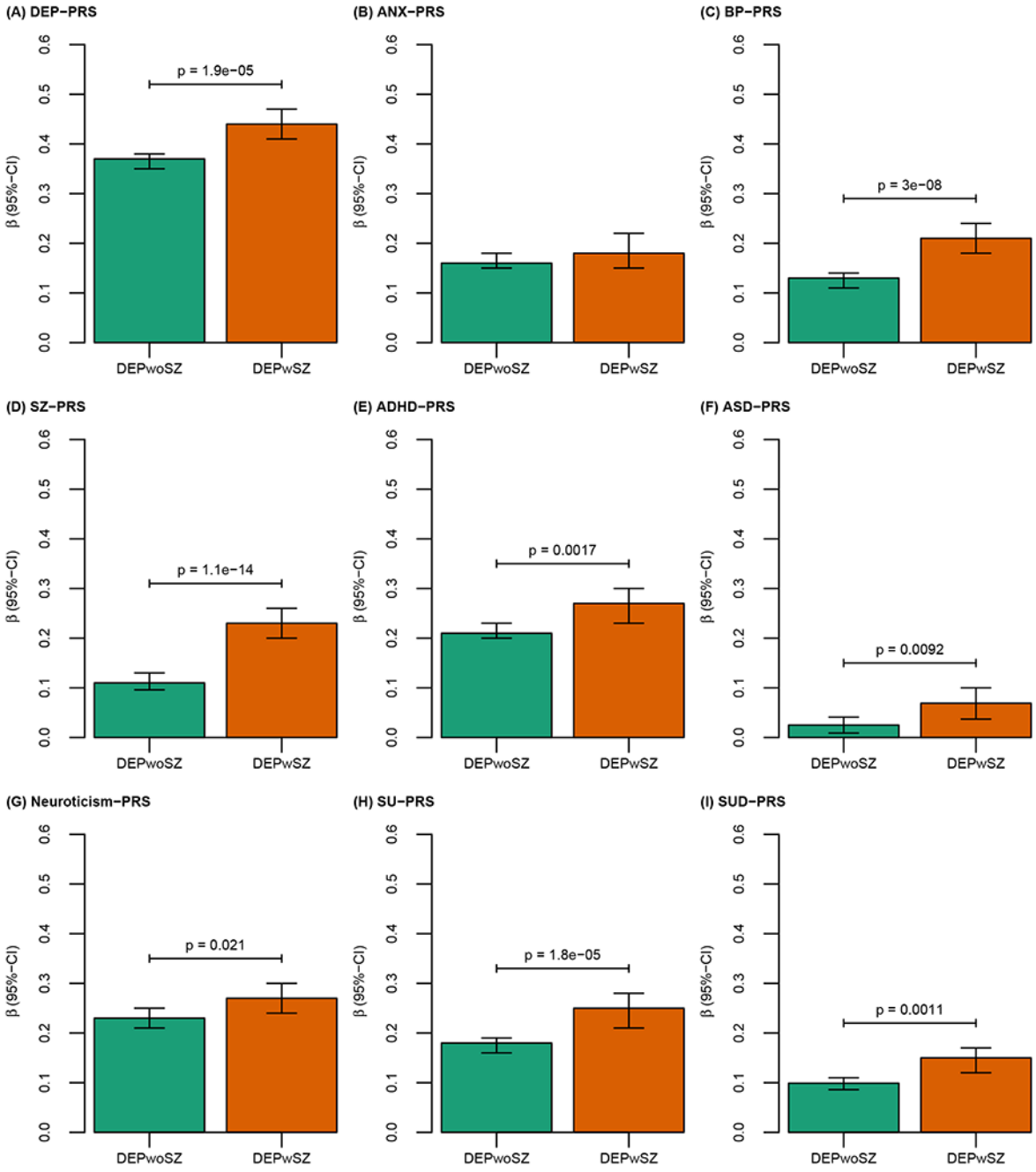
MvPRS analyses of depression cases with/without anxiety (ANX). Depression-subphenotype is shown on the x-axis ($N_{DEPwoANX}$=22114, $N_{DEPwoANX}$=7044 and N_{ctrls}=38142). The slope (β) of the linear regression (95% CI) for each depression subphenotype is shown on the y-axis. Significant difference between β for depression without/with an additional diagnosis is indicated with horizontal line with nominal two-sided P-value above, i.e. the Wald test of equal group effect (see Supplementary Table 12A). Overall two-sided P-value=1.2e-21. Cases with bipolar disorder were excluded. The polygenic risk scores analyzed are (A) PRS for depression (DEP-PRS), (B) PRS for anxiety (ANX-PRS), (C) PRS for bipolar disorder (BP-PRS), (D) PRS for schizophrenia (SZ-PRS), (E) PRS for ADHD (ADHD-PRS), (F) PRS for autism (ASD-PRS), (G) PRS for neuroticism (Neuroticism-PRS), (H) PRS for substance use (SU-PRS), and (I) PRS for substance use disorder (SUD-PRS). See Supplementary Table S12B, S12C, Supplementary Figure S14-1 and S14-2 for sex-stratified analyses.



Extended Display Figure 3:

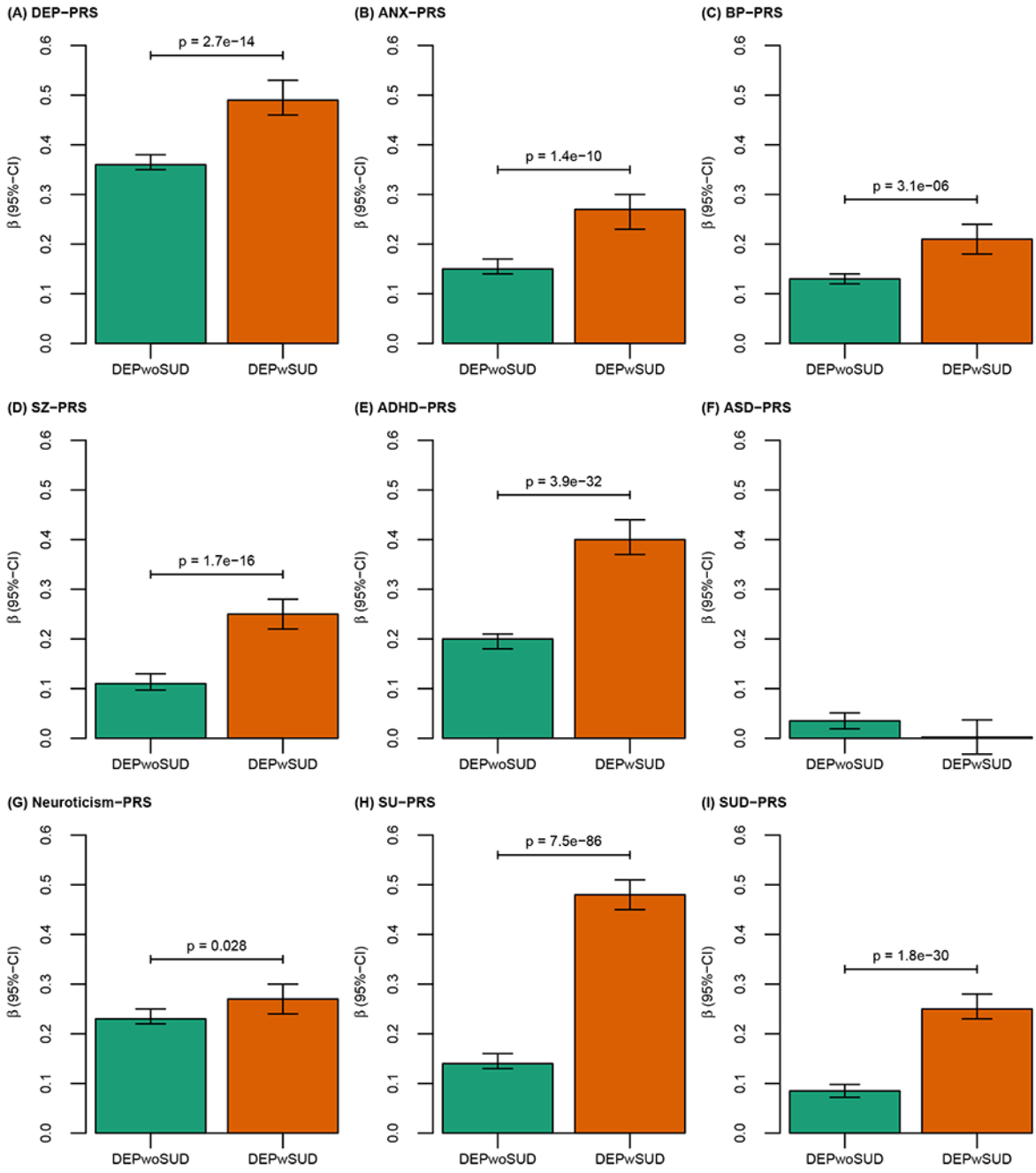
MvPRS analyses of depression cases with/without bipolar disorder (BP). Depression-subphenotype is shown on the x-axis ($N_{DEPwoBP}$=29158, DEPwBP, N=1460 and Nctrls=38200). The slope (β) of the linear regression (95% CI) for each depression subphenotype is shown on the y-axis. Significant difference between β for depression without/with an additional diagnosis is indicated with horizontal line with nominal two-sided P-value above, i.e. the Wald test of equal group effect (See Supplementary Table 12D). Overall two-sided P-value=1.5e-15. The polygenic risk scores analyzed are (A) PRS for

depression (DEP-PRS), (B) PRS for anxiety (ANX-PRS), (C) PRS for bipolar disorder (BP-PRS), (D) PRS for schizophrenia (SZ-PRS), (E) PRS for ADHD (ADHD-PRS), (F) PRS for autism (ASD-PRS), (G) PRS for neuroticism (Neuroticism-PRS), (H) PRS for substance use (SU-PRS), and (I) PRS for substance use disorder (SUD-PRS). See Supplementary Table S12E, S12F, Supplementary Figure S14–1 and S14-2 for sex-stratified analyses.



