

1 ***Identification of 64 new risk loci for major depression, refinement of the genetic***
2 ***architecture and risk prediction of recurrence and comorbidities***

3

4 Thomas D. Als^{1,2,3}, Mitja Kurki⁴, Jakob Grove^{1,2,3,5}, Georgios Voloudakis^{6,7,8,9,10,11}, Karen
5 Therrien^{6,7,8,9,10,11,12}, Elisa Tasanko¹³, Trine Tollerup Nielsen^{1,2,3}, Joonas Naamanka¹³, Kumar
6 Veerapen^{14,15,16}, Daniel Levey^{17,18}, Jaroslav Bendl^{6,7,8,9,10}, Jonas Bybjerg-Grauholm^{2,19}, Biao
7 Zheng^{6,7,8,9,10}, Ditte Demontis^{1,2,3}, Anders Rosengren^{2,20}, Georgios Athanasiadis^{2,20}, Marie Bækved-
8 Hansen^{2,19}, Per Qvist^{1,2,3}, Bragi Walters²¹, Thorgeir Thorgeirsson²¹, Hreinn Stefánsson²¹, Katherine L
9 Musliner^{2,22,23,24}, Veera Manikandan^{1,2,3}, Leila Farajzadeh^{1,2,3}, Janne Thirstrup^{1,2,3}, Bjarni J.
10 Vilhjálmsson^{2,5,25}, John J. McGrath^{25,26,27}, Manuel Mattheisen², Sandra Meier², iPSYCH-Broad
11 Consortium, Esben Agerbo^{2,22,23}, Kári Stefánsson²¹, Merete Nordentoft^{2,28}, Thomas Werge^{2,20}, David
12 M. Hougaard^{2,19}, Preben B. Mortensen^{2,22,23}, Murray Stein^{29,30}, Joel Gelernter^{17,18}, Iris Hovatta¹³,
13 Panos Roussos^{6,7,8,9,10,11,31}, Mark J. Daly^{4,14,15,16}, Ole Mors^{2,32}, Aarno Palotie⁴, and Anders D.
14 Børglum^{1,2,3}

15

16 ¹ Department of Biomedicine, Aarhus University, Aarhus, Denmark

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

18 ³ Center for Genomics and Personalized Medicine, Aarhus, Denmark

19 ⁴ Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

20 ⁵ Bioinformatics Research Centre, Aarhus University, 8000 Aarhus C, Denmark

21 ⁶ Center for Disease Neurogenetics, Icahn School of Medicine at Mount Sinai, NY, USA

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

23 ⁸ Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New

24 York, NY, USA

25 ⁹ Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount

26 Sinai, NY, USA

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

28 ¹¹ Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA

29 Medical Center, Bronx, NY, USA

30 ¹² Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York,

31 NY, USA

32 ¹³ Department of Psychology and Logopedics, SleepWell Research Program, University of Helsinki,

33 Finland

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

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

38 ¹⁶ Department of Medicine, Harvard Medical School, Boston, MA, USA

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

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

42 ¹⁹ Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut,
43 Copenhagen, Denmark

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

46 ²¹ deCODE genetics / Amgen, Reykjavik, Iceland

47 ²² NCCR - National Centre for Register-Based Research, Business and Social Sciences, Aarhus
48 University, Aarhus, Denmark

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

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

51 ²⁵ NCCR - National Centre for Register-Based Research, Aarhus University, 8210 Aarhus V, Denmark

1 ²⁶ Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Brisbane, QLD
2 4076, Australia

3 ²⁷ Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia

4 ²⁸ Mental Health Centre Copenhagen, Capital Region of Denmark, Copenhagen University Hospital,
5 Copenhagen, Denmark

6 ²⁹ Psychiatry Service, VA San Diego Healthcare System, San Diego, CA, USA

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

9 ³¹ Center for Dementia Research, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY
10 10962, USA

11 ³² Psychosis Research Unit, Aarhus University Hospital-Psychiatry, Denmark

12 Corresponding authors: Thomas D Als tda@biomed.au.dk and Anders D Børghlum

13 anders@biomed.au.dk

14

15 **Abstract**

16 Major depression (MD) is a common mental disorder and a leading cause of disability

17 worldwide. We conducted a GWAS meta-analysis of more than 1.3 million individuals,

18 including 371,184 with MD, identifying 243 risk loci. Sixty-four loci are novel, including

19 glutamate and GABA receptors that are targets for antidepressant drugs. Several biological

20 pathways and components were enriched for genetic MD risk, implicating neuronal

21 development and function. Intersection with functional genomics data prioritized likely

22 causal genes and revealed novel enrichment of prenatal GABAergic neurons, astrocytes and

23 oligodendrocyte lineages.

24 We found MD to be highly polygenic, with around 11,700 variants explaining 90% of the

25 SNP heritability. Bivariate Gaussian mixture modeling estimated that > 97% of risk variants

26 for other psychiatric disorders (anxiety, schizophrenia, bipolar disorder and ADHD) are

27 influencing MD risk when both concordant and discordant variants are considered, and nearly

28 all MD risk variants influence educational attainment. Additionally, we demonstrated that

29 MD genetic risk is associated with impaired complex cognition, including verbal reasoning,

30 attention, abstraction and mental flexibility.

1 Analyzing Danish nation-wide longitudinal data, we dissected the genetic and clinical
2 heterogeneity, revealing distinct polygenic architectures across case subgroups of MD
3 recurrency and psychiatric comorbidity and demonstrating two- to six-fold increases in
4 absolute risks for developing comorbid psychiatric disorders among MD cases with the
5 highest versus the lowest polygenic burden.

6 The results deepen the understanding of the biology underlying MD and its progression and
7 inform precision medicine approaches in MD.

8

9 **Introduction**

10 Major depression (MD) is a genetic and phenotypic complex disorder with a lifetime
11 prevalence of 15-20%¹⁻³. It is often recurrent and accompanied by considerable morbidity and
12 co-morbidity, excess mortality, increased risk of suicide and substantial costs worldwide⁴⁻⁸.

13 Individuals diagnosed with MD have an increased risk of developing practically all other
14 types of mental disorders, particularly anxiety (ANX), bipolar disorder (BD), schizophrenia
15 (SZ) and substance use disorder (SUD)^{9,10}.

16 Heritability estimates based on twin studies ($h^2 = 0.37$) have indicated that familial
17 aggregation of MD is influenced by additive genetic effects¹¹, and several recent studies have
18 documented a considerable genetic overlap between MD and multiple psychiatric as well as
19 somatic disorders and traits¹²⁻¹⁷. Despite the substantial heritability of MD and other mental
20 disorders, the potential for translating genetic insights into precision psychiatry has yet to be
21 fulfilled, including demonstrating clinical utility of polygenic risk scores (PRS)¹⁸⁻²⁰.

22

23 Major advancement in understanding the genetic architecture of MD has only recently been
24 achieved, primarily via genome-wide association studies (GWAS) led by the Psychiatric
25 Genomics Consortium (PGC). Most recently, GWAS results from the PGC¹², UK Biobank

1 (UKB)^{14,21}, FinnGen and 23andMe, Inc.²² were combined with data from the Million Veteran
2 Program (MVP) in a large meta-analysis²³, identifying 178 risk loci. The identified loci
3 explain a small fraction of the overall heritability of MD^{12,21,23} and even larger GWASs are
4 needed to further elucidate genetic factors contributing to the risk of developing MD and
5 advance genetically informed patient stratification and outcome prediction towards clinical
6 utility.
7 Here, we present the largest MD GWAS to date, including expanded iPSYCH²⁴ and FinnGen
8 cohorts, meta-analyzed with PGC, UKB, 23andMe and MVP data²³, revealing numerous
9 novel risk loci and characterizing functional implications of associated variants by
10 intersecting with functional genomics data. We refine the genetic architecture of MD and
11 case subgroups, and demonstrate the impact of MD genetic risk on domains of cognitive
12 performance. Leveraging nation-wide longitudinal health data on the Danish iPSYCH cohort,
13 we dissect the genetic architecture of single-episode and recurrent MD as well as cases who
14 have developed ANX, BD, SZ and SUD. Furthermore, to inform precision psychiatry
15 approaches, we calculate time-dependent absolute risks and hazard rate ratios for developing
16 recurrent depression, ANX, BP, SZ and SUD depending on different polygenic burdens of
17 MD cases.

18

19 **Results**

20 *Genome-wide association*

21 We analyzed data from the large population-based case-cohort of iPSYCH^{24,25}, which include
22 genotypes from all individuals born in Denmark between 1981 and 2008 who have received
23 treatment for MD in hospitals and outpatient clinics (ICD-10 codes F32-F33 in the Danish
24 Psychiatric Central Research Register²⁶). Compared to the latest GWAS of MD²³, which
25 included samples from the initial iPSYCH2012 cohort^{12,25}, we added 11,710 cases and 18,410

1 controls from the expanded iPSYCH2015 cohort^{24,25}, summarizing to a total of 30,618 cases
2 and 38,200 controls after relatedness pruning and removal of ancestry outliers. In addition,
3 we included an updated dataset consisting of 28,098 cases and 228,817 controls from the
4 FinnGen study. When combining these with data from previously published samples from the
5 PGC, UKB, 23andMe and MVP, the number of samples added up to a total of 371,184 MD
6 cases and 978,703 controls (Supplementary Table S1).

7 We performed a variance weighted fixed effects meta-analysis using METAL²⁷, testing the
8 effects of 6,037,120 SNPs common across all the data sets. This revealed a total of 303
9 genome-wide significant LD-independent ($r^2 < 0.1$) lead variants located in 243 distinct loci.
10 A conditional association analysis using GCTA-COJO^{28,29} retained 251 independent SNPs in
11 the 243 loci. Manhattan plots are shown in Figure 1, regional plots are provided in
12 Supplementary Figure S2 and details on lead variants are provided in Supplementary Table
13 S2A, while GCTA-COJO independent variants are listed in Supplementary Table S2B. No
14 statistically significant heterogeneity was observed between the datasets (Supplementary
15 Figure S3). Sixty-four (64) of the 243 loci are novel, i.e., not overlapping with the two most
16 recent MD meta-analyses^{14,23} (Supplementary Table S2A). The three most significant loci
17 were located near *NEGR1*, in *SORCS3* and in the *HIST1* histone cluster, respectively. Among
18 the novel loci, the three strongest associations were in *BPTF*, *LINGO1* and *GRIAI* (Table 1).
19 All three genes have been associated with monogenic forms of neurodevelopmental
20 disorders³⁰⁻³² and *GRIAI*, encoding glutamate ionotropic receptor AMPA type subunit 1
21 (GluA1), is the first genome-wide significant locus implicating an AMPA receptor subunit in
22 MD. We also note that the seventh-strongest novel locus is located in *GABRA1*, which is the
23 first time a GABA receptor locus has been identified in GWAS of MD. Both AMPA and
24 GABA receptors are targets for antidepressants^{33,34}.

25

1 *Genetic correlations and heritability*

2 Analyzing the GWAS meta-analysis by LD score regression (LDsc)³⁵ produced a genomic
3 inflation factor (λ_{GC}) estimate of 1.89 with an intercept of 1.06 (SE = 0.01) and an
4 attenuation ratio of 0.047 (SE = 0.012), indicating that 95% of the observed inflation of the
5 test statistics (Supplementary Figure S1) is due to polygenic signal rather than population
6 structure. The individual GWASs (iPSYCH2015, Howard et al. 2019¹⁴, Wray et al. 2018¹²,
7 Levey et al. 2020²³ and FinnGen) showed significant pairwise genetic correlations, ranging
8 from $r_G = 0.77$ to $r_G = 0.95$ (Supplementary Table S4 and Figure S6), thus corroborating
9 that the GWAS results are combined in a meta-analysis. We note that the SNP-heritability
10 estimate³⁵ for the iPSYCH cohort ($h_{SNP}^2 = 0.167$, SE = 0.014, prevalence = 0.2) was
11 significantly higher (around double) compared to the other cohorts (Supplementary Table S5
12 and Figure S7). This may reflect that the iPSYCH sample is more homogeneous and includes
13 relatively severe cases who have been treated in hospitals.

14

15 We investigated genetic correlations with other phenotypes available at LD Hub³⁶ and in-
16 house, including published GWASs (N = 258) and GWAS results for UKB traits (N = 597).
17 MD was significantly correlated ($P < 2 \times 10^{-4}$) with 364 phenotypes representing several
18 domains and overall confirming previous observations^{12-15,37-42} (Supplementary Table S6A
19 and Figure S8A). Among psychiatric disorders, MD showed significant correlation with e.g.,
20 ADHD ($r_G = 0.56$, SE = 0.022, $P = 1 \times 10^{-135}$), autism spectrum disorder (ASD) ($r_G = 0.35$,
21 SE = 0.033, $P = 6.5 \times 10^{-24}$), BP ($r_G = 0.31$, SE = 0.033, $P = 3 \times 10^{-18}$), SZ ($r_G = 0.33$, SE =
22 0.021, $P = 4 \times 10^{-53}$), ANX ($r_G = 0.79$, SE = 0.017, $P = 3.2 \times 10^{-193}$), alcohol dependence (r_G
23 = 0.65, SE = 0.097, $P = 4.3 \times 10^{-9}$) and cannabis use disorder ($r_G = 0.44$, SE = 0.036, $P = 2.2$
24 $\times 10^{-31}$). In UKB data, the three strongest genetic correlations were seen for “Seen doctor

1 (GP) for nerves _ anxiety _ tension or depression”, “Mood swings” and “Miserableness” ($r_G =$
2 0.96, 0.71, 0.71, respectively; Supplementary Table S6B and Figure S8B).

3
4 To dissect the observed genetic overlap further, we used uni- and bivariate gaussian mixture
5 modeling, as implemented in MiXeR⁴³, to quantify the actual number of variants that (i)
6 explain 90% of the SNP heritability of MD, and (ii) overlap between MD and MD-correlated
7 phenotypes. Approximately 11,700 (SE = 310) common variants were estimated to confer
8 liability to MD, suggesting that MD is the most polygenic of the major psychiatric disorders
9 evaluated, showing number of risk variants ranging between 6-10,000^{43,44} (Figure 3,
10 Supplementary Table S7). MiXeR considers all variants irrespective of the direction of
11 correlation (i.e. both variants with same and opposite direction of effects, hereafter
12 collectively referred to as “influencing” variants). Strikingly, the vast majority of variants
13 conferring risk to the other psychiatric disorders investigated were found to influence MD
14 (range 85-98%; Figure 3; Supplementary Table S7), most pronounced for ANX, SZ, BP and
15 ADHD with 97-99%⁴⁴ of their risk variants also influencing MD. The other investigated traits
16 showed also substantial overlap with MD (range 85-98%). Notably, the fraction of
17 concordant variants within the shared part varied considerably; lowest for educational
18 attainment (42%) and highest for SUD (88%) and ANX (90%).