Extended Display Figure 4:

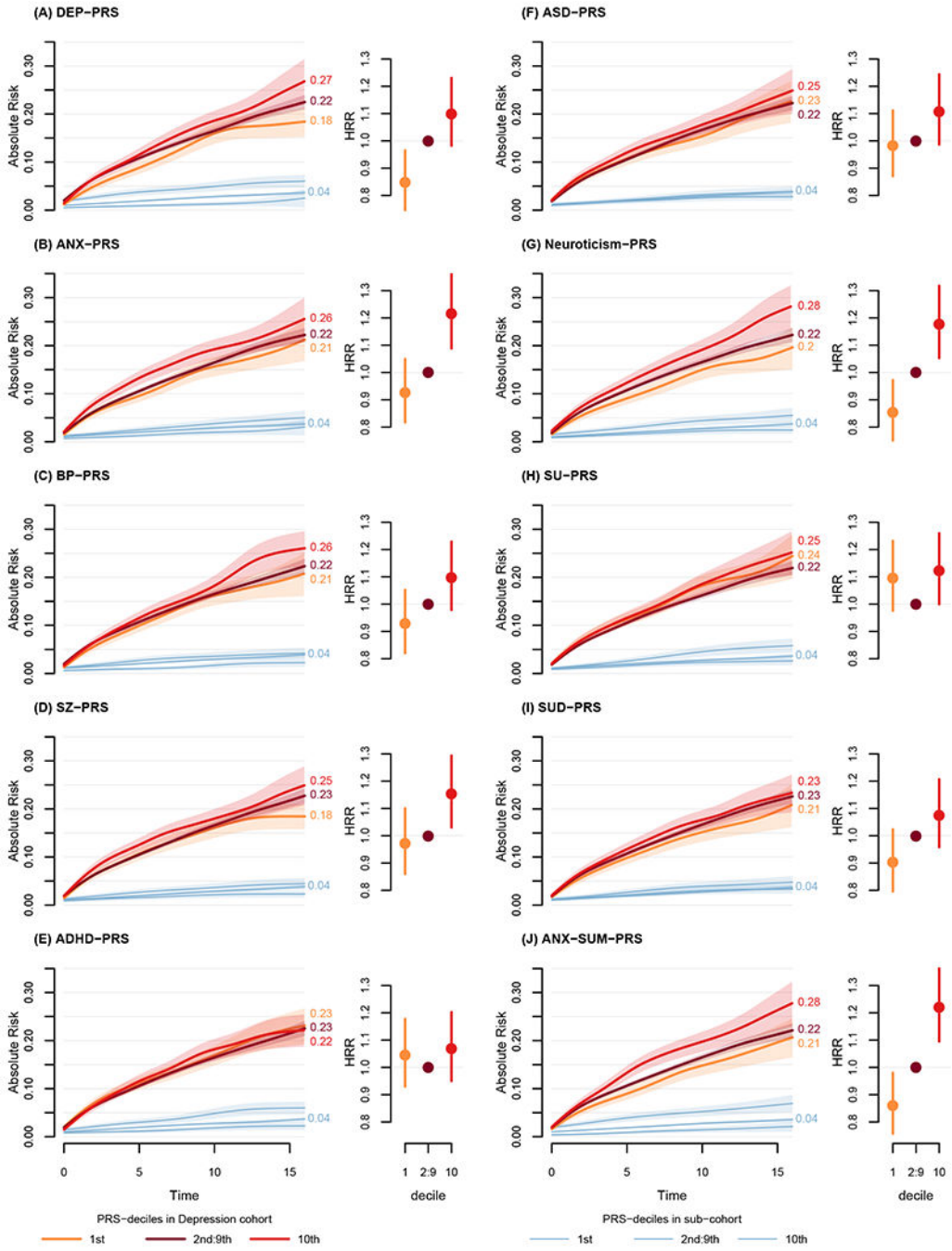
MvPRS analyses of depression cases with/without schizophrenia (SZ). Depression-subphenotype is shown on the x-axis ($N_{DEPwoSZ}$=25253, N_{DEPwSZ}=3905 and N_{ctrls}=38142). The slope (β) of the linear regression (95% CI) for each depression subphenotype is shown on the y-axis. Significant difference between β for depression without/with an additional diagnosis is indicated with horizontal line with nominal two-sided P-value above, i.e. the Wald test of equal group effect (see Supplementary Table 12G). Overall two-sided P-value=1.7e-15. Cases with bipolar disorder were excluded. The polygenic risk scores analyzed are (A) PRS for depression (DEP-PRS), (B) PRS for anxiety (ANX-PRS), (C) PRS for bipolar disorder (BP-PRS), (D) PRS for schizophrenia (SZ-PRS), (E) PRS for ADHD (ADHD-PRS), (F) PRS for autism (ASD-PRS), (G) PRS for neuroticism (Neuroticism-PRS), (H) PRS for substance use (SU-PRS), and (I) PRS for substance use disorder (SUD-PRS). See Supplementary Table S12H, S12I, Supplementary Figure S14-1 and S14-2 for sex-stratified analyses.



Extended Display Figure 5:

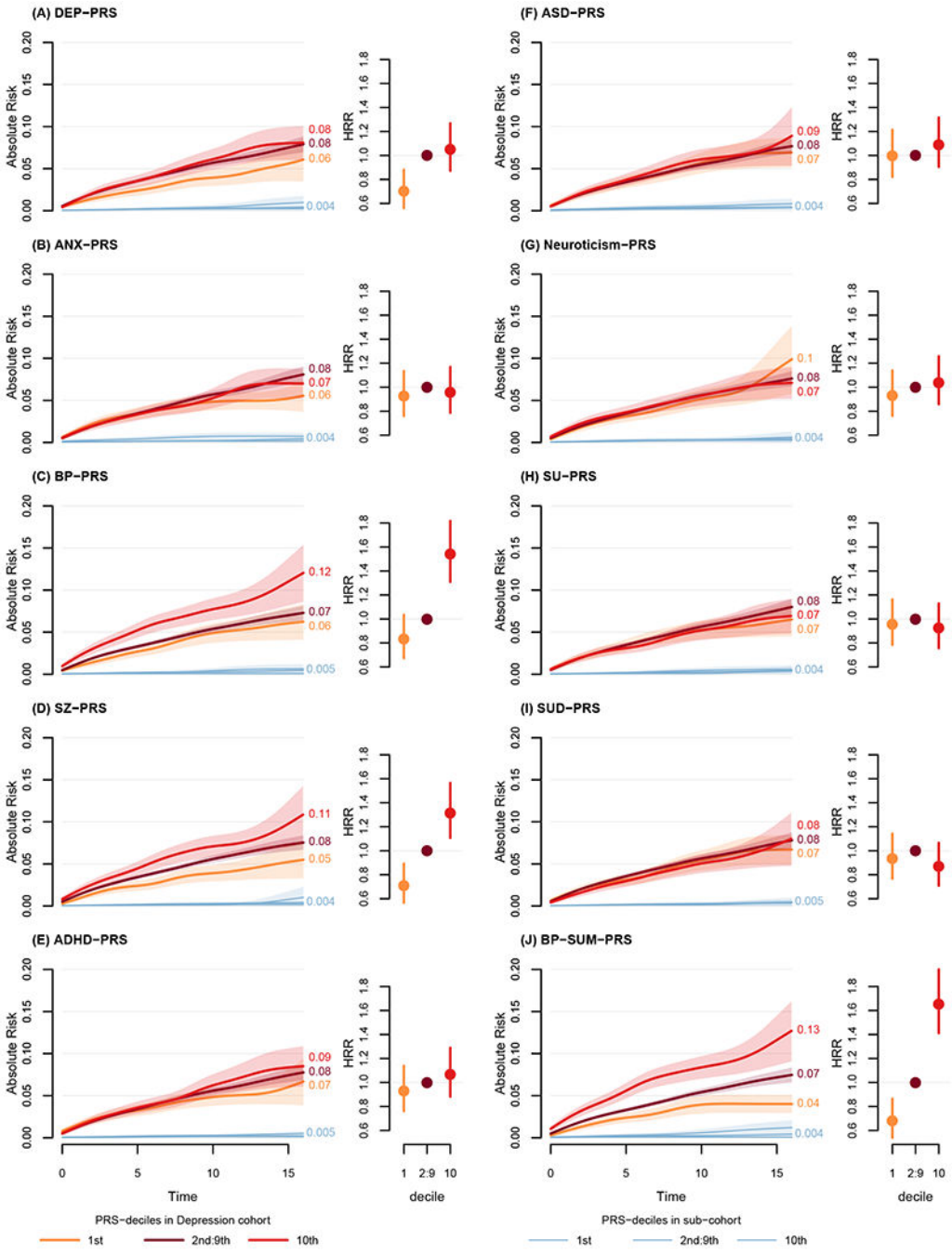
MvPRS analyses of depression cases with/without substance use disorder (SUD). Depression-subphenotype is shown on the x-axis ($N_{DEPwoSUD}$=25620, $N_{DEPwSUD}$=3538 and N_{ctrls}=38142). The slope (β) of the linear regression (95% CI) for each depression subphenotype is shown on the y-axis. Significant difference between β for depression without/with an additional diagnosis is indicated with horizontal line with nominal two-sided P-value above, i.e. the Wald test of equal group effect (see Supplementary Table 12J). Overall two-sided P-value=6.4e-97. Cases with bipolar disorder were excluded. The

polygenic risk scores analyzed are (A) PRS for depression (DEP-PRS), (B) PRS for anxiety (ANX-PRS), (C) PRS for bipolar disorder (BP-PRS), (D) PRS for schizophrenia (SZ-PRS), (E) PRS for ADHD (ADHD-PRS), (F) PRS for autism (ASD-PRS), (G) PRS for neuroticism (Neuroticism-PRS), (H) PRS for substance use (SU-PRS), and (I) PRS for substance use disorder (SUD-PRS). See Supplementary Table S12K, S12L, Supplementary Figure S14-1 and S14-2 for sex-stratified analyses.



Extended Display Figure 6:

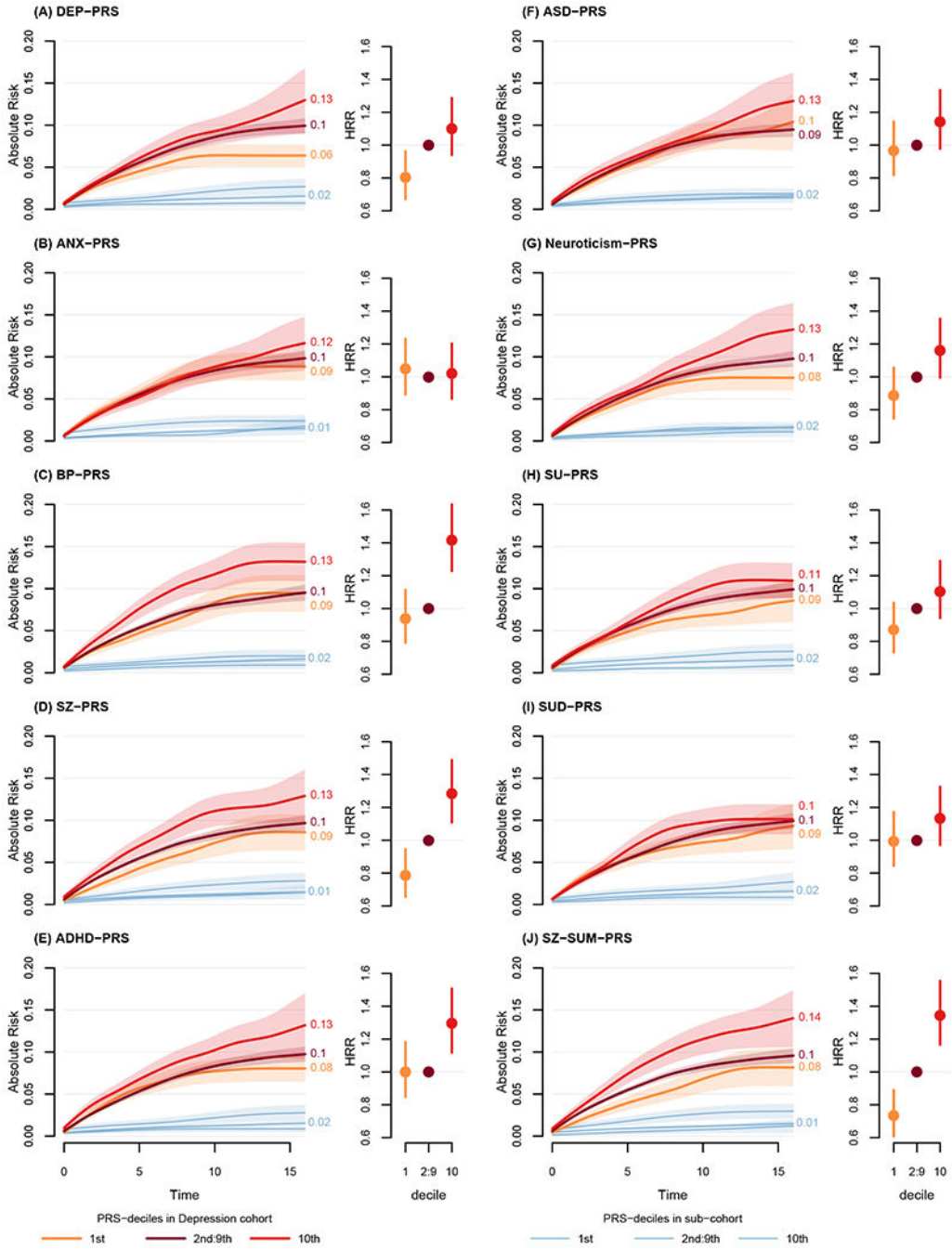
Left subpanels: Absolute risk (95%-CI) of developing anxiety since first depression episode for three groups of PRS deciles (1st, 2nd-to-9th and 10th) among 25124 depression cases with/without ($N_{DEPwANX}$=3010, $N_{DEPwoANX}$=221 14) anxiety. The absolute risk (95%-CI) of anxiety in light blue for the iPSYCH2015 subcohort (random population sample) excluding all depression cases, aligned to match the endpoint for the depression-cohort. **Right subpanels:** HRRs (95%) for 1st and 10th decile using 2nd-to-9th decile as reference (see Supplementary Table S13A) in colors matching the absolute risks curves. For deciles of (A) Depression DEP-PRS, (B) Anxiety ANX-PRS, (C) Bipolar disorder BP-PRS, (D) Schizophrenia SZ-PRS, (E) ADHD-PRS, (F) Autism ASD-PRS, (G) Neuroticism-PRS, (H) Substance Use SU-PRS, (I) Substance Use Disorder SUD-PRS and (J) sum of PRSs. The SUM-PRS was calculated by adding PRSs for multiple phenotypes weighted by log(OR) with the aim of optimizing prediction (see methods for details). See Supplementary Figure S15-1 – S15-2 and Supplementary Table S13B - S13C for sex-stratified analyses.



Extended Display Figure 7:

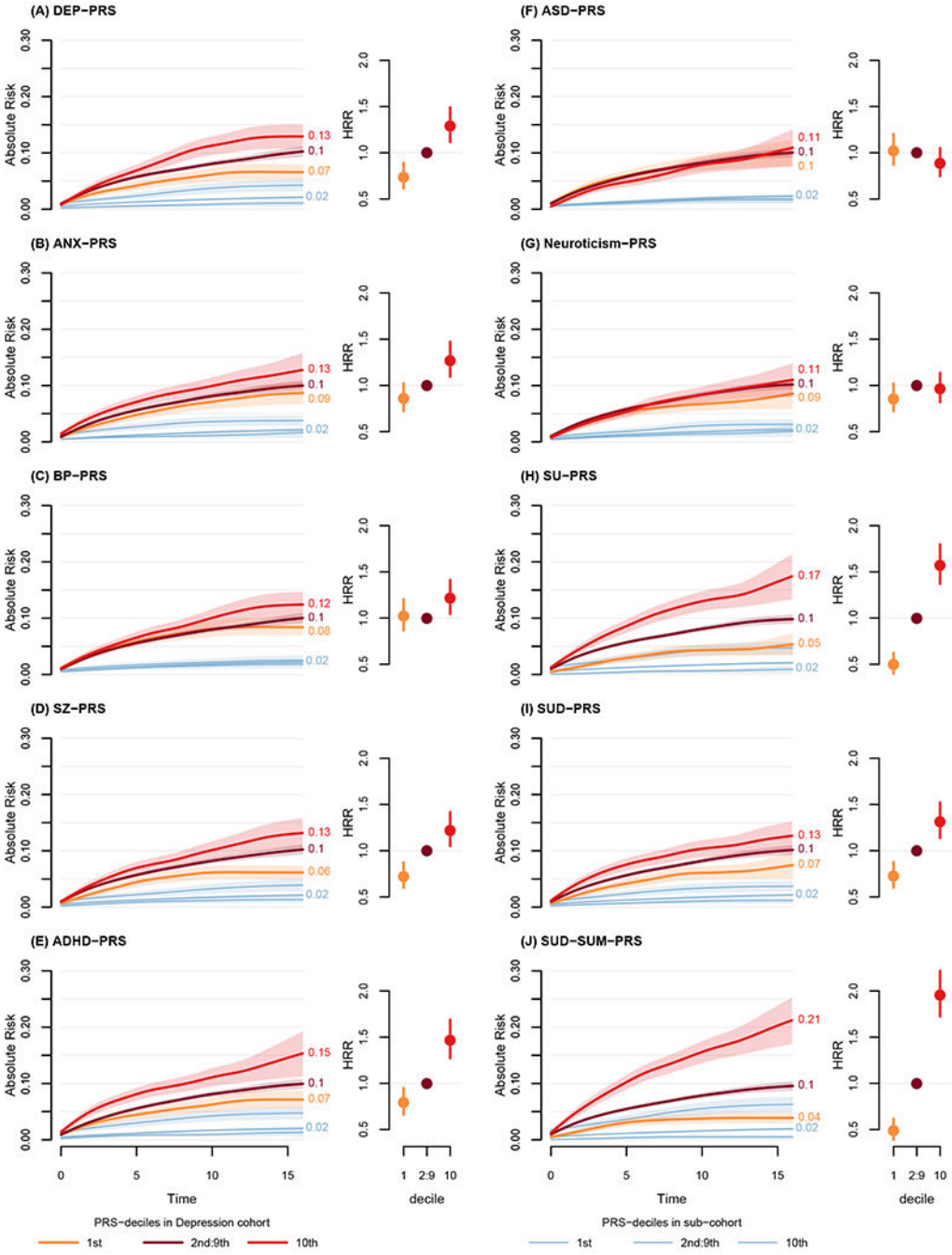
Left subpanels: Absolute risk (95%-CI) of transitioning into bipolar disorder since first depression episode for three groups of PRS deciles (1st, 2nd-to-9th and 10th) among 30300 depression cases with/without (N_{DEPwBP}=1142, $N_{DEPwoBP}$=29158) anxiety. The absolute risk (95%-CI) of anxiety in light blue for the iPSYCH2015 subcohort (random population sample) excluding all depression cases, aligned to match the endpoint for the depression-cohort. **Right subpanels:** HRRs (95%) for 1st and 10th decile using 2nd-to-9th decile as reference (see Supplementary Table S14A) in colors matching the absolute risks curves.

For deciles of (A) Depression DEP-PRS, (B) Anxiety ANX-PRS, (C) Bipolar disorder BP-PRS, (D) Schizophrenia SZ-PRS, (E) ADHD-PRS, (F) Autism ASD-PRS, (G) Neuroticism-PRS, (H) Substance Use SU-PRS, (I) Substance Use Disorder SUD-PRS and (J) sum of PRSs. The SUM-PRS was calculated by adding PRSs for multiple phenotypes weighted by log(OR) with the aim of optimizing prediction (see methods for details). See Supplementary Figure S16-1 – S16-2 Supplementary Tables S14B – S14C for sex-stratified analyses.



Extended Display Figure 8:

Left subpanels: Absolute risk (95%-CI) of developing schizophrenia since first depression episode for three groups of PRS deciles (1st, 2nd-to-9th and 10th) among 28714 depression cases with/without (N_{DEPwSZ}=1606, $N_{DEPwoSZ}$=27108) anxiety. The absolute risk (95%-CI) of anxiety in light blue for the iPSYCH2015 subcohort (random population sample) excluding all depression cases, aligned to match the endpoint for the depression-cohort. **Right subpanels:** HRRs (95%) for 1st and 10th decile using 2nd-to- 9th decile as reference (see Supplementary Table S15A) in colors matching the absolute risks curves. For deciles of (A) Depression DEP-PRS, (B) Anxiety ANX-PRS, (C) Bipolar disorder BP-PRS, (D) Schizophrenia SZ-PRS, (E) ADHD-PRS, (F) Autism ASD-PRS, (G) Neuroticism-PRS, (H) Substance Use SU-PRS, (I) Substance Use Disorder SUD-PRS and (J) sum of PRSs. The SUM-PRS was calculated by adding PRSs for multiple phenotypes weighted by log(OR) with the aim of optimizing prediction (see methods for details). See Supplementary Figure S17-1 – S17-2 and Supplementary Tables S15B – S15C for sex-stratified analyses.



Extended Display Figure 9:

Left subpanels: Absolute risk (95%-CI) of developing SUD since first depression episode for three groups of PRS deciles (1st, 2nd-to-9th and 10th) among 27249 depression cases with/without ($N_{DEPwSUD}$=1629, $N_{DEPwoSUD}$=25620) anxiety. The absolute risk (95%-CI) of anxiety in light blue for the iPSYCH2015 subcohort (random population sample) excluding all depression cases, aligned to match the endpoint for the depression-cohort. **Right subpanels:** HRRs (95%) for 1st and 10th decile using 2nd-to-9th decile as reference (see Supplementary Table S16A) in colors matching the absolute risks curves. For deciles

of (A) Depression DEP-PRS, (B) Anxiety ANX-PRS, (C) Bipolar disorder BP-PRS, (D) Schizophrenia SZ-PRS, (E) ADHD-PRS, (F) Autism ASD-PRS, (G) Neuroticism-PRS, (H) Substance Use SU-PRS, (I) Substance Use Disorder SUD-PRS and (J) sum of PRSs. The SUM-PRS was calculated by adding PRSs for multiple phenotypes weighted by log(OR) with the aim of optimizing prediction (see methods for details). See Supplementary Figure S18-1 – S18-2 and Supplementary Tables S16B – S16C for sex-stratified analyses.

Extended Display Table 1:

Number of comorbid diagnoses among depression subtypes: single-episode and recurrent depression. Numbers are after relatedness pruning / removal of ancestry outliers. Number in brackets are fraction of depression cases being comorbid with bipolar disorder (BP), schizophrenia (SZ), anxiety (ANX), autism (ASD), ADHD, substance use disorder (SUD) or having depression as the only diagnosis or being the fraction of depression cases that do not have bipolar disorder. Except for population-sample where numbers in brackets indicate phenotype fractions in the population-based iPSYCH2015 sample. All numbers are without bipolar disorder except for the BP column. Whether recurrent depression was associated with increased comorbidity compared to single-episode depression was tested using logistic regression for each additional diagnosis while adjusting for age. Since the age-distribution for cases was skewed compared to controls individuals were grouped in age bins in five-year intervals to construct dummy variables for the age adjustment. Odds-Ratios (OR=log(β)) and nominal two-sided P-values based on logistic regression z statistics are shown.

Phenotype	**BP**	**SZ**	**ANX**	**ASD**	**ADHD**	**SUD**	**DEP only**	**DEP wo BP**	**DEP total**
Depression									
No. samples	1460	3905	7044	2255	2571	3538	15226	29158	30618
F(*subtype*/*DEP*)	(0.0501)	(0.134)	(0.242)	(0.0773)	(0.0882)	(0.121)	(0.522)	(0.95)	-
Single-episode									
No. samples	961	2773	5104	1758	1933	2545	12765	23140	24101
F(*subtype*/*single*)	(0.0415)	(0.12)	(0.221)	(0.076)	(0.0835)	(0.11)	(0.552)	(0.96)	-
Recurrent									
No. samples	499	1132	1940	497	638	993	2461	6018	6517
F(*subtype*/*recurrent*)	(0.0829)	(0.188)	(0.322)	(0.0826)	(0.106)	(0.165)	(0.409)	(0.92)	-
Population-sample									
No. samples	106	401	888	636	774	502	556	1019	1067
F(*subtype*/*pop*)	(0.0027)	(0.01)	(0.023)	(0.016)	(0.02)	(0.013)	(0.014)	(0.026)	(0.027)
Test for difference in comorbidity between singel-episode and recurrent									
OR	1.86	1.76	1.64	1.48	1.39	1.64	0.53	-	-
p – value	3.1e-27	9.6e-48	2e-53	4.2e-12	1.3e-11	3.7e-32	5.8e-102	-	-

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003), the EU H2020 Program (Grant No. 667302, “CoCA”), NIH/NIMH (1U01MH109514-01 and 1R01MH124851-01 to ADB) and the Universities and University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to ADB).