19
20 *Association with cognitive performance*

21 Educational attainment was among the most prominent negatively correlated traits. To further
22 evaluate the impact of MD genetic risk on cognition, we analyzed the association of MD
23 polygenic scores (MD-PRS) with 15 cognitive measures in the Philadelphia
24 Neurodevelopmental Cohort (PNC, N = 4,973)^{45,46}. Cognitive performance was measured by
25 the Computerized Neurocognitive Battery⁴⁷, including 14 tests in 5 domains: executive-

1 control, episodic memory, complex cognitive processing, social cognition, and sensorimotor
2 speed. In addition, the Wide Range Achievement Test (WRAT-4)⁴⁸ was used as a proxy
3 measure for overall IQ⁴⁶. MD-PRS was negatively associated with attention ($\beta = -0.036$, SE
4 = 0.012, $P = 0.032$) and abstraction and mental flexibility ($\beta = -0.036$, SE = 0.013, $P = 0.032$)
5 in the executive-control domain, and verbal reasoning ($\beta = -0.029$, SE = 0.011, $P = 0.032$) in
6 the reasoning domain, while e.g. the WRAT-4 test did not show association (Figure 2,
7 Supplementary Table S8). The results demonstrate, for the first time, that genetic MD risk is
8 associated with attenuated functioning in specific cognitive domains. This is consistent with
9 observations that individuals suffering from MD show lower performance in cognitive
10 domains such as executive function, memory, language and attention⁴⁹⁻⁵¹.

11

12 *Gene-wise and pathway analysis*

13 A genome-wide gene-based association study conducted in Multi-Marker Analysis of
14 GenoMic Annotation (MAGMA)⁵², mapped the GWAS SNPs to 17,840 protein coding genes
15 and revealed 411 genes significantly associated with MD after Bonferroni correction for the
16 number of genes tested ($P < 2.8 \times 10^{-6}$). A total of 314 significant genes were located in 141
17 of the 242 identified GWAS loci, while the remaining 97 significant genes were located
18 outside these loci (Supplementary Table S9D). The most significant gene within each of the
19 141 genomic risk loci are labeled in Figure 1B.

20 To investigate enrichment of biological pathways we analyzed 8,664 gene-sets derived from
21 GO Biological Process ($N = 7,658$) and GO Cellular Components ($N = 1,006$) ontology in the
22 MSigDB database, identifying 479 significant gene-sets after correction for multiple testing
23 ($FDR < 0.05$, Supplementary Table S10B), including several gene-sets that have not
24 previously shown significant enrichment^{14,23}. The majority of the top-ranking sets related to
25 neuronal development and function including the five most significant gene-sets ($P_{adj} <$

1 10^{-14}): GO_SYNAPSE, GO_NEURON_PART, GO_SYNAPSE_ORGANIZATION,
2 GO_SYNAPSE_ASSEMBLY and GO_NEURON_DIFFERENTIATION.

3

4 *Transcriptome-wide association and GWAS-eQTL prioritisation*

5 To identify and prioritize putative causal genes, we performed a transcriptome-wide
6 association study (TWAS), imputing the genetically regulated gene expression using
7 EpiXcan⁵³ and models trained on expression data from the PsychENCODE Consortium^{54,55}
8 for genes and isoforms detected in the dorsolateral prefrontal cortex (DLPFC). Among
9 34,646 transcripts (genes and isoforms) tested, we identified 2,541 transcripts at FDR < 0.05
10 and, after Bonferroni correction of all transcripts tested, 324 transcripts from 201 genes
11 showing significant differential imputed gene expression in DLPFC between MD cases and
12 controls (Supplementary Table S11). The Bonferroni significant transcripts were located in
13 88 independent⁵⁶ regions. The top gene/isoform is labeled in Figure 1A and regional
14 TWAS/GWAS Miami-plots for each of the 88 regions are shown in Figure S9. In 38 of the
15 88 regions, the top transcript was > 100 times more significant than the second-most
16 associated gene/isoform in the region (Supplementary Table S11B), appointing those as
17 plausible causal candidates.

18 To further prioritize likely causal genes and variants, we performed co-localization analyses
19 integrating fine-mapped GWAS results, using the CAUSALdb pipeline (Methods), and eQTL
20 data from a meta-analysis of three brain datasets⁵⁷ applying a fixed-effect model. First, we
21 adopted the Coloc method^{58,59}, which revealed 13 genes with strong evidence for both
22 GWAS-association, eQTL-association and co-localization (i.e., with a posterior probability of
23 PPH4 > 0.8; Supplementary Table S12). The three top-ranked genes were: *FURIN*, *NEGR1*
24 and *CKS2*. Secondly, we conducted the eQTL and GWAS CAusal Variants Identification in
25 Associated Regions (eCAVIAR) approach⁶⁰, in which both eQTL and GWAS were fine-

1 mapped and the product of posterior probability (CLPP) was calculated, prioritizing variants
2 with at least a single variant with $CLPP \geq 0.01$. The eCAVIAR approach revealed five
3 prioritized variants located in three genes: *FURIN*, *GPR27* and *TCTA* (Supplementary Table
4 S13 and Figure S10).

5

6 *Tissue and cell type enrichment*

7 We next tested whether the GWAS results were enriched with respect to the transcriptomic
8 profiles of human tissues. At the specific tissue level, we found significant enrichment
9 exclusively in brain tissues, including all the brain tissues analyzed (Supplementary Figure
10 S11). Cell-type enrichment analyses revealed experiment-wide significant association (across
11 all 13 datasets tested) of primarily neuronal cell-types, including dopaminergic and
12 GABAergic neurons (Supplementary Figure S12 and S13). Interestingly, both GABAergic
13 neurons and oligodendrocyte progenitor cells of the human prefrontal cortex were enriched
14 already at prenatal stages.

15 To further evaluate cell-type enrichment, we intersected the GWAS results with two recent
16 epigenomic maps of cell-specific open chromatin^{61,62} using an LD score partitioned heritability
17 approach⁶³. Again, we observed a clear contrast between the enrichments in the brain and non-
18 brain tissue (Supplementary Figure S14). Consistent with our FUMA-based results and prior
19 reports^{12,14}, the strongest associations were measured for neuronal cell types, phenotypically
20 manifested by severe synaptic loss and deficits in functional connectivity^{64,65}. Conversely, to
21 our knowledge, the reported association of MD with astrocytes and oligodendrocyte lineages
22 have not yet been described by genetic data, albeit having support in behavioral and
23 postmortem studies⁶⁶⁻⁶⁸.

24

25 *Polygenic architecture and co-morbidity rates for single-episode and recurrent depression*

1 To dissect the polygenic architecture of single-episode and recurrent depression, we used
2 dates of diagnosis in the Danish Psychiatric Central Research Register to group the 30,618
3 iPSYCH MD cases into 24,101 cases with single-episode and 6,517 with recurrent depression
4 (see Methods and Supplementary Table S14 for details). We conducted three GWASs: 1)
5 single-episode vs controls, 2) recurrent depression vs. controls and 3) recurrent vs. single-
6 episode depression (excluding all controls). SNP-heritability estimates were similar for
7 recurrent and single-episode depression and, for the case-only analysis, not significantly
8 different from zero (Supplementary Table S5 and Figure S7). Likewise, the genetic
9 correlations with other phenotypes showed similar patterns for single-episode and recurrent
10 depression (Supplementary Figure S15 and Table S15).

11 We next investigated the polygenic load for MD in the recurrent and single-episode case
12 groups, using a multivariate polygenic risk score (mvPRS) approach¹⁵ and the MD GWAS
13 meta-analysis^{14,23} without iPSYCH samples for training. This analysis showed significant
14 association of single-episode ($\beta = 0.36, SE = 0.0081, P < 10^{-7}$) and recurrent ($\beta =$
15 $0.43, SE = 0.014, P < 10^{-7}$) subgroups with PRS for MD (MD-PRS), but with a
16 significantly larger effect size for recurrent than for single-episode depression ($P = 4.8 \times 10^{-6}$,
17 Supplementary Table S16A and Figure S16A). This increased polygenic load among
18 recurrent cases reinforces previous observations^{11,12,69,70}.

19 To further dissect the genetic architecture of the two subgroups we investigated the PRS load
20 for psychiatric disorders and traits showing strong genetic correlation with MD (Figure
21 S16B-I, Table S16A) including ANX, BD, SZ, ASD, ADHD, SUD, substance use (SU) and
22 neuroticism. The results showed an overall pattern of increased PRS among recurrent cases
23 compared to single-episode cases ($P = 0.00075$), primarily driven by significantly different
24 burdens of MD, ANX, BP and neuroticism PRS (Figure S16 and Table S16A).

25

1 To complement these analyses, we compared the rates of comorbid psychiatric disorders
2 between single-episode and recurrent cases, using data from the Danish Psychiatric Central
3 Research Register, and found highly significant increase in comorbidity among recurrent
4 cases (Supplementary Table S17). This is consistent with well-established correlations
5 between MD symptom severity, recurrence, chronicity and comorbidities (e.g.^{10,71}) and, to
6 our knowledge, it is the first quantification of differences in psychiatric comorbidity rates
7 between single-episode and recurrent case groups. The most compelling differences were
8 observed for ANX, BP and SZ.
9 Summarizing, compared to single-episode cases, recurrent cases showed significantly
10 increased polygenic load of particularly MD-, BP- and neuroticism-PRS as well as increased
11 rates of comorbid psychiatric disorders.

12

13 *Hazard rate ratio and absolute risk of recurrent depression*

14 To investigate whether PRS can prospectively predict recurrence among individuals with
15 first-onset depression, we performed Cox regression analyses estimating hazard rate ratio
16 (HRR) and absolute risk of developing a second episode of depression over time among
17 individuals having a first diagnosis of MD. This was done for groups of increasing polygenic
18 load, in PRS deciles. We found that the HRR of developing a second-episode generally
19 increased with increasing MD-PRS, most significantly for the 10th MD-PRS decile compared
20 to the first decile ($HRR_{10} = 1.33, SE = 0.06, P = 2.5 \times 10^{-6}$; Supplementary Figure S18A
21 and Table S18). Similarly, the absolute risk of a second diagnosis of depression increased
22 with time since first diagnosis for all MD-PRS deciles, with the trajectories for the 10th decile
23 reaching an absolute risk of 0.35 compared to 0.24 for the lowest decile (Supplementary
24 Figure S18A).

1 Prompted by the results of the mvPRS analysis, we performed the same analysis for PRSs
2 from other psychiatric disorders, but without yielding significant results (Figure S18, Table
3 S18). The HRR for developing single-episode and recurrent MD in the general population
4 can be found in Supplementary Figure S17.

5

6 *Dissecting the polygenic architecture of psychiatric comorbidity among MD cases*

7 Individuals diagnosed with MD are at an increased risk of developing other mental disorders
8 as documented in Supplementary Table S17 and a large body of studies (e.g.^{9,10,72}), but little
9 is known about the genetic constitution of those that develop these conditions. We examined
10 the polygenic architecture of MD cases who have developed ANX, BP, SZ and SUD, using
11 mvPRS analysis and PRSs from MD and eight psychiatric phenotypes genetically correlated
12 with MD (ANX, BD, SZ, ASD, ADHD, SUD, SU and neuroticism; Supplementary Table
13 S6A and Figure S8). This analysis showed that the subgroup of MD cases who had
14 transitioned to BP, had a significantly increased PRS for three out of the eight psychiatric
15 phenotypes tested compared to MD cases without a later BP diagnosis, including BP-PRS (P
16 $= 5.2 \times 10^{-17}$), SZ-PRS ($P = 3.2 \times 10^{-12}$) and MD-PRS ($P = 8.3 \times 10^{-5}$) (Figure S19 (B), Table
17 S19B). MD cases with a co-diagnosis of SZ showed increased PRS load for all eight
18 psychiatric phenotypes except ASD (Figure S19 (C), Table S19C), most significantly for SZ-
19 PRS ($P = 2.9 \times 10^{-14}$), BP-PRS ($P = 5.5 \times 10^{-8}$), substance use (SU) PRS ($P = 2.4 \times 10^{-5}$) and
20 MD-PRS ($P = 2.5 \times 10^{-5}$). The most compelling results were revealed for SUD comorbidity,
21 showing highly increased PRS loads for most of the psychiatric phenotypes (Figure S19 (D),
22 Table S19D), most significantly for SU-PRS ($P = 2.3 \times 10^{-84}$), ADHD-PRS ($P = 1.2 \times 10^{-31}$)
23 and SUD-PRS ($P = 2.4 \times 10^{-30}$).

24 Overall, we demonstrated that all three comorbid subgroups (MD-BP, MD-SZ, MD-SUD)
25 have increased polygenic burdens of common risk variants for several psychiatric phenotypes

1 compared to the non-comorbid MD (MD-only) case groups (overall mvPRS p-values: P_{MD-}
2 $ANX = 1.2 \times 10^{-21}$, $P_{MD-BP} = 3.7 \times 10^{-16}$, $P_{MD-SZ} = 3.2 \times 10^{-14}$, $P_{MD-SUD} = 1 \times 10^{-93}$), revealing an
3 overall PRS pattern that distinguishes the comorbid case groups from their non-comorbid
4 counterparts. We also note that all three comorbid subgroups showed an increased load of
5 MD-PRS compared to the MD-only case groups.

7 *Hazard rate ratio and absolute risk of comorbidity among MD cases*

8 To assess the use of PRS in prediction of developing comorbid disorders, we performed Cox
9 regression analysis on MD cases with and without additional diagnosis of ANX, BP, SZ and
10 SUD, using the same eight PRSs as in the above-mentioned mvPRS analysis. In addition, we
11 included four aggregate scores combining all PRSs into a single score weighted by their
12 association with ANX, BP, SZ and SUD respectively (see Methods).

13 We found that the HRR for developing BP increased with increasing polygenic loads
14 (Supplementary Figure S21 and Table S20B), most significantly for BP-PRS ($HRR_{10} = 1.84$,
15 $SE = 0.12$, $P = 9.3 \times 10^{-7}$) and the aggregate score (BP-SUM-PRS; $HRR_{10} = 2.32$, $SE = 0.14$,
16 $P = 7.5 \times 10^{-10}$) with the absolute risk trajectory reaching 0.11 for the 10th BP-SUM-PRS
17 decile compared to 0.04 for the 1st decile as shown in Figure 4.

18 A similar pattern was observed for MD cases developing SZ (Supplementary Figure S22,
19 Table S20C) where the highest HRR was seen for the aggregate SZ-SUM-PRS ($HRR_{10} =$
20 1.82 , $SE = 0.11$, $P = 1 \times 10^{-7}$), with the absolute risk curve for the 10th decile reaching 0.14,
21 well separated from the 1st decile trajectory approaching 0.067 (Figure 4).