We would like to thank the research participants and employees of 23andMe, Inc. for making this work possible.

We thank all members of the iPSYCH-Broad Consortium for their efforts and collaborative spirit. The team at the Center for Disease Neurogenomics at Icahn School of Medicine at Mount Sinai was supported by the NIH (K08MH122911 to GV; T32MH087004 to KT; R01MH125246, R01AG067025, U01MH116442 and R01MH109677 to PR).

The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing Interests Statement

B.J.V. is a member of the advisory board for Allelica. D.D. has received a speaker fee from Takeda. B.W., T.T., H.S. and K.S. are employed at deCODE/Amgen. M.J.D. is a founder of Maze Therapeutics and on the Scientific Advisory Board of RBNC Therapeutics. The other authors declare no competing interests.

Data availability

Supplementary Figures S1 – S15 and S24 – S28 are available at https://doi.org/10.6084/m9.figshare.22139849.

The GWAS meta-analysis summary statistics from this publication, not including 23andMe, are available at https://ipsych.dk/en/research/downloads/. To access the summary statistics from the meta-analysis of all cohorts, including 23andMe, a data transfer agreement is required from 23andMe (dataset-request@23andMe.com) before a request is made to the corresponding authors. See https://research.23andme.com/collaborate/#dataset-access/ for more information and to apply for access to the data.

All relevant iPSYCH data are available from the authors after approval by the iPSYCH Data Access Committee and can only be accessed on the secured Danish server (GenomeDK, https://genome.au.dk) as the data are protected by Danish legislation. For data access please contact: Anders D. Børglum (anders@biomed.au.dk).

The downloadable data of The Haplotype Reference Consortium was used for imputation:http://www.haplotype-reference-consortium.org/

Data used for brain transcriptome model generation are available from PsychENCODE (http://resource.psychencode.org/); genotypes are controlled data and access instructions are provided at https://www.synapse.org/#!Synapse:syn4921369/wiki/477467.

Resaources for the co-localization analysis are available at PsychENCODE (http://resource.psychencode.org/), ROSMAP

Note that some datasets have been indirectly accessed at the FUMA website. In general PsychENCODE (http://resource.psychencode.org/) was used for SNP annotations (enhancer, H3K27ac markers), eQTLs and HiC based enhancer-promoter interactions. GTEx v6/v7/v8 eQTLs and gene expression used in the pipeline were obtained from GTEx (http://www.gtexportal.org/home/). The following eQTL datasets in FUMA were used for gene mapping: BrainSeq (http://eqtl.brainseq.org/), PsychENCODE eQTLs (http://resource.psychencode.org/), Common Mind Consortium (https://www.synapse.org//#!Synapse:syn5585484), BRAINEAC (http://www.braineac.org/), GTEx/V8/Brain (https://www.gtexportal.org/home/datasets/). Chromatin interaction datasets in FUMA used for gene mapping: PsychENCODE eQTLs and HiC based enhancer-promoter interactions (http://resource.psychencode.org/), HiC (https://doi.org/10.1101/406330, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112) and Roadmap – brain (https://egg2.wustl.edu/roadmap/web_portal/DNase_reg.html). The following single cell RNA-sequencing data sets were used in the cell-type specific analyses in FUMA (https://fuma.ctglab.nl/tutorial#datasets): PsychENCODE Human developmental and adult brain samples (http://resource.psychencode.org/), Allen Brain Atlas Cell Type (http://celltypes.brain-map.org/api/v2/well_known_file_download/694416667), DroNc Human brain samples (hippocampus) (https://portals.broadinstitute.org/single_cell#study-dronc-seq-single-nucleus-rna-seq-on-human-archived-brain, https://www.gtexportal.org/), GSE104276 Human Prefrontal cortes brain samples (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104276), GSE67835 Human Cortex brain samples (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67835), Linarsson GSE101601 Human temporal cortex brain samples (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101601) and GSE76381 Human midbrain samples (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76381).

Gene ontology of Biological Process and Cellular Components data sets of MsigDB v7.0 (https://www.gsea-msigdb.org/gsea/msigdb) were used for the gene-set enrichment analysis in FUMA's GENE2FUNC module. Please refer to https://fuma.ctglab.nl/links and https://fuma.ctglab.nl/tutorial#datasets for additional information on availability of datasets.

References

1. Kessler RC & Bromet EJ The epidemiology of depression across cultures. Annual review of public health 34, 119–138 (2013). 10.1146/annurev-publhealth-031912-114409
2. Hasin DS et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. JAMA psychiatry 75, 336–346 (2018). 10.1001/iamapsvchiatry.2017.4602 [PubMed: 29450462]
3. Pedersen CB et al. A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. JAMA psychiatry 71, 573–581 (2014). 10.1001/iamapsvchiatrv.2014.16 [PubMed: 24806211]
4. Lopez AD, Mathers CD, Ezzati M, Jamison DT & Murray CJ Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 367, 1747–1757 (2006). 10.1016/S0140-6736(06)68770-9 [PubMed: 16731270]

5. Wittchen HU et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology 21, 655–679 (2011). 10.1016/i.euroneuro.2011.07.018 [PubMed: 21896369]
6. Gustavsson A. et al. Cost of disorders of the brain in Europe 2010. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology 21, 718–779 (2011). 10.1016/i.euroneuro.2011.08.008 [PubMed: 21924589]
7. Angst F, Stassen HH, Clayton PJ & Angst J Mortality of patients with mood disorders: follow-up over 34-38 years. J Affect Disord 68, 167–181 (2002). [PubMed: 12063145]
8. Ferrari AJ et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS medicine 10, e1001547 (2013). 10.1371/journal.pmed.1001547 [PubMed: 24223526]
9. Plana-Ripoll O. et al. Exploring Comorbidity Within Mental Disorders Among a Danish National Population. JAMA psychiatry 76, 259–270 (2019). 10.1001/iamapsvchiatrv.2018.3658 [PubMed: 30649197]
10. McGrath JJ et al. Comorbidity within mental disorders: a comprehensive analysis based on 145 990 survey respondents from 27 countries. Epidemiol Psychiatr Sci 29, e153 (2020). 10.1017/S2045796020000633 [PubMed: 32782057]
11. Sullivan PF, Neale MC & Kendler KS Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 157, 1552–1562 (2000). 10.1176/appi.aip.157.10.1552 [PubMed: 11007705]
12. Wray NR et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of maj'or depression. Nat Genet 50, 668–681 (2018). 10.1038/s41588-018-0090-3 [PubMed: 29700475]
13. Demontis D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature Genetics (2018). 10.1038/s41588-018-0269-7
14. Howard DM et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 22, 343–352 (2019). 10.1038/s41593-018-0326-7 [PubMed: 30718901]
15. Grove J. et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 51, 431–444 (2019). 10.1038/s41588-019-0344-8 [PubMed: 30804558]
16. Quello SB, Brady KT & Sonne SC Mood disorders and substance use disorder: a complex comorbidity. Sci Pract Perspect 3, 13–21 (2005). 10.1151/spp053113 [PubMed: 18552741]
17. Davis L, Uezato A, Newell JM & Frazier E Major depression and comorbid substance use disorders. Curr Opin Psychiatry 21, 14–18 (2008). 10.1097/YCQ.0b013e3282f32408 [PubMed: 18281835]
18. Murray GK et al. Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review. JAMA psychiatry 78, 210–219 (2021). 10.1001/iamapsvchiatry.2020.3042 [PubMed: 33052393]
19. Rees E & Qwen MJ Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. Genome Med 12, 43 (2020). 10.1186/s13073-020-00734-5 [PubMed: 32349784]
20. Perkins DO et al. Polygenic Risk Score Contribution to Psychosis Prediction in a Target Population of Persons at Clinical High Risk. Am J Psychiatry 177, 155–163 (2020). 10.1176/appi.aip.2019.18060721 [PubMed: 31711302]
21. Howard DM et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nature communications 9, 1470 (2018). 10.1038/s41467-018-03819-3
22. Kurki MI et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 613, 508–518 (2023). 10.1038/s41586-022-05473-8 [PubMed: 36653562]
23. Hyde CL et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet 48, 1031–1036 (2016). 10.1038/ng.3623 [PubMed: 27479909]

24. Levey DF et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat Neurosci 24, 954–963 (2021). 10.1038/s41593-021-00860-2 [PubMed: 34045744]
25. Bybjerg-Grauholm J. et al. The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. medRxiv, 2020.2011.2030.20237768 (2020). 10.1101/2020.11.30.20237768
26. Pedersen CB et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry (2017). 10.1038/mp.2017.196
27. Mors O, Perto GP & Mortensen PB The Danish Psychiatric Central Research Register. Scand J Public Health 39, 54–57 (2011). 10.1177/1403494810395825 [PubMed: 21775352]
28. Willer CJ, Li Y & Abecasis GR METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010). 10.1093/bioinformatics/btq340 [PubMed: 20616382]
29. Yang J, Lee SH, Goddard ME & Visscher PM GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88, 76–82 (2011). 10.1016/i.aihg.2010.11.011 [PubMed: 21167468]
30. Yang J, Lee SH, Goddard ME & Visscher PM Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. Methods in molecular biology 1019, 215–236 (2013). 10.1007/978-1-62703-447-0_9 [PubMed: 23756893]
31. Stankiewicz P. et al. Haploinsufficiency of the Chromatin Remodeler BPTF Causes Syndromic Developmental and Speech Delay, Postnatal Microcephaly, and Dysmorphic Features. Am J Hum Genet 101, 503–515 (2017). 10.1016/i.aihg.2017.08.014 [PubMed: 28942966]
32. Ansar M. et al. Biallelic variants in LINGO1 are associated with autosomal recessive intellectual disability, microcephaly, speech and motor delay. Genetics in medicine : official journal of the American College of Medical Genetics 20, 778–784 (2018). 10.1038/gim.2017.113 [PubMed: 28837161]
33. de Ligt J. et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 367, 1921–1929 (2012). 10.1056/NEJMoa1206524 [PubMed: 23033978]
34. Duman RS, Shinohara R, Fogaca MV & Hare B Neurobiology of rapid-acting antidepressants: convergent effects on GluA1-synaptic function. Mol Psychiatry 24, 1816–1832 (2019). 10.1038/s41380-019-0400-x [PubMed: 30894661]
35. Hara H. et al. TAK-653, an AMPA receptor potentiator with minimal agonistic activity, produces an antidepressant-like effect with a favorable safety profile in rats. Pharmacol Biochem Behav 211, 173289 (2021). 10.1016/i.pbb.2021.173289 [PubMed: 34655652]
36. Bulik-Sullivan BK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 47, 291–295 (2015). 10.1038/ng.3211 [PubMed: 25642630]
37. Verduijn J. et al. Using Clinical Characteristics to Identify Which Patients With Major Depressive Disorder Have a Higher Genetic Load for Three Psychiatric Disorders. Biol Psychiatry 81, 316–324 (2017). 10.1016/i.biopsych.2016.05.024 [PubMed: 27576130]
38. Zheng J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33, 272–279 (2017). 10.1093/bioinformatics/btw613 [PubMed: 27663502]
39. Johnson EC et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. Lancet Psychiatry 7, 1032–1045 (2020). 10.1016/S2215-0366(20)30339-4 [PubMed: 33096046]
40. Zhou H. et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. Nat Neurosci 23, 809–818 (2020). 10.1038/s41593-020-0643-5 [PubMed: 32451486]
41. Deak JD & Johnson EC Genetics of substance use disorders: a review. Psychol Med, 1–12 (2021). 10.1017/S0033291721000969
42. Gelernter J & Polimanti R Genetics of substance use disorders in the era of big data. Nat Rev Genet (2021). 10.1038/s41576-021-00377-1

43. Olvera RL et al. Common genetic influences on depression, alcohol, and substance use disorders in Mexican-American families. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 156B, 561–568 (2011). 10.1002/aimg.b.31196 [PubMed: 21557468]
44. Adams MJ et al. Genetic stratification of depression by neuroticism: revisiting a diagnostic tradition. Psychol Med 50, 2526–2535 (2020). 10.1017/S0033291719002629 [PubMed: 31576797]
45. Frei O. et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nature communications 10, 2417 (2019). 10.1038/s41467-019-10310-0
46. Demontis D. et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. medRxiv, 2022.2002.2014.22270780 (2022). 10.1101/2022.02.14.22270780
47. Werme J, van der Sluis S, Posthuma D & de Leeuw CA An integrated framework for local genetic correlation analysis. Nat Genet 54, 274–282 (2022). 10.1038/s41588-022-01017-y [PubMed: 35288712]
48. Watanabe K, Taskesen E, van Bochoven A & Posthuma D Functional mapping and annotation of genetic associations with FUMA. Nature communications 8, 1826 (2017). 10.1038/s41467-017-01261-5
49. Zhang W. et al. Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. Nature communications 10, 3834 (2019). 10.1038/s41467-019-11874-7
50. Wang D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362 (2018). 10.1126/science.aat8464
51. Gandal MJ et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362 (2018). 10.1126/science.aat8127
52. Berisa T & Pickrell JK Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics 32, 283–285 (2016). 10.1093/bioinformatics/btv546 [PubMed: 26395773]
53. Zeng B. et al. Trans-ethnic eQTL meta-analysis of human brain reveals regulatory architecture and candidate causal variants for brain-related traits. medRxiv, 2021.2001.2025.21250099 (2021). 10.1101/2021.01.25.21250099
54. Giambartolomei C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. Plos Genet 10, e1004383 (2014). 10.1371/iournal.pgen.1004383 [PubMed: 24830394]
55. Dobbyn A. et al. Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. Am J Hum Genet 102, 1169–1184 (2018). 10.1016/i.aihg.2018.04.011 [PubMed: 29805045]
56. Hormozdiari F. et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. Am J Hum Genet 99, 1245–1260 (2016). 10.1016/i.aihg.2016.10.003 [PubMed: 27866706]
57. Meuleman W. et al. Index and biological spectrum of human DNase I hypersensitive sites. Nature 584, 244–251 (2020). 10.1038/s41586-020-2559-3 [PubMed: 32728217]
58. Corces MR et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nat Genet 52, 1158–1168 (2020). 10.1038/s41588-020-00721-x [PubMed: 33106633]
59. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228–1235 (2015). 10.1038/ng.3404 [PubMed: 26414678]
60. Holmes SE et al. Lower synaptic density is associated with depression severity and network alterations. Nature communications 10, 1529 (2019). 10.1038/s41467-019-09562-7
61. Wang YT et al. Novel rapid-acting glutamatergic modulators: Targeting the synaptic plasticity in depression. Pharmacol Res 171, 105761 (2021). 10.1016/i.phrs.2021.105761 [PubMed: 34242798]

62. Szebeni A. et al. Shortened telomere length in white matter oligodendrocytes in major depression: potential role of oxidative stress. Int J Neuropsychopharmacol 17, 1579–1589 (2014). 10.1017/S1461145714000698 [PubMed: 24967945]

63. Zhou B, Zhu Z, Ransom BR & Tong X Oligodendrocyte lineage cells and depression. Mol Psychiatry 26, 103–117 (2021). 10.1038/s41380-020-00930-0 [PubMed: 33144710]

64. Leng L. et al. Menin Deficiency Leads to Depressive-like Behaviors in Mice by Modulating Astrocyte-Mediated Neuroinflammation. Neuron 100, 551–563 e557 (2018). 10.1016/i.neuron.2018.08.031 [PubMed: 30220511]

65. Satterthwaite TD et al. Neuroimaging of the Philadelphia neurodevelopmental cohort. Neuroimage 86, 544–553 (2014). 10.1016/i.neuroimage.2013.07.064 [PubMed: 23921101]

66. Calkins ME et al. The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. J Child Psychol Psychiatry 56, 1356–1369 (2015). 10.1111/jcpp.12416 [PubMed: 25858255]

67. Gur RC et al. Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. Neuropsychology 26, 251–265 (2012). 10.1037/a0026712 [PubMed: 22251308]

68. Wilkinson GS & Robertson GJ Wide range achievement test (WRAT4). (Pearson).

69. Cullen B. et al. Cognitive function and lifetime features of depression and bipolar disorder in a large population sample: Cross-sectional study of 143,828 UK Biobank participants. Eur Psychiatry 30, 950–958 (2015). 10.1016/i.eurpsy.2015.08.006 [PubMed: 26647871]

70. Lim J. et al. Sensitivity of cognitive tests in four cognitive domains in discriminating MDD patients from healthy controls: a meta-analysis. Int Psychogeriatr 25, 1543–1557 (2013). 10.1017/S1041610213000689 [PubMed: 23725644]

71. Snyder HR Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. Psychol Bull 139, 81–132 (2013). 10.1037/a0028727 [PubMed: 22642228]

72. Musliner KL et al. Polygenic liability, stressful life events and risk for secondary- treated depression in early life: a nationwide register-based case-cohort study. Psychol Med, 1–10 (2021). 10.1017/S0033291721001410

73. Musliner KL et al. Polygenic Liability and Recurrence of Depression in Patients With First-Onset Depression Treated in Hospital-Based Settings. JAMA psychiatry (2021). 10.1001/iamapsvchiatry.2021.0701

74. Steffen A, Nubel J, Jacobi F, Batzing J & Holstiege J Mental and somatic comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data. BMC Psychiatry 20, 142 (2020). 10.1186/s12888-020-02546-8 [PubMed: 32228541]

75. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev 31, 1117–1125 (2011). 10.1016/i.cpr.2011.07.004 [PubMed: 21820991]

76. Green EK et al. Variation at the GABAA receptor gene, Rho 1 (GABRR1) associated with susceptibility to bipolar schizoaffective disorder. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 153B, 1347–1349 (2010). 10.1002/aimg.b.31108 [PubMed: 20583128]

77. Gulsuner S. et al. Genetics of schizophrenia in the South African Xhosa. Science 367, 569–573 (2020). 10.1126/science.aay8833 [PubMed: 32001654]

78. Brennand KJ et al. Modelling schizophrenia using human induced pluripotent stem cells. Nature 473, 221–225 (2011). 10.1038/nature09915 [PubMed: 21490598]

79. Duman RS, Sanacora G & Krystal JH Altered Connectivity in Depression: GABA and Glutamate Neurotransmitter Deficits and Reversal by Novel Treatments. Neuron 102, 75–90 (2019). 10.1016/i.neuron.2019.03.013 [PubMed: 30946828]

80. Selten M, van Bokhoven H & Nadif Kasri N Inhibitory control of the excitatory/inhibitory balance in psychiatric disorders. F1000Res 7, 23 (2018). 10.12688/f1000research.12155.1 [PubMed: 29375819]

81. Gao R & Penzes P Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. Curr Mol Med 15, 146–167 (2015). 10.2174/1566524015666150303003028 [PubMed: 25732149]
82. Bruining H. et al. Measurement of excitation-inhibition ratio in autism spectrum disorder using critical brain dynamics. Scientific reports 10, 9195 (2020). 10.1038/s41598-020-65500-4 [PubMed: 32513931]
83. Nelson SB & Valakh V Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. Neuron 87, 684–698 (2015). 10.1016/i.neuron.2015.07.033 [PubMed: 26291155]
84. Sequeira A. et al. Global brain gene expression analysis links glutamatergic and GABAergic alterations to suicide and major depression. PLoS One 4, e6585 (2009). 10.1371/iournal.pone.0006585 [PubMed: 19668376]
85. Merali Z. et al. Dysregulation in the suicide brain: mRNA expression of corticotropin-releasing hormone receptors and GABA(A) receptor subunits in frontal cortical brain region. J Neurosci 24, 1478–1485 (2004). 10.1523/JNEUROSCI.4734-03.2004 [PubMed: 14960621]
86. Vinkers CH, Mirza NR, Olivier B & Kahn RS The inhibitory GABA system as a therapeutic target for cognitive symptoms in schizophrenia: investigational agents in the pipeline. Expert Opin Investig Drugs 19, 1217–1233 (2010). 10.1517/13543784.2010.513382
87. Engin E, Liu J & Rudolph U alpha2-containing GABA(A) receptors: a target for the development of novel treatment strategies for CNS disorders. Pharmacol Ther 136, 142–152 (2012). 10.1016/i.pharmthera.2012.08.006 [PubMed: 22921455]
88. Braat S & Kooy RF The GABAA Receptor as a Therapeutic Target for Neurodevelopmental Disorders. Neuron 86, 1119–1130 (2015). 10.1016/i.neuron.2015.03.042 [PubMed: 26050032]
89. Benasi G. et al. Benzodiazepines as a Monotherapy in Depressive Disorders: A Systematic Review. Psychother Psychosom 87, 65–74 (2018). 10.1159/000486696 [PubMed: 29466801]
90. Fasipe OJ, Agede OA & Enikuomehin AC Announcing the novel class of GABAA receptor selective positive allosteric modulator antidepressants. Future Sci OA 7, FSO654 (2020). 10.2144/fsoa-2020-0108 [PubMed: 33437518]
91. Hindley G. et al. The shared genetic basis of mood instability and psychiatric disorders: A cross-trait genome-wide association analysis. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 189, 207–218 (2022). 10.1002/aimg.b.32907 [PubMed: 35841185]
92. Hujoel MLA, Loh P-R, Neale BM & Price AL Incorporating family history of disease improves polygenic risk scores in diverse populations. bioRxiv, 2021.2004.2015.439975 (2021). 10.1101/2021.04.15.439975
93. Torkamani A, Wineinger NE & Topol EJ The personal and clinical utility of polygenic risk scores. Nat Rev Genet 19, 581–590 (2018). 10.1038/s41576-018-0018-x [PubMed: 29789686]
94. Vieta E. et al. Early Intervention in Bipolar Disorder. Am J Psychiatry 175, 411–426 (2018). 10.1176/appi.aip.2017.17090972 [PubMed: 29361850]
95. Lieberman JA, Small SA & Girgis RR Early Detection and Preventive Intervention in Schizophrenia: From Fantasy to Reality. Am J Psychiatry 176, 794–810 (2019). 10.1176/appi.aip.2019.19080865 [PubMed: 31569988]
96. Carney T & Myers B Effectiveness of early interventions for substance-using adolescents: findings from a systematic review and meta-analysis. Subst Abuse Treat Prev Policy 7, 25 (2012). 10.1186/1747-597X-7-25 [PubMed: 22697269]
97. Kessler RC & Price RH Primary prevention of secondary disorders: a proposal and agenda. Am J Community Psychol 21, 607–633 (1993). 10.1007/BF00942174 [PubMed: 8192124]
98. Polygenic Risk Score Task Force of the International Common Disease, A. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat Med 27, 1876–1884 (2021). 10.1038/s41591-021-01549-6 [PubMed: 34782789]
99. Hollegaard MV et al. Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. BMC Genet 12, 58 (2011). 10.1186/1471-2156-12-58 [PubMed: 21726430]