22
23 The most striking results were obtained for SUD showing substantial differences in HHR and
24 absolute risk estimates across PRS deciles for several phenotypes (Supplementary Figure
25 S23, Table S20D). Again, the most significant prediction was obtained for the aggregate

1 SUD-SUM-PRS ($HRR_{10} = 3.92$, $SE = 0.13$, $P = 3.6 \times 10^{-27}$) with the absolute risk trajectory
2 reaching 0.17 at 15 years after first episode for the 10th decile in contrast to the first decile
3 reaching a maximum risk of 0.03 (Figure 4), indicating that the absolute risk for developing
4 SUD is almost 6 times higher for the most genetically burdened (top decile) MD case group
5 compared to the least burdened group in the lowest decile at 15 years after first episode.

6

7

8 **Discussion**

9 We performed an MD GWAS meta-analysis of more than 1.3 million individuals, identifying
10 251 independent risk variants in 243 genomic loci, of which 64 are novel. Among the novel
11 loci, we highlight *GRI1* and *GABRA1*, encoding a glutamate AMPA receptor subunit
12 (GluA1) and a GABA receptor subunit ($\alpha 1$), respectively. Neither of these two loci have
13 previously been identified in GWAS of MD or other mental disorders (Supplementary Table
14 S2). The two genes showed also significant imputed differential expression in the DLPFC (at
15 Bonferroni and $FDR < 0.05$ significance, respectively; Supplementary Table S11) in our
16 TWAS. We note that another GABA receptor subunit gene, *GABRB1*, was also significant (at
17 $FDR < 0.05$) in the TWAS (Supplementary Table S11). *GABRB1*, although not significantly
18 associated with MD in the current GWAS, has previously been implicated in schizoaffective
19 disorder⁷³ and SZ^{74,75}. Our pathway analysis reinforced previous reports^{12,14,23} and extended
20 the enrichment of glutamatergic and GABAergic synapses and functions, further indicating
21 that glutamatergic and GABAergic dysfunctions are key etiologic components in MD. Along
22 with the observed enrichment of GABAergic cell-types present already prenatally, this
23 supports the accumulating multidisciplinary evidence that implicate excitatory/inhibitory
24 (E/I) imbalance with MD⁷⁶ as well as other psychiatric disorders⁷⁷⁻⁸⁰.

25

1 Our TWAS found that *GRI1* expression in DLPFC was decreased in cases compared to
2 controls. This is consistent with findings in preclinical pharmacological studies of mood
3 disorders showing promising results of AMPA receptor potentiators³⁴, as well as the
4 convergent rapid and sustained increase in GluA1 and other synaptic proteins associated with
5 most fast-acting antidepressants³³. Thus, our results provide human genetic evidence that
6 emphasizes positive modulation of the AMPA receptor as an interesting pharmacological
7 approach.

8 In contrast to a reported decrease in *GABRA1* expression in the DLPFC of depressed suicide
9 cases^{81,82}, our TWAS pointed to an increased expression in MD cases. GABA_A receptors
10 provide critical inhibitory control of the firing of glutamatergic excitatory neurons and they
11 are the binding partner of several drugs in mood disorders and potential drug targets for other
12 mental disorders⁸³⁻⁸⁵. Inhibitory synapse formation depends on the expression and selective
13 postsynaptic clustering of GABA_A receptor subunits, which is a combinatorial process in
14 which pentamers are assembled from nineteen different subunits (α 1-6, β 1-3, γ 1-3, δ , ϵ , π ,
15 and θ). The subunit composition varies according to brain region and neuronal subtype, and
16 confers unique physiological and pharmacological properties. Cortical expression of
17 *GABRA1* increases drastically in the first few years of postnatal human development⁸⁶, and in
18 the adult brain, α 1 contributes to the majority of both synaptic and extra-synaptic GABA_A
19 receptors complexes. Antidepressant effects have been associated with pharmacological
20 modulation of both synaptic⁸⁷ and, more promising, extra-synaptic GABA_A receptors⁸⁸.

21 Although a recent study indicated that reversing the E/I imbalance alone only induces short-
22 term antidepressant effects, but not long-term clinical response⁸⁹, our findings support the use
23 of pharmacological modulators of the α 1 β x δ subset of extra-synaptic GABA_A complexes.
24

1 The MiXeR bivariate Gaussian mixture modeling revealed compelling genetic overlap
2 between MD and other traits when considering both concordant and discordant variants,
3 suggesting that psychiatric disorders and correlated traits are substantially more intertwined
4 than indicated by their genetic correlations. Perhaps the most surprising results were seen for
5 educational attainment (EA), showing that 98% of the MD risk variants are also influencing
6 EA and, vice-versa, 88% of EA variants are influencing MD. This almost complete overlap in
7 influencing variants, with concordant (42% of the shared variants) and discordant effects, is
8 consistent with the traits' overall negative genetic correlation ($r_G = -0.23$) and refines the
9 understanding of their polygenic architecture appreciably. Although generally less
10 pronounced, a similar picture was observed for the other mental and behavioural traits
11 examined. Furthermore, the notable overlap suggests that future fine-mapping efforts to
12 pinpoint the causal variants could increase power by combining data from traits with
13 overlapping influencing variants when accounting for the directional effects.

14
15 In the Danish iPSYCH cohort, we assessed PRS-based predictions for developing recurrent
16 depression and psychiatric comorbidity over time and found significant differences in hazard
17 rates and absolute risks between MD cases in the highest and lowest PRS deciles. The
18 iPSYCH cohort includes relatively young individuals (mean age 23.37 years, SE = 6.9, at
19 follow-up December 31st 2016) and thus, a substantial number of the MD cases will develop
20 recurrent MD and/or comorbid disorders later in life. This indicates that the observed risks,
21 and likely also the differences across PRS deciles, will increase further over time. Moreover,
22 the prediction could probably be improved by including other risk factors such as e.g. family
23 history⁹⁰ and clinical/phenotypic variables, as it has been shown in other complex disorders⁹¹.
24 In clinical settings, a targeted effort could be envisaged that offer intensified monitoring for
25 development of e.g. BP, SZ or ANX among MD cases with the highest PRS burden (or

1 combined genetic and clinical risk) to obtain early diagnosis and initiate early treatment,
2 which may have beneficial effects^{92,93}. Similarly, identifying high risk for developing SUD
3 could potentially be an informative point of attention for both the physician and the patient⁹⁴.
4 However, assessment of the clinical utility at the individual level of such potential
5 applications is warranted⁹⁵.

6
7

8 **Methods**

9 *Samples*

10 The iPSYCH2015 sample is a nation-wide population sample extracted from a baseline
11 cohort consisting of all children born in Denmark between May 1, 1981, and December 31,
12 2008, who were alive and resided in Denmark on their one-year birthday, and who have a
13 known mother^{24,25}. Individuals diagnosed with one (or more) of six major psychiatric
14 disorders (ADHD, ASD, MDD, BP, SZ, Anorexia) were identified via the Danish Psychiatric
15 Central Research Register²⁶, which includes data on all individuals treated in Denmark at
16 psychiatric hospitals (from 1969 onward) as well as at outpatient psychiatric clinics (from
17 1995 onward). A random sample from the same birth cohort were chosen as control group.
18 The iPSYCH2015 cohort consists of the initial case-cohort sample iPSYCH2012²⁵ and the
19 recent extension iPSYCH2015i²⁴.

20 Individuals with a ICD10 F32-F33 diagnosis in 2016 or earlier were considered as MD cases
21 and the random population-based sample excluding individuals with an MD diagnosis were
22 used as controls. All individuals diagnosed with BP were excluded, except for the analyses
23 involving co-occurrence of MD and BP diagnosis. MD cases were divided into two groups;
24 1) individuals diagnosed with a single episode of depression, and 2) those fulfilling the
25 criteria for recurrent depression. The date of the first depression episode for each individual

1 was defined as the start date of the first contact in the Danish Psychiatric Central Research
2 Register with an ICD-10 code of F32-F33 diagnoses at an age of 10 years or older. When
3 defining later depression episodes only contacts with the ICD-10 codes: F32, F33.0-F33.3,
4 F33.8-F33.9 were considered. Individuals with the ICD-10 code F33.4 “Major depressive
5 disorder, recurrent, in remission” were omitted because these contacts are unlikely to indicate
6 a new episode of depression. Only ICD-10: F32, F33.0-F33.3, F33.8-F33.9 diagnosis at dates
7 later than 60 days after the end date for all previous ICD10: F32-F33 diagnosis were
8 considered as new episodes of depression, thus categorizing that particular individual as
9 having recurrent depression.

10 In addition to depression phenotypes, individuals with the following diagnoses were recorded
11 in the iPSYCH sample: BP (F30-F31), SZ (F20), SUD (F10-F19, excluding F1X.0 acute
12 intoxication), ANX (F40-F43), ASD (F84.0, F84.1, F84.5, F84.8 or F84), ADHD (F90.0).

13 In the iPSYCH2015 cohort²⁴, we have added 11,710 MD cases and 18,410 controls to the
14 iPSYCH2012 sample^{12,25}, summarizing to a total of 34,095 cases and 45,393 controls prior to
15 relatedness pruning and removal of ancestry outliers (Supplementary Table S14). A total of
16 30,618 cases and 38,200 controls were retained after relatedness pruning and removal of
17 ancestry outliers (Supplementary Table S14). Of these, 6,517 had recurrent episodes of
18 depression (see definition in Methods), while 24,101 were classified as single-episode
19 depression cases (Supplementary Table S14).

20 The number of cases with an additional psychiatric diagnosis (BP, SZ, ANX, ASD, ADHD,
21 SUD) are shown in Supplementary Table S17. Individuals with a diagnosis of BP were
22 excluded from all subsequent analyses except for analyses involving BP.

23 Whether recurrent depression was associated with increased comorbidity compared to single-
24 episode depression was tested using logistic regression for each additional diagnosis while

1 adjusting for age. Individuals were grouped in five-year age bins in order to construct dummy
2 variables for the age adjustment (Supplementary Table S17).

3
4 All analyses of the iPSYCH sample were performed at the secured national GenomeDK high-
5 performance computing cluster in Denmark. The study was approved by the Scientific Ethics
6 Committee in the Central Denmark Region and the Danish Data Protection Agency. The study was
7 approved by the appropriate local scientific ethics committees and IRBs.

8
9

10 *Genotyping, QC and imputation*

11 iPSYCH2015 samples were linked via the unique national personal identification number to
12 the Danish Neonatal Screening Biobank (DNSB) at Statens Serum Institute (SSI), where
13 DNA was extracted from Guthrie cards, and whole-genome amplification was performed in
14 triplicate, as described previously^{96,97}.

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

19 Genotyping of the iPSYCH2012 sample was performed at the Broad Institute of Harvard and
20 MIT (Cambridge, MA, USA) with PsychChip arrays from Illumina according to the
21 manufacturer's instructions. Genotype calling of markers with $MAF \geq 0.01$ was performed
22 by merging call sets from GenCall⁹⁸ and Birdseed⁹⁹, and less frequent variants ($MAF < 0.01$)
23 were called with zCall¹⁰⁰. Genotyping and data processing were carried out in 23 waves.

24 Genotyping of the iPSYCH2015i sample was performed at Statens Serum Institut (SSI,
25 Copenhagen, Denmark) using the Global Screening Array v2 with a Multi disease drop in

1 (Illumina, San Diego, California) according to the manufacturer's instructions. Genotype
2 calling of markers was performed using GenTrain V3. Preimputation quality control was
3 performed using the Ricopili¹⁰¹ pipeline with the specified parameters in the following order.
4 Initially SNPs with a call rate < 0.95 were removed, and subsequently all individuals with a
5 call rate in cases or controls of < 0.95 or an autosomal heterozygosity deviation F_{HET} outside
6 the interval $[-0.2;0.2]$ were removed. Individuals where stated sex was not consistent with sex
7 derived from genotypes were flagged. Subsequently QC was conducted at the marker level,
8 keeping markers with call rate ≥ 0.98 , missing difference ≤ 0.02 between cases and controls,
9 with MAF ≥ 0.01 , Hardy-Weinberg equilibrium (HWE) in controls p value $\geq 1 \times 10^{-06}$ and
10 Hardy-Weinberg equilibrium (HWE) in cases p value $\geq 1 \times 10^{-10}$ (See
11 <https://sites.google.com/a/broadinstitute.org/ricopili/preimputation-qc> for further details).
12 The effects of three batch variables on marker genotypes were tested in iPSYCH2012
13 (ArrayPlate.ID, PreProc.Plate and wave) and iPSYCH2015i separately (Array.Batch,
14 ArrayPlate.ID and PreProc.Plate). This was done using relatedness-pruned dataset with
15 ancestry outliers removed in order to avoid removal of markers where batch effects were
16 caused by population structure or cryptic relatedness rather than genotype artefacts.
17 Pairwise relatedness coefficients ($\hat{\pi}$) were estimated with plink using a LD pruned and MAF
18 filtered set of SNPs (snps-only, window size = 5,000, step size = 300, $r^2 < 0.05$, MAF $>$
19 0.05). Principal Component Analysis was conducted using the same set of LD pruned and
20 MAF filtered SNPs, with random removal of one member of each pair with a relatedness
21 coefficient ($\hat{\pi}$) higher than 0.2. Eigenvectors were inferred using EIGENSOFT version 6.1.4
22 on the relatedness-pruned set of individuals, and subsequently projecting all individuals onto
23 those eigenvectors based on their genotypes. Individuals with all four grandparents born in
24 Denmark were used as a reference for constructing a 3-dimensional ellipsoid using principal

1 components 1, 2 and 3 with a radius of 5 standard deviations from the mean. Individuals
2 located outside this ellipsoid were removed prior to the testing for batch effects.
3 Each genotyped marker was tested for association with each batch versus
4 the remaining batches pooled. This was done for the following batch-variables:
5 ArrayPlate.ID, PreProc.Plate and wave for iPSYCH2012 and Array.Batch, ArrayPlate.ID and
6 PreProc.Plate for iPSYCH2015i. Batch-association testing was conducted using plink
7 v1.90b4.
8 The exclusion of SNPs strongly associated with any of the batch-variables were based on
9 their minimum P-value across all associations per variable. The cut-off for the wave and
10 Array.Batch was $\min(p) < 2 \times 10^{-10}$ and for PreProc.Plate and ArrayPlate.ID $\min(p) < 2 \times 10^{-$
11 12 , based on a Bonferroni correction for the number of markers tested and the number of
12 associations done per batch-variable trying to incorporate the degree of nestedness of these.
13 After removing SNPs falling for any of the above cut-off the remaining distribution was
14 evaluated using QQ-plots. The expected minimum p-distribution was calculated using the
15 inverse cumulative distribution of N independent distributions as suggested in supplementary
16 of Schork et al.¹⁰², N being the number batch-variable values. Reviewing these QQ-plots it is
17 evident that despite filtering by p-value some signal from the batch-variable remains in the
18 dataset.
19 iPSYCH2012 and iPSYCH2015i were imputed using the ricopili pipeline¹⁰¹. Prephasing was
20 done using Eagle v2.3.5¹⁰³ and the subsequent imputation was conducted using Minimac3¹⁰⁴,
21 using the downloadable version of the Haplotype Reference Consortium (HRC) (accession
22 number: EGAD00001002729)¹⁰⁵ as reference.
23
24 *Relatedness pruning and removal of ancestry outliers*

1 Best guest genotypes from iPSYCH2012 and iPSYCH2015i were merged, filtered and LD-
2 pruned down to a set of roughly 30K markers, with imputation INFO score > 0.8 , $r^2 < 0.075$,
3 located outside regions of long-range LD as defined by Price et al.¹⁰⁶, minor allele frequency
4 > 0.05 and no deviation from Hardy-Weinberg proportions ($P > 1 \times 10^{-4}$). Relatedness
5 coefficients, based on “identity-by-state”, were estimated using plink v1.9, in order to
6 identify related (and duplicated samples), with $\hat{\pi} > 0.2$ and one individual of each such pair
7 was subsequently excluded at random. PCA was carried out using the same set of filtered and
8 LD-pruned SNPs as implemented in the ricopili pipeline¹⁰¹. A subsample of European
9 ancestry was selected as an ellipsoid in the space of PC1-3 centred and scaled using the mean
10 and 8 standard deviation of the subsample whose grandparents were all known to be born in
11 Denmark.