100. Borglum AD et al. Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. Mol Psychiatry (2013). 10.1038/mp.2013.2
101. Illumina Gencall Data Analysis Software (Illumina, Inc., San Diego, 2005).
102. Korn JM et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 40, 1253–1260 (2008). 10.1038/ng.237 [PubMed: 18776909]
103. Goldstein JI et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics 28, 2543–2545 (2012). 10.1093/bioinformatics/bts479 [PubMed: 22843986]
104. Lam M. et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. Bioinformatics 36, 930–933 (2020). 10.1093/bioinformatics/btz633 [PubMed: 31393554]
105. Price AL et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–909 (2006). 10.1038/ng1847 [PubMed: 16862161]
106. Schork AJ et al. A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. Nat Neurosci 22, 353–361 (2019). 10.1038/s41593-018-0320-0 [PubMed: 30692689]
107. Loh PR et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48, 1443–1448 (2016). 10.1038/ng.3679 [PubMed: 27694958]
108. Howie B, Fuchsberger C, Stephens M, Marchini J & Abecasis GR Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44, 955–959 (2012). 10.1038/ng.2354 [PubMed: 22820512]
109. Iglesias AI et al. Haplotype reference consortium panel: Practical implications of imputations with large reference panels. Hum Mutat 38, 1025–1032 (2017). 10.1002/humu.23247 [PubMed: 28493391]
110. Price AL et al. Long-range LD can confound genome scans in admixed populations. Am J Hum Genet 83, 132–135; author reply 135-139 (2008). 10.1016/i.aihg.2008.06.005 [PubMed: 18606306]
111. Zhou W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 50, 1335–1341 (2018). 10.1038/s41588-018-0184-y [PubMed: 30104761]
112. Levey DF et al. Reproducible Genetic Risk Loci for Anxiety: Results From approximately 200,000 Participants in the Million Veteran Program. Am J Psychiatry 177, 223–232 (2020). 10.1176/appi.aip.2019.19030256 [PubMed: 31906708]
113. Bulik-Sullivan B. et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236–1241 (2015). 10.1038/ng.3406 [PubMed: 26414676]
114. Pasman JA et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat Neurosci 21, 1161–1170 (2018). 10.1038/s41593-018-0206-1 [PubMed: 30150663]
115. Walters RK et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci 21, 1656–1669 (2018). 10.1038/s41593-018-0275-1 [PubMed: 30482948]
116. Karlsson Linner R et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 51, 245–257 (2019). 10.1038/s41588-018-0309-3 [PubMed: 30643258]
117. Liu M. et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet 51, 237–244 (2019). 10.1038/s41588-018-0307-5 [PubMed: 30643251]
118. Mullins N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet (2021). 10.1038/s41588-021-00857-4
119. Lee JJ et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50, 1112–1121 (2018). 10.1038/s41588-018-0147-3 [PubMed: 30038396]

120. Nagel M, Watanabe K, Stringer S, Posthuma D & van der Sluis S Item-level analyses reveal genetic heterogeneity in neuroticism. Nature communications 9, 905 (2018). 10.1038/s41467-018-03242-8

121. Trubetskoy V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature (2022). 10.1038/s41586-022-04434-5

122. Bellenguez C et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet 54, 412–436 (2022). 10.1038/s41588-022-01024-z [PubMed: 35379992]

123. International League Against Epilepsy Consortium on Complex, E. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nature communications 9, 5269 (2018). 10.1038/s41467-018-07524-z

124. Yengo L. et al. A Saturated Map of Common Genetic Variants Associated with Human Height from 5.4 Million Individuals of Diverse Ancestries. bioRxiv, 2022.2001.2007.475305 (2022). 10.1101/2022.01.07.475305

125. Schaid DJ, Chen W & Larson NB From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19, 491–504 (2018). 10.1038/s41576-018-0016-z [PubMed: 29844615]

126. Kichaev G. et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. Bioinformatics 33, 248–255 (2017). 10.1093/bioinformatics/btw615 [PubMed: 27663501]

127. Kichaev G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. Plos Genet 10, e1004722 (2014). 10.1371/iournal.pgen.1004722 [PubMed: 25357204]

128. Kichaev G & Pasaniuc B Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. Am J Hum Genet 97, 260–271 (2015). 10.1016/i.aihg.2015.06.007 [PubMed: 26189819]

129. Benner C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32, 1493–1501 (2016). 10.1093/bioinformatics/btw018 [PubMed: 26773131]

130. Bennett DA et al. Religious Orders Study and Rush Memory and Aging Project. Journal of Alzheimer's disease : JAD 64, S161–S189 (2018). 10.3233/JAD-179939 [PubMed: 29865057]

131. Consortium, G. T. et al. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017). 10.1038/nature24277 [PubMed: 29022597]

132. Prive F, Arbel J & Vilhjalmsson BJ LDpred2: better, faster, stronger. Bioinformatics (2020). 10.1093/bioinformatics/btaa1029

133. Therneau TM & Grambsch PM Modeling Survival Data: Extending the Cox Model. (Springer, 2000).

134. A Package for Survival Analysis in R, R package version 3.2-13 (2021).

135. Link CL Confidence-Intervals for the Survival Function Using Cox Proportional- Hazard Model with Covariates. Biometrics 40, 601–610 (1984). 10.2307/2530904 [PubMed: 6394063]

136. Lee AB, Luca D, Klei L, Devlin B & Roeder K Discovering genetic ancestry using spectral graph theory. Genet Epidemiol 34, 51–59 (2010). 10.1002/gepi.20434 [PubMed: 19455578]

137. Klei L, Kent BP, Melhem N, Devlin B & Roeder K Gemtools: a fast and efficient approach to estimating genetic ancestry. *arXiv* **1104.1162.** (2011).

138. Ge T, Chen CY, Ni Y, Feng YA & Smoller JW Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nature communications 10, 1776 (2019). 10.1038/s41467-019-09718-5

139. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7 (2015). 10.1186/s13742-015-0047-8 [PubMed: 25722852]

140. Hoffman GE & Schadt EE variancePartition: interpreting drivers of variation in complex gene expression studies. Bmc Bioinformatics 17, 483 (2016). 10.1186/s12859-016-1323-z [PubMed: 27884101]

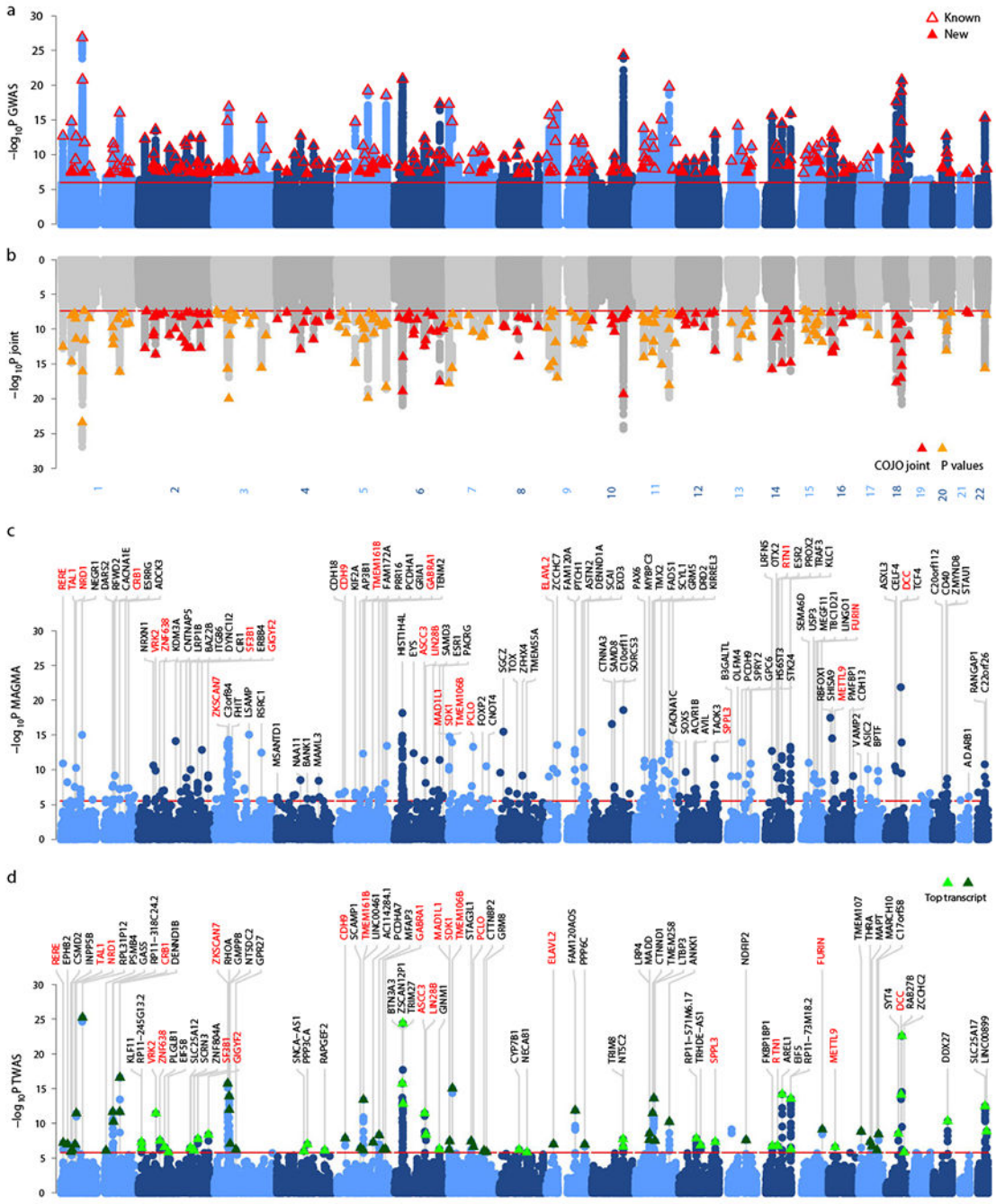


Figure 1:

The x-axes show genomic position, and the y-axes shows statistical significance as $-\log_{10}(P)$. (A) Manhattan plot of the primary meta-analysis (371184 cases and 978703 controls) using an inverse variance-weighted fixed-effects model, the y-axis shows statistical significance as $-\log_{10}(P)$ of z statistics (two-sided nominal P-values). The red horizontal line represents the threshold for genome-wide significant association ($P=5\times10^{-8}$), QQ-plot in Supplementary figure S5. Data on chromosome X were only available for iPSYCH and FinnGen, see Supplementary Figure S2). (B) Manhattan plot of the primary

meta-analysis from A, the y-axis shows $-\log_{10}$(P) of z statistics (two-sided nominal P-values), with highlighted joint p-values from GCTA-COJO plotted on top. The red horizontal line represents the threshold for genome-wide significant association ($P=5\times10^{-8}$). (C) Gene-based analysis: y-axis shows $-\log_{10}$(P) of F statistics (two-sided nominal P-values) implemented in MAGMA. A red line indicates Bonferroni corrected genome-wide significance; P<2.8×10–6 (testing 17,840 protein coding genes). A total 411 significant genes of which 268 were in 93 of the 243 genomic risk loci (Supplementary Table S6). The most significant gene, overlapping with a top gene in the TWAS D, within each of the 93 genomic loci are highlighted. (D) TWAS: The y-axis shows significance as $-\log_{10}$(P) of association z statistics (Wald test; two-sided P-values). Genes are represented by both gene and isoform expression. A red line indicates Bonferroni corrected genome-wide significance; P<1.44×10–6 (34,646 tests for all reliably imputed genes and isoforms). The top transcript is labelled and the corresponding $-\log_{10}$(P) is highlighted for each independent linkage disequilibrium block.