12

13 *FinnGen*

14 The FinnGen (<https://www.finngen.fi/en>) study combines genotype data with longitudinal
15 health register data of Finland, including the causes of death, inpatient, outpatient, and drug
16 reimbursement registers.

17 A GWAS for depression was conducted in FinnGen release 6 samples containing 28098
18 cases (ICD-10/9: F32 or F33) and 228817 controls without manic episodes (ICD10 F30),
19 bipolar affective disorder (ICD-10: F31, ICD-9: 296[2-7], ICD8: 296), persistent mood
20 disorders (ICD10: F34) or other or unspecified mood disorder (ICD10: F38,F39, ICD-8:
21 29699). Variants with an imputation info score < 0.6 were excluded. GWAS was run with
22 SAIGE logistic mixed model (version 0.39.1) using age, sex, 10 PCs and genotyping batch
23 as covariates.

24 A GWAS for anxiety disorders was conducted in the FinnGen release 6 sample consisting of
25 7,671 cases with a generalized anxiety disorder (ICD-10: F41.1, ICD-9: 3000C, ICD-8:

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

10 Subjects were genotyped with Illumina and Affymetrix arrays (Illumina Inc., San Diego, and
11 Thermo Fisher Scientific, Santa Clara, CA, USA) as described
12 (<https://www.finnngen.fi/en/researchers/genotyping>). Genotyping and imputation with the
13 Finnish population-specific SISu v3 reference panel were conducted, as described
14 (<https://www.protocols.io/view/genotype-imputation-workflow-v3-0-xbgfijw>). SNPs were
15 pruned for minor allele frequency (MAF) ≥ 0.01 and imputation info score > 0.8 . GWAS was
16 performed using the Scalable and Accurate Implementation of GEneralized mixed model
17 (SAIGE) v0.20¹⁰⁷ with a kinship matrix as a random effect and age, sex, the first 10 principal
18 components (PCs), and genotyping batch as fixed effects.

19
20 *GWAS analyses*

21 Genome-wide-association analyses within iPSYCH were conducted using the RICOPILI
22 pipeline¹⁰¹, applying an additive logistic regression model using dosages of the imputed
23 genotypes. Analyses were adjusted for PC 1-10 from a PCA using the remaining subsample
24 after removal of ancestry outliers and pruning for relatedness. In addition, MD cases with a
25 diagnosis of bipolar disorder were excluded.

1 Summary statistics from iPSYCH2015 and the following external samples were included in a
2 fixed-effects variance-weighted meta-analysis in METAL²⁷: 1) 230,118 broadly defined MD
3 cases and 545,339 controls from Howard et al. 2019¹⁴, including self-reported MD of
4 23andMe²² and the broadly defined MD phenotype of UKB; 2) 35,077 narrowly defined MD
5 cases and 95,406 controls from Wray et al. 2018¹², excluding 23andMe²²; 3) depression
6 phenotypes from the Million Veteran Program (MVP)²³ based on ICD codes derived from
7 electronic health records (83,810 cases and 166,405 controls); 4) 28,098 cases with ICD-10
8 F32 and/or F33 diagnoses and 228,817 controls from FinnGen.
9 For comparison, we also conducted a GWAS and several downstream analyses using a
10 narrower definition of MD that excluded self-reported MD of 23andMe^{12,22} and the broadly
11 defined MD phenotype of UKB^{14,21} (Supplementary Figure S5, Table S3). The results were
12 very similar to the primary analyses without noteworthy differences.

13

14 *SNP-heritability, genetic correlations and overlap with other phenotypes*

15 We estimated SNP-heritability (h_{SNP}^2) for iPSYCH2015 cases with MD, single episode and
16 recurrent depression and for the external GWAS summary statistics outlined above using LD
17 score regression³⁵. Genetic correlations within iPSYCH and between iPSYCH and external
18 GWAS summary statistics was estimated using LD score regression³⁵.

19 In addition, the genetic correlations of MD meta-analysis, single-episode and recurrent
20 depression with other phenotypes were, evaluated using LD score regression¹⁰⁸ at the LD
21 Hub³⁶ website. We used all available phenotypes on LD Hub, but we performed analyses for
22 the UKB traits (N = 597) and the remaining individual phenotypes (N = 258) separately.
23 Levels of experiment-wide significance (Bonferroni correction for number of tests applied)
24 were also established separately within the two groups, i.e. in the UKB traits ($P < 8.38 \times 10^{-5}$)
25 and the remaining individual phenotypes ($p < 0.00019$), respectively. LD Hub traits were

1 supplemented with LD score regression³⁵ analyses performed locally using updated or in-
2 house GWAS summary statistics for traits not available at LD-hub. These include ADHD,
3 ASD¹⁵, cannabis use disorder³⁷, cannabis use¹⁰⁹, alcohol dependence¹¹⁰, drinks per week,
4 smoking ever¹¹¹, age of initiation as a regular smoker, current versus a former smoker
5 (smoking cessation), number of cigarettes per day, smoking initiation¹¹²

6
7 We applied MiXeR⁴³ on our MD GWAS summary statistics and a selection of additional traits
8 (Supplementary Table S7 and Figure 3) to estimate (i) the number of variants explaining 90%
9 of the SNP heritability of each trait and (ii) the genetic overlap between MD and each trait.
10 MiXeR analysis were conducted with default settings (<https://github.com/precimed/mixer>) in
11 a two-step process: 1) a univariate model for each trait to produce estimates of the proportion
12 of variants with non-zero additive genetic effect on the trait (i.e. “polygenicity”) and the
13 variance of effect sizes of these non-zero variants (i.e. “discoverability”). 2) the variance
14 estimates obtained in the univariate analysis were applied in the bivariate model (i.e. MD vs.
15 each of the additional traits) to obtain four estimates representing (i) zero-effect SNPs in both
16 traits; (ii) SNPs with a specific non-zero effect on one of the two traits; and (iii) SNPs with a
17 non-zero effect on both traits. Estimates of polygenic overlap and genetic correlation between
18 pairs of traits were obtained by combining these four parameters.

19

20 *Conditional analysis and finemapping*

21 We identified potential independent genome-wide significant lead variants for each of the
22 broadly defined genome-wide-significant locus identified pipeline from our GWAS meta-
23 analysis results¹⁰¹: Neighboring index SNPs were considered independent when $r^2 < 0.1$ in a
24 sliding 3 Mb window. For each index SNP we defined the associated LD-region by recording
25 the left and rightmost variant with $r^2 < 0.1$. To define GWAS loci, a 50kb window was added

1 on each side of the LD-region and overlapping LD-regions were combined into a single locus.
2 Only a single SNP was kept from within the MHC region, due to extended linkage
3 disequilibrium and a strong association signal of the MHC region (chr 6:25-35 Mb).
4 In order to identify additional independent index variants, we performed a stepwise model
5 selection procedure to select independently associated SNPs also implemented in GCTA-
6 COJO^{28,29}.
7 We assigned posterior probabilities (PP) of being causal to SNPs and constructed credible
8 sets of SNPs that cumulatively capture 95% of the regional posterior probability¹¹³ using
9 PAINTOR¹¹⁴⁻¹¹⁶, CAVIARBF and FINEMAP¹¹⁷, using the CUASALdb pipeline
10 (<https://github.com/mulinlab/CAUSALdb-finemapping-pip>). We applied a conservative
11 approach and assumed one causal variant for each locus.
12 Co-localization analyses were performed in order to evaluate the extent of overlap between
13 eQTL and GWAS signatures in MD and to identify putative causal genes from GWAS
14 associations. Considering only the 95% credible SNPs from the fine-mapped MD-meta-
15 analysis, we integrated GWAS results and eQTL data from a previous meta-analysis
16 integrating signals among three brain datasets⁵⁷ applying a fixed-effect model. The eQTL
17 data originates from eQTL meta-analysis on RNA-sequenced gene expression data from the
18 dorsolateral prefrontal cortex from PsychENCODE⁵⁴ and ROSMAP¹¹⁸, and from 13 brain
19 regions from GTEx¹¹⁹.
20 Using the Coloc method^{58,59}, we extracted eQTL signals of genes within 200 kb distance to
21 significant GWAS variants ($P < 5 \times 10^{-8}$) using effect sizes (β -values) and standard errors
22 from eQTL and GWAS as input. Four hierarchical hypotheses⁵⁹ were tested: H0, no
23 association; H1, GWAS association only; H2, eQTL association only; H3, both GWAS and
24 eQTL association but no co-localization; H4, both GWAS and eQTL association and co-
25 localization, considering $PPH4 > 0.8$ as strong evidence from both GWAS, eQTL and co-

1 localization. In addition, we conducted the eQTL and GWAS CAusal Variants Identification
2 in Associated Regions (eCAVIAR) approach⁶⁰, in which, both eQTL and GWAS were fine-
3 mapped, and the product of posterior probability (CLPP) was calculated, prioritizing genes
4 with at least a single variant with $CLPP \geq 0.01$.

5

6 *Gene-wise and pathway analysis*

7 We used a number of different approaches and data including those available via the FUMA
8 v1.3.6a⁵² website (<http://fuma.ctglab.nl>) for downstream annotation and functional
9 characterization of significant loci.

10 We performed a genome-wide gene-based association study using the Multi-Marker Analysis
11 of GenoMic Annotation (MAGMA) tool, as implemented in FUMA version 1.3.6a⁵². Using
12 this we mapped SNPs from the MD GWAS meta-analysis to 17,840 protein coding genes, and
13 performed Bonferroni correction for the total number of protein-coding genes ($P < 2.8 \times 10^{-6}$,
14 Supplementary Table S9D).

15 Protein coding genes were mapped if they were located with a distance of 10Kb up- or
16 downstream index variants or if a credible variant was annotated to the gene based on eQTL
17 data or chromatin interaction data from human brain (data sets used in the mapping can be
18 found in the Supplementary Note). No additional variant filtering by functional annotation was
19 applied in the eQTL and chromatin interaction mapping. This analysis identified 411 MD risk
20 genes, which were used in a gene-set enrichment analysis within the GENE2FUNC module of
21 FUMA, where we analyzed 8,664 gene-sets derived from GO Biological Process ($N = 7,658$;
22 393 gene-sets) and GO Cellular Components ($N = 1,006$; 86 gene-sets) ontology in the
23 MSigDB database.

24

25 *Tissue and cell type enrichment*

1 We also used FUMA to perform tissue expression analyses on data available through their
2 website, by tested whether the identified genetic associations for both the primary and narrow
3 MD phenotype definitions were enriched regarding transcriptome profiles of human tissues
4 using summary statistics based on all SNPs. Finally, we used FUMA to perform cell-type
5 enrichment analyses⁵² based the MD GWAS summary statistics. We use MAGMA gene-
6 property analysis to test cell type specificity of phenotype with GWAS summary statistics.
7 MAGMA gene-property analysis with scRNA-seq: The gene-property analysis aims to test
8 relationships between cell specific gene expression profiles and disease-gene associations.
9 In all above-mentioned analyses implemented in FUMA default settings were applied.
10 To further evaluate whether the genomic loci implicated in MD are enriched in any particular
11 cell type, we intersected common MD risk variants with two recent epigenomic maps of cell-
12 specific open chromatin^{61,62} using LD score partitioned heritability approach⁶³(Supplementary
13 Figure S14).

15 *Transcriptome-wide association study*

16 In addition, we performed a transcriptome-wide association study (TWAS), imputing the
17 genetically regulated gene expression using EpiXcan⁵³ and using models trained on
18 PsychENCODE Consortium (PEC)^{54,55} expression data for genes and transcripts detected in
19 the dorsolateral prefrontal cortex (DLPFC), with the aim of identifying and prioritizing
20 putative causal loci for the broad MD phenotype definition. A total of 34,646
21 genes/transcripts were tested (transcripts with prediction performance $R^2 > 0.01$ and
22 prediction performance q value < 0.05 with the Benjamini-Hochberg method were retained),
23 and applying a significance threshold of $P < 1.44 \times 10^{-6}$ (corresponding to Bonferroni
24 correction of all genes and isoforms tested; Figure 1A, S9 and Table S11), using information
25 on approximately independent linkage disequilibrium blocks in human populations⁵⁶ to

1 identify independent genomic regions with genes/transcripts showing significant differential
2 gene expression between MD cases and controls.

3
4 *Polygenic risk scores*
5 Using available summary statistics from published GWAS as training datasets, we calculated
6 polygenic risk scores (PRS) for individuals in the iPSYCH2015 sample using LDpred2¹²⁰. Summary
7 statistics were filtered for an imputation info score $\text{INFO} \geq 0.9$ if available. When using external
8 summary stats not processed using RICOPILI or imputed using different imputation references, we
9 excluded all ambiguous markers to avoid potential strand conflicts. To improve performance of the
10 scores and avoid including artefacts from batch effects, we restricted the summary stats to include
11 only SNPs known to be present in both the iPSYCH2012 and iPSYCH2015i data at a reasonable
12 quality (info score 0.6 and MAF 0.01). This step also checked for allele flips.
13 To derive a PRS for MD within the iPSYCH2015 sample we used the MD meta-analysis^{12,14,22,23}
14 excluding all iPSYCH samples as training. We also calculated PRS based on GWAS summary
15 statistics from BP¹²¹, SZ¹²², ADHD¹³ and ASD¹⁵. The PRSs for ADHD and ASD were based on a
16 combination of external GWAS sumstats and internal training. A meta-analysis of two GWAS for
17 anxiety, one based on self-report of physician diagnosis of ANX in the MVP¹²³ and the other being
18 core anxiety in the FinnGen cohort (see FinnGen description above), served as weights for
19 generating an ANX-PRS summarizing genetic risk of anxiety. To derive a PRS for neuroticism we
20 used GWAS summary stats of a weighted neuroticism sum-score, constructed by adding up ten
21 individual item responses (Fed-up, Guilt, Irr, Miss, Mood, Tense, Nerv feel, Suf Nerv, worry emb
22 and worry) by Nagel et al.¹²⁴. In addition, we calculated PRSs using the following GWAS summary
23 stats for traits related to substance-use and substance-use-disorder as training datasets. These include
24 *cannabis use disorder*³⁷(excluding iPSYCH samples, i.e., training only based on data from PGC and
25 data from decode genetics), *cannabis use*¹⁰⁹, *alcohol dependence*¹¹⁰, *drinks per week*, *smoking*

1 *ever*¹¹¹, *age of initiation* as a regular smoker, current versus a former smoker (*smoking cessation*),
2 *number of cigarettes per day*, *smoking initiation*¹¹². All summary statistics were combined in a
3 pseudo-meta-analysis, subsequently used for generating a SU-PRS summarizing genetic liability for
4 substance use while summary stats for *alcohol dependence*¹¹⁰ and *cannabis use disorder*³⁷ were
5 pseudo-meta-analyzed to generate a SUD-PRS summarizing genetic liability of substance use
6 disorder.

7
8 With the aim at improving the PRS prediction we attempted to exploit the genetic overlap of MD
9 with other phenotypes. We therefore combined all of the above PRSs into a single score as a
10 weighted sum¹⁵. We chose here to use the log(OR) for the logistic regression of sub-phenotype of
11 interest on each score as a continuous factor as a measure of ‘importance’ in context of the sub-
12 phenotype of interest (either recurrent vs. single-episode or MD with additional diagnosis of BP, SZ
13 or SUD) and adjusting for PCs. We added a PRS weighted by its log(OR) and the standardized one
14 PRS at the time, starting with the phenotype with highest abs(log(OR)) and ending with the lowest.
15 This way we ended up with a sequence of scores
16 starting with S_0 and continuing with:

$$17 \quad S_k = \frac{\sum_{i=1}^k \log(OR_{P_i}) S_{P_i} - \mu(\sum_{i=1}^k \log(OR_{P_i}) S_{P_i})}{\sigma(\sum_{i=1}^k \log(OR_{P_i}) S_{P_i})}$$

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

21 To examine potential polygenic heterogeneity across MD sub-phenotypes, we investigated how PRS
22 trained on the different phenotypes described above were distributed across MD sub-phenotypes in
23 iPSYCH using a multivariate PRS approach. The method is described in detail in Grove et al.¹⁵, and
24 is a regression of multiple outcome variables, and in principle a linear regression for each PRS on the

1 MD sub-phenotypes allowing for comparisons on the average PRS across sub-phenotypes for PRS
2 from a number of phenotypes while accounting for the inherent correlation between scores and
3 adjusting for necessary covariates. MD-BP cases were excluded from the analyses not involving BP.
4 In addition, we applied a Cox proportional hazard model to estimate hazard rate ratios and absolute
5 risk of developing 1) a second episode of MD, 2) BP, 3) SZ and 4) SUD, among individuals already
6 being diagnosed with their first episode of MD. Analyses were stratified by PRS deciles of the
7 phenotypes described above and using functions from the R-packages survival ([https://CRAN.R-](https://CRAN.R-project.org/package=survival)
8 [project.org/package=survival](https://CRAN.R-project.org/package=survival))^{125,126} and survminer ([https://CRAN.R-](https://CRAN.R-project.org/package=survminer)
9 [project.org/package=survminer](https://CRAN.R-project.org/package=survminer)), while correcting for batch (iPSYCH2012 and iPSYCH2015i) and
10 PCs using an in-house pipeline.

11
12 *MD-PRS in the Philadelphia Neurodevelopmental Cohort (PNC)*
13 MD-PRS were calculated for 4,973 individuals of European ancestry from the Philadelphia
14 Neurodevelopmental Cohort (PNC)^{45,46} using the primary MD GWAS meta-analysis of the current
15 study as training. Genotypes from the first PNC release (dbGaP phs000607.v1.p1) and
16 neurocognitive phenotypes from the third release (dbGaP phs000607.v3.p2) were utilized.
17 Individuals whose genotypically inferred and phenotypically reported sex did not match, those who
18 did not meet the identity by descent (IBD) filter ($\hat{\pi} > 0.185$), those who did not meet the individual-
19 level missingness filter of 0.05, and those with heterozygosity rates +/- three standard deviations
20 from the mean were removed. SNPs that failed to meet the Hardy-Weinberg proportions (HWP),
21 minor allele frequency (MAF), and SNP-level missingness filters of 0.00001, 0.01, and 0.05,
22 respectively, were also removed. Genotype imputation was performed using the reference panel
23 HRC r1.1 2016 on the Michigan Imputation Server
24 (<https://imputationserver.sph.umich.edu/index.html#!>), selecting the "Mixed population" option.
25 Imputed SNPs with an imputation R^2 of < 0.03 and those who failed to meet the missingness, MAF,

1 and HWP thresholds stated above were removed. The first two Multi-Dimensional Scaling (MDS)
2 dimensions were plotted to identify and remove outlier genotypes. The GemTools package in R
3 (<http://www.compgen.pitt.edu/GemTools/GEM%20Documentation.pdf>) and Ward's hierarchical
4 clustering methods were used to identify individuals of European ancestry.
5 PRS-CS¹²⁷ was used to assign per-allele posterior SNP effect sizes to MD GWAS summary
6 statistics. A European LD reference panel provided by the developers of PRS-CS was utilized
7 (<https://github.com/getian107/PRScs>), which draws from the 1000 Genomes Project data. The
8 following PRS-CS default settings were used: parameter a in the γ - γ prior = 1, parameter b in the γ - γ
9 prior = 0.5, MCMC iterations = 1,000, number of burn-in iterations = 500, and thinning of the
10 Markov chain factor = 5. The global shrinkage parameter phi was set using a fully Bayesian
11 determination method. Individual-level MD-PRS were calculated using Plink v2.0¹²⁸. The
12 associations between scaled (mean = 0, SD = 1) MD-PRS and 15 scaled neurocognitive phenotypes
13 in the PNC were assessed using linear regression. Covariates included in the analysis were age at
14 neurocognitive testing, age squared, the first 10 MDS dimensions, sex, and genotyping batch.
15 Adjusted R² was used to report the total variance explained by MD-PRS and model covariates for the
16 15 tested neurocognitive phenotypes. Additionally, a variance partitioning tool
17 (https://github.com/GabrielHoffman/misc_vp/blob/master/calcVarPart.R) was used to determine the
18 variance explained by MD-PRS and each covariate individually. FDR-adjusted p-values were
19 reported.

20

21 **Data availability**

22 Summary statistics from this publication are available at <http://ipsych.au.dk/downloads/>. All
23 relevant iPSYCH data are available from the authors after approval by the iPSYCH Data
24 Access Committee and can only be accessed on the secured Danish server (GenomeDK,

1 <https://genome.au.dk>) as the data are protected by Danish legislation. For data access please
2 contact: Anders D. Børglum.

3 Data used for brain transcriptome model generation are available from PsychENCODE
4 (overview of available data sets at <http://resource.psychencode.org/>); genotypes are
5 controlled data and access instructions are provided at
6 <https://www.synapse.org/#!Synapse:syn4921369/wiki/477467>. Note that some datasets have
7 been indirectly accessed at the respective analytical websites (through the FUMA website).
8 Please refer to these websites (e.g., for FUMA <https://fuma.ctglab.nl/links> and
9 <https://fuma.ctglab.nl/tutorial#datasets>) for availability of datasets used in the respective
10 follow-up analyses / lookups.

11

12 **Code availability**

13 Please refer to individual sections of the methods above for published code (e.g., Ricopili,
14 LDpred2, MiXeR). Because the in-house scripts used for data processing and analysis of the
15 iPSYCH data on the GenomeDK HPC infrastructure is highly dependent on that context, they
16 can only be obtained from the authors upon request. This way we can ensure the proper
17 context is explained in dialogue with the interested parties.

18

19 **Inclusion and ethics**

20 The research has included local researchers throughout the research process, including study design,
21 study implementation, data ownership, intellectual property and authorship of publications. Roles
22 and responsibilities were agreed amongst collaborators ahead of the research.

23 This research does not result in stigmatization, incrimination, discrimination or any personal risk of
24 the participants. Local and regional research relevant to the study have been considered in the listed
25 citations.

1 **Acknowledgements**

2 The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-
3 2014-1724, and R248-2017-2003), the EU H2020 Program (Grant No. 667302, “CoCA”),
4 NIH/NIMH (1U01MH109514-01 and 1R01MH124851-01 to ADB) and the Universities and
5 University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was
6 supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and
7 statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for
8 Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus
9 University, Denmark (grant to ADB).

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

12 The team at the Center for Disease Neurogenomics at Icahn School of Medicine at Mount Sinai was
13 supported by the NIH (K08MH122911 to GV; T32MH087004 to KT; R01MH125246,
14 R01AG067025, U01MH116442 and R01MH109677 to PR).

15

16 **Author contributions**

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

18 **Sample and/or data provider and processing:** J.B.Grauholm, D.D., A.R., G.A., M.B.Hansen, P.Q.,
19 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,
20 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.

21 **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.

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

23 **Study design and direction:** A.D.B.

24 All authors contributed with critical revision of the manuscript.

25

1 **Competing interests**

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

6

7

References

8

- 9 1 Kessler, R. C. & Bromet, E. J. The epidemiology of depression across cultures. *Annual*
10 *review of public health* **34**, 119-138, doi:10.1146/annurev-publhealth-031912-
11 114409 (2013).
- 12 2 Hasin, D. S. *et al.* Epidemiology of Adult DSM-5 Major Depressive Disorder and Its
13 Specifiers in the United States. *JAMA psychiatry* **75**, 336-346,
14 doi:10.1001/jamapsychiatry.2017.4602 (2018).
- 15 3 Pedersen, C. B. *et al.* A comprehensive nationwide study of the incidence rate and
16 lifetime risk for treated mental disorders. *JAMA psychiatry* **71**, 573-581,
17 doi:10.1001/jamapsychiatry.2014.16 (2014).
- 18 4 Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J. Global and
19 regional burden of disease and risk factors, 2001: systematic analysis of population
20 health data. *Lancet* **367**, 1747-1757, doi:10.1016/S0140-6736(06)68770-9 (2006).
- 21 5 Wittchen, H. U. *et al.* The size and burden of mental disorders and other disorders of
22 the brain in Europe 2010. *European neuropsychopharmacology : the journal of the*
23 *European College of Neuropsychopharmacology* **21**, 655-679,
24 doi:10.1016/j.euroneuro.2011.07.018 (2011).
- 25 6 Gustavsson, A. *et al.* Cost of disorders of the brain in Europe 2010. *European*
26 *neuropsychopharmacology : the journal of the European College of*
27 *Neuropsychopharmacology* **21**, 718-779, doi:10.1016/j.euroneuro.2011.08.008
28 (2011).
- 29 7 Angst, F., Stassen, H. H., Clayton, P. J. & Angst, J. Mortality of patients with mood
30 disorders: follow-up over 34-38 years. *J Affect Disord* **68**, 167-181 (2002).
- 31 8 Ferrari, A. J. *et al.* Burden of depressive disorders by country, sex, age, and year:
32 findings from the global burden of disease study 2010. *PLoS medicine* **10**, e1001547,
33 doi:10.1371/journal.pmed.1001547 (2013).
- 34 9 Plana-Ripoll, O. *et al.* Exploring Comorbidity Within Mental Disorders Among a
35 Danish National Population. *JAMA psychiatry* **76**, 259-270,
36 doi:10.1001/jamapsychiatry.2018.3658 (2019).
- 37 10 Steffen, A., Nubel, J., Jacobi, F., Batzing, J. & Holstiege, J. Mental and somatic
38 comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis
39 groups using German nationwide ambulatory claims data. *BMC Psychiatry* **20**, 142,
40 doi:10.1186/s12888-020-02546-8 (2020).

- 1 11 Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major
2 depression: review and meta-analysis. *Am J Psychiatry* **157**, 1552-1562,
3 doi:10.1176/appi.ajp.157.10.1552 (2000).
- 4 12 Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and
5 refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681,
6 doi:10.1038/s41588-018-0090-3 (2018).
- 7 13 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for
8 attention deficit/hyperactivity disorder. *Nature Genetics*, doi:10.1038/s41588-018-
9 0269-7 (2018).
- 10 14 Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102
11 independent variants and highlights the importance of the prefrontal brain regions.
12 *Nat Neurosci* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).
- 13 15 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum
14 disorder. *Nat Genet* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).
- 15 16 Quello, S. B., Brady, K. T. & Sonne, S. C. Mood disorders and substance use disorder:
16 a complex comorbidity. *Sci Pract Perspect* **3**, 13-21, doi:10.1151/spp053113 (2005).
- 17 17 Davis, L., Uezato, A., Newell, J. M. & Frazier, E. Major depression and comorbid
18 substance use disorders. *Curr Opin Psychiatry* **21**, 14-18,
19 doi:10.1097/YCO.0b013e3282f32408 (2008).
- 20 18 Murray, G. K. *et al.* Could Polygenic Risk Scores Be Useful in Psychiatry?: A Review.
21 *JAMA psychiatry* **78**, 210-219, doi:10.1001/jamapsychiatry.2020.3042 (2021).
- 22 19 Rees, E. & Owen, M. J. Translating insights from neuropsychiatric genetics and
23 genomics for precision psychiatry. *Genome Med* **12**, 43, doi:10.1186/s13073-020-
24 00734-5 (2020).
- 25 20 Perkins, D. O. *et al.* Polygenic Risk Score Contribution to Psychosis Prediction in a
26 Target Population of Persons at Clinical High Risk. *Am J Psychiatry* **177**, 155-163,
27 doi:10.1176/appi.ajp.2019.18060721 (2020).
- 28 21 Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK
29 Biobank identifies variants in excitatory synaptic pathways. *Nature communications*
30 **9**, 1470, doi:10.1038/s41467-018-03819-3 (2018).
- 31 22 Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major
32 depression in individuals of European descent. *Nat Genet* **48**, 1031-1036,
33 doi:10.1038/ng.3623 (2016).
- 34 23 Levey, D. F. *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and
35 meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat*
36 *Neurosci* **24**, 954-963, doi:10.1038/s41593-021-00860-2 (2021).
- 37 24 Bybjerg-Grauholm, J. *et al.* The iPSYCH2015 Case-Cohort sample: updated directions
38 for unravelling genetic and environmental architectures of severe mental disorders.
39 *medRxiv*, 2020.2011.2030.20237768, doi:10.1101/2020.11.30.20237768 (2020).
- 40 25 Pedersen, C. B. *et al.* The iPSYCH2012 case-cohort sample: new directions for
41 unravelling genetic and environmental architectures of severe mental disorders. *Mol*
42 *Psychiatry*, doi:10.1038/mp.2017.196 (2017).
- 43 26 Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research
44 Register. *Scand J Public Health* **39**, 54-57, doi:10.1177/1403494810395825 (2011).
- 45 27 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
46 genomewide association scans. *Bioinformatics* **26**, 2190-2191,
47 doi:10.1093/bioinformatics/btq340 (2010).