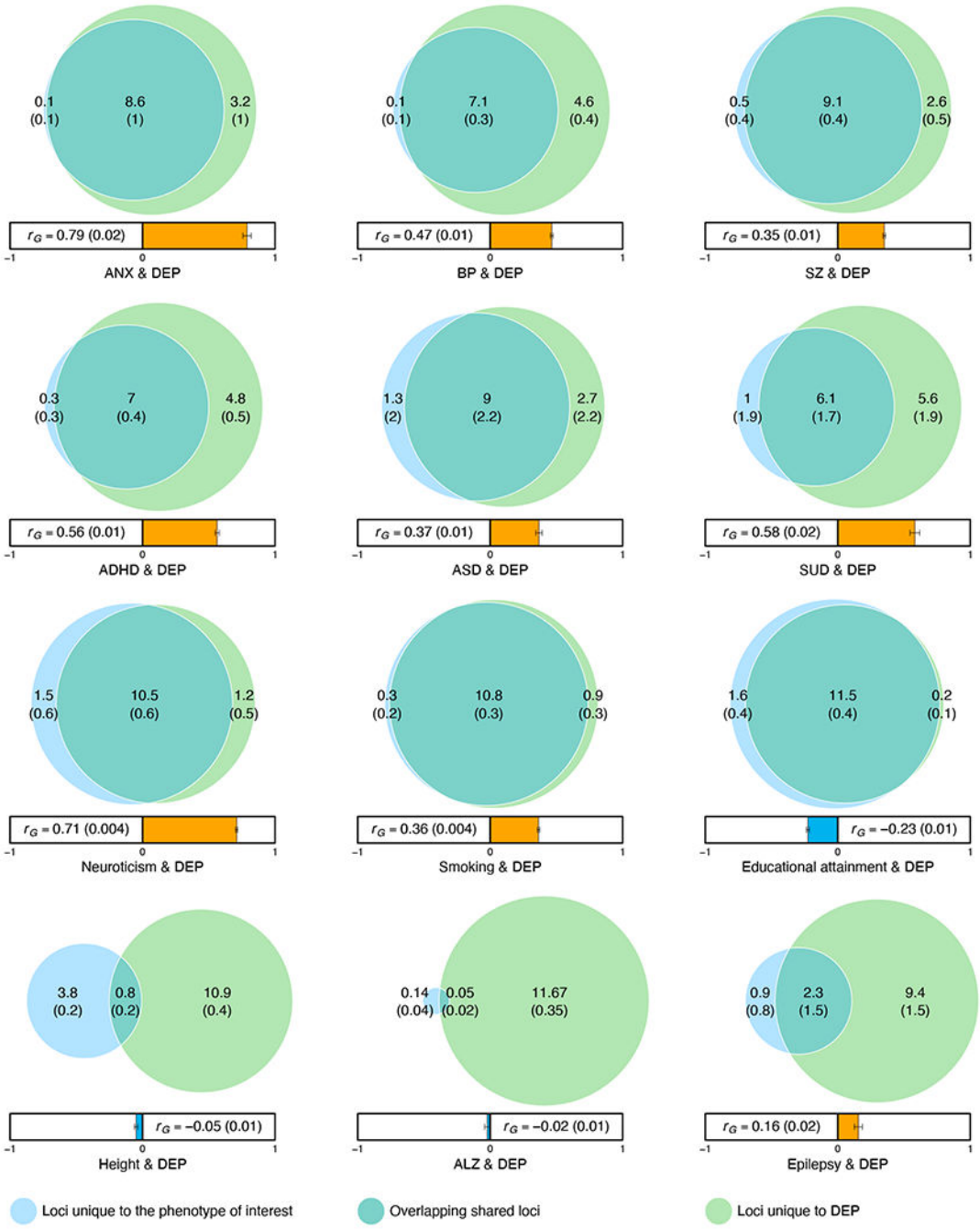


Figure 2:

Venn diagrams showing MiXeR results of the estimated number of variants shared between depression (N=1349887) and psychiatric disorders (anxiety (N= 361365), bipolar disorder (N= 405771), schizophrenia (N=153808), ADHD (N=225534), autism (N=46350, SUD (N=46568), neuroticism (N=380506)) and genetically correlated phenotypes (smoking initiation (N=632803), educational attainment (N=3037499)) and other phenotypes (height (N=1597374), Alzheimer's (N=788989), epilepsy (N=44889)) not expected to show high genetic correlation with depression. Circles represent loci unique to depression and unique

to the phenotype of interest and corresponding overlapping shared loci. The number of shared variants +/− SE are shown in thousands. The size of the circles reflects the polygenicity of each phenotype, with larger circles corresponding to greater polygenicity. Point estimates of genetic correlation (r_G) and 95% CI between depression and each phenotype is shown at the bottom with an accompanying scale (−1 to +1) (see also Supplementary Table S6A and Figure S9-1).

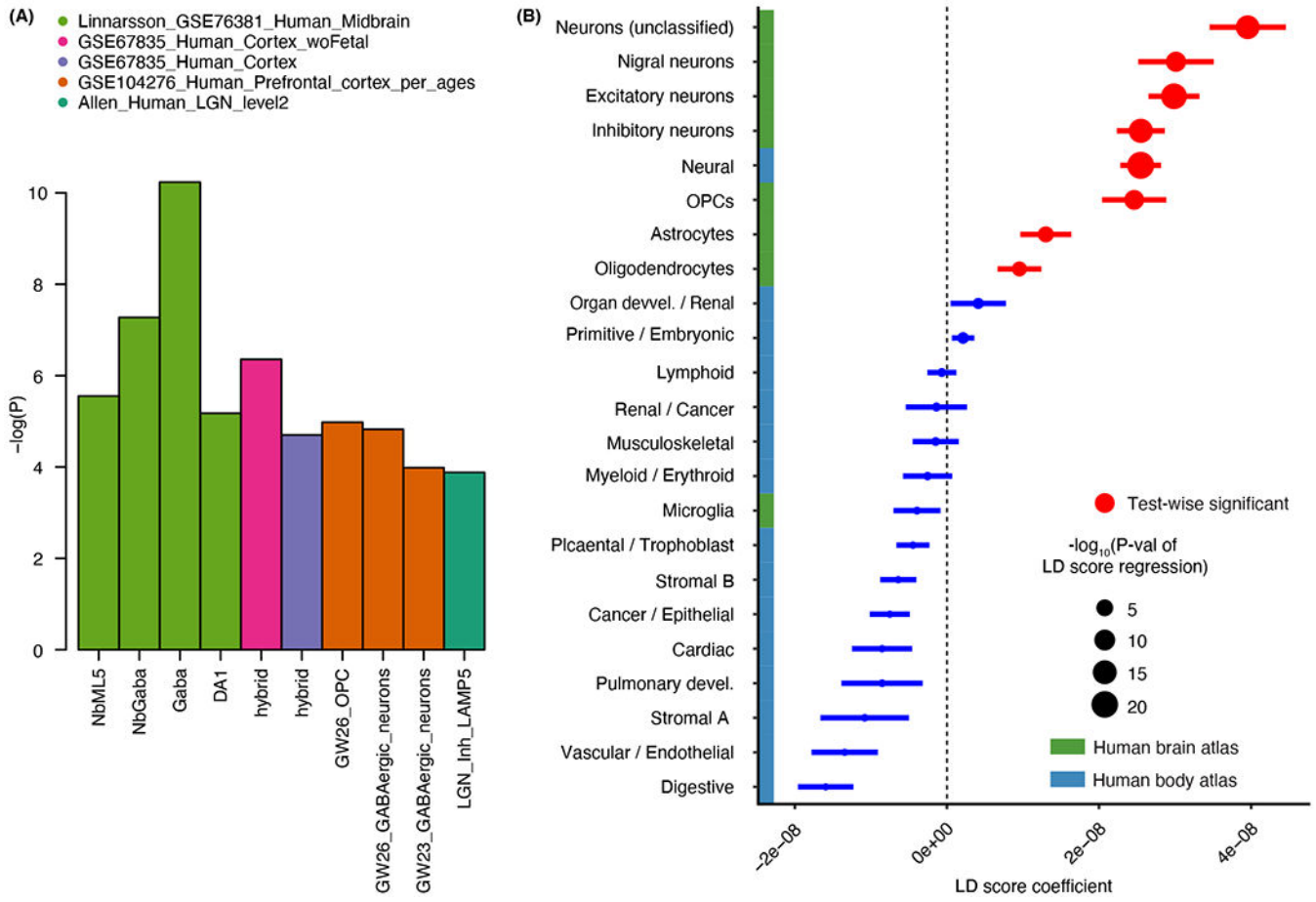


Figure 3:

(A) Significant cell types across datasets in MAGMA analysis implemented in FUMA. The y-axis shows $-\log_{10}$ of one-sided nominal P-values based on z statistics for cell-types significant across datasets (x-axis), i.e. after experiment wide Bonferroni correction. The primary depression GWAS (N_{cases}=371184 and N_{ctrls}=978703) was used as input for this analysis. **(B)** Enrichment of depression risk variants with cell-specific open chromatin regions. Dots represents the LD score coefficients and horizontal bars reflect standard error (x-axis) for various cell-types (y-axis). A positive LD score coefficient signifies enrichment in heritability. Dot size reflects two-sided P-value of LD score z statistics, and color code indicate test-wide significance (BH-corrected P-value<0.05). The primary depression GWAS (Ncases = 371184 and Nctrls = 978703) was used as input for this analysis.

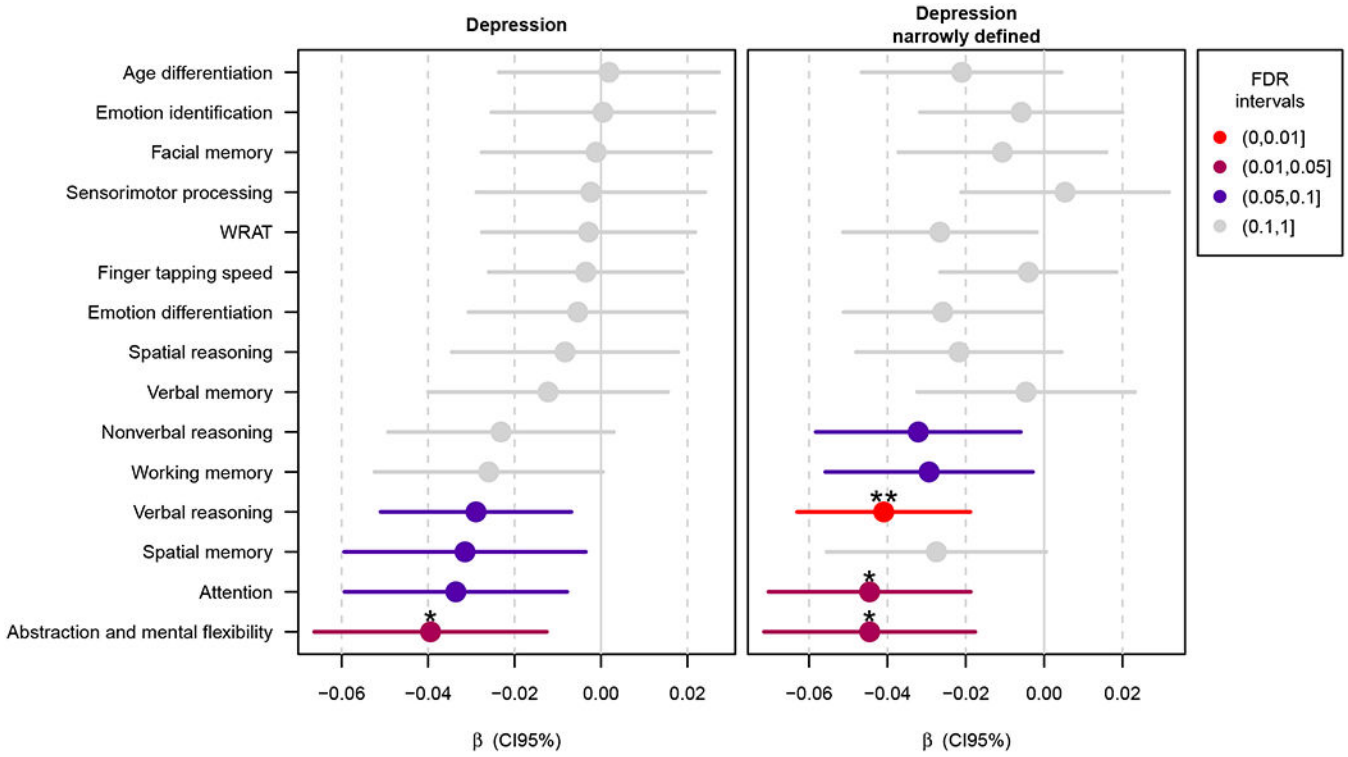


Figure 4:

Association of DEP-PRS (broadly and narrowly defined) with measures of cognitive abilities in the PNC cohort (N=4,973). Regression coefficients (β) and 95% confidence interval) from linear regression testing for the association of DEP-PRS with the 15 neurocognitive measures listed on the y-axis. Significance is based on two-sided FDR adjusted P-values from linear regression t statistics. Colors indicate FDR-adjusted (to account for 30 total tests) P-value intervals (See supplementary table S12 for details) and ** corresponds to FDR-adjusted P-value (q-value) <0.01, * corresponds to 0.01<=q-value<0.05. The primary DEP-PRS was significantly associated with Abstraction and mental flexibility (q-value=0.03). The narrow DEP-PRS was significantly associated with Verbal reasoning (q-value=0.009), Attention (q-value=0.011) and Abstraction and mental flexibility (q-value=0.012). The neurocognitive phenotypes included performance on the Computerized Neurocognitive Battery (CNB: age differentiation, emotion identification, facial memory, sensorimotor processing, finger tapping speed, emotion differentiation, spatial reasoning, verbal memory, nonverbal reasoning, working memory, verbal reasoning, spatial memory, attention, abstraction and mental flexibility)[68], as well as results from the Wide Range Achievement Test (WRAT-4)[69].

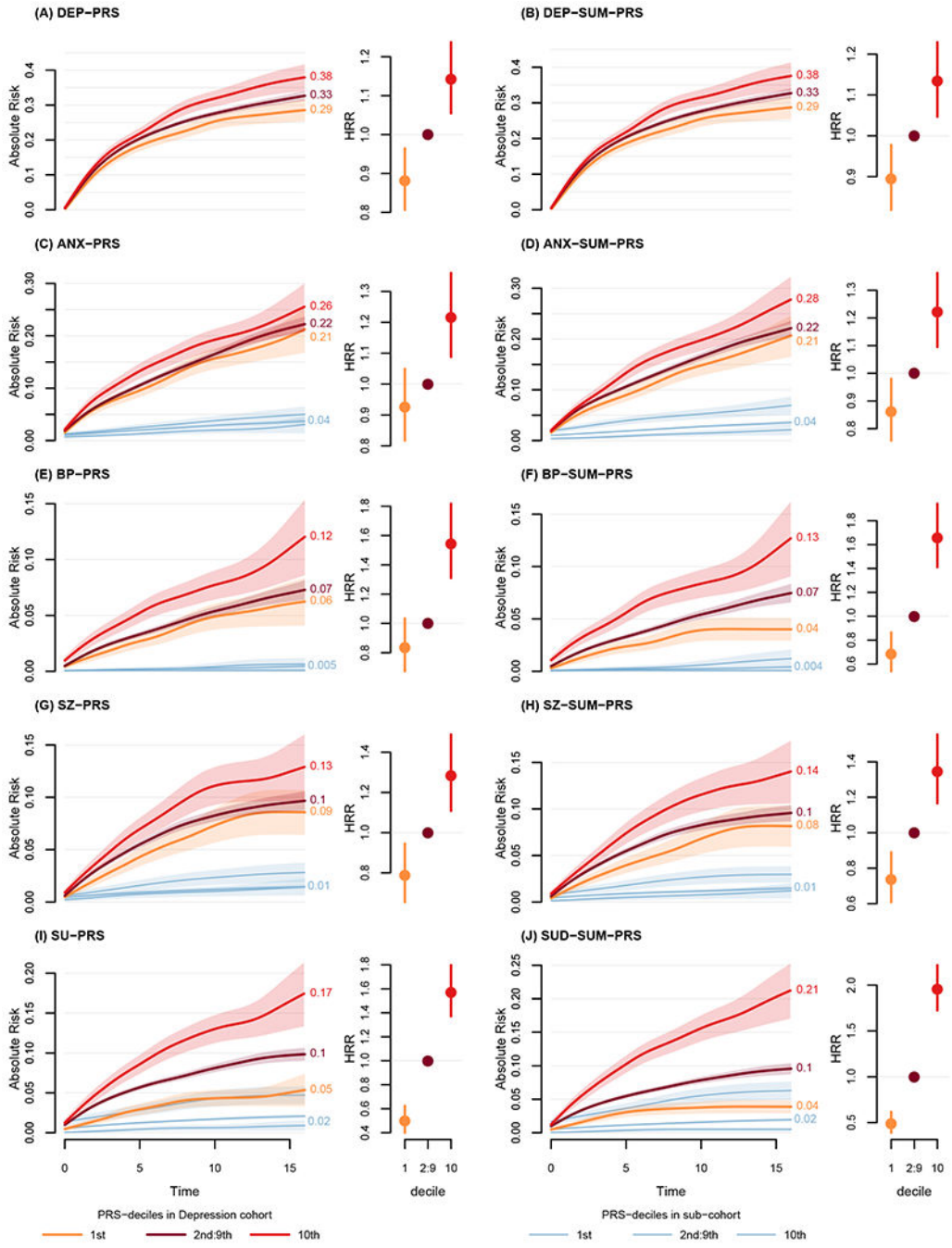


Figure 5:

Absolute risk (95% CI) over time since first depression episode and HRR (95% CI) of (A,B) a second episode of depression stratified by three groups of DEP-PRS and DEP-SUM-PRS deciles (N=29158); (C,D) developing anxiety stratified by ANX-PRS and ANX-SUM-PRS (N=25124); (E,F) transitioning into bipolar disorder stratified by BP-PRS and BP- SUM-PRS (N=30300); (G,H) developing schizophrenia stratified by SZ-PRS and SZ-SUM-PRS (N=28714); (I,J) developing substance-use-disorder stratified by SU-PRS and SUD- SUM-PRS (N=18856). SUM-PRSs were calculated by adding PRSs for multiple phenotypes

weighted by log(OR) aiming to optimize prediction. HRR (95% CI) for 1st, 2nd-to-9th and 10th decile are shown as large dots in different colors, using the middle (80% prediction interval) of the PRS as reference. Absolute risk (95% CI) of anxiety (C,D), bipolar disorder (E,F), schizophrenia (G,H) and SUD (I,J) is shown for the iPSYCH2015 random population (sub-cohort) in less bright colors.

Table 1:

Top 5 index SNPs for previous and top 15 index SNPs for novel depression GWAS loci. Results for the top 5 among previously identified loci and top 15 novel genome-wide significant index variants identified in the GWAS meta-analysis of 371,184 cases with depression and 978,703 controls. Locus number (in bold for novel loci) The location (chromosome (Chr) base position (BP)), alleles (A1/A2), frequency of A1-allele (f(A1)) in cases and controls, odds ratio (OR) of the effect with respect to A1, standard error (SE) and two-sided unadjusted association P-values (P) of z statistics from a inverse-variance weighted fixed effects model of the index variants are given. Gene names (in italic) of genes located within 50 kb from index variants are listed (See supplementary table S2 for details on all 303 index SNPs).

Locus	Chr	SNP	BP	A_1/A_2	$f(A_1)$		OR	SE	P	Nearby genes
					cases	ctrls				
8	1	rs7531118	72837239	T/C	0.473	0.461	0.97	0.0029	1.50E-27	*NEGR1*
9	1	rs10890020	73668836	A/G	0.508	0.51	0.97	0.0029	1.70E-21	*LINC01360*
38	2	rs1320138	144158287	T/C	0.428	0.435	1.02	0.0028	4.80E-09	*ARHGAP15*
45	2	rs6715105	198445601	T/C	0.326	0.341	1.02	0.003	1.30E-09	*ANKRD44, ANKRD44-IT1, SF3B1, COQ10B, HSPD1, SNORA105A, SNORA105B, HSPE1, HSPE1-MOB4, MOB4, RFTN2, MARS2, BOLL, PLCL1*
53	3	rs56029819	43478295	T/C	0.853	0.849	0.98	0.004	2.50E-09	*SNRK, SNRK-AS1, ANO10*
80	5	rs4262121	31078958	G/C	0.528	0.546	0.98	0.0029	4.30E-09	-
89	5	rs62379847	120109119	A/C	0.655	0.652	0.98	0.003	1.80E-09	*PRR16*
93	5	rs2964003	153216733	A/G	0.829	0.82	1.02	0.0038	8.10E-10	*GRIA1, LINC01861*
94	5	rs4596421	161269895	C/T	0.708	0.693	0.98	0.0031	2.40E-09	*GABRA*
98	6	rs35883476	28368508	G/C	0.91	0.918	1.05	0.0052	1.30E-21	*HIST1 histone cluster, BTN3A2, BTN2A2, BTN3A1, BTN2A3P,BTN3A3, BTN2A1, LOC285819,BTN1A1, HCG11, HMGN4,LOC105374988, ABT1, ZNF322, GUSBP2, LINC00240, LOC100270746, MIR3143, PRSS16, POM121L2, VN1R10P,ZNF204P, ZNF391, ZNF184, LINC01012, LOC100131289, OR2B2, OR2B6, ZNF165, ZSCAN12P1, ZSCAN16-AS1, ZSCAN16,ZKSCAN8, ZNF192P1, TOB2P1,ZSCAN9, ZKSCAN4, NKAPL,ZSCAN26, PGBD1, ZSCAN31,ZKSCAN3, ZSCAN12, ZSCAN23*
99	6	rs10947690	37631768	A/G	0.751	0.756	0.98	0.0033	6.30E-09	*MDGA1*
114	7	rs957360	3660918	C/G	0.714	0.702	1.02	0.0031	1.30E-09	*SDK1*
124	7	rs1986692	133743393	A/G	0.612	0.61	1.02	0.0029	2.80E-09	*EXOC4*
156	10	rs1909696	77582203	G/T	0.337	0.328	0.98	0.003	3.30E-09	*LRMDA, LOC105378367*
158	10	rs1021363	106610839	A/G	0.349	0.335	1.03	0.0029	5.30E-25	*SORCS3*
191	13	rs9561331	94017476	G/A	0.869	0.873	0.98	0.0042	3.70E-09	*GPC6*
199	14	rs7141014	98667928	T/C	0.796	0.806	0.98	0.0035	3.30E-09	-
210	15	rs4886915	78075023	A/G	0.421	0.423	0.98	0.0028	6.10E-10	*LINGO1*
225	17	rs60856912	65892343	G/T	0.825	0.817	0.97	0.0038	1.80E-11	*BPTF, C17orf58, KPNA2*
233	18	rs12967143	53099012	G/C	0.301	0.292	1.03	0.0031	2.00E-21	*TCF4, TCF4-AS1, MIR4529*