- 1 28 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide
2 complex trait analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011
3 (2011).
- 4 29 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. Genome-wide complex trait
5 analysis (GCTA): methods, data analyses, and interpretations. *Methods in molecular
6 biology* **1019**, 215-236, doi:10.1007/978-1-62703-447-0_9 (2013).
- 7 30 Stankiewicz, P. *et al.* Haploinsufficiency of the Chromatin Remodeler BPTF Causes
8 Syndromic Developmental and Speech Delay, Postnatal Microcephaly, and
9 Dysmorphic Features. *Am J Hum Genet* **101**, 503-515,
10 doi:10.1016/j.ajhg.2017.08.014 (2017).
- 11 31 Ansar, M. *et al.* Biallelic variants in LINGO1 are associated with autosomal recessive
12 intellectual disability, microcephaly, speech and motor delay. *Genetics in medicine :
13 official journal of the American College of Medical Genetics* **20**, 778-784,
14 doi:10.1038/gim.2017.113 (2018).
- 15 32 de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual
16 disability. *N Engl J Med* **367**, 1921-1929, doi:10.1056/NEJMoa1206524 (2012).
- 17 33 Duman, R. S., Shinohara, R., Fogaca, M. V. & Hare, B. Neurobiology of rapid-acting
18 antidepressants: convergent effects on GluA1-synaptic function. *Mol Psychiatry* **24**,
19 1816-1832, doi:10.1038/s41380-019-0400-x (2019).
- 20 34 Hara, H. *et al.* TAK-653, an AMPA receptor potentiator with minimal agonistic
21 activity, produces an antidepressant-like effect with a favorable safety profile in rats.
22 *Pharmacol Biochem Behav* **211**, 173289, doi:10.1016/j.pbb.2021.173289 (2021).
- 23 35 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
24 polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295,
25 doi:10.1038/ng.3211 (2015).
- 26 36 Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score
27 regression that maximizes the potential of summary level GWAS data for SNP
28 heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279,
29 doi:10.1093/bioinformatics/btw613 (2017).
- 30 37 Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of
31 cannabis use disorder. *Lancet Psychiatry* **7**, 1032-1045, doi:10.1016/S2215-
32 0366(20)30339-4 (2020).
- 33 38 Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563
34 individuals yields insights into biology and relationships with other traits. *Nat
35 Neurosci* **23**, 809-818, doi:10.1038/s41593-020-0643-5 (2020).
- 36 39 Deak, J. D. & Johnson, E. C. Genetics of substance use disorders: a review. *Psychol
37 Med*, 1-12, doi:10.1017/S0033291721000969 (2021).
- 38 40 Gelernter, J. & Polimanti, R. Genetics of substance use disorders in the era of big
39 data. *Nat Rev Genet*, doi:10.1038/s41576-021-00377-1 (2021).
- 40 41 Olvera, R. L. *et al.* Common genetic influences on depression, alcohol, and substance
41 use disorders in Mexican-American families. *American journal of medical genetics.
42 Part B, Neuropsychiatric genetics : the official publication of the International Society
43 of Psychiatric Genetics* **156B**, 561-568, doi:10.1002/ajmg.b.31196 (2011).
- 44 42 Adams, M. J. *et al.* Genetic stratification of depression by neuroticism: revisiting a
45 diagnostic tradition. *Psychol Med* **50**, 2526-2535, doi:10.1017/S0033291719002629
46 (2020).

- 1 43 Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between
2 complex traits beyond genetic correlation. *Nature communications* **10**, 2417,
3 doi:10.1038/s41467-019-10310-0 (2019).
- 4 44 Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the
5 genetic architecture and implicate several cognitive domains. *medRxiv*,
6 2022.2002.2014.22270780, doi:10.1101/2022.02.14.22270780 (2022).
- 7 45 Satterthwaite, T. D. *et al.* Neuroimaging of the Philadelphia neurodevelopmental
8 cohort. *Neuroimage* **86**, 544-553, doi:10.1016/j.neuroimage.2013.07.064 (2014).
- 9 46 Calkins, M. E. *et al.* The Philadelphia Neurodevelopmental Cohort: constructing a
10 deep phenotyping collaborative. *J Child Psychol Psychiatry* **56**, 1356-1369,
11 doi:10.1111/jcpp.12416 (2015).
- 12 47 Gur, R. C. *et al.* Age group and sex differences in performance on a computerized
13 neurocognitive battery in children age 8-21. *Neuropsychology* **26**, 251-265,
14 doi:10.1037/a0026712 (2012).
- 15 48 Wilkinson, G. S. & Robertson, G. J. *Wide range achievement test (WRAT4)*.
16 (Pearson).
- 17 49 Cullen, B. *et al.* Cognitive function and lifetime features of depression and bipolar
18 disorder in a large population sample: Cross-sectional study of 143,828 UK Biobank
19 participants. *Eur Psychiatry* **30**, 950-958, doi:10.1016/j.eurpsy.2015.08.006 (2015).
- 20 50 Lim, J. *et al.* Sensitivity of cognitive tests in four cognitive domains in discriminating
21 MDD patients from healthy controls: a meta-analysis. *Int Psychogeriatr* **25**, 1543-
22 1557, doi:10.1017/S1041610213000689 (2013).
- 23 51 Snyder, H. R. Major depressive disorder is associated with broad impairments on
24 neuropsychological measures of executive function: a meta-analysis and review.
25 *Psychol Bull* **139**, 81-132, doi:10.1037/a0028727 (2013).
- 26 52 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping
27 and annotation of genetic associations with FUMA. *Nature communications* **8**, 1826,
28 doi:10.1038/s41467-017-01261-5 (2017).
- 29 53 Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and
30 shared biological mechanisms mediating susceptibility to complex traits. *Nature*
31 *communications* **10**, 3834, doi:10.1038/s41467-019-11874-7 (2019).
- 32 54 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model
33 for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).
- 34 55 Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD,
35 schizophrenia, and bipolar disorder. *Science* **362**, doi:10.1126/science.aat8127
36 (2018).
- 37 56 Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in
38 human populations. *Bioinformatics* **32**, 283-285, doi:10.1093/bioinformatics/btv546
39 (2016).
- 40 57 Zeng, B. *et al.* Trans-ethnic eQTL meta-analysis of human brain reveals regulatory
41 architecture and candidate causal variants for brain-related traits. *medRxiv*,
42 2021.2001.2025.21250099, doi:10.1101/2021.01.25.21250099 (2021).
- 43 58 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic
44 association studies using summary statistics. *Plos Genet* **10**, e1004383,
45 doi:10.1371/journal.pgen.1004383 (2014).

- 1 59 Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex
2 and Co-localization with Schizophrenia GWAS. *Am J Hum Genet* **102**, 1169-1184,
3 doi:10.1016/j.ajhg.2018.04.011 (2018).
- 4 60 Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes.
5 *Am J Hum Genet* **99**, 1245-1260, doi:10.1016/j.ajhg.2016.10.003 (2016).
- 6 61 Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive
7 sites. *Nature* **584**, 244-251, doi:10.1038/s41586-020-2559-3 (2020).
- 8 62 Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal
9 variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* **52**,
10 1158-1168, doi:10.1038/s41588-020-00721-x (2020).
- 11 63 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-
12 wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404
13 (2015).
- 14 64 Holmes, S. E. *et al.* Lower synaptic density is associated with depression severity and
15 network alterations. *Nature communications* **10**, 1529, doi:10.1038/s41467-019-
16 09562-7 (2019).
- 17 65 Wang, Y. T. *et al.* Novel rapid-acting glutamatergic modulators: Targeting the
18 synaptic plasticity in depression. *Pharmacol Res* **171**, 105761,
19 doi:10.1016/j.phrs.2021.105761 (2021).
- 20 66 Szebeni, A. *et al.* Shortened telomere length in white matter oligodendrocytes in
21 major depression: potential role of oxidative stress. *Int J Neuropsychopharmacol* **17**,
22 1579-1589, doi:10.1017/S1461145714000698 (2014).
- 23 67 Zhou, B., Zhu, Z., Ransom, B. R. & Tong, X. Oligodendrocyte lineage cells and
24 depression. *Mol Psychiatry* **26**, 103-117, doi:10.1038/s41380-020-00930-0 (2021).
- 25 68 Leng, L. *et al.* Menin Deficiency Leads to Depressive-like Behaviors in Mice by
26 Modulating Astrocyte-Mediated Neuroinflammation. *Neuron* **100**, 551-563 e557,
27 doi:10.1016/j.neuron.2018.08.031 (2018).
- 28 69 Musliner, K. L. *et al.* Polygenic liability, stressful life events and risk for secondary-
29 treated depression in early life: a nationwide register-based case-cohort study.
30 *Psychol Med*, 1-10, doi:10.1017/S0033291721001410 (2021).
- 31 70 Musliner, K. L. *et al.* Polygenic Liability and Recurrence of Depression in Patients
32 With First-Onset Depression Treated in Hospital-Based Settings. *JAMA psychiatry*,
33 doi:10.1001/jamapsychiatry.2021.0701 (2021).
- 34 71 Richards, D. Prevalence and clinical course of depression: a review. *Clin Psychol Rev*
35 **31**, 1117-1125, doi:10.1016/j.cpr.2011.07.004 (2011).
- 36 72 Etchecopar-Etchart, D. *et al.* Comorbid Major Depressive Disorder in Schizophrenia:
37 A Systematic Review and Meta-Analysis. *Schizophr Bull* **47**, 298-308,
38 doi:10.1093/schbul/sbaa153 (2021).
- 39 73 Green, E. K. *et al.* Variation at the GABAA receptor gene, Rho 1 (GABRR1) associated
40 with susceptibility to bipolar schizoaffective disorder. *American journal of medical*
41 *genetics. Part B, Neuropsychiatric genetics : the official publication of the*
42 *International Society of Psychiatric Genetics* **153B**, 1347-1349,
43 doi:10.1002/ajmg.b.31108 (2010).
- 44 74 Gulsuner, S. *et al.* Genetics of schizophrenia in the South African Xhosa. *Science* **367**,
45 569-573, doi:10.1126/science.aay8833 (2020).
- 46 75 Brennand, K. J. *et al.* Modelling schizophrenia using human induced pluripotent stem
47 cells. *Nature* **473**, 221-225, doi:10.1038/nature09915 (2011).

- 1 76 Duman, R. S., Sanacora, G. & Krystal, J. H. Altered Connectivity in Depression: GABA
2 and Glutamate Neurotransmitter Deficits and Reversal by Novel Treatments. *Neuron*
3 **102**, 75-90, doi:10.1016/j.neuron.2019.03.013 (2019).
- 4 77 Selten, M., van Bokhoven, H. & Nadif Kasri, N. Inhibitory control of the
5 excitatory/inhibitory balance in psychiatric disorders. *F1000Res* **7**, 23,
6 doi:10.12688/f1000research.12155.1 (2018).
- 7 78 Gao, R. & Penzes, P. Common mechanisms of excitatory and inhibitory imbalance in
8 schizophrenia and autism spectrum disorders. *Curr Mol Med* **15**, 146-167,
9 doi:10.2174/1566524015666150303003028 (2015).
- 10 79 Bruining, H. *et al.* Measurement of excitation-inhibition ratio in autism spectrum
11 disorder using critical brain dynamics. *Scientific reports* **10**, 9195,
12 doi:10.1038/s41598-020-65500-4 (2020).
- 13 80 Nelson, S. B. & Valakh, V. Excitatory/Inhibitory Balance and Circuit Homeostasis in
14 Autism Spectrum Disorders. *Neuron* **87**, 684-698, doi:10.1016/j.neuron.2015.07.033
15 (2015).
- 16 81 Sequeira, A. *et al.* Global brain gene expression analysis links glutamatergic and
17 GABAergic alterations to suicide and major depression. *PLoS One* **4**, e6585,
18 doi:10.1371/journal.pone.0006585 (2009).
- 19 82 Merali, Z. *et al.* Dysregulation in the suicide brain: mRNA expression of corticotropin-
20 releasing hormone receptors and GABA(A) receptor subunits in frontal cortical brain
21 region. *J Neurosci* **24**, 1478-1485, doi:10.1523/JNEUROSCI.4734-03.2004 (2004).
- 22 83 Vinkers, C. H., Mirza, N. R., Olivier, B. & Kahn, R. S. The inhibitory GABA system as a
23 therapeutic target for cognitive symptoms in schizophrenia: investigational agents in
24 the pipeline. *Expert Opin Investig Drugs* **19**, 1217-1233,
25 doi:10.1517/13543784.2010.513382 (2010).
- 26 84 Engin, E., Liu, J. & Rudolph, U. alpha2-containing GABA(A) receptors: a target for the
27 development of novel treatment strategies for CNS disorders. *Pharmacol Ther* **136**,
28 142-152, doi:10.1016/j.pharmthera.2012.08.006 (2012).
- 29 85 Braat, S. & Kooy, R. F. The GABAA Receptor as a Therapeutic Target for
30 Neurodevelopmental Disorders. *Neuron* **86**, 1119-1130,
31 doi:10.1016/j.neuron.2015.03.042 (2015).
- 32 86 Duncan, C. E. *et al.* Prefrontal GABA(A) receptor alpha-subunit expression in normal
33 postnatal human development and schizophrenia. *J Psychiatr Res* **44**, 673-681,
34 doi:10.1016/j.jpsychires.2009.12.007 (2010).
- 35 87 Benasi, G. *et al.* Benzodiazepines as a Monotherapy in Depressive Disorders: A
36 Systematic Review. *Psychother Psychosom* **87**, 65-74, doi:10.1159/000486696
37 (2018).
- 38 88 Fasipe, O. J., Agedo, O. A. & Enikuomihin, A. C. Announcing the novel class of GABA-
39 A receptor selective positive allosteric modulator antidepressants. *Future Sci OA* **7**,
40 FSO654, doi:10.2144/fsoa-2020-0108 (2020).
- 41 89 ten Doesschate, F. *et al.* Neural excitation/inhibition imbalance and the treatment of
42 severe depression. *bioRxiv*, 2021.2007.2009.451784,
43 doi:10.1101/2021.07.09.451784 (2021).
- 44 90 Hujoel, M. L. A., Loh, P.-R., Neale, B. M. & Price, A. L. Incorporating family history of
45 disease improves polygenic risk scores in diverse populations. *bioRxiv*,
46 2021.2004.2015.439975, doi:10.1101/2021.04.15.439975 (2021).

- 1 91 Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of
2 polygenic risk scores. *Nat Rev Genet* **19**, 581-590, doi:10.1038/s41576-018-0018-x
3 (2018).
- 4 92 Vieta, E. *et al.* Early Intervention in Bipolar Disorder. *Am J Psychiatry* **175**, 411-426,
5 doi:10.1176/appi.ajp.2017.17090972 (2018).
- 6 93 Lieberman, J. A., Small, S. A. & Girgis, R. R. Early Detection and Preventive
7 Intervention in Schizophrenia: From Fantasy to Reality. *Am J Psychiatry* **176**, 794-
8 810, doi:10.1176/appi.ajp.2019.19080865 (2019).
- 9 94 Carney, T. & Myers, B. Effectiveness of early interventions for substance-using
10 adolescents: findings from a systematic review and meta-analysis. *Subst Abuse Treat*
11 *Prev Policy* **7**, 25, doi:10.1186/1747-597X-7-25 (2012).
- 12 95 Polygenic Risk Score Task Force of the International Common Disease, A. Responsible
13 use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med*
14 **27**, 1876-1884, doi:10.1038/s41591-021-01549-6 (2021).
- 15 96 Hollegaard, M. V. *et al.* Robustness of genome-wide scanning using archived dried
16 blood spot samples as a DNA source. *BMC Genet* **12**, 58, doi:10.1186/1471-2156-12-
17 58 (2011).
- 18 97 Borglum, A. D. *et al.* Genome-wide study of association and interaction with
19 maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol Psychiatry*,
20 doi:10.1038/mp.2013.2 (2013).
- 21 98 Illumina Gencall Data Analysis Software (Illumina, Inc., San Diego, 2005).
- 22 99 Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs,
23 common copy number polymorphisms and rare CNVs. *Nat Genet* **40**, 1253-1260,
24 doi:10.1038/ng.237 (2008).
- 25 100 Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics
26 and population analysis. *Bioinformatics* **28**, 2543-2545,
27 doi:10.1093/bioinformatics/bts479 (2012).
- 28 101 Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**,
29 930-933, doi:10.1093/bioinformatics/btz633 (2020).
- 30 102 Schork, A. J. *et al.* A genome-wide association study of shared risk across psychiatric
31 disorders implicates gene regulation during fetal neurodevelopment. *Nat Neurosci*
32 **22**, 353-361, doi:10.1038/s41593-018-0320-0 (2019).
- 33 103 Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium
34 panel. *Nat Genet* **48**, 1443-1448, doi:10.1038/ng.3679 (2016).
- 35 104 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and
36 accurate genotype imputation in genome-wide association studies through pre-
37 phasing. *Nat Genet* **44**, 955-959, doi:10.1038/ng.2354 (2012).
- 38 105 Iglesias, A. I. *et al.* Haplotype reference consortium panel: Practical implications of
39 imputations with large reference panels. *Hum Mutat* **38**, 1025-1032,
40 doi:10.1002/humu.23247 (2017).
- 41 106 Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations.
42 *Am J Hum Genet* **83**, 132-135; author reply 135-139, doi:10.1016/j.ajhg.2008.06.005
43 (2008).
- 44 107 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample
45 relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341,
46 doi:10.1038/s41588-018-0184-y (2018).

- 1 108 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and
2 traits. *Nat Genet* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).
- 3 109 Paskan, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic
4 overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci*
5 **21**, 1161-1170, doi:10.1038/s41593-018-0206-1 (2018).
- 6 110 Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common
7 genetic underpinnings with psychiatric disorders. *Nat Neurosci* **21**, 1656-1669,
8 doi:10.1038/s41593-018-0275-1 (2018).
- 9 111 Karlsson Linner, R. *et al.* Genome-wide association analyses of risk tolerance and
10 risky behaviors in over 1 million individuals identify hundreds of loci and shared
11 genetic influences. *Nat Genet* **51**, 245-257, doi:10.1038/s41588-018-0309-3 (2019).
- 12 112 Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights
13 into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244,
14 doi:10.1038/s41588-018-0307-5 (2019).
- 15 113 Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate
16 causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504,
17 doi:10.1038/s41576-018-0016-z (2018).
- 18 114 Kichaev, G. *et al.* Improved methods for multi-trait fine mapping of pleiotropic risk
19 loci. *Bioinformatics* **33**, 248-255, doi:10.1093/bioinformatics/btw615 (2017).
- 20 115 Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical
21 fine-mapping studies. *Plos Genet* **10**, e1004722, doi:10.1371/journal.pgen.1004722
22 (2014).
- 23 116 Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic
24 Fine-Mapping Studies. *Am J Hum Genet* **97**, 260-271, doi:10.1016/j.ajhg.2015.06.007
25 (2015).
- 26 117 Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from
27 genome-wide association studies. *Bioinformatics* **32**, 1493-1501,
28 doi:10.1093/bioinformatics/btw018 (2016).
- 29 118 Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project.
30 *Journal of Alzheimer's disease : JAD* **64**, S161-S189, doi:10.3233/JAD-179939 (2018).
- 31 119 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues.
32 *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).
- 33 120 Prive, F., Arbel, J. & Vilhjalmsson, B. J. LDpred2: better, faster, stronger.
34 *Bioinformatics*, doi:10.1093/bioinformatics/btaa1029 (2020).
- 35 121 Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar
36 disorder cases provides new insights into the underlying biology. *Nat Genet*,
37 doi:10.1038/s41588-021-00857-4 (2021).
- 38 122 Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
39 schizophrenia. *Nature*, doi:10.1038/s41586-022-04434-5 (2022).
- 40 123 Levey, D. F. *et al.* Reproducible Genetic Risk Loci for Anxiety: Results From
41 approximately 200,000 Participants in the Million Veteran Program. *Am J Psychiatry*
42 **177**, 223-232, doi:10.1176/appi.ajp.2019.19030256 (2020).
- 43 124 Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level
44 analyses reveal genetic heterogeneity in neuroticism. *Nature communications* **9**, 905,
45 doi:10.1038/s41467-018-03242-8 (2018).
- 46 125 Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox*
47 *Model.* (Springer, 2000).

1 126 A Package for Survival Analysis in R, R package version 3.2-13 (2021).
2 127 Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via
3 Bayesian regression and continuous shrinkage priors. *Nature communications* **10**,
4 1776, doi:10.1038/s41467-019-09718-5 (2019).
5 128 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and
6 richer datasets. *GigaScience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
7

Table 1: Top 5 index SNPs for previous and top 15 index SNPs for novel MD 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 Major Depression and 978,703 controls. The location (chromosome (Chr)) base position (BP)), alleles (A1/A2), odds ratio (OR) of the effect with respect to A1, standard error (SE) and association P-values (P) from inverse-variance weighted fixed effects model of the index variants are given. “Novel” indicates if the locus is a novel MD risk locus i.e. not identified previously. Genes located within 50 kb from index variants are listed (See supplementary table 2 for details on all 303 index SNPs)

| Loc | Chr | indexSNP | BP | A1/A2 | cases | ctrls | OR | SE | P | Nearby genes | Novel |
|-----|-----|------------|-----------|-------|-------|-------|------|--------|---------|---|-------|
| 8 | 1 | rs7531118 | 72837239 | T/C | 0.473 | 0.461 | 0.97 | 0.0029 | 1.5e-27 | <i>NEGR1</i> | no |
| 9 | 1 | rs10890020 | 73668836 | A/G | 0.508 | 0.510 | 0.97 | 0.0029 | 1.7e-21 | <i>LINC01360</i> | no |
| 38 | 2 | rs1320138 | 144158287 | T/C | 0.428 | 0.435 | 1.02 | 0.0028 | 4.8e-09 | <i>ARHGAP15</i> | yes |
| 45 | 2 | rs6715105 | 198445601 | T/C | 0.326 | 0.341 | 1.02 | 0.0030 | 1.3e-09 | <i>ANKRD44, ANKRD44-IT1, SF3B1, COQ10B, HSPD1, SNORA105A, SNORA105B, HSPE1, HSPE1-MOB4, MOB4, RFTN2, MARS2, BOLL, PLCL1</i> | yes |
| 53 | 3 | rs56029819 | 43478295 | T/C | 0.853 | 0.849 | 0.98 | 0.0040 | 2.5e-09 | <i>SNRK, SNRK-AS1, ANO10</i> | yes |
| 80 | 5 | rs4262121 | 31078958 | G/C | 0.528 | 0.546 | 0.98 | 0.0029 | 4.3e-09 | - | yes |
| 89 | 5 | rs62379847 | 120109119 | A/C | 0.655 | 0.652 | 0.98 | 0.0030 | 1.8e-09 | <i>PRR16</i> | yes |
| 93 | 5 | rs2964003 | 153216733 | A/G | 0.829 | 0.820 | 1.02 | 0.0038 | 8.1e-10 | <i>GRIA1, LINC01861</i> | yes |
| 94 | 5 | rs4596421 | 161269895 | C/T | 0.708 | 0.693 | 0.98 | 0.0031 | 2.4e-09 | <i>GABRA1</i> | yes |
| 98 | 6 | rs35883476 | 28368508 | G/C | 0.910 | 0.918 | 1.05 | 0.0052 | 1.3e-21 | <i>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</i> | no |
| 99 | 6 | rs10947690 | 37631768 | A/G | 0.751 | 0.756 | 0.98 | 0.0033 | 6.3e-09 | <i>MDGA1</i> | yes |
| 114 | 7 | rs957360 | 3660918 | C/G | 0.714 | 0.702 | 1.02 | 0.0031 | 1.3e-09 | <i>SDK1</i> | yes |
| 124 | 7 | rs1986692 | 133743393 | A/G | 0.612 | 0.610 | 1.02 | 0.0029 | 2.8e-09 | <i>EXOC4</i> | yes |
| 156 | 10 | rs1909696 | 77582203 | G/T | 0.337 | 0.328 | 0.98 | 0.0030 | 3.3e-09 | <i>LRMDA, LOC105378367</i> | yes |
| 158 | 10 | rs1021363 | 106610839 | A/G | 0.349 | 0.335 | 1.03 | 0.0029 | 5.3e-25 | <i>SORCS3</i> | no |
| 191 | 13 | rs9561331 | 94017476 | G/A | 0.869 | 0.873 | 0.98 | 0.0042 | 3.7e-09 | <i>GPC6</i> | yes |
| 199 | 14 | rs7141014 | 98667928 | T/C | 0.796 | 0.806 | 0.98 | 0.0035 | 3.3e-09 | - | yes |
| 210 | 15 | rs4886915 | 78075023 | A/G | 0.421 | 0.423 | 0.98 | 0.0028 | 6.1e-10 | <i>LINGO1</i> | yes |
| 225 | 17 | rs60856912 | 65892343 | G/T | 0.825 | 0.817 | 0.97 | 0.0038 | 1.8e-11 | <i>BPTF, C17orf58, KPNA2</i> | yes |
| 233 | 18 | rs12967143 | 53099012 | G/C | 0.301 | 0.292 | 1.03 | 0.0031 | 2.0e-21 | <i>TCF4, TCF4-AS1, MIR4529</i> | no |

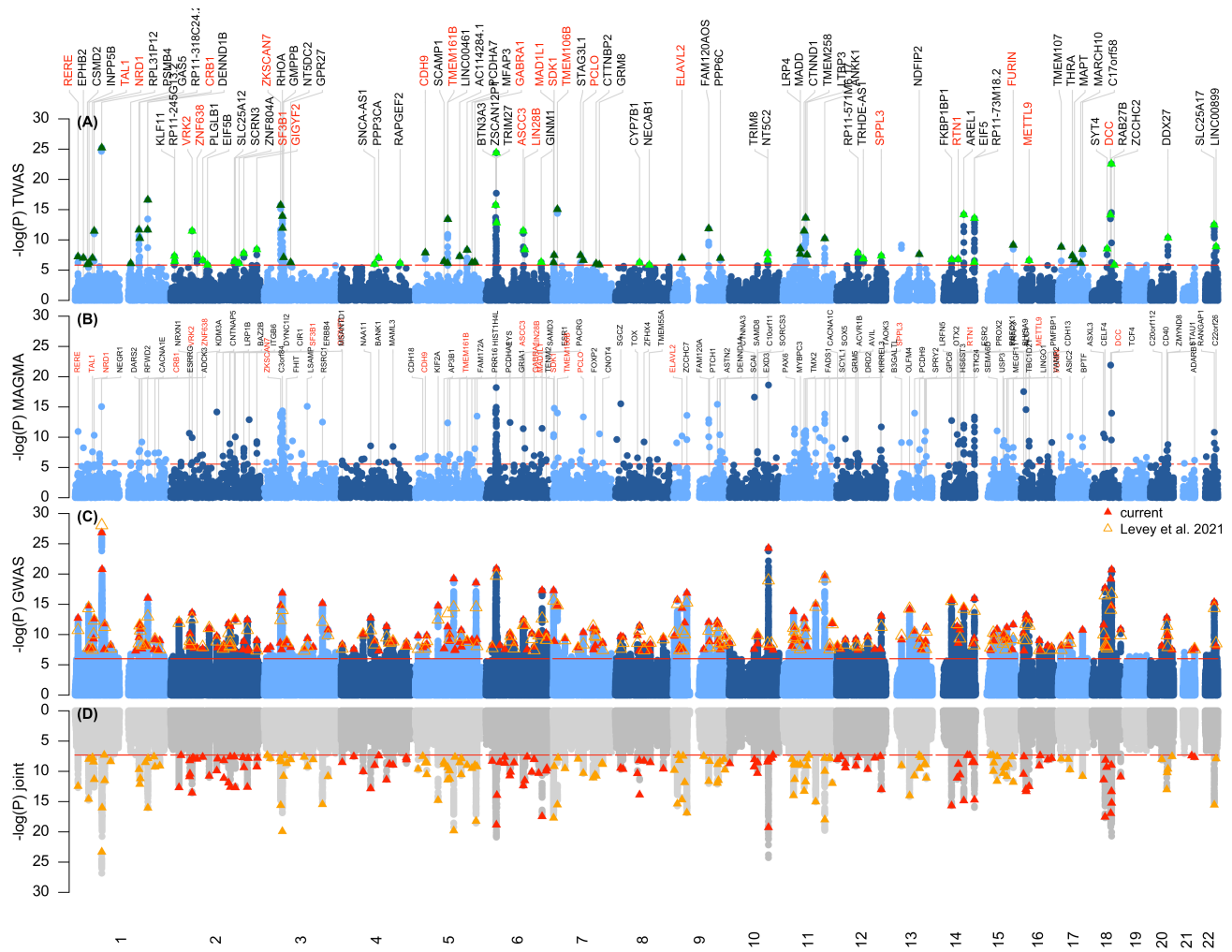


Figure 1: Miami plots... (A) TWAS, genes are represented by both gene expression and isoform expression (= features, represented by the dots). A red line indicates Bonferroni corrected genome-wide significance within analyses; $p < 1.44 \times 10^{-6}$; corresponding to Bonferroni correction of all the 34,646 features. The top transcript is labeled for each independent linkage disequilibrium block as previously described (PMID: 26395773). (B) Gene-based MAGMA analysis: A total of 268 significant genes were located in 93 of the 238 genomic risk loci identified using FUMA, while the remaining 143 significant genes were located outside these loci (table S7). The most significant gene within each of the 93 genomic risk loci are labeled in this plot. Gene names are labelled red if they overlap with the top gene within each independent locus of the TWAS in (A). (C) Manhattanplot of the meta-analysis GWAS of major depression (broad definition). Neighbouring index SNPs were considered independent when $r^2 < 0.1$ in a sliding 3 Mb window. For each index SNP we defined the associated LD-region by recording the left and rightmost variant with $r^2 < 0.1$. 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. 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). (D) Joint p-values as implemented in GCTA-COJO ingreen, while original p-values are in shades of grey.

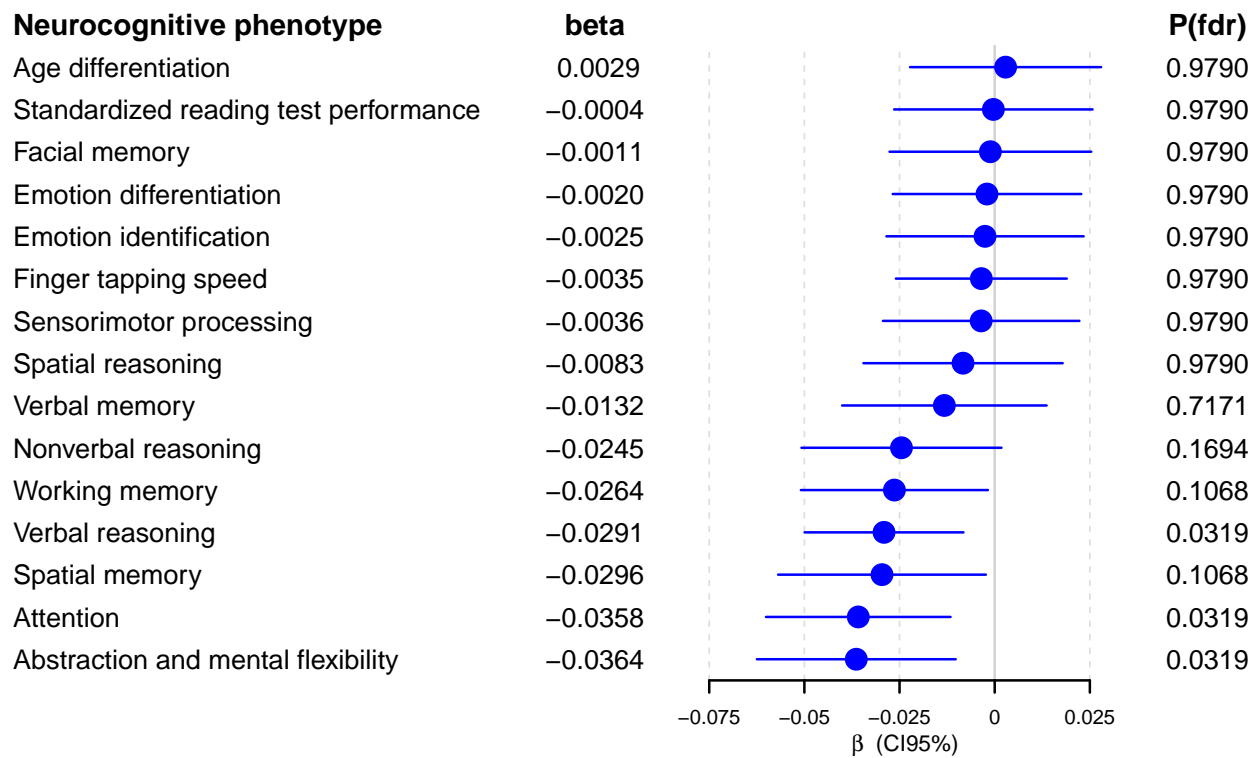


Figure 2: Association of MD-PGS with measures of cognitive abilities in the PNC cohort (N=4,973). Beta values (and 95% confidence intervals indicated as horizontal bars) from linear regression testing for the association of MD-PGS with the 15 neurocognitive measures listed on the y-axis. β values and FDR adjusted p-values are indicated.

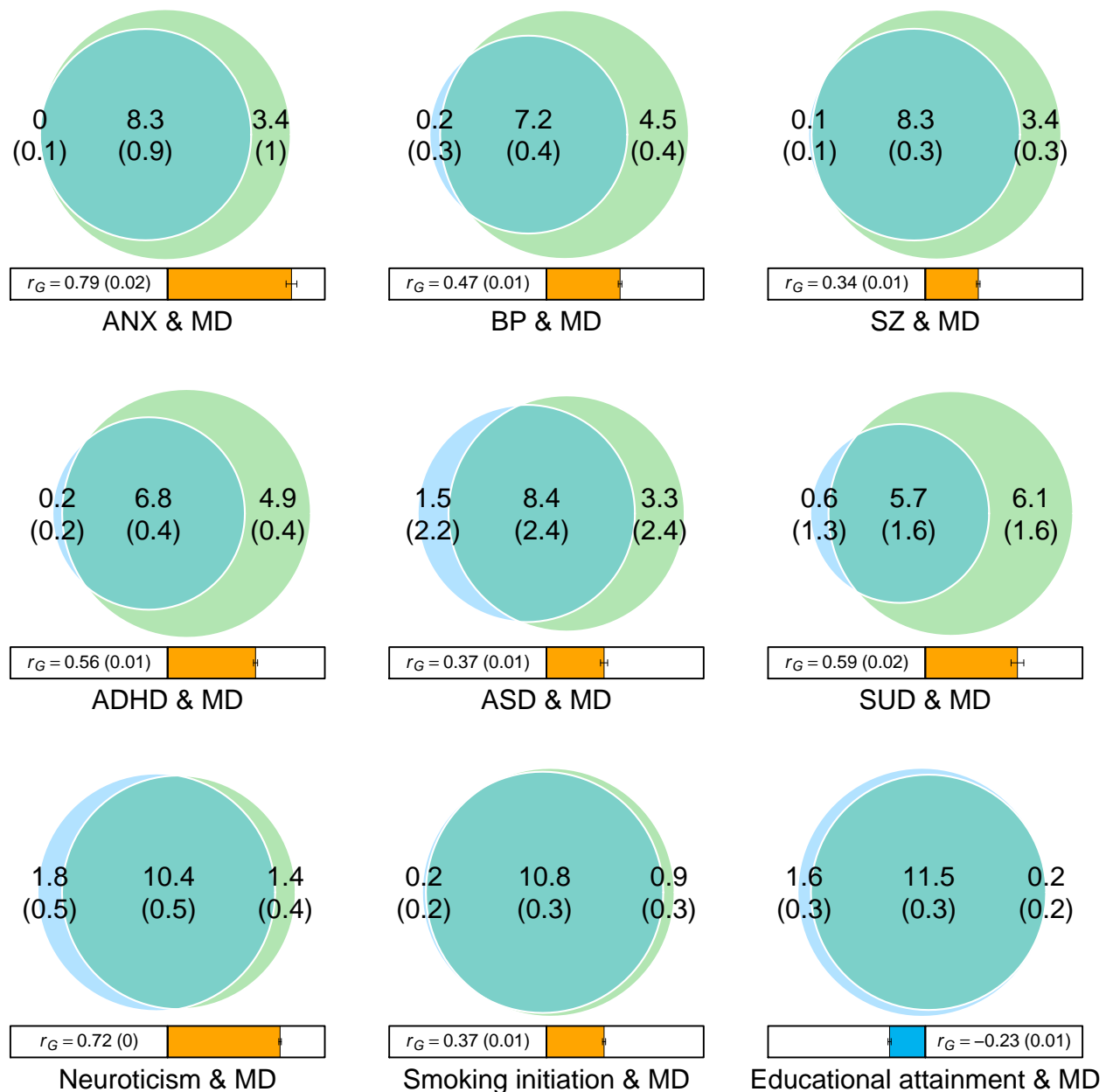


Figure 3: MiXeR bivariate analysis. Venn diagrams showing MiXeR results of the estimated number of variants shared between MD and psychiatric disorders (with significant genetic correlations with MD) and phenotypes representing other domains with high genetic correlation with MD. Circles represent loci unique to MD (green) and unique to the phenotype of interest (blue) and corresponding overlapping (shared) loci. The number of shared variants (and standard errors) are shown in thousands. The size of the circles reflects the polygenicity of each phenotype, with larger circles corresponding to greater polygenicity. The estimated genetic correlation (r_g and 95% confidence intervals) between MD and each phenotype from LDSC is shown below the corresponding Venn diagram, with an accompanying scale (-1 to +1) with blue and orange representing negative and positive genetic correlations, respectively. Bivariate results for the following comparisons are shown: BIPxMDD , SCZxMDD , ADHDxMDD , ASDxMDD , ALDCUDxMDD , neuroxMDD , SmoInixMDD , EAxMDD , (see also Supplementary Table XXX).

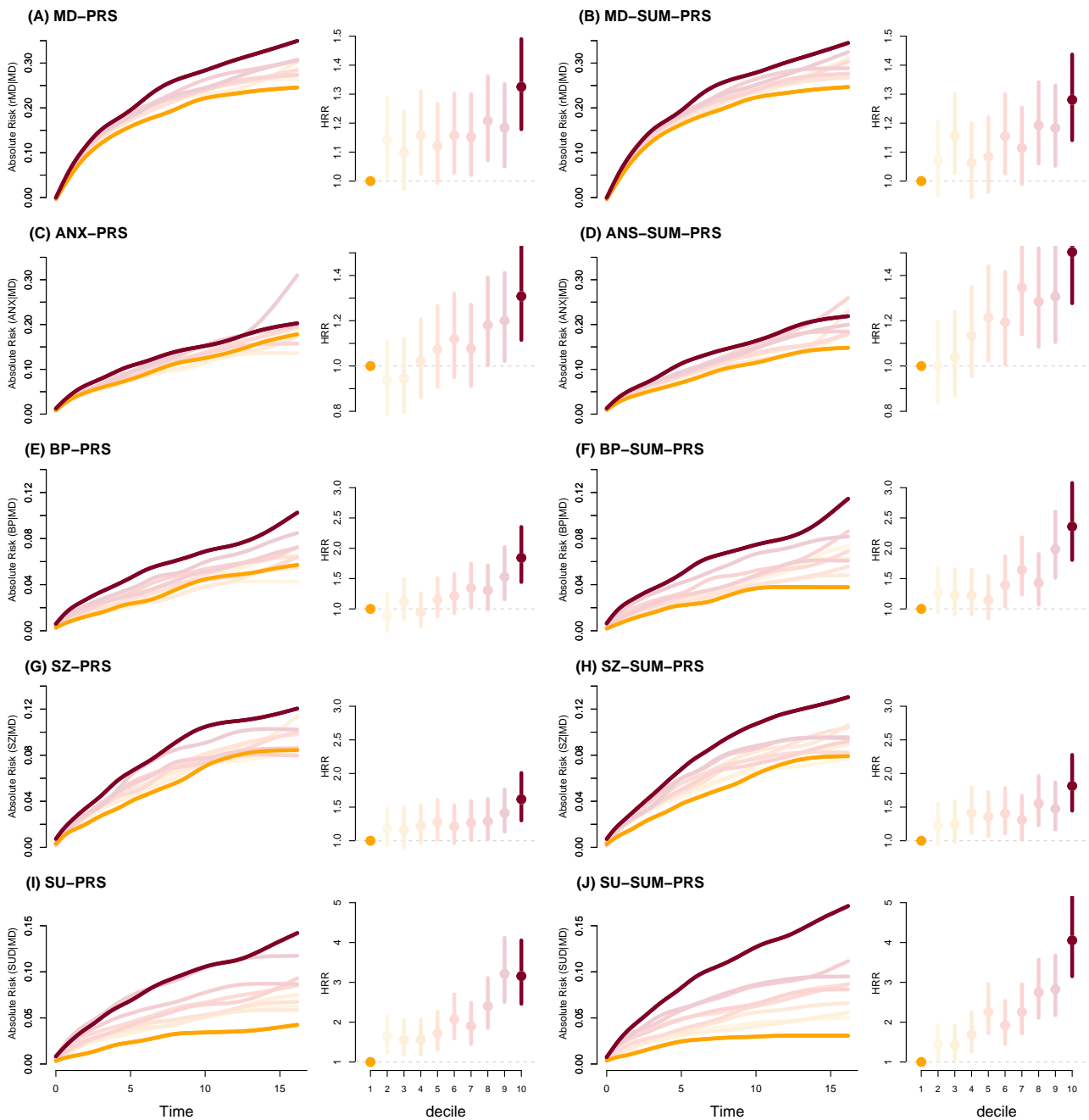


Figure 4: Absolute risk and HRR of developing a second episode of depression stratified by MD-PRS deciles (A) and MD-SUM-PRS (B). Absolute risk and HRR of transitioning into bipolar disorder stratified by BP-PRS deciles (C) and BP-SUM-PRS (D). Absolute risk and HRR of transitioning into schizophrenia stratified by SZ-PRS deciles (C) and SZ-SUM-PRS (D). Absolute risk and HRR of transitioning into substance-use-disorder stratified by SU-PRS deciles (C) and SUD-SUM-PRS (D). MD-SUM-PRS, BP-SUM-PRS, SZ-SUM-PRS and SUD-SUM-PRS were calculated by adding PRSs for multiple phenotypes weighted by $\log(OR)$ with the aim of optimising prediction (see methods for details). See supplementary figures S15, S17, S18 and S19 for results based on all PRSs analysed in the current study